5G in Practice: Measuring Emerging Wireless Technology in Rural Iowa for Edge Devices in Distributed Computation Workloads

Zack Murry [1] Alicia Esquivel Morel [1 ‡] Kate Keahey [2 3 ‡]

[1]University of Missouri [2]University of Chicago [3]Argonne National Laboratory ‡Advisor

Background

Fifth-generation (5G) wireless technologies offer an opportunity to **improve internet access in rural communities**. However, there is a lack of infrastructure to measure the effective speeds of existing and emerging broadband radios for rural consumers.

Figure 1. 5G radio infrastructure.

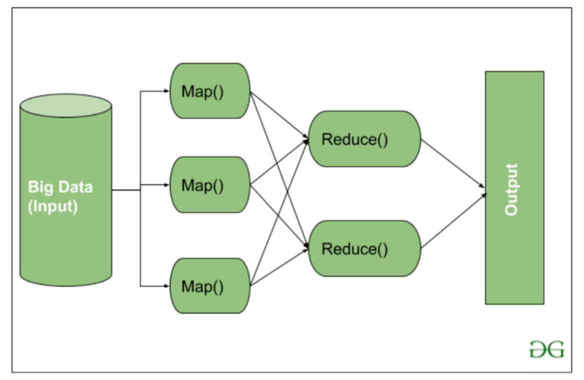


Figure 2. MapReduce data flow diagram.

In addition, we consider the role of 5G networks in distributed computing at the edge. While distributed computing paradigms like **MapReduce** [1] have traditionally been used on a cluster of high-performance servers, wireless edge computing offers a potential for such data processing in a **peer-to-peer, cost-effective, and power-effective** manner.

Problem Statement

We investigated the following research question: how do 5G wireless technologies compare to wired connections and existing rural infrastructure in speed (throughput and latency) and in supporting advanced distributed computation workloads?

Specifically, we aimed to address the following concerns:

- **Precision**: How can we accurately measure one-way 5G latency?
- **Speed**: What is the effect of using 5G wireless networks in distributed computation workloads on speed in comparison to wired connections?
- **Scalability**: How does distributed computation scale when connected over 5G links?

Solution Approach

In collaboration with the **Agricultural and Rural (ARA)** Wireless Living Lab, we deployed **six** Raspberry Pi 4 single-board computers (SBCs) in a **six-mile radius** around Ames, Iowa.

Each Raspberry Pi was connected to ARA's 5G radio infrastructure, including **millimeter-wave** (mmWave) 5G links. Additionally, several devices were connected to a wired fiberoptic link.

- Four devices were equipped with a **pulse-per-second GPS** for precise time synchronization
- The devices were managed by **FLOTO** [2], a open-source fleet management tool based on k3s and OpenBalena and created by the University of Chicago
- **Containerized applications** were deployed to measure latency and bandwidth to local devices and remote servers, and to construct **ad-hoc Apache Hadoop clusters** [4]

The data measured by the experiments was streamed to a server on Chameleon cloud for collection, aggregation, and visualization.

Architecture

Our Raspberry Pis were deployed in university buildings, city buses, and farms equipped with 5G radios.

