Fault Tolerance

Definition: Fault tolerance

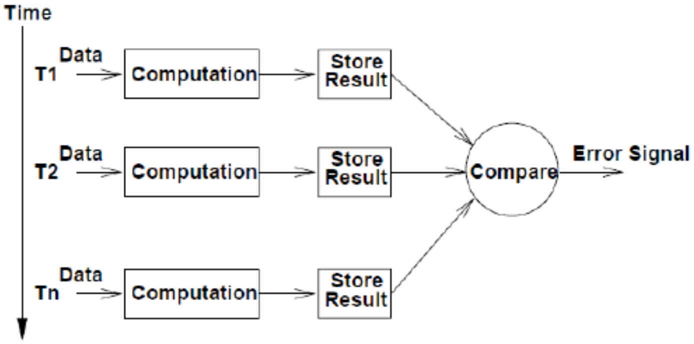
Ability of a system to continue error-free operation even in the presence of unexpected fault

1 Approaches based on redundancy

- Apply duplication to increase system reliability
- System architecture approach
  - Incorporate Active or Passive replication
  - Design server configuration and number of replicated servers
  - Could be expensive due to requiring extra hardware
- Operational approach
  - Replicate system operations to offer fault tolerance
    * Time redundancy
    * Component redundancy
    * Information redundancy
    * Communication redundancy

1.1 Time redundancy

- Perform the same operation multiple times
- No fault if getting the same result each time
- Detect temporary faults but not permanent ones
- Impact system performance



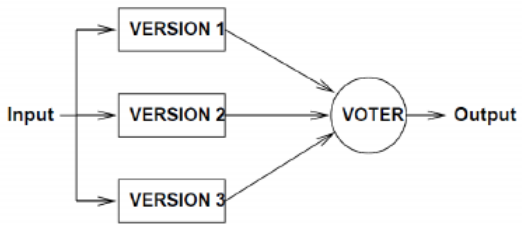
1.2 Component redundancy

Replicate component and compare outputs:

- Introduce two or more independent running components which provide the same functionalities
- Impose little or no performance impact

N-Version Programming (NVP):

- Design diversity - implementing multiple versions of the program
- Tolerate hardware and software faults, but not correlated faults



1.3 Information Redundancy

Encode outputs with error detection or correcting code
Advantage:

- Less hardware is required than replicating module
- Support fault detection

Drawback

- Added complexity in design
- Fault recovery capability may be limited

2 Communication Failures

Client is unable to locate server

- Use an exception handler (programming language dependent)
- Check out available/update servers from a directory service

Client request to server is lost:

- Apply timeout to await server reply then re-send
- If multiple requests appear to get lost assume ”can’t locate server” error

Server crashes after receiving client request

- Server may stop before or after returning the info, or before ACK
- Store user request in the FE
- Rebuild or use alternate server to retry request
- Give up and report failure

Server reply to client is lost

- Apply timeout to await server reply