Measuring Technical Debt in Mission Critical Trading Systems A Service Level Quantitative Framework

Research Proposal for MSc Computing Project

Tobias Zeier

2026-02-23

University of Essex Online

Supervisor: Zahid Ullah

Co-Supervisor: Doug Millward

Project title

Measuring Technical Debt in Mission-Critical Trading Systems using a Service Level Quantitative Framework

Significance, contribution and research problem

Technical debt describes the long term cost of short term technical decisions and is widely discussed in software engineering and information systems research. Existing approaches and tools focus mainly on code level metrics or qualitative classifications, such as technical debt ratio and quadrant models, which are useful for code quality assessment but less suited to complex enterprise environments. In mission-critical trading information technology services, operational reliability depends on tightly coupled services, vendor platforms and infrastructure components rather than internally developed code alone, which existing frameworks do not adequately capture.

Trading services in regulated financial institutions operate under strict availability and performance requirements, where failures can cause immediate financial loss, regulatory exposure and reputational damage. Despite this, remediation decisions about technical debt are often made without an empirical understanding of how technical debt at service level relates to operational risk. As a result, there is a gap between technical debt theory and its practical application in service management for trading information technology environments. This project will address that gap by designing and evaluating a service level quantitative framework that combines multiple indicators into a composite technical debt score and by examining how this score relates to operational performance.

The proposed work contributes to computing and enterprise information technology management by providing a quantitative decision support framework tailored to mission-critical trading services. It also responds to accreditation expectations for Masters level computing projects by delivering a non-trivial artefact that demonstrates advanced technical and research skills.

Aims and objectives

The aim of this research is to design and evaluate a quantitative framework for measuring and prioritising technical debt at the service level within mission-critical trading information technology environments.

The objectives are:

1. To conduct a structured literature review on technical debt concepts, measurement approaches and the operational and regulatory context of trading information technology services.
2. To identify measurable service level indicators that represent different aspects of technical debt and operational impact in mission-critical trading services.
3. To design a multi criteria decision model, based on an established method such as TOPSIS, that combines these indicators into a composite technical debt score for each service.
4. To implement the decision model as a prototype artefact using the Python programming language.
5. To apply the artefact to a single organisational case study using anonymised operational service management data in order to examine the correlation between the technical debt scores and operational metrics such as incident frequency, mean time to repair and change failure rate.
6. To evaluate the practical usefulness, interpretability and limitations of the framework through empirical analysis and expert feedback.

Research question

How can service level technical debt in mission-critical trading information technology services be quantified using a multi criteria decision model, and how do the resulting technical debt scores relate to operational performance metrics such as incident frequency, mean time to repair and change failure rate.

Methodology and development strategy

The project will adopt a design science research methodology, which is appropriate for computing projects that aim to design and evaluate artefacts addressing real world problems. Design science emphasises purposeful artefact creation, iterative refinement and rigorous evaluation, which align with the goals of this project. The research will follow a staged, plan driven structure with clearly defined phases and limited iteration. Major stages include project initiation and proposal, literature review and gap analysis, research design and data strategy, framework and model design, data preparation and analysis design, artefact implementation, empirical evaluation and dissertation writing.

A case study strategy will be used for empirical evaluation. The framework will be applied to anonymised operational service management data from a single trading information technology organisation. This allows detailed analysis within a realistic context while remaining feasible within the time and resource constraints of the Masters project.

Development processes and resources

The development processes consist of three main streams.

First, indicator and model design. Based on the literature review, a set of service level technical debt indicators will be defined. These may include measures related to operational performance, architectural complexity and regulatory criticality. Each indicator will be classified as a benefit or cost criterion for inclusion in the multi criteria decision model, and weights will be specified to reflect their relative importance.

Second, multi criteria model specification. A specific multi-criteria decision-making method such as TOPSIS will be selected to combine the indicators into a single technical debt score per service. The method will be configured to compare each service with an ideal best and ideal worst solution, producing scores that support ranking and prioritisation. The mathematical formulation, including normalisation, weighting and distance calculation, will be documented to ensure transparency and reproducibility.

Third, artefact implementation in Python. The framework will be implemented as a Python based prototype. The artefact will load anonymised indicator data, apply normalisation and weights, execute the TOPSIS model and produce scores and rankings. Python is chosen due to its strong support for numerical computation and data analysis and its wide use in both academic and industrial settings. Supporting resources include an appropriate Python environment, version control, office tools for documentation and plotting and secure storage for the anonymised operational data.