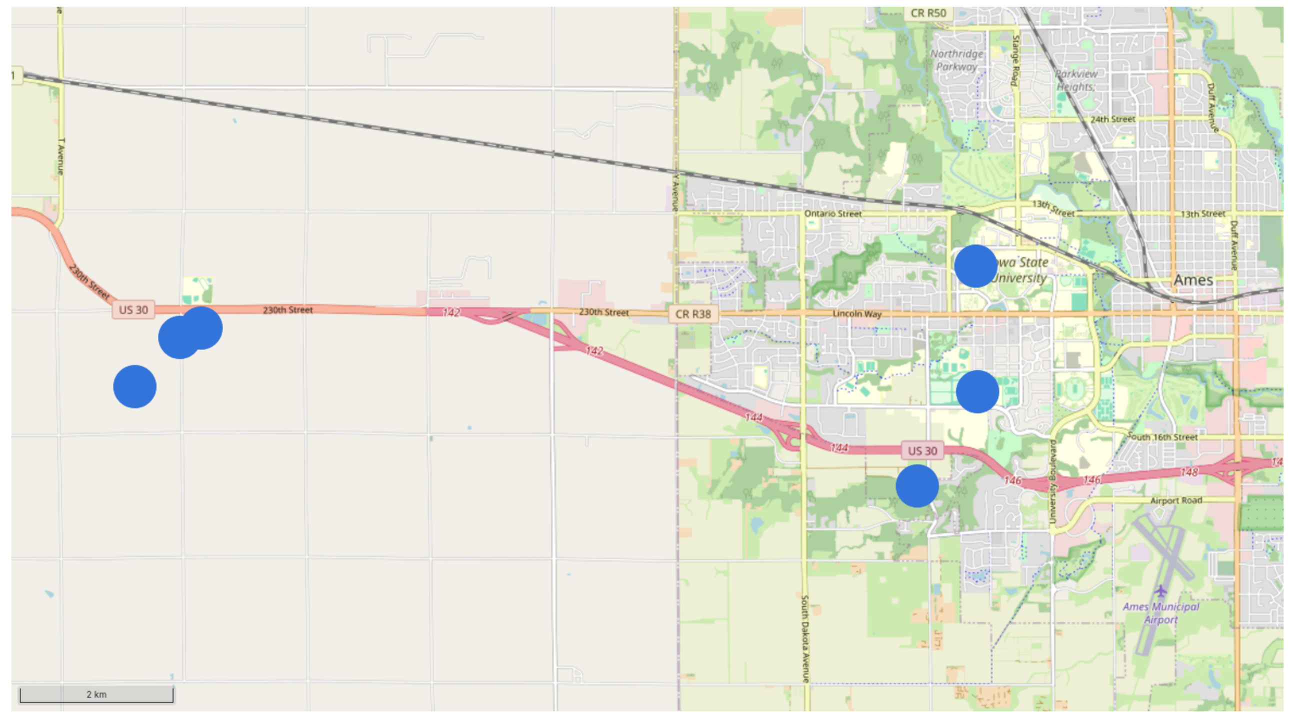
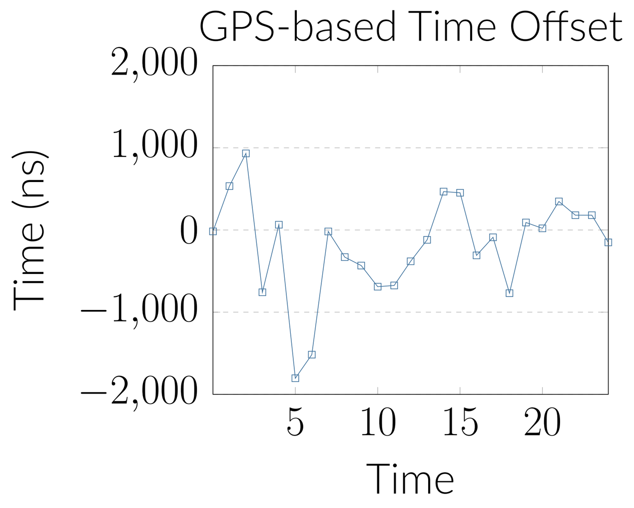
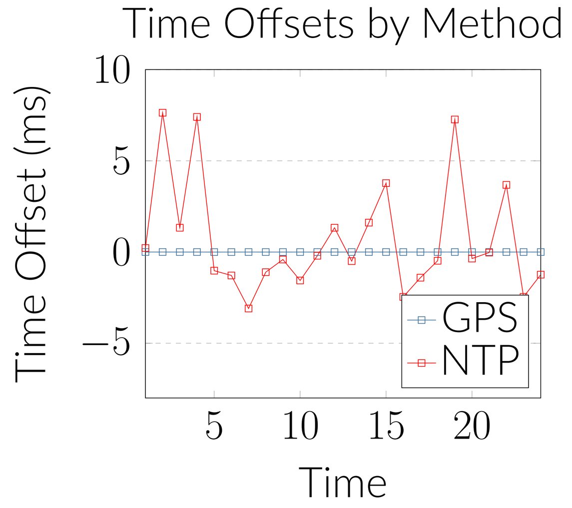


Figure 3. Map of deployed devices in Ames, Iowa.

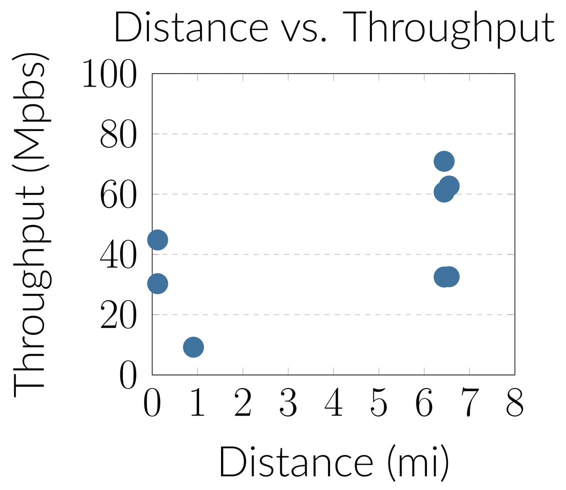
GPS-based Time Synchronization

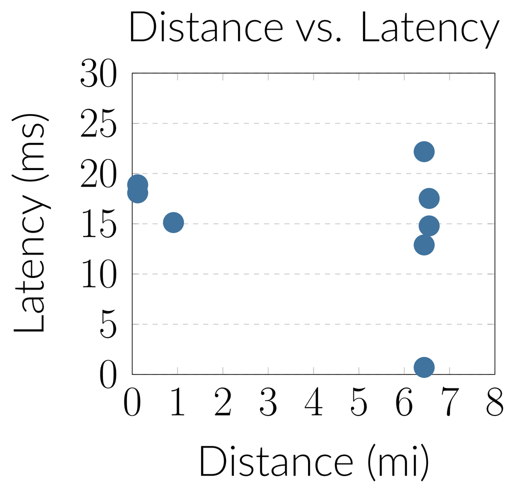
In order to precisely measure one-way latency, we implemented precise time synchronization using a GT-U7 GPS receiver on each device. We found that the GPS synchronization was **over 4,000x more precise** than Network Time Protocol (NTP) over 5G.





