Fault Tolerance

- **Part 1:** Agreement in presence of faults
  - Two army problem
  - Byzantine generals problem
- **Part 2:** Reliable communication
- **Part 3:** Distributed commit
  - Two phase commit
  - Three phase commit
- Next class:
  - Paxos and RAFT
  - Failure recovery
    - Checkpointing
    - Message logging

Fault Tolerance

- **Single machine systems**
  - Failures are all or nothing
    - OS crash, disk failures
- **Distributed systems:** multiple independent nodes
  - Partial failures are also possible (some nodes fail)
- **Question:** Can we automatically recover from partial failures?
  - Important issue since probability of failure grows with number of independent components (nodes) in the systems
  - \( \text{Prob(failure)} = \text{Prob(Any one component fails)} = 1 - \text{P(no failure)} \)
A Perspective

- Computing systems are not very reliable
  - OS crashes frequently (Windows), buggy software, unreliable hardware, software/hardware incompatibilities
  - Until recently: computer users were “tech savvy”
    - Could depend on users to reboot, troubleshoot problems
  - Growing popularity of Internet/World Wide Web
    - “Novice” users
    - Need to build more reliable/dependable systems
  - Example: what is your TV (or car) broke down every day?
    - Users don’t want to “restart” TV or fix it (by opening it up)
- Need to make computing systems more reliable
  - Important for online banking, e-commerce, online trading, webmail…

Basic Concepts

- Need to build *dependable* systems
- Requirements for dependable systems
  - Availability: system should be available for use at any given time
    - 99.999% availability (five 9s) \(\Rightarrow\) very small down times
  - Reliability: system should run continuously without failure
  - Safety: temporary failures should not result in a catastrophic
    - Example: computing systems controlling an airplane, nuclear reactor
  - Maintainability: a failed system should be easy to repair
Basic Concepts (contd)

• Fault tolerance: system should provide services despite faults
  – Transient faults
  – Intermittent faults
  – Permanent faults

Failure Models

<table>
<thead>
<tr>
<th>Type of failure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crash failure</td>
<td>A server halts, but is working correctly until it halts</td>
</tr>
<tr>
<td>Omission failure</td>
<td></td>
</tr>
<tr>
<td>Receive omission</td>
<td>A server fails to respond to incoming requests</td>
</tr>
<tr>
<td>Send omission</td>
<td>A server fails to receive incoming messages</td>
</tr>
<tr>
<td>Timing failure</td>
<td>A server's response lies outside the specified time interval</td>
</tr>
<tr>
<td>Response failure</td>
<td></td>
</tr>
<tr>
<td>Value failure</td>
<td>The server's response is incorrect</td>
</tr>
<tr>
<td>State transition failure</td>
<td>The value of the response is wrong</td>
</tr>
<tr>
<td>Arbitrary failure</td>
<td>A server may produce arbitrary responses at arbitrary times</td>
</tr>
</tbody>
</table>

• Different types of failures.
Failure Masking by Redundancy

• Triple modular redundancy: can handle one failure in circuit

Agreement in Faulty Systems

• How should processes agree on results of a computation?
• *K-fault tolerant*: system can survive k faults and yet function
• Assume processes fail silently
  – Need (k+1) redundancy to tolerant k faults
• *Byzantine failures*: processes run even if sick
  – Produce erroneous, random or malicious replies
    • Byzantine failures are most difficult to deal with
  – Need ? Redundancy to handle Byzantine faults
Byzantine Faults

- Simplified scenario: two perfect processes with unreliable channel
  - Need to reach agreement on a 1 bit message
- **Two Generals Problem:** Two armies waiting to attack
  - Each army coordinates with a messenger
  - Messenger can be captured by the hostile army
  - Can generals reach agreement?
  - Property: Two perfect process can never reach agreement in presence of unreliable channel
  - Concept of **Common knowledge**
- **Byzantine generals problem:** Can N generals reach agreement with a perfect channel?
  - M generals out of N may be traitors

