AVIONIC SYSTEMS STANDARDISATION COMMITTEE

ASSC ARCHITECTURE WORKING GROUP

Doc No : ASSC/510/2/9 *

Issue 5- March 1997

THE USE OF METRICS FOR ARCHITECTURE EVALUATION*

0. EXECUTIVE SUMMARY

One of the activities of the ASSC Architecture Working Group is to consider methods for the assessment of candidate avionic architectures. Metrics offer a potentially useful tool for this purpose, and this report examines their possibilities and provides practical guidance for their application.

It covers the reasons for using metrics, typical metrics and details of how they might be used as well as their applications. It includes a section on metrics being used during the ASAAC Phase 1 military programme and another on their use during the civil Control Technology Programme (CTP).

The conclusion is that metrics can provide a means of assessing architectures and are also useful in detailed system engineering tasks such as selecting data network standards or choosing between subsystem options. In each case, however, the methods and metrics need to be selected carefully.

Blank page

Table of contents **Page no.**

Figures and tables

1. INTRODUCTION

The ASSC Architecture Working Group was established to address system architecture issues relating to future avionic system technology and standards. The activities of the Working Group include the assessment of candidate architectures and supporting standards for applicability to the avionic environment. This work requires some means of assessing effectiveness and applicability and the WG has considered an approach based on a reference model (ASSC/510/2/10) and a set of metrics.

The document has been produced by the Architecture WG as a guide to the potential use of metrics for:-

a) Assisting the WG itself in the assessment of candidate architectures and supporting standards.

b) Assisting system engineers in architecture assessment and selection.

c) Assisting in the broader system engineering process.

The document considers:-

i) The reasons for using metrics.

ii) Typical metrics and how they might be used.

iii) The application of metrics in support of a), b) and c) above and an example of their usage.

It is intended that this document will be expanded at a later date to include a preferred list of metrics to be used by the Architecture WG and which could, potentially, be used for other purposes.

2. WHY USE METRICS?

The System Engineering process, including the assessment of architectures and supporting standards, inevitably involves the definition of candidate solutions, review of these potential solutions through trade-off analysis, and selection of preferred options. In the past, much of this activity was intuitive, relying on largely subjective assessments supported by some objective information. It is arguable that the intuitive, unstructured approach has led to “less-than-optimal” decision making. The growing complexity of integrated systems also makes an intuitive approach less tenable.

Alternative approaches are available which attempt to bring more objectivity into the process. These approaches are based on defining a set of metrics which can be used to assess and quantify the “goodness” of potential options. These metrics can be truly measurable quantities or parameters such as mass or volume, or less tangible properties, such as flexibility, which can be given a degree of objectivity through a scoring process. An approach based on metrics offers a number of potential advantages, including:-

a) Improved quality of decision making due to less reliance on individual’s subjective options.

b) Improved traceability in the assessment and selection process, allowing decisions to be revisited if circumstances change or initial assumptions prove invalid.

c) The ability to conduct meaningful sensitivity analyses on decisions and selections, to isolate or highlight dominant factors and to refine options to produce a more optimal solution.

d) The possibility of building up an increasing expertise or knowledge base by continually reviewing the success and validity of past assessments throughout the life of systems.

The use of metrics is consistent with the recent general trends towards more structured and objective approaches to all business processes.

3. TYPICAL METRICS AND SCORING AND COMBINING METHODS

There are two basic categories of metrics:-

a) Those that are based upon measurable properties, such as cost, mass, volume or power consumption.

b) Those that are based upon less tangible properties, such as upgradeability, flexibility and operational effectiveness.

Table 1 provides lists of metrics in each category that may be relevant to the assessment of candidate system architectures and supporting standards. The lists are not exhaustive, but are based upon the experience of ASSC Architecture Working Group members with the use of metrics.

The assessment process starts with the selection of an appropriate set of metrics. Two sets of rules for the use of the metrics must also be defined. Firstly, rules are needed for "scoring" the "goodness" of each option against each metric. Secondly, a method is needed for combining the individual metric scores into an overall "goodness" metric for each option.

