Fault-tolerant Computing

Frans Kaashoek
6.033 Spring 2007
April 4, 2007
Where are we in 6.033?

• Modularity to control complexity
  • Names are the glue to compose modules
• Strong form of modularity: client/server
  • Limit propagation of errors
• Implementations of client/server:
  • In a single computer using virtualization
  • In a network using protocols
• Compose clients and services using names
  • DNS
How to respond to failures?

• Failures are contained; they don’t propagate
  • Benevolent failures
• Can we do better?
  • Keep computing despite failures?
  • Defend against malicious failures (attacks)?
• Rest of semester: handle these “failures”
  • Fault-tolerant computing
  • Computer security
Fault-tolerant computing

- General introduction: today
  - Replication/Redundancy
- The hard case: transactions
  - updating permanent data in the presence of concurrent actions and failures
- Replication revisited: consistency
A fatal exception 0E has occurred at 0028:C00068F8 in PPT.EXE<01> + 000059F8. The current application will be terminated.

* Press any key to terminate the application.
* Press CTRL+ALT+DEL to restart your computer. You will lose any unsaved information in all applications.

Press any key to continue
Availability in practice

- **Carrier airlines (2002 FAA fact book)**
  - 41 accidents, 6.7M departures
  - ✓ 99.9993% availability

- **911 Phone service (1993 NRIC report)**
  - 29 minutes per line per year
  - ✓ 99.994%

- **Standard phone service (various sources)**
  - 53+ minutes per line per year
  - ✓ 99.99+%

- **End-to-end Internet Availability**
  - ✓ 95% - 99.6%
PRODUCT OVERVIEW

Cheetah 15K.4
Mainstream enterprise disc drive

Simply the best price/performance, lowest cost of ownership disc drive ever

KEY FEATURES AND BENEFITS

• The Cheetah® 15K.4 is the highest-performance drive ever offered by Seagate®, delivering maximum IOPS with fewer drives to yield lower TCO.
• The Cheetah 15K.4 price-per-performance value united with the breakthrough benefits of serial attached SCSI (SAS) make it the optimal 3.5-inch drive for rock solid enterprise storage.
• Proactive, self-initiated background management functions improve media integrity, increase drive efficiency, reduce incidence of integration failures and improve field reliability.
• The Cheetah 15K.4 shares its electronics architecture and firmware base with Cheetah 10K.7 and Savvio® to ensure greater factory consistency and reduced time to market.

KEY SPECIFICATIONS

• 146-, 73- and 36-GBByte capacities
• 3.3-msec average read and 3.8-msec average write seek times
• Up to 96-MBytes/sec sustained transfer rate
• 1.4 million hours full duty cycle MTBF
• Serial Attached SCSI (SAS), Ultra320 SCSI and 1 Gbits/sec Fibre Channel interfaces
• 5-year warranty

For more information on why 15K is the industry’s best price/performance disc drive for use in mainstream storage applications, visit http://specials.seagate.com/15k
Disk failure conditional probability distribution

- Infant mortality
- Burn out
- Stable failure period
- Expected operating lifetime

1 / (reported MTTF)
Data from http://www.mortality.org
Fail-fast disk

```c
failfast_get (data, sn) {
    get (s, sn);
    if (checksum(s.data) = s.cksum) {
        data ← s.data;
        return OK;
    } else {
        return BAD;
    }
}
```
Careful disk

careful_get (data, sn) {
    r ← 0;
    while (r < 10) {
        r ← failfast_get (data, sn);
        if (r = OK) return OK;
        r++;
    }
    return BAD;
}
Durable disk (RAID 1)

durable_get (data, sn) {
    r ← disk1.careful_get (data, sn);
    if (r = OK) return OK;
    r ← disk2.careful_get (data, sn);
    signal(repair disk1);
    return r;
}