CMPSCI 630 Systems **Spring 2023**

Lecture 13

Lecturer: Emery Berger *Scribe: Fadhil I. Kurnia*

In this lecture, we are discussing three main topics: Jim Gray's paper on fault-tolerant systems, RAID, and Bitcoin. The common goal of those topics is building a fault-tolerant system.

13.1 Fault-Tolerant Systems

As the name suggest, a fault-tolerant system should handle failures. Generally failures can be random (e.g random failure model) or coordinated (e.g coordinated attack). The coordinated failures, in particular, is related to system security that we learn in the previous lectures.

Each system has their own failure model, i.e failure that the system tries to mitigate. For example, the machine described by Jim Gray assumes fail-stop failures, RAID assumes disk failures, and the Bitcoin paper assumes integrity failures. That is, each system has different design based on the failure model it assumes.

Next, we are discussing some failure models and impossibility regarding some failures.

Byzantine Failure. This failure model assume arbitrary failures, such as message drop, long-time wait, duplicate messages, and even forged messages. However, the adversary still can not break the law of physics or computation; for example, an adversary still can not break an encryption or hash function. That being said, encryption could help mitigate some failures in this failure model. For instance, RSA algorithm – which uses public & private key and becomes the basis for all encryption that we use today in HTTPS – can be used to prevent adversary from changing the messages. To function properly, a good encryption techniques should use *one-way functions*, as illustrated below.

```
f(x) can be computed easily.
f'(f(x)) is hard to be computed.
```

CAP: Consistency, Availability, and Partition-tolerant. CAP defines a fundamental rule for system that operate under availability and partition-tolerant failures. The theorem says that among the three properties in CAP, we can only pick two. The consistency property ensures that every client gets the latest (the most up to date) values from the system, the availability property ensures that the system *always* responds to client's requests, while the partition-tolerant property ensures that the system can handle any network partition. The theorem is obvious, for example if we have a partitioned system then the system can not synchronize the latest data, thus the system can either stop serving requests (unavailable) or keep serving requests with stale data (no consistency).

FLP Impossibility. This theorem says that in an asynchronous system, it is impossible to design a deterministic consensus algorithm that can satisfy agreement, termination, and fault tolerance. Thus, the system can not even agree on a single bit (achieving consensus for a single bit) under certain failures. This impossibility stems from the fact that the system can not differentiate between failures and timeout (long delayed response).

The Byzantine General Problem. This problem was proposed by Leslie Lamport using an analogy of parliament which tries to reach a consensus in an ancient city of Paxos. The protocol is so complex, making multiple papers trying to explain the protocol differently. In particular, Raft, simplifies the explanation by using leader election and logging abstraction. To keep the system available under f failures machines, the protocol needs $3f + 1$ nodes under byzantine model or $2f + 1$ under fail-stop model.

Problem with Crypto Finance. Crypto, arguably, solves no actual problem that the society should care about. It uses wasteful proof-of-work and append-only ledger abstraction. The proof-of-work (PoW) mechanism requires the system to solve a cryptographic puzzle first before it can do anything. People use CPU, GPU, and even ASIC to do the PoW, specifically to mine a new block in Bitcoin; this requires a lot of energy. Moreover, some researchers start to find issue with how the protocol works. For instance, in the selfish miner attack, a group of machines can hold the newly created block so that they have higher chance of getting the next blocks.