There are many options available for scoring and metric combination. Scoring methods can vary from simple linear scales to complex functions that embody concepts such as minimum acceptable values and "diminishing returns" at extreme values. Figure 1 illustrates some simple and complex scoring methods for an easily measured or estimated metric, such as mass. Assigning scores to the less tangible metrics is a far more difficult issue and basically relies upon converting a subjective judgement into an objective score. Figure 2 illustrates a simple scoring method for "updateability" based on a subjective assessment, in this case the constraints on updates imposed by technology level.

There are also many options for combining individual metric scores into an overall score of "goodness" or figure-of-merit. The most widely used method is probably weighted summation of the scores in which each metric is given a predetermined weighting appropriate to its importance and individual metric scores are multiplied by this weighting prior to summation. It should be noted that the preferred option need not be that with the highest figure-of-merit. In some applications it may be more appropriate to select an option that exceeds certain score thresholds for the major metrics or the metric that has the most balanced individual metric scores (ie a good score for each metric rather than very high scores for some and very low for others).

The discussion above is restricted to the so-called "hard methodologies" which attempt to provide a quantified assessment of the quality of options. These methods are supported by well-established theories and indeed, a whole science of "decision theory" has evolved around them in recent years to address decision making issues. In addition to the hard methodologies a number of "soft" methodologies also exist which attempt to help the user to make incremental progress towards decisions using previous experience rather than metrics. The choice of methodologies should depend upon the particular application, with hard methodologies being most applicable to problems where quantifiable supporting data exists. Soft methodologies are more applicable to problems where there is difficulty in providing quantifiable data and greater uncertainty concerning input information. There are software-based tools available to support the hard methodologies which have been widely used for a number of applications. Some of these tools provide considerable flexibility in scoring and weighting methods and support sensitivity and robustness analysis.

4. PROBLEMS WITH USING METRICS

While the advantages of using metrics in the manner described in the preceding sections are considerable, there are a number of problems worthy of discussion.

One major problem results from the basic principle of quantifying the qualitative data that underpins the hard methodologies. Tangible properties such as mass volume, cost etc can be easily quantified and included in a scoring process. The less tangible properties, such as risk, flexibility, updateability and some aspects of operational effectiveness which are difficult to measure, can only be quantified through some arbitrary scoring process, such as that illustrated in Figure 2. Finding a satisfactory process presents one problem and then combining scores arrived at in this way with more rigorously developed scores for the tangible properties presents further difficulties.

A second major problem arises from the concept of producing a single figure-of-merit for each option. This implies the combination of scores for a range of disparate properties such as:-

- physical properties
- intangible properties
- cost-related properties
- risk-related properties

In practice, it is often very difficult to arrive at a satisfactory method of combining scores. Weighting metrics prior to summation provides a means to emphasise the properties considered most important but reaching agreement on a set of weighting is often problematic.

One way around this problem is to avoid the use of a single figure-of-merit covering all of the disparate properties and have a figure for each major set. Thus each option considered would have separate figures-of-merit for, say, physical properties, cost and risk. The selection of the best overall option is then left to a subjective judgement with these figures as objective inputs to the decision making process.

The single figure-of-merit approach has other drawbacks, mentioned briefly in section 3. Unless the scoring process is carefully constructed and sophisticated, a method that produces a single “goodness” figure can favour a solution that has some outstanding features (and thus has high scores against some metrics) and some poor ones. This is often not the best solution because the weak areas can cause significant problems. Some methods for combining scores attempt to overcome this problem by penalising options that have extreme scores against some metrics and favouring those that score well against them all.

A third concern in using metrics is that the process can, without safeguards, be misused to produce a desired result, favourable to a particular option or solution. The most obvious form of misuse is the manipulation of scores and weightings by individuals to favour their preferred option. Proponents of a particular option will propose inflated scores for that option or seek to impose a weighting system that favours the metrics where their option scores most highly. This type of misuse or abuse can be either conscious or unconscious. The use of intangible metrics perhaps makes this problem worse because scores for intangible properties are always arguable. Methods have been developed to address this problem. These include the elimination of extreme inputs during the scoring process and frameworks for producing consensus values for scores and weightings where several individuals have an input.

