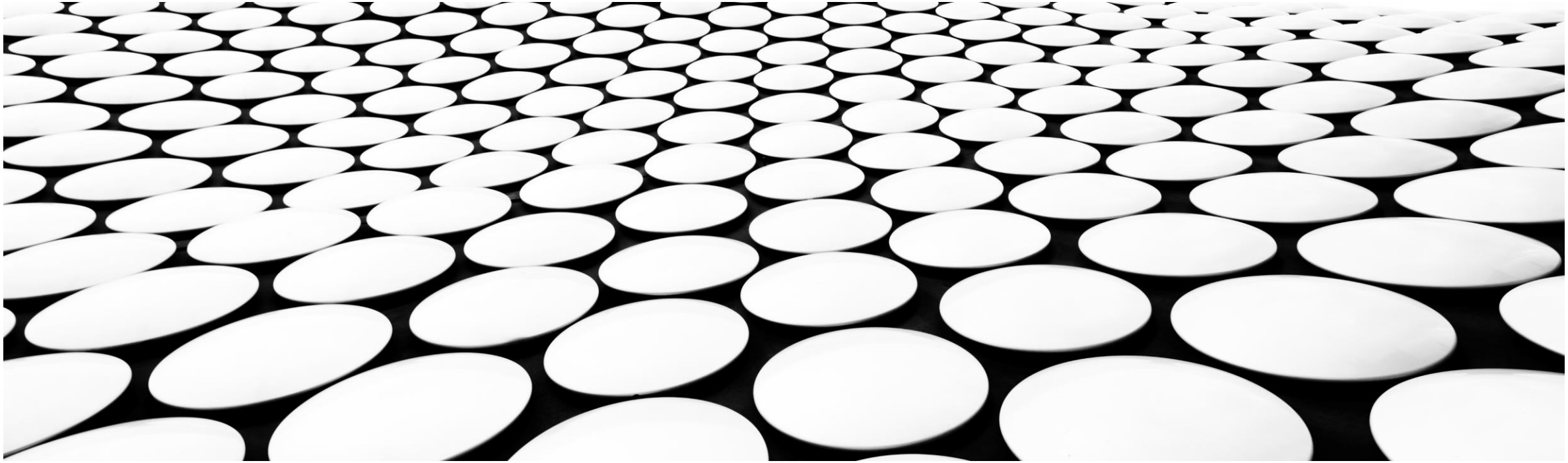
Implementations of distributed systems

COEN-317: Distributed Systems
Robert Bruce
Department of Computer Science and Engineering
Santa Clara University



Distributed systems: industries

Industries that use distributed systems:

- "Telecommunications networks (including cellular networks and the fabric of the internet)" [1].
- "Graphical and video-rendering systems" [1].
- "Scientific computing, such as protein folding and genetic research" [1].
- "Airline and hotel reservation systems" [1].
- "Multiuser video conferencing systems" [1].
- "Cryptocurrency processing systems (e.g. Bitcoin)" [1].
- "Peer-to-peer file-sharing systems (e.g. BitTorrent)" [1].
- "Distributed community compute systems (e.g. Folding@Home)" [1].
- "Multiplayer video games" [1].
- "Global, distributed retailers and supply chain management (e.g. Amazon)" [1].

1. https://www.splunk.com/en_us/data-insider/what-are-distributed-systems.html

Distributed systems: implementations

Examples:

- OpenStack.
- Elastic Utility Computing Architecture, Linking Your Programs To Useful Systems (EUCALYPTUS).
- Apache Storm
- Berkeley Open Infrastructure for Network Computing (BOINC).
- Great Internet Mersenne Prime Search (GIMPS).
- Climate Prediction.
- Protein folding.
- Software distribution (Git)

Openstack

OpenStack:

- "OpenStack is a cloud operating system that controls large pools of compute, storage, and networking resources throughout a datacenter, all managed and provisioned through APIs with common authentication mechanisms" [1]
- Open source software repository: https://releases.openstack.org/zed/

1. https://www.openstack.org/software/

Eucalyptus

Elastic Utility Computing Architecture, Linking Your Programs To Useful Systems (EUCALYPTUS):