Ethical considerations and risk assessments

Ethical approval is required for all research projects, including those that use human data and records from external organisations. Approval must be granted before any data are collected or used. This project uses operational service data from an external trading organisation, which will be treated as human data in a broad sense and will therefore undergo ethical review.

Key ethical considerations include organisational permission, confidentiality, anonymity and data security. Written permission will be obtained from the organisation, confirming which data may be used and under what conditions, and ethical approval will be sought through the university process before analysis begins. Operational data will be anonymised at source, with service names, system identifiers and any personal identifiers removed or replaced with codes. The dissertation will refer to services only in generic terms and will not disclose vendor names, product names or internal identifiers.

Data will be stored on secure, password protected storage controlled by the student and retained only for as long as required to complete the dissertation and associated assessment. The raw dataset will then be securely deleted, while the dissertation will contain only non identifying summary statistics. The project plan also recognises risks such as limited data availability or quality, scope creep, over complexity, time constraints due to full-time employment, misalignment between research question and analysis and ethical or confidentiality constraints. Mitigation strategies include using multiple indicators, prioritising interpretability,

freezing artefact scope early, following a structured plan, continuously mapping methods to objectives and using anonymised and aggregated data only.

Evaluation methods

Evaluation will focus on both the internal behaviour of the model and its empirical performance with real data. Internal evaluation will include logical checks to confirm that services with worse indicators receive higher technical debt scores and that the model behaves as expected under changes in indicator values and weights. Simple sensitivity analyses will be performed to assess the effect of different weight configurations on service rankings.

Empirical evaluation will apply the Python artefact to the anonymised service level dataset from the case study organisation. For each service, the decision model will produce a composite technical debt score and ranking. These scores will then be compared with observed operational metrics such as incident frequency, mean time to repair and change failure rate using descriptive statistics and correlation analysis, while recognising that these analyses identify associations rather than causal relationships. Where feasible, feedback will be sought from practitioners in the case study organisation about the interpretability and practical usefulness of the scores and rankings.

Description of artefact

The artefact will be a Python based analytical tool that implements the service level technical debt scoring framework for mission-critical trading information technology services. Its purpose is to support more informed decisions about technical debt remediation by providing a transparent, repeatable method for calculating and comparing technical debt scores across services.

Functionally, the artefact will ingest anonymised service level indicator data from structured files, normalise indicators, apply weights, execute the selected multi criteria decision model and generate outputs including a composite technical debt score and ranking for each service. It will also produce simple tables and figures for use in the dissertation and in the artefact

presentation. The prototype will emphasise correctness, clarity and interpretability rather than complex user interfaces.

Timeline of proposed activities

The timeline follows the project plan and Gantt chart already defined for the project. Weeks one to six are allocated to project initiation and proposal, including refinement of the topic, confirmation of aims and objectives and completion of the research proposal and ethical approval application. Weeks seven to ten focus on the literature review and gap analysis, covering technical debt, service management and operational resilience literature and identifying the research gap.

Weeks eleven to thirteen are dedicated to research design and data strategy, including confirmation of the design science approach, definition of the case study design and agreement of data access arrangements. Weeks fourteen to eighteen cover framework and model design, during which service level indicators are defined and the multi criteria model is specified. Weeks nineteen to twenty-one are used for data preparation and analysis design, including obtaining anonymised data and defining the analysis steps. Weeks twenty-two to twenty-four are allocated to artefact implementation in Python, followed by week twenty-five for empirical evaluation using the case study data. Weeks twenty-six to twenty-eight are reserved for writing and submission of the dissertation, and weeks twenty-nine to thirty for preparation and delivery of the artefact presentation and oral defence.

Key literature related to the project

The literature review will cover several strands of work. First, conceptual work on technical debt, including definitions, types and critiques of the metaphor, and studies proposing frameworks and metrics such as technical debt ratio, SonarQube based measures, quadrant models and prioritisation heuristics, with a focus on their limitations in enterprise contexts. Second, literature on technical debt in financial and regulated environments, including legacy systems in capital markets, regulatory constraints, risk appetite, auditability and the impact

of technical debt on operational risk and service continuity.

Methodological literature on design science research in computing and information systems will guide the research design, including problem identification, artefact construction and evaluation. Project management literature comparing staged, iterative and hybrid approaches will support the choice of a plan driven but iterative management strategy. Finally, literature on multi criteria decision-making and related methods, including applications of TOPSIS and similar techniques to service prioritisation and risk assessment, will inform the design of the technical debt scoring model. All sources will be cited and referenced using the Harvard style required by the University of Essex Online.