All of the problems discussed above need to be carefully considered before and during the process of using metrics for assessment and evaluation. Metrics should be used with caution and should, perhaps, generally form only a part of the evaluation process, supported by judgement and experience. Wherever possible, analysis of the results of the process should be carried out to gain a greater insight into the main contributory factors and sensitivity analysis should be undertaken where there are doubts about the validity of input data.

5. USE OF METRICS FOR ARCHITECTURE ASSESSMENT

5.1 General Points

Metrics can clearly be used in support of any evaluation or decision making process. The assessment and selection of an appropriate architecture for an avionics system is just such a process where metrics can be applied with the benefits, noted in Section 2, of objectivity, traceability and “analysability”. Architecture evaluation or selection is, however, usually carried out, initially at least, during the very early stages of the system engineering process with the aim of defining the major building block that underpins all subsequent activities. At this early stage in the process, the quality of information available is often poor and this increases the difficulty of using metrics as much of the input data will inevitably be speculative. It is important, therefore, that the potential problems highlighted in Section 4 are carefully considered.

Metrics can be used to support architecture assessment and selection in two different applications:-

i) The selection of an architecture for a particular system application.

ii) The selection of an architecture for more generic application, such as for use in all avionic systems for UK platforms for the next 20 years.

The first case is the most straightforward because implementation-related factors can be easily quantified and used in the scoring process. The second case is more difficult because many potential implementations must be considered. It is therefore less easy to produce objective data and a great deal more effort is required.

The activities of the Architecture WG fall mainly in the latter category because the work is aimed at identifying suitable architectures and supporting standards for generic application.

Whatever the application, the key to the successful use of metrics for architecture assessment is the selection of an appropriate set of metrics and a method for combining metric scores. Members of the Architecture Working Group have made available information relating to the use of metrics in two major programmes, the ASAAC Phase 1 Programme and the Civil Control Technology Programme. This information is provided in the sections below to illustrate some of the issues involved in using metrics for architecture assessment.

5.2 The Use of Metrics During the ASAAC Phase 1 Programme

The ASAAC programme was aiming to produce generically applicable architecture concepts and so recognised the need to consider multiple implementations. Three basic categories of metrics were identified:-

- Mission related metrics
  - functionality
  - dependability

- minimum dispatch capability
- minimum full function capability
- physical aspects (mass, volume, power, cooling)
- enhancement

• Operational metrics

- availability
- sustainability
- system time to minimum dispatch
- maintenance man hours/flight hour
- maintenance period

• Life cycle cost metrics

- development cost
- production cost
- initial support cost
- operational support cost

Some potential metrics, such as battlefield survivability, were rejected because they were considered too subjective. Implementation specific parameters, such as process power, memory size and reconfigurability were also rejected to maintain the generic approach to architecture definition. Weightings were assigned to each metric to reflect relative importance.

The assessment process was carried out in the later stages of the ASAAC Phase 1 Programme. It was acknowledged that, although the assessment process had been useful, further work is needed in subsequent phases to further refine the architecture concepts. The ASAAC experience with metrics seems to have been favourable and certainly illustrates some of the points raised earlier in this document.

Since the completion of ASAAC Phase 1, the DRA have funded GEC-Marconi Avionics to investigate the use of metrics for architecture selection. A hierarchical set of metrics has been defined during the programme and the hierarchy and the definition of the metrics is included as Annex A of this document.

5.3 The Use of Metrics During the Control Technology Programme (CTP)

The second phase of the Control Technology Programme was aimed at examining the application of civil Integrated Modular Avionics (IMA) concepts (as embodied in the ARINC 651 Design Guidelines) to fixed-wing, rotary wing and engine platforms. Westland Helicopters Ltd (WHL) used a metric-based approach to select a preferred IMA concept for a nominal future helicopter avionics system update. Three IMA architecture concepts were considered alongside a baseline, conventional technology, option.