- "Eucalyptus is a set of web services, modeled after and compatible with Amazon Web Services (AWS). Written mostly in Java, Eucalyptus integrates components from over 100 open source projects, tested and packaged into a single easy-to-install and easy-to-use product. Eucalyptus runs on virtualized infrastructure (Linux+KVM)" [1].
- Open source under the BSD license model.
- Description of the project at https://github.com/eucalyptus/eucalyptus/wiki

1. https://github.com/eucalyptus/eucalyptus/wiki

Apache Storm

Apache Storm:

- "free and open source distributed realtime computation system" [1].
- "Apache Storm is fault-tolerant: when workers die, Apache Storm will automatically restart them. If a node dies, the worker will be restarted on another node" [2].
- Downloads at https://storm.apache.org/index.html

1. https://storm.apache.org/index.html
2. https://storm.apache.org/about/fault-tolerant.html

What is volunteer computing?

Volunteer computing:

- A form of distributed computing in which a user donates CPU or GPU computing to a given project.
- The user downloads then configures the volunteer client computing software.
- The client software is usually configured to run when the user's computer remains idle.
- The client software receives computation requests from the server.
- The client software sends computation results to the server.

Examples of volunteer computing:

- Berkeley Open Infrastructure for Network Computing (BOINC)
- Great Internet Mersenne Prime Search (GIMPS)
- Climate Prediction
- Protein folding

Volunteer computing: BOINC

BOINC:

- Web-based interface makes it easy to monitor and configure the software.
- Can be modified to meet one's specific volunteer computing needs.
- The main landing page for BIONC is at https://boinc.berkeley.edu/
- The BOINC Project cookbook provides extensive documentation on how to setup your own BOINC server: at https://github.com/marius311/boinc-server-docker/blob/master/docs/cookbook.md#boinc-project-cookbook-with-boinc-server-docker

Volunteer computing: GIMPS

Great Internet Mersenne Prime Search (GIMPS):

- A distributed computing project to find Mersenne prime numbers [1].
- A Mersenne prime, *N*, is of the form $N = 2^P - 1$ where *N* is only divisible by *N* and 1.
- More information about GIMPS is at https://www.mersenne.org/

1. https://www.mersenne.org/

Volunteer computing: Climate Prediction

Climate Prediction:

- Goal: "simulate the climate for the next century, producing predictions of temperature, rainfall and the probability of extreme weather events" [1].
- Climate prediction runs on BOINC software [2].
- More information about Climate Prediction application is at https://www.climateprediction.net/

1. https://www.climateprediction.net/
2. https://www.climateprediction.net/about/

Volunteer computing: folding@home

Folding@home:

- "A distributed computing project for simulating protein dynamics, including the process of protein folding and the movements of proteins implicated in a variety of diseases" [1].
- More information about Folding@home at https://foldingathome.org/

1. https://foldingathome.org/about-2/?lng=en

Git version control software

Git:

- An open-source distributed system for version control [1]
- More information about Git at https://git-scm.com/

1. https://git-scm.com/

The future of distributed systems?

Given the ubiquity of mobile phones, IoT (Internet of Things), and cellular 5G communication, I see potential for projects involving:

- Mobile phone computation using a wide array of smartphones for computation of a given task via an ad-hoc network (MANET). MANET is an acronym for Mobile Ad hoc Network.
- Vehicle-to-vehicle communication for autonomous cars (VANET). VANET is an acronym for Vehicular Ad hoc Network.

For further reading

- N. Cardona, E. Coronado, S. Latré, R. Riggio and J. M. Marquez-Barja, "Software-Defined Vehicular Networking: Opportunities and Challenges," *IEEE Access*, vol. 8, pp. 219971-219995, 2020, doi:10.1109/ACCESS.2020.3042717.
- Z. Nurlan, T. Zhukabayeva, M. Othman, A. Adamova and N. Zhakiyev, "Wireless Sensor Network as a Mesh: Vision and Challenges," *IEEE Access*, vol. 10, pp. 46-67, 2022, doi:10.1109/ACCESS.2021.3137341.