Comparing several candidate architectures variants : An Industrial Case Study

Sébastien Madelénat,
Christophe Labreuche
and Jérome Le Noir
Thales Research & Technology, France
firstname.name@thalesgroup.com

Grégory Gailliard
Thales Communications & Security, France
firstname.name@thalesgroup.com

***Abstract*—In system and software engineering, the analysis of architectural variants is most of time irrational and manual. The most common approach for comparing variants is comparing results for each variant evaluation. Most advanced approaches available in architecture evaluation are suffering from three principal weaknesses: the absence of criteria elicitation method, no representation of real-life strategies and no explaination of the outcomes. This paper relates experiments of a MCDA tooled method addressing these weaknesses. The experimentation is supported by an industrial use case consisting in selecting the best platform for an handheld Software-defined Radio. Its architecture description is formalised with model-based design tools. As a result we conclude the experimented approach provides sharper results than classic approach on the class of decision problem exposed by avoiding false positives. This approach seems to be promising to improve the confidence in our Decision Analysis Report and their quality in terms of argumenting the reasons of a decision.**

Keywords: System engineering, Decision model, Tradeoff, Multi Criteria Decision Analysis

I. Introduction

In system and software engineering, the analysis of architectural variants is most of the time subjective and manual. The justification of a variant is seldom based on the assets, the flaws and strengths of the different options. Ideally, assessing or comparing several candidate architectures (variants) should be based on some decision criteria – corresponding to a Multi- Criteria Decision Aiding (MCDA) problem. Among the classic hand-made "Decision Analysis Report", widely used in the industry, the industrial methods and tools state-of-the-art presents several weaknesses.

This paper presents an industrial practical experience based on a MCDA tooled method for evaluating and comparing several candidate deployments of a Software-defined Radio application on hardware platforms[1] in a Model-based system engineering (MBSE) context.

A. Needs for decision aid to select the best candidate alternative

In [11] the empirical demonstration has been made that variability in the description model of the architecture allows to easily construct a set of candidate architecture descriptions. Then one aims at finding the options that best fits with the needs of the various stakeholders. This choice problem can be formulated as the maximization of a set of decision criteria. The difficulty is that the decision criteria are usually numerous and conflicting. One may indeed have performance criteria versus cost criteria which cannot be met both at the same time. The difficulty is manifold. First of all, the decision criteria are described by metrics and one first needs to identify a set of relevant metrics. Then, the selection of one alternative among several on the basis of a set of metrics is complex since there are commensurateness issues (combine "apples with oranges" as the metrics are given in different units), and one aims at making arbitrage between the metrics.

Nowadays, in system and software engineering, practices make this analysis most of the time empiric and irrational ([9], [8], [1], [5] and [7]) as described in subsection II-A. One may say deciding is an essentially irrational activity [12]. Therefore it is not very surprising to find irrational analysis practices. Rational analysis requires providing arguments about choices. MCDA approaches allow this. Moreover the justification of the choice of an alternative is very seldom based on the assets and flaws of the different options. Assessing or comparing several candidate alternatives on the basis of some decision criteria corresponds to a Multi-Criteria Decision Aiding (MCDA) problem. The approach consists in constructing an explicit multi-criteria decision model. The main benefits of explicitly constructing such a model are to reach objectivity in the analysis, come up with a recommendation from a well-established methodology, and the possibility to justify the results. The model is elicited from interviews with the stakeholders and expresses the preference of these actors.

The Section 2 presents an analysis of existing approaches in engineering and the flaws induced, the Section 3 presents our approach, the Section 4 describes the experimental use case and the evaluation preparation, finally the Section 5 concludes on the results and discusses about the proposed approach applicability.

[1]The research activities were conducted in the context of ITEA2-MERgE (Multi-Concerns Interactions System Engineering, ITEA2 11011), a European collaborative project with a focus on safety and security

II. Analysis of the existing Multi-criteria Approaches in Engineering

The problem of identifying a set of relevant metrics has been performed in different domains such as performance ([16]), availability ([14]), modifiability ([13]). Moreover, the Software Engineering Institute (SEI) has developed the Architectural Trade-off Analysis Method SM (ATAM SM) and validated its usefulness in practice ([15]). ATAM aims at making rational choices among competing architectures. It bases its criteria elicitation on stakeholders scenario over the system in consideration. The ATAM uses stakeholders' perspectives to produce a collection of scenarios that define the qualities of interest for the particular system under consideration. Scenarios give specific instances of usage, performance and growth requirements, types of failures, and possible threats and modifications. Once the important quality attributes are identified, the architectural decisions relevant to each one can be analyzed with respect to their appropriateness. The ATAM was designed to make rational choices among competing architectures, based upon well-documented analyses of system attributes, concentrating on the identification of trade-off points. The ATAM also serves as a vehicle for the early clarification of requirements. As a result of performing an architecture trade-off analysis, enacting ATAM allows an enhanced understanding of, and confidence in, a system's ability to meet its requirements. It helps also eliciting a documented rationale for the architectural choices made, consisting of both the scenarios used to motivate the attribute-specific analyses and the results of those analyses. Compared to traditional MCDA methods, ATAM does not perform multicriteria analysis. It qualifies the link between the description of the architectures (e.g. through a feature model) and the criteria, thanks to concepts such as sensitivity points (a feature that has a large influence on at least one criterion) or trade-offs (a feature which has a positive impact on a criterion but at the same time a negative impact on another criterion). ATAM and MCDA approaches are complementary: ATAM can be used as a complement to a MCDA approach as well as ATAM can be completed by a MCDA approach.