Three categories of metrics were used, as listed below:-

- Costs incurred over the system life cycle
  - Development
  - Production
  - In-service
- Risk
  - Certification
  - Commonality of Components
  - Multiple Sources
  - Advanced Technology
- Technical
  - Reliability
  - Maintainability
  - Mass
  - Installation
  - Future Growth

The Saaty (or Analytical Hierarchy Process) method was used for the assessment. This method involves comparing each option against all others using a scale of relative merit (ie equal to, slightly better than, much better than, etc) for each metric.

The scale used was:-

1 = Equal

2 = Slightly Better	1/2 = Slightly Worse
3 = Moderately Better	1/3 = Moderately Worse
4 = Much Better	1/4 = Much Worse
5 = Very Much Better	1/5 = Very Much Worse

The individual pair comparison scores are entered in a matrix and the combination of scores for each metric is by means of:-

- Calculating the product of the scores for each option
- Calculating the nth root of the product (where n is the number of options)
- Summation of the roots
- Division of the option root by the total

The example shown below is the matrix for mass.

	Baseline	Arch 1	Arch 2	Arch 3	Product	4th Root	Score
Baseline	1.00	0.50	0.33	0.25	0.04	0.45	0.09
Arch 1	2.00	1.00	0.33	0.25	0.17	0.64	0.13
Arch 2	3.00	3.00	1.00	0.50	4.50	1.46	0.30
Arch 3	4.00	4.00	2.00	1.00	32.00	2.38	0.48
Total						**4.93**	**1.00**

The same process was used during the WHL CTP activities to arrive at weightings for:-

- each metric within the three metric categories noted earlier
- the three metric categories themselves

The metric scores for each option were then multiplied by each metric weighting and summed to give an overall score for each metric category. These metric category scores for each option were then, in turn, multiplied by the category weighting and summed to give the overall figure-of-merit for each architecture option.

This process was used by WHL to select an architecture for more detailed definition and evaluation during the programme. Interestingly, although the scores for each metric and each category of metric differed significantly between the options (total ranges of 100%+ difference were common), the final figure-of-merit for the options differed only very slightly (the best score was only 10% better than the worst). The option with the highest overall score was not the highest scorer in any of the three categories. WHL carried out an analysis of the outcome of the assessment to ensure that no particular weighting or metric had an undue impact upon the assessment. The overall conclusion was that the best "all round" option had been selected and this was used as the basis for further work.

5.4 Use of Metrics Throughout the Life Cycle

It is important that metrics are not used in a "one-shot" manner and then discarded. One of the principle advantages of using metrics is that the evaluation can be revisited as more "hard" information becomes available and the evaluation reconsidered as appropriate. By continually refining the data inputs as downstream system engineering activities are undertaken, it is possible to build up a knowledge base that can be used in subsequent evaluations using the same metrics, and to set targets for future architectures and system implementations.

6. USE OF METRICS IN THE SYSTEM ENGINEERING PROCESS

In concluding this document, it is worth emphasising that the use of metrics need not, of course, be limited to architecture evaluation. The system engineering process embodies a continuous series of trade-off analyses and decision points, each of which would be supported by a metric-based approach. Indeed, most of the experience gained with the use of metrics has been at the more detailed levels of system engineering for selecting appropriate data network standards or choosing between sub-system options. Clearly, for each application the methods and metrics used must be carefully chosen but much of the discussion contained in the early sections of this document is relevant and some of the metrics highlighted may themselves be of use.

TABLE 1 TYPICAL METRICS FOR ARCHITECTURE ASSESSMENT

Tangible Metrics	Intangible Metrics
Physical Properties	Upgradeability
	Flexibility
- environmental tolerance	Reconfigurability
- volume	Safety
- mass	Technology dependancy
- power consumption	Risk
- cooling requirements	Modularity
- ballistic tolerance	Scaleability
Operational Properties	
- reliability	
- maintenance man hours	
- availability	
- maintenance period	
- interchangeability	
Costs	
- development cost	
- production cost	
- in-service cost	
Functionality	
- processing capability	
- memory capacity	
- data throughput	

