

Unlocking the Power of Edge Computing for CPS

Friday Nov 16, 10:15 AM to 12:45 PM (2.5 hours)

Location: **NSF CPS PI meeting**

URL for the entire conference video recording: <https://bluejeans.com/s/3jd7q>

10:15 AM– 10:25 AM

Welcome and overview of the day ([slides](#))

10:25 AMto 11:05 AM: Session A (Chair: Anish Arora)

(Three talks each 8 mins, followed by 15 min Q&A for the presenters)

Bob Iannucci, CMU-Silicon Valley, “Network Edge Considered Harmful” ([slides](#))

Mung Chiang, Purdue/Princeton, “Fog/Edge and Dispersive AI” ([slides](#))

Mahadev Satyanarayanan, CMU, “Research Challenges in IoT” ([slides](#))

11:05 to 11:10

Stretch time

11:10 AMto 11:50 AM: Session I (Chair: Bob Iannucci)

(Three talks each 8 mins, followed by 15 min Q&A for the presenters)

Prashant Shenoy, UMass, “Edge-enabled Utility-preserving Privacy for Data-driven CPS Systems” ([slides](#))

Bharath Balasubramanian, AT&T, “State Management for Telco’s Edge” ([slides](#))

Vladimir Kolesnikov, Georgia Tech, “Efficient Crypto Techniques for the Edge” ([slides](#))

11:50 to 11:55 AM

Stretch time

11:55 AMto 12:20 AM: Session 1 (Chair: Kishore Ramachandran)

(Two talks each 8 mins, followed by 10 min Q&A for the presenters)

Aakanksha Chowdhery, Google Brain, “From Cloud to Edge: Advances in Mobile AI” ([slides](#))

Sanjiv Doshi, CISCO, “Practical approaches to managing, orchestrating and securing cyber-physical systems” ([slides](#))

12:20 to 12:45 PM: Wrap up (Chairs: Anish Arora and Kishore Ramachandran)

12:45 PM: Adjourn

3:00 – 4:00 PM: Report back to PI meeting of the Workshop Summary (Anish Arora and Kishore Ramachandran) ([slides](#))

Titles and Abstract of Talks

Session A

Bob Iannucci, CMU-Silicon Valley

Title: "Network Edge Considered Harmful"

Abstract: The so-called edge of the network has become a focus of interest and activity. This has been motivated "from below" — by compute- and energy-impooverished devices seeking to offload computing (the Cloudlet model)... and "from above" — seeking to reduce latency by moving computing from the cloud to the network edge (the Fog Computing model). The benefits are clear, and the costs (e.g., investments in hardware at the edge, the need for security- and privacy-preserving mechanisms, resource allocation and QoS challenges, the issues that come with migratory devices) are now being addressed. This is giving rise to thinking of the edge of the network as a foundational concept — one worthy of a name and a place in the hierarchy of computing. Should it be so?

We argue that this so-called edge of the network is more of an artifact of time and technology than a well-justified abstraction. It is true that the vagaries and energy taxes of making the wireless hop from a device to the network demand special attention. But such concerns should only rise to the attention of programmers once all alternatives have been exhausted and, hopefully, only in abstract terms. The cost of computing remains such a concern only in a declining sense. What can be done on a modern smartphone is truly impressive, and the capabilities of IoT devices will follow suit. Coin-cell-sized devices boot and run Linux today and are rapidly becoming full-fledged multitasking, multi-context systems running neural networks. We conclude that the characteristics that make the network edge distinct may vanish or should be hidden from view.

A more fundamental concern, in our view, is the mental burden that comes from compelling programmers to develop and verify apps that consist of separately-created programs for each of (a) the devices, (b) the cloud, and now (c) the edge. In each generation of computing, it has been the enabling of a vast ecosystem of programmers that has ushered in economic transformation. The edge, as an abstraction, may be a disabler rather than an enabler. We argue that the revolution promised by the Internet of Things can only take place when the same millions of programmers who made mobile apps a phenomenon can do the same to, say, the smart city with comparable effort. As such, we would do well to consider not how to create a new programming nexus (the edge) but rather how to erase it and others — a concept we call Edgelessness. From a hardware standpoint, it is certainly true that adding computing resources at points in the network as close to the devices is beneficial. But programming this hardware needs to be transparent. The objective of edgeless computing is to make such resources available but to mask, to the maximum extent possible, the programming complexities of having done so. We must think of entities like smart cities as integrated computing platforms in which a single programmer can easily write an app, determine if it is logically correct, and separately address questions of what-computing-goes-where, how synchronization happens and to what precision, how processes are triggered to migrate, how routing/switching/naming works, and what to do when all else fails. It must enable the programmer to mask inevitable failures with redundancy.

To this end, we consider the sketch of a meta-programming language called TickTalk along with foundational hardware support that could move us a bit closer to the economic transformation we seek.