A. Flaws of the existing MCDA methods used in system or software engineering

If one wishes to interpret the metrics and be able to assess the architectures or compare architectures, one needs to add further information. We have identified four levels of information that can be added:

Level 0. This is the case when one has no further information than the list of metrics, except the sense of preference on each metric. When one has only the list of metrics and the sense of variation on each metric, e.g. for a cost metric, the smaller the better. This ordering is not discriminating and does not allow selecting the best option. The Pareto ordering is useful when the number of criteria is small (usually 2 or 3), and when the number of options is large. These two assumptions are false in our context.

Level 1. One is assumed here to provide at least one reference value on each metric. The most usual reference value in engineering is the "target value" which is the budget value of the metric for which the associated requirement is met. Additionally, one may also add the "threshold value" which is the maximal or minimal value of the metric for which the utility of the system is seriously questioned. At this level, one may say whether a given criterion is either completely satisfied (comparison of the metric with the target value) or not satisfied at all (comparison of the metric with the threshold value) but no interpretation of the other values of the metric can be given. The methods on the engineering of requirements belong to this level but as the criteria are not combined, these approaches do not allow to perform a trade-off analysis between the attributes. This is however necessary when not all requirements can be met all together and some compromise shall be reached.

Level 2. On top of level 1, one assumes that a satisfaction function that provides the degree to which the criterion is satisfied for every value of the metric. This is usually described as a piece-wise affine function. This does not specify how to weigh up the values of the various criteria to provide an overall assessment of the alternatives.

Level 3. On top of level 2, one assumes that the way the different criteria shall be combined, is specified. This is classically done through weights assigned to the criteria. In order to make recommendations over the options that go beyond the Pareto ordering, one needs to go to Level 3. The most elaborate MCDA methods available in system engineering is the use of a weighted sum in a quantitative utility method. This allows computing an overall assessment of the candidate options. The existing approaches suffer from three main limitations.

No elicitation method. In the existing tools, the stakeholders enter directly the thresholds and the weights. The understanding of a stakeholder on the parameters of the model is always limited when they are out his expertise domain. The conclusions of the evaluation cannot be justified if the stakeholder provides directly the values of the parameters. Rather one shall ask the stakeholder to provide information that are confident with, e.g. example of decision or assessment, and deduce the values of the parameters. Hence some elicitation methods to construct the values of the parameters shall be used.

No representation of real-life decision strategies. Most of the aggregation functions used are the weighted sum. A weight is assigned to each criterion, quantifying the importance of criteria. The main asset of this model is to be easily understandable. However, this simple model suffers from several limitations. It assumes preferential independence between the criteria: the contribution of one criterion to the overall evaluation does not depend on the marks with respect to the other criteria. This independence is often not met due to the presence of interaction between criteria. A typical example is the presence of a veto criterion. In system analysis, one has often to consider high level consequences (such as operational, financial, human factor, and so on) at the highest level of analysis. The most important criterion concerns surely the operational aspects since the other ones are more or less non-functional. As a consequence, if the operational part is not well-satisfied there is no other well-satisfied criterion that can save the solution. There is no analyst that is happy of a solution that completely misses the mission but costs very little. This means that the operational aspects behave like a veto. This cannot be modeled by a weighted sum.

No explanation of the outcomes. The decision maker is usually the person who is responsible for the decision. He often has to explain his decision to other actors - for instance, peers, managers, executive board or shareholders. These actors have often no time to go into the technicality of the decision model. In order to convince them on the merits of the decision, a synthetic explanation needs to be given such as an argumentation about metrics interpretation based on criteria thresholds, weights and interactions.

B. Existing Metric Analysis Approaches in Engineering

In existing approaches, levels 2 and 3 are often put together. We will consider these levels together in the following. We present the review of the existing approaches used in architecture trades studies in model based system engineering among the levels of information described above.

Methods at Level 0: The construction of the Pareto frontier among the set of options is proposed in the ASDL (Aerospace Systems Design Laboratory) developed at Georgia Tech [9][8], and also in the ESTECO tool for multi-domain engineering[1].

Methods at Level 1: The methods on the engineering of requirements belong to this level but use only one level. The two objective and threshold levels are used in the capability-based approach developed by the US Department of Defense. In order to meet the future needs, the force transformation shall be analyzed according to seven concerns summarized under the acronym DOTMLPF (Doctrine, Organization, Training, Materiel, Leadership, Personnel and Facilities)[5].