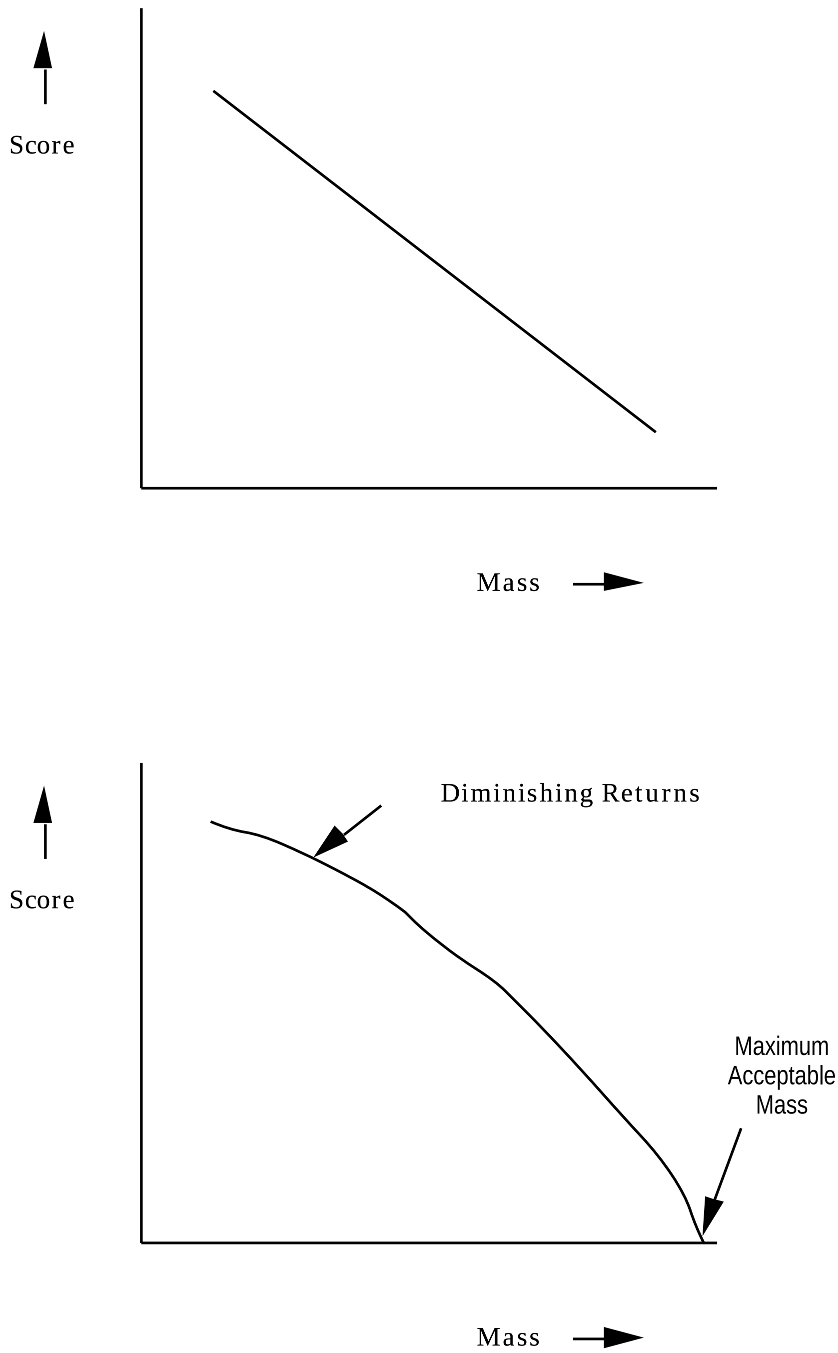

Figure 1 Typical Scoring Methods

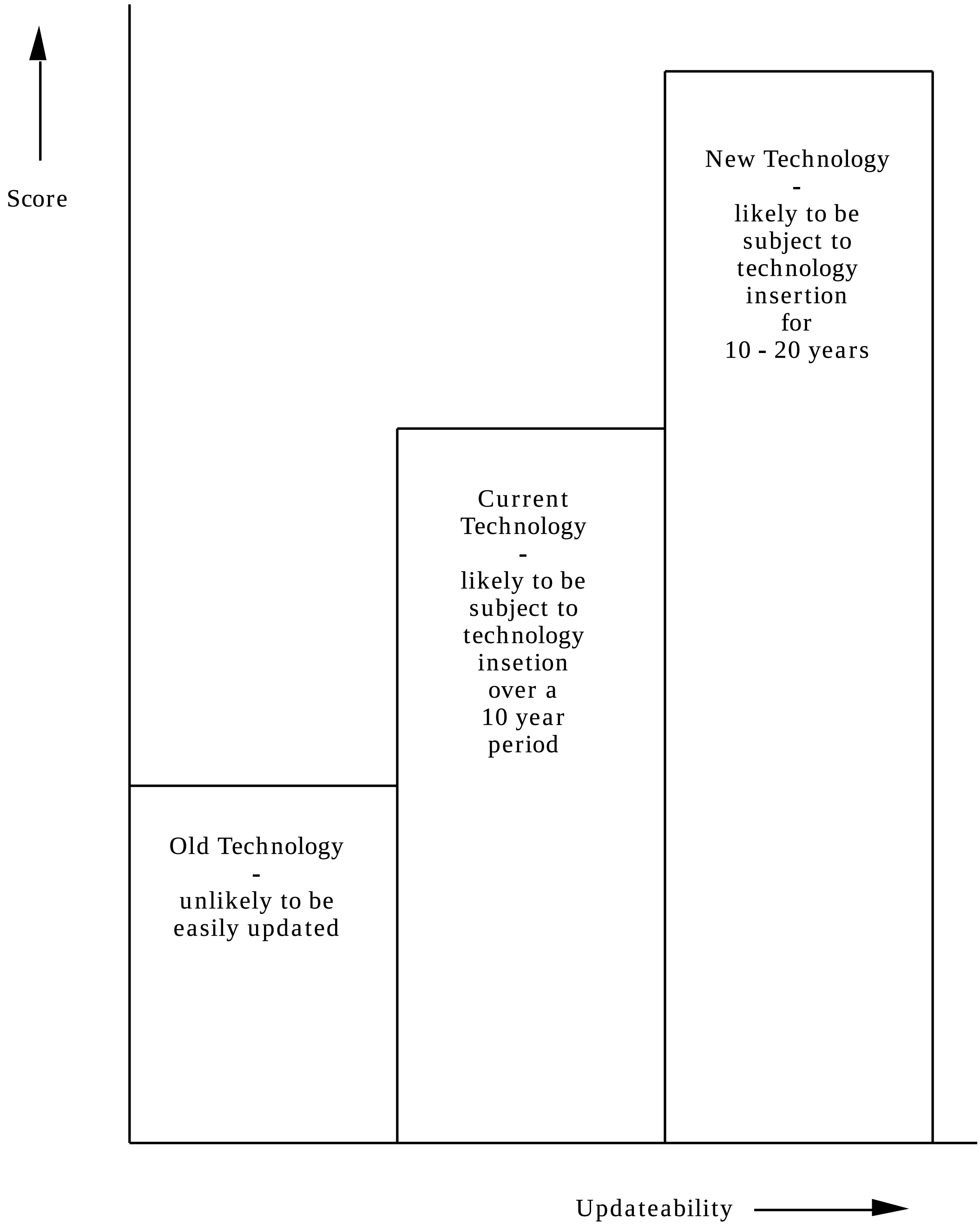

Figure 2 Scoring Method for an Intangible Metric

ANNEX A METRICS FOR ARCHITECTURE ASSESSMENT

The following are the set of metrics that were evaluated under the DRA funded programme "Metrics for Architecture Assessment" (ASF/2067/E). These are shown hierarchically in Figure A.

A.1. Capability

- **System Engineering**

Ease Of Element Implementation: The ease of converting architecture element specifications to implementations. A subjective metric.

Ease Of Element Verification And Validation: The ease of proving architecture element requirements are met. A subjective metric.

Ease Of System Development: The ease of converting system requirements into a design using the architecture. A subjective metric.

Ease Of System Verification And Validation: The ease of proving system requirements are met. A subjective metric.

