FACULTY OF ENGINEERING & TECHNOLOGY

Brajesh Mishra
Assistant Professor
Department of Computer Science & Engineering

Topics Covered

Security in Mobile Networks. Fault-tolerance
Fault tolerance vs. high availability
Load balancing and failover: fault tolerance for web applications

Security in Mobile Networks. Fault-tolerance

Fault tolerance refers to the ability of a system (computer, network, cloud cluster, etc.) to continue operating without interruption when one or more of its components fail.

The objective of creating a fault-tolerant system is to prevent disruptions arising from a single point of failure, ensuring the high availability and business continuity of mission-critical applications or systems.

Fault-tolerant systems use backup components that automatically take the place of failed components, ensuring no loss of service.

Security in Mobile Networks. Fault-tolerance

These include:
Hardware systems that are backed up by identical or equivalent systems. For example, a server can be made fault tolerant by using an identical server running in parallel, with all operations mirrored to the backup server.
Software systems that are backed up by other software instances. For example, a database with customer information can be continuously replicated to another machine. If the primary database goes down, operations can be automatically redirected to the second database.
Power sources that are made fault tolerant using alternative sources. For example, many organizations have power generators that can take over in case main line electricity fails.

Fault tolerance vs. high availability

High availability refers to a system's ability to avoid loss of service by minimizing downtime. It's expressed in terms of a system's uptime, as a percentage of total running time. Five nines, or 99.999% uptime, is considered the "holy grail" of availability.
In most cases, a business continuity strategy will include both high availability and fault tolerance to ensure your organization maintains essential functions during minor failures, and in the event of a disaster.
While both fault tolerance and high availability refer to a system's functionality over time, there are differences that highlight their individual importance in your business continuity planning.

Fault tolerance vs. high availability

Some important considerations when creating fault tolerant and high availability systems in an organizational setting include:

Downtime – A highly available system has a minimal allowed level of service interruption. For example, a system with “five nines” availability is down for approximately 5 minutes per year. A fault-tolerant system is expected to work continuously with no acceptable service interruption.

Scope – High availability builds on a shared set of resources that are used jointly to manage failures and minimize downtime. Fault tolerance relies on power supply backups, as well as hardware or software that can detect failures and instantly switch to redundant components.

Cost – A fault tolerant system can be costly, as it requires the continuous operation and maintenance of additional, redundant components. High availability typically comes as part of an overall package through a service provider (e.g., load balancer provider).

Load balancing and failover: fault tolerance for web applications

In the context of web application delivery, fault tolerance relates to the use of load balancing and failover solutions to ensure availability via redundancy and rapid disaster recovery.
Load balancing solutions allow an application to run on multiple network nodes, removing the concern about a single point of failure. Most load balancers also optimize workload distribution across multiple computing resources, making them individually more resilient to activity spikes that would otherwise cause slowdowns and other disruptions.