---

**Byzantine Generals Problem**

![Diagram of Byzantine Generals Problem]

1. Got(1, 2, x, 4)
2. Got(1, 2, y, 4)
3. Got(1, 2, 3, 4)
4. Got(1, 2, z, 4)

Faulty process

(a) 

1. Got(1, 2, x, 4)  
2. Got(1, 2, y, 4)  
3. Got(1, 2, 3, 4)  
4. Got(1, 2, z, 4)

(b) 

1. Got(1, 2, x, 4)  
2. Got(1, 2, x, 4)  
3. Got(1, 2, x, 4)  
4. Got(1, 2, x, 4)

(c) 

1. Got(1, 2, x, 4)  
2. Got(1, 2, x, 4)  
3. Got(1, 2, x, 4)  
4. Got(1, 2, x, 4)

- Recursive algorithm by Lamport
- The Byzantine generals problem for 3 loyal generals and 1 traitor.
  a) The generals announce their troop strengths (in units of 1 kilosoldiers).
  b) The vectors that each general assembles based on (a)
  c) The vectors that each general receives in step 3.
Byzantine Generals Problem Example

The same as in previous slide, except now with 2 loyal generals and one traitor.

Property: With $m$ faulty processes, agreement is possible only if $2m+1$ processes function correctly out of $3m+1$ total processes. [Lamport 82]
  - Need more than two-thirds processes to function correctly (for $m=1$, 3 out of 4 processes)

Byzantine Fault Tolerance

- Detecting a faulty process is easier
  - $2k+1$ to detect $k$ faults

- Reaching agreement is harder
  - Need $3k+1$ processes ($2/3$rd majority needed to eliminate the faulty processes)

- Implications on real systems:
  - How many replicas?
  - Separating agreement from execution provides savings
Reaching Agreement

- If message delivery is unbounded,
  - No agreement can be reached even if one process fails
  - Slow process indistinguishable from a faulty one

- BAR Fault Tolerance
  - Until now: nodes are byzantine or collaborative
  - New model: Byzantine, Altruistic and Rational
  - Rational nodes: report timeouts etc

Reliable One-One Communication

- Issues were discussed in Lecture 3
  - Use reliable transport protocols (TCP) or handle at the application layer
- RPC semantics in the presence of failures
- Possibilities
  - Client unable to locate server
  - Lost request messages
  - Server crashes after receiving request
  - Lost reply messages
  - Client crashes after sending request
Reliable One-Many Communication

- Reliable multicast
  - Lost messages => need to retransmit
- Possibilities
  - ACK-based schemes
    - Sender can become bottleneck
  - NACK-based schemes

Broadcast Ordering

- Broadcast (or multicast) ordered important for replication

- FIFO broadcast: if a process sends m1 and then m2, all other processes receive m1 before m2
- Totally ordered: If a process receives m1 before m2 (regardless of sender), all processes receive m1 before m2
  - Does not imply FIFO, all processes just agree on order
- Causally ordered: if send(m1)->send(m2) => recv(m1)->recv(m2)
- State machine replication (SMR)
  - Broadcast requests to all replicas using totally ordered broadcast; replicas apply requests in order.
Atomic Multicast

- Atomic multicast: a guarantee that all process received the message or none at all
  - Replicated database example
  - Need to detect which updates have been missed by a faulty process
- Problem: how to handle process crashes?
- Solution: group view
  - Each message is uniquely associated with a group of processes
    - View of the process group when message was sent
    - All processes in the group should have the same view (and agree on it)

Virtually Synchronous Multicast

Implementing Virtual Synchrony in Isis

(a) Process 4 notices that process 7 has crashed, sends a view change
(b) Process 6 sends out all its unstable messages, followed by a flush message
(c) Process 6 installs the new view when it has received a flush message from everyone else
Implementing Virtual Synchrony