- **Flexibility**

System To Mission: The capability to change rôle entirely; from interceptor to strike for example. A subjective metric.

Mission Context Switch: The capability to change between mission modes; from offensive to defensive for example. A subjective metric.

Reusability: The proportion of elements usable on other platforms with no modification.

System Enhancement: The capability to expand system functionality while maintaining the characteristics of the defined system.

- **Evolutivity**

Architecture Instance Technology: The level of technology used in an architecture instance. A subjective metric.

Architecture Technology: The minimum technology required for an architecture to be implemented. A subjective metric.

Scalability: The ability to incorporate changes to the architecture with minimal impact on unchanged parts. A subjective metric.

- **Fault Tolerance**

Sustainability: The average time for which an architecture instance will not experience a critical failure when starting with the full installed system resources and no maintenance being carried out. A critical failure is when the system can no longer provide all the system application functions required for the mission (i.e. falls below the minimum resources).

Mean Time Between Critical Failure: The average time for which an architecture instance will not experience a critical failure when normal maintenance activities are being carried out to restore the full installed system resources. A critical failure is when the system can no longer provide all the system application functions required for the mission (i.e. falls below the minimum resources).

- **Physical Characteristics**

Mass: The mass of the equipment.

Volume: The volume of the equipment.

Consumption: The electrical power consumption of the equipment.

Dissipation: The proportion of the engine power which is taken up in order to remove heat from the equipment.

- **System Functionality**

Minimum Resources: The minimum proportion of the installed resources required to provide all the system application functions required for a specific mission.

Performance: The ability to provide all system application function requirements (specific to individual system missions).

A.2. Operational

- **Operational Infrastructure**

Maintenance Provision: The means by which maintenance is provided. A subjective metric.

Maintenance Facilities: The requirements for maintenance facilities. A subjective metric.

STTE Resources: The level of Special to Type Test Equipment required to support an architecture or architecture instance. A subjective metric.

- **Reliability**

Availability: A measure of the degree to which an item is in an operable and commitable state at the start of a mission when the mission is called for at an unknown (random) time.

Dependability: The probability of completing a designated mission successfully upon choosing to use the aircraft for the mission with only the minimum resources available.

Graceful Degradation: The ability to continue a level of service in the presence of a critical failure according to the priority of the functions. A critical failure is when the system can no longer provide all the system application functions required for the mission (i.e. falls below the minimum resources). A subjective metric.

- **Maintenance Effort**

Maintenance Interval: The time between maintenance tasks for first line maintenance.

Maintenance Manpower: The labour required to carry out maintenance tasks at all maintenance levels.

Maintenance Skills: The need for additional training or skill acquisition to carry out specific maintenance tasks at all maintenance levels. A subjective metric.

A.3. Logistics

- **Logistics Infrastructure**

Spares Provision: The means by which spares are provided. A subjective metric.

Spares Facilities: The requirements for spares facilities. A subjective metric.

- **Replenishment**

Technology Obsolescence: The availability of the technology to provide spares replenishment. A subjective metric.

Off The Shelf'ness: The proportion of components utilised within an architecture which are special-to-type. A subjective metric.

Long Lead Times: The time taken to get hold of replacement components. A subjective metric.

- **Architecture Elements**

Number of Element Types: The number of differing element types in an architecture.

No Fault Found Removals: The proportion of first line maintenance activities where a defect is thought to be present during operation, resulting in equipment replacement, but subsequent inspection reveals no fault.

A.4. Cost

- **System Costs**

Development: The cost of translating an architecture design into a specific architecture instance to meet the system requirements.

Manufacture: The cost of manufacturing a copy of a specific architecture instance.

Initial Support: The cost of initial support activities for a manufactured architecture instance.

Operational Support: The cost of operational support for an architecture instance.

Disposal: The cost of disposing of an architecture, including hazardous material disposal, cost of disposal personnel and the offset of costs from the disposal of assets.

- **Architecture Costs**

Technology R&D: The cost of the R & D studies required in the development of an architecture.

Proof of Concept: The cost of the proof of concept studies required in the development of an architecture.