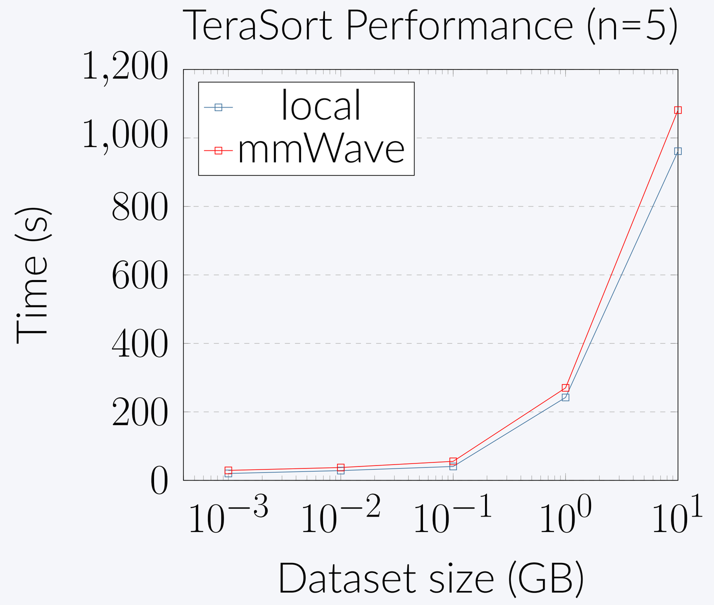
5G Network Results

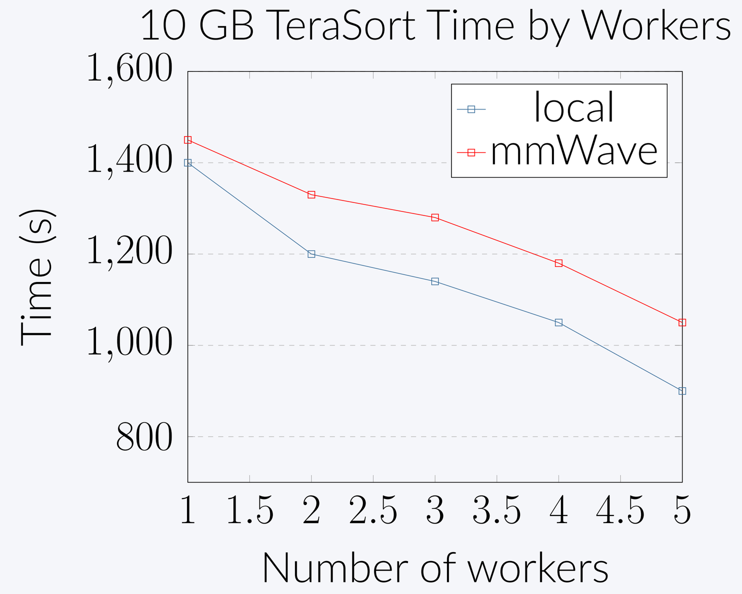




Distributed Computation Results

We measured the performance of the 5G distributed network using TeraSort, a standard Hadoop benchmark that sorts a dataset with arbitrary size.





Methodology

We created a custom Docker image that automatically installs and configures a Hadoop master or worker node on each device, which is available on our GitHub [3]. The devices negotiate SSH key permissions, discover peer IP addresses, and conduct benchmarks without human intervention.

We **do not use a central server**, instead electing a device to serve as a "master" node in a quasi-peer-to-peer manner to take full advantage of the low-latency 5G connections.

Conclusion

- Our experimental results indicate that the use of 5G in edge-based distributed computation **does not significantly hinder the performance** of Hadoop, as measured using the TeraSort benchmark
- The **network overhead** in this architecture **may be greater with more worker nodes due to additional data syncronization** across devices using the Hadoop Distributed File System (HDFS)
- We **measured the latency and throughput** of wireless connectivity over millimeter radio wave connections and **compared the efficacy of network- and GPS-based time synchronization methods**
- Future work: compare different 5G link technologies for this use case and modify our architecture to work with other distributed computation tools and paradigms

References

[1] Jeffrey Dean and Sanjay Ghemawat. Mapreduce: simplified data processing on large clusters. *Commun. ACM*, 51(1):107–113, jan 2008.

[2] Alicia Esquivel Morel, Mark Powers, Kate Keahey, Zack Murry, Tomas Javier Sitzmann, Jianfeng Zhou, and Prasad Calyam. Ibis: An infrastructure mananagement framework for adaptable, multi-sensor data collection in scientific research. *WOCC'24 - Second Workshop on Converged Computing on Edge, Cloud, and HPC*, 2024.

[3] Zack Murry. Hadoop Docker image for Raspberry Pi 4. `https://github.com/ZackMurry/hadoop-docker-raspberry-pi`, 2024.

[4] Konstantin Shvachko, Hairong Kuang, Sanjay Radia, and Robert Chansler. The hadoop distributed file system. In *2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*, pages 1–10, 2010.