<table>
<thead>
<tr>
<th>Multicast</th>
<th>Basic Message Ordering</th>
<th>Total-Ordered Delivery?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliable multicast</td>
<td>None</td>
<td>No</td>
</tr>
<tr>
<td>FIFO multicast</td>
<td>FIFO-ordered delivery</td>
<td>No</td>
</tr>
<tr>
<td>Causal multicast</td>
<td>Causal-ordered delivery</td>
<td>No</td>
</tr>
<tr>
<td>Atomic multicast</td>
<td>None</td>
<td>Yes</td>
</tr>
<tr>
<td>FIFO atomic multicast</td>
<td>FIFO-ordered delivery</td>
<td>Yes</td>
</tr>
<tr>
<td>Causal atomic multicast</td>
<td>Causal-ordered delivery</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Distributed Commit

- Atomic multicast example of a more general problem
  - All processes in a group perform an operation or not at all
  - Examples:
    - Reliable multicast: Operation = delivery of a message
    - Distributed transaction: Operation = commit transaction
- Problem of distributed commit
  - All or nothing operations in a group of processes
- Possible approaches
  - Two phase commit (2PC) [Gray 1978 ]
  - Three phase commit
Two Phase Commit

- Coordinator process coordinates the operation
- Involves two phases
  - Voting phase: processes vote on whether to commit
  - Decision phase: actually commit or abort

Implementing Two-Phase Commit

actions by coordinator:

```
while START_2PC to local log;
multicast VOTE_REQUEST to all participants;
while not all votes have been collected {
    wait for any incoming vote;
    if timeout {
        while GLOBAL_ABORT to local log;
multicast GLOBAL_ABORT to all participants;
exit;
    } record vote;
}
if all participants sent VOTE_COMMIT and coordinator votes COMMIT{
    write GLOBAL_COMMIT to local log;
multicast GLOBAL_COMMIT to all participants;
} else {
    write GLOBAL_ABORT to local log;
multicast GLOBAL_ABORT to all participants;
}
```

- Outline of the steps taken by the coordinator in a two phase commit protocol
Implementing 2PC

actions by participant:
write INIT to local log;
wait for VOTE_REQUEST from coordinator;
if timeout {
    write VOTE_ABORT to local log;
    exit;
}
if participant votes COMMIT {
    write VOTE_COMMIT to local log;
    send VOTE_COMMIT to coordinator;
    wait for DECISION from coordinator;
    if timeout {
        multicast DECISION_REQUEST to other participants;
        wait until DECISION is received; /* remain blocked */
        write DECISION to local log;
    }
    if DECISION == GLOBAL_COMMIT
        write GLOBAL_COMMIT to local log;
    else if DECISION == GLOBAL_ABORT
        write GLOBAL_ABORT to local log;
    } else {
        write VOTE_ABORT to local log;
        send VOTE_ABORT to coordinator;
    }

actions for handling decision requests: / *executed by separate thread */
while true {
    wait until any incoming DECISION_REQUEST is received; /* remain blocked */
    read most recently recorded STATE from the local log;
    if STATE == GLOBAL_COMMIT
        send GLOBAL_COMMIT to requesting participant;
    else if STATE == INIT or STATE == GLOBAL_ABORT
        send GLOBAL_ABORT to requesting participant;
    else
        skip; /* participant remains blocked */
}

Recovering from a Crash

- If INIT: abort locally and inform coordinator
- If Ready, contact another process Q and examine Q’s state

<table>
<thead>
<tr>
<th>State of Q</th>
<th>Action by P</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMMIT</td>
<td>Make transition to COMMIT</td>
</tr>
<tr>
<td>ABORT</td>
<td>Make transition to ABORT</td>
</tr>
<tr>
<td>INIT</td>
<td>Make transition to ABORT</td>
</tr>
<tr>
<td>READY</td>
<td>Contact another participant</td>
</tr>
</tbody>
</table>
Three-Phase Commit

Two phase commit: problem if coordinator crashes (processes block)
Three phase commit: variant of 2PC that avoids blocking