Research Statement

The Internet of Things (IoT), a network in which everyday objects automatically communicate to computers and each other to provide services, is considered to hugely impact our everyday lives and the US National Intelligence Council lists it as one of six “Disruptive Civil Technologies” with potential impacts on national interests [1]. IoT is growing at a rapid pace; some estimates put a vision of a trillion devices by 2035 [2], generating zettabytes of data [3]. This scale introduces unprecedented challenges due to three key issues. First, devices require frequent battery replacements and also become obsolete as their aging hardware is unable to meet requirements to support quickly-evolving AI and other cutting-edge software, resulting in a severe impact on the carbon footprint. Second, an over reliance on the cloud results in IoT applications which are slow to react, limits non-Internet use cases, and doesn’t scale well due to high bandwidth and cloud operating costs. Finally, users face a lack of transparency and control of their data privacy with the scale at which data is being collected and used in their everyday lives.

My work addresses these challenges by using “ambient compute”, compute power which is available at nearby, powerful devices. Compared to resource-constrained IoT devices, ambient compute is offered by resource-capable IoT devices, which have: i) better computing resources, ii) considerable idle time, implying computational resources to spare, and iii) a steady source of energy. Using ambient compute reduces the carbon footprint of IoT devices, reduces reliance on cloud, and provides users with better privacy controls over their IoT data.

For low power devices, nearby gateway devices like Raspberry Pis are present to collect data from these devices and send to the cloud. We utilize these gateway devices as the ambient compute which can execute applications with reasonable workload using data from multiple devices. For example, this can be an application which triggers alerts for anomalies in plug-level appliance power data in a factory. But combining gateways into a cohesive computational unit is challenging as devices can be heterogeneous, gateways could fail, and networking complexity can complicate writing applications. To address these challenges, we build NexusEdge [4], a distributed middleware, a software which runs on each of the gateways enabling them to work as a cohesive platform. Compared to prior works, the middleware provides resilience to gateway failures by migrate applications from one gateway to another, and optimizes which gateway to run on based on device requirements. When compared to Amazon’s AWS IoT Greengrass, NexusEdge shows a 10x improvement in application latency, and a 2.5x reduction in network traffic, indicating better scalability and responsiveness.

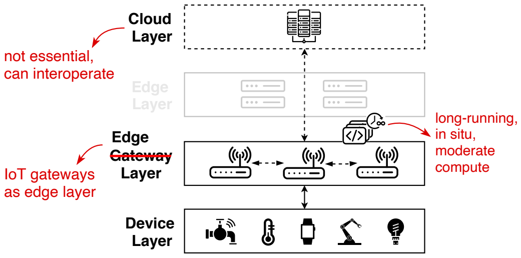
Cloud Layer
not essential, can interoperate
Edge Layer
long-running, in situ, moderate compute
Edge ~~Gateway~~ Layer
IoT gateways as edge layer
Device Layer

Figure 1: Ambient Compute from Gateways

Resource-constrained IoT devices like AR/VR headsets can improve the quality of service (QoS) by relying on the type of ambient compute provided by another class of smart home and office devices like gaming consoles. For example, an application on a smart watch which detects if its user is washing their hands using AI could save watch battery power and improve the prediction quality by utilizing the ambient compute provided by a more powerful GPU on a nearby Playstation 5. Prior works have studied task offloading to dedicated and expensive “edge computing” servers, and require prior profiling of tasks to meet deadlines which is impractical as IoT environments consist of a variety of tasks as well as devices. In our work, we leverage existing devices available in the user’s space and avoid task profiling and note that tasks typically repeat in user environments. We use this to train a reinforcement learning based task scheduler, fReeLoaders [5], which learns over time the best resource-capable device to execute a given task. Compare to other baseline task schedulers, fReeLoaders show 17.9% improvement in the task QoS and a 61.8% reduction in energy usage, and is adaptable and customizable.

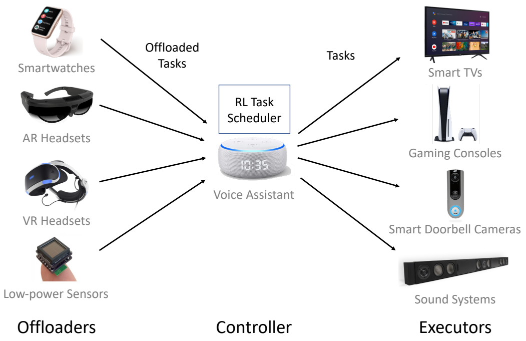

Figure 2: Ambient Compute from Resource-capable Devices

Privacy mechanisms currently for IoT are based on aggregating data at a single point (cloud), and allowing users to filter data later. However, this can be problematic for users in shared spaces like offices where users potentially are not aware of what data of theirs is being collected or which applications are using their data. I plan to explore this challenge and I see this as another opportunity for utilizing ambient compute to build a system in which data flow from devices to IoT applications is controlled by users based on their privacy rules. These rules can be very elaborate; for example, a user can allow access to their occupancy data to applications only during their working hours, they can opt to share occupancy data to an emergency evacuation application, but not to a room scheduling application and so on. Prior works have demonstrated building policy rule languages, and collecting policy rules from users, but do not describe an actual infrastructure to enforce policies or do not do so on streaming data from IoT devices. The NexusEdge middleware has some existing routing mechanisms that allow only specific device data to be provided to applications, and our insight is to extend this to enforce privacy rules which can be used to allow or deny data streams to flow to edge applications, similar to how firewall rules operate on network traffic in routers. I plan to explore this topic in more detail for my latter part of research.

References

[1] National Intelligence Council. Disruptive civil technologies: Six technologies with potential impacts on us interests out to 2025, 2023. URL https://apps.dtic.mil/sti/citations/ADA519715.

[2] Arm Limited (or its affiliates). White paper: The economics of a trillion connected devices, 2023. URL https://community.arm.com/arm-community-blogs/b/internet-of-things-blog/posts/white-paper-the-route-to-a-trillion-devices.

[3] Larry Dignan. Iot devices to generate 79.4zb of data in 2025, says idc, 2023. URL https://www.zdnet.com/article/iot-devices-to-generate-79-4zb-of-data-in-2025-says-idc/.

[4] Nabeel Nasir, Victor Ariel Leal Sobral, Li-Pang Huang, and Bradford Campbell. Nexusedge: Leveraging iot gateways for a decentralized edge computing platform. In *2022 IEEE/ACM 7th Symposium on Edge Computing (SEC)*, pages 82–95. IEEE, 2022.

[5] Nabeel Nasir and Bradford Campbell. Enabling elasticity on the edge using heterogeneous gateways. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, pages 395–396, 2021.