Mung Chiang, Purdue/Princeton

Title: "Fog/Edge and Dispersive AI"

Abstract: The cloud-to-things continuum offers promises and pitfalls of placing functionalities. I hope to spend a few minutes on potential architectures for edge/fog that matches the unique advantages/challenges, and then a few minutes on its application to dispersive AI and collaborative computing.

Mahadev Satyanarayanan, CMU

Title: "Research Challenges in IoT"

Abstract: Edge computing is a new paradigm in which access to the cloud is mediated by small multi-tenant "data centers in boxes" called cloudlets that are placed in close network proximity to mobile devices and sensors in the IoT. This lowers end-to-end latency for highly immersive human-in-the-loop applications such as augmented reality, and tightly-constrained device-in-the-loop applications such as drone control. It also improves bandwidth scalability for applications such as real-time analytics on a large array of video cameras. In this talk, I will briefly summarize the research challenges that emerge in three areas: (a) Security and Privacy; (b) Limited Elasticity; and (c) Dynamic Data Aggregation and Reconfiguration.

Session I

Prashant Shenoy, UMass

Title: "Edge-enabled Utility-preserving Privacy for Data-driven CPS Systems"

Abstract: Data-driven CPS systems in domains such as energy, transportation, and health are becoming increasingly common. Such systems subject sensed data to sophisticated analytics to glean insights, which may also reveal private user information. Data obfuscation methods proposed for CPS systems have the potential to preserve user privacy but also remove any utility present in the data for performing useful analytics. In this talk, I will argue for the need for utility-preserving privacy techniques for CPS systems that maintain user privacy while also retaining the ability to perform analytics on the data to provide useful smart services to users. I will describe how edge computing can simplify the deployment of such utility-preserving privacy transformations for data-driven CPS systems and applications.

Bharath Balasubramanian, AT&T

Title: "State Management for Telco's Edge"

Abstract: A fundamental component of AT&T's Edge Cloud is a state-management service that provides the ability to federate, replicate and manage state across 1000s of sites across the world. Designing such a state-management service is a challenging problem due to the following reasons (among many): (i) As the edge use-cases become richer with AR/VR and 360 video, the amount of data generated at the edge will be massive and the state-management service needs to deal with this volume of data. (ii) To deal with wide-area latencies and the need for partitioned operation across the world, the state-management service needs to leverage flexible consistency semantics for data replication to provide much better performance without compromising on service correctness. For example, authentication information maintained by the VNF-management plane will need much stronger semantics than some of the application data. Rather than provide a one-size-fits-all data replication solution, we can use this diversity. (iii) The state-management service has to adapt to a wide range of access patterns ranging from mobile edge users to relatively static VNF-management services. (iv) Finally, it also has to deal with data privacy and isolation needs, especially for use-cases that dictate that data never leaves the edge. Existing solutions fail to address one or more of these challenges and hence, we as a research community need to explore alternate solutions.

Vladimir Kolesnikov, Georgia Tech

Title: "Efficient Crypto Techniques for the Edge"

Abstract: In this brief talk I will discuss several efficient cryptographic techniques which may be effectively used at the network core and the edge, where low latency and high performing security is needed. Time permitting, I will review secure computation, zero-knowledge proofs, and aggregated authentication.

Session 1

Aakanksha Chowdhery, Google Brain

Title: "From Cloud to Edge: Advances in Mobile AI"

Abstract: Smartphone applications and embedded platforms are ubiquitous today enabling data analytics to drive new insights and automation in the next-generation of cyber-physical systems. Neural-network based techniques promise analytics with high-accuracy on image, video and text data but they require novel ways of deploying them on the mobile devices because the computational cost can be prohibitive and we need to protect consumer data. In this talk, I briefly describe some of the key advances in mobile AI enabling machine learning on such devices. These advances include a) the lightweight machine learning platforms for e.g., TensorFlow Lite, b) compression techniques to reduce the inference latency of the over-parametrized neural networks, c) the emergence of new hardware architectures such as edgeTPUs to accelerate neural-network computations, and d) the availability of differential privacy tools to retrain models without sending all the data from mobile devices to the cloud.

Sanjiv Doshi and Russ Gyurek, CISCO

Title: "Practical approaches to managing, orchestrating and securing cyber-physical systems"

Abstract: A cyber-physical system (CPS) is not a monolithic environment where one size fits all. Every use case is unique and dependent on several attributes like scale, connectivity, physical attributes etc. For example requirements for an outdoor deployment in bandwidth constrained utility setting is different from roadside intersection deployment, to a manufacturing plant. One thing that is certain though, is the need for network elements to enable CPS. This gives a network device manufacturer a front row seat in this increasingly important ecosystem. Being at this desirable location enables us to secure the "thing" devices helping with secure onboarding, end-to-end segmentation, adaptive deception, DDoS attacks as well as transport agnostic things visibility. By adopting cloud native semantics we are creating an ecosystem of distributed apps that our customers and third party developers can develop for a custom CPS environment. Cisco is leveraging its unique position by transforming a set of network elements to a platform tuned to this environment. This talk will focus on deployment scenarios, present approaches and future direction in a CPS ecosystem.