Methods at Levels 2 and 3: Levels 2 and 3 are often put together. The most elaborate MCDA methods one can find in system and software engineering is the use of a weighted sum in a quantitative utility method. Some workflow simulation tools such as SIMUL-8 [2] are based on a MCDA weighted sum model. SIMUL-8 integrates a MCDA tool called VISA[3]. A simple MCDA approach based on a weighted sum has recently been integrated in the IBM Rhapsody tool to perform trade studies[4]. The Canadian Department of Defense has also developed its approach for Capability Based Planning. The evaluation for each capability domain is performed at the strategic, operational and tactical levels against six concerns summarized under the acronym PRICIE (Personnel, Research, Infrastructure, Concepts, Information et Equipment)[21]. An evaluation in the numerical scale [0,100] is performed on each concern and a relative importance between the concerns is expressed. This allows computing an overall assessment of the candidate options, using a simple weighted sum model.

III. Our approach

Myriad[17] is an experimental tool for MCDA developped at Thales Research and Technology France. It proposes advanced methods and tools for decision making. It has been successfully experimented in several operational cases in aiding in decision making.

Myriad proposes criteria elicitation tooled method for defining criteria relative weight and thresholds based on a technique among which an extension of MACBETH[10] to account for interacting criterion. Its aggregation function, the Choquet integral, is capable of taking into account important cases (veto, favor, complementarity among criteria) for simulating real decision strategies. Myriad proposes key features such as production of evaluation reports. These reports are about evaluation results explanation based on the evaluation model analysis, augmented with improvement recommendations sorted by potential score impact for each criteria. Moreover, when multiple evaluation are done using the same evaluation model, the report proposes evaluations comparison argumentation. Regarding ATAM, the presented method does not addresses the metrics identification or architecture description updates recommendations for improvement. For this latter part, in [20] Montmain et al. proposed an extension of the method presented in this paper. It consists in an optimization algorithm recommending feature variables updates that maximizes the overall criteria satisfaction based on an influence model relating the feature model to criteria. For these reasons we propose to experiment the capability of Myriad to go beyond limits of existing approaches.

A. Basic concepts

The three basic concepts in the evaluation tree are:

Metric "U". Usually, a metric is a numerical quantity to assess the level of one objective achievement. We will use the word "metric" in a broader sense (like attribute)[22]: instrument which synthesizes in qualitative or quantitative terms, certain information which should lay the foundation for a judgment of an alternative relative to certain of its characteristics, attributes or effects (consequence).

Criterion "C". A criterion is a specification of the preference that an individual has on the values of a metric relatively to a concern. This specification amounts to construct a function – called **utility function** – which returns for each value of the metric the relative performance level (goodness) which positions it on a preference scale. The underlying scale is often a numerical scale, such as the [0,1] interval in which the value 0 is judged unacceptable relatively to the concern of the criterion, and value 1 is judged perfectly satisfactory relatively to the concern of the criterion. One can give an absolute judgment on an alternative according to a criterion.

Aggregation[22] "A". This is a procedure that produces an evaluation of any alternative by taking into account, in a comprehensive way, the performance levels of the alternative according to the criteria corresponding to a set of concerns. There are often nested aggregations. The hierarchical organization of criteria in nested aggregations is due to done to group criteria according to similar concerns.

B. Concept related to utility functions

The construction of a utility function requires two elements. Firstly, an **interval scale** is needed. The evaluation

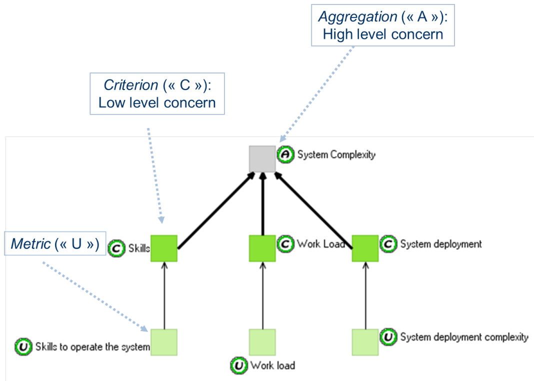

Fig. 1. The three main concepts.

is not simply ordinal since the scales coming from different criteria are combined trough arithmetic operators to allow compensation among criteria. The underlying scale is often a numerical scale, such as the [0,1] interval in which the value 0 is judged unacceptable, and value 1 is judged perfectly satisfactory relatively to the concern of the criterion. An interval scale is a scale in which the notion of difference makes sense. In the scale [0,1], going from utility 0.1 to 0.2 is equivalent in terms of satisfaction gain to going from 0.8 to 0.9. Secondly, **commensurability** between the different criteria is required. Commensurability means that a same evaluation on different criteria has the same meaning. For instance, evaluation 0.3 shall have the same interpretation whatever the criterion. Commensurability is very complex to obtain for ordinal scales, as the user needs to compare all elements of all metrics together. By contrast, interval scales can be made commensurate more easily. An interval scale has two degrees of freedom. In order to fix entirely an interval scale, it is sufficient to fix the utility for two values of the metric. These particular elements will have the same utility on the different criteria, and are called reference elements. A reference level is an abstract level for which one can identