Availability: Resilience by Design

Summary

The BCS ITLF Availability Working Group[1] is focussed on how to improve the resilience of operational systems[2].

This paper focusses on

- why resilience by design is important for both new and existing systems;
- characteristics of resilient systems,
- resilience by design applied to existing operational systems
- new tools e.g. AI based, which help to increase resilience.

Resilience is an important non-functional requirement[3] when designing or procuring a system. The principles of resilience by design are understood for new system development, and taught in some computer science courses,[4]. It is widely recognised that critical to improving resilience are the approaches to how software is designed, built, tested, and deployed.

Many of these principles can be applied to the maintenance and upgrade of existing operational systems, with improvement in resilience and hence availability of these.

Characteristics of Resilience

Resiliency describes the system's ability to withstand and recover from disruptions, including failures and attacks; it focuses on the speed and effectiveness of response and recovery after a disruption or outage.[5].

[1] In this text and others in the series of ITLF Availability Papers, "we" refers to the Availability/Service Resilience Working Group of the IT Leaders Forum of the BCS – the Chartered Institute for IT and co-opted colleagues who have provided additional insight.

[2] A company's "operational systems" are those systems needed to execute the company's daily commercial business and all subsequent data processing.

[3] See for instance https://digital.nhs.uk/services/gp-connect/develop-gp-connect-services/development/non-functional-requirements

[4] https://pdf.ncl.ac.uk/ug/2025/g400.pdf

[5] See Availability: IT is a utility in https://www.bcs.org/membership-and-registrations/member-communities/bcs-it-leaders-forum/papers/

The effectiveness of service resilience measures is measured by the outage time of the service. We are particularly informed by the data supplied to the Treasury Select Committee[6] which found that 43% of outages were due to failed upgrades[7]. Reducing this failure rate is particularly relevant to systems which are critical in terms of impact to the business, (IBS)[8].

So, in this paper we particularly look for approaches to operational systems change that incorporate design for resilience thinking at a time of operational system upgrade and could reduce the number of outages.

Key characteristics of resilient systems:

- *Recovery from failure:* The system should be able to quickly return to normal operations after a disruption, ideally automatically, and without significant degradation of service.

- *Minimizing downtime:* Resilience requirements often specify the maximum allowable downtime (e.g., through Recovery Time Objectives, or RTOs) and may require mechanisms like redundancy, failover, and automated recovery processes.

- *Continued operation during faults:* A resilient system is designed to continue functioning, possibly at a reduced level, even when some components fail. This is known as fault tolerance and may involve techniques such as load balancing, clustering, and error handling.

- *Handling both anticipated and unanticipated failures:* Resilience extends beyond traditional fault tolerance by aiming to handle not only known failure modes but also unforeseen disruptions, maintaining dependability under changing conditions.

- *Adaptability and Flexibility:* A resilient system should flexible and be able to continue functioning when gradual and abrupt changes/stresses occur.

[6] https://committees.parliament.uk/committee/158/treasury-committee/news/205611/more-than-one-months-worth-of-it-failures-at-major-banks-and-building-societies-in-the-last-two-years/

[7] https://www.bcs.org/media/rxdmjr5h/availability-6-report-nine-banks-data-roundtable-220425.pdf

[8] https://www.bcs.org/media/2ucjm0ch/availability-the-fs-process.pdf

Resilience by Design - Principles

Resilience as a non-functional requirement[9] has been understood and applied by both enterprise and solution architects, and is included in frameworks for system development such as the Open Group's TOGAF/O-AA[10].

Many of the architectural and design principles specified for new systems development can contribute to improving resilience[11] in operational systems. In the list below, the approaches useful for operational systems maintenance and upgrade are in bold.

- *Holistic architectural view*: for and end-to-end business service, including supply chain elements, such as common or third party services; evaluation of the entire IT ecosystem view to ensure dependencies are accurately identified;

- *Redundancy*: elimination of single points of failure through redundancy, clustering, geographical distribution, and multiple service providers;

- *Thinking modularity e.g. stateless applications*: designing for (introducing) independence of applications and services;

- *Automation*: implementing automated failover and switchover processes to minimise downtime, reduce manual intervention and streamline incident response;

- *Data protection and recovery*: key to any resilience approach is the ability to safeguard against data loss, and to ensure known levels of data coherence across multiple sites or services;

- *Agility and Scalability*: designing systems to scale rapidly and elastically in response to changes to load;

- *Loose coupling*: which includes the implementation of bulkheads, timeouts, circuit breakers and data separation;

[9] Non Functional Requirement are the criteria that define how a system should behave, rather than what it is supposed to do. https://www.perforce.com/blog/alm/what-are-non-functional-requirements-examples

[10] www.opengroup.org/togaf

[11] https://londonpublishingpartnership.co.uk/books/resilience-of-services-reducing-the-impact-of-it-failures/ page 53

- *Elegant degradation*: designing systems such that when they are stressed, the more important functions get priority;

- *Situational awareness*: designing systems to ensure that they don't make the situation worse e.g. retrying transactions when a service is overloaded;

- *Design to deal with (unexpected) issues (unhappy flow)*: instead of crashing when there is an (unexpected) error or abnormality, ensure that the system flags the error for remediation and proceeds.

- *Design for detection and management of failure in the implementation process*: either through the use of concurrent versions (where multiple versions can be live at the same time for different customers), canary services (where new versions are released to a small subset of users first) or alpha/beta releases.

In addition, operational management should consider

- *Proactive Risk Management*: identifying and mitigating vulnerabilities before they lead to incidents using techniques such as failure simulation, scenario planning, automated controls;

- *Continuous monitoring of IT operations*: using AI and advanced analytics to predict, detect outages and enable rapid response/recovery;

- *Continuous improvement and testing*: regular review of resilience processes, automation where possible and frequent testing;

- *Forensic analysis of incidents, near misses and disruptions*: performing analysis such as Root Cause Analysis to learn and implement lessons identified. [a no blame culture enhances the effectiveness of this process];

- *Chaos engineering*: proactively stressing the system and/or trying to induce failures in order to determine weaknesses and areas for improvement;

- *Engagement with SRE[12] engineers, Business Continuity and Information Security*: given that resilience needs to be a focus across the whole system life cycle;

[12] Site Reliability Engineers

Use of AI to improve operational resilience

New technologies such as GenAI, Agentic AI and Machine Learning have the capability to enhance resilience as a core and robust IT capability. Some potential applications are[13]:

- *Predictive Analytics and Early Warning:* using AI models to forecast potential disruptions—such as hardware failures, cyber threats, or environmental hazards—before they escalate. This allows organizations to act pre-emptively.

- *Data and System Resilience:* AI can be used to support robust data governance, ensuring data accuracy, integrity, and availability. It can also help organizations test and validate their digital strategies across multiple dimensions of resilience: system, cyber, informational, organizational, operational, and people.

- *Resource Optimization:* AI can dynamically allocate resources (such as compute, storage, or network bandwidth) based on real-time demand and predicted needs, ensuring optimal performance and minimizing bottlenecks during peak loads or disruptive events.

Recommendations

We recommend that we engage with the BCS SIG on Enterprise Architecture to explore how design for resilience can become a core capability, and the approaches updated to support the advent of new technologies.

We recommend that we explore BCS links to relevant bodies (e.g. The Open Group as above) to pursue the development of consistent approaches to design for resilience across both new and existing operational systems.

[13] See also Use of AI in Improving Service Resilience in https://www.bcs.org/membership-and-registrations/member-communities/bcs-it-leaders-forum/papers/