- **Management Costs**

Activity Management: The cost of managing the development and use of an architecture and its instances.

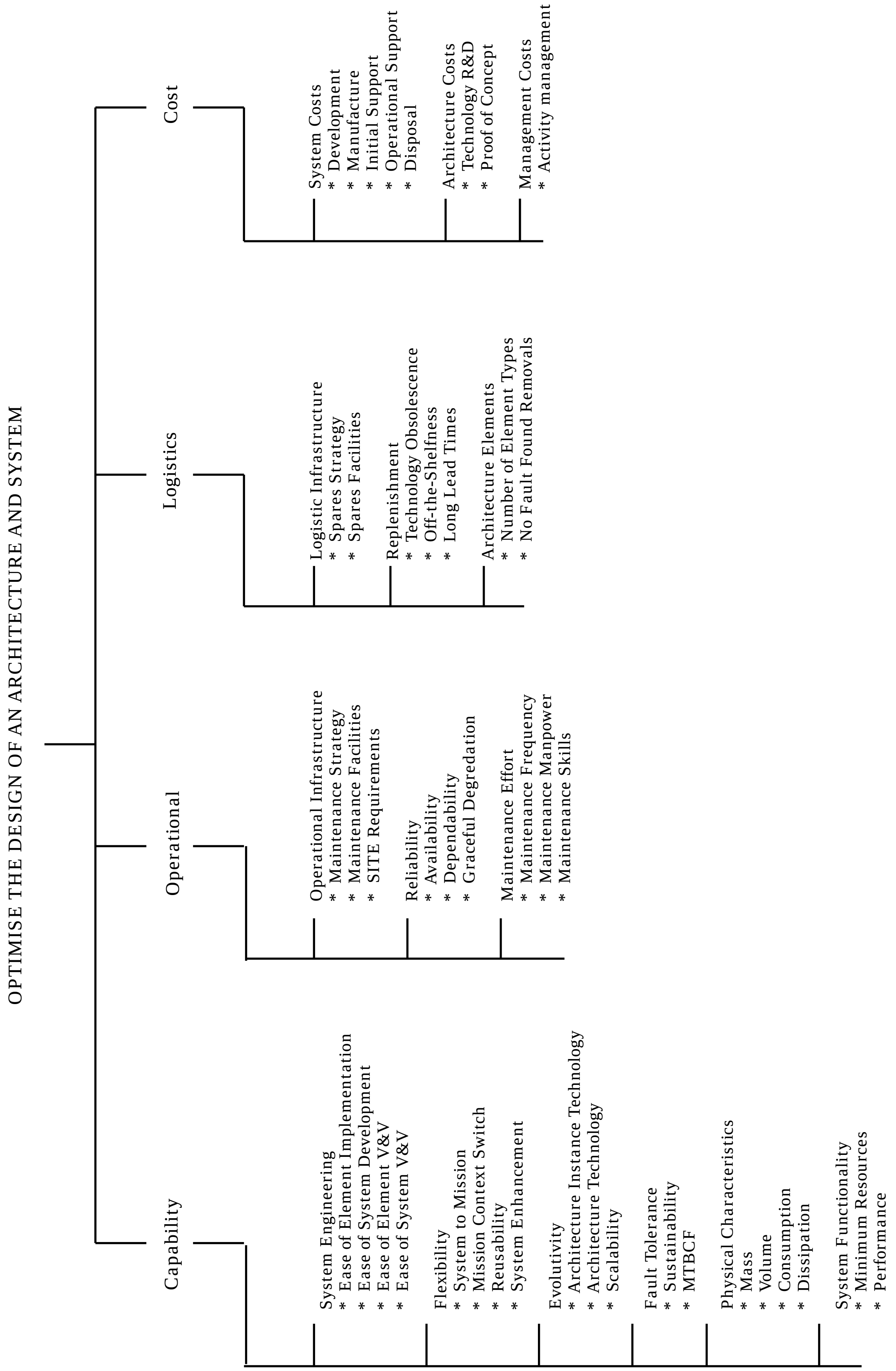

Figure A **Metrics for Architecture Assessment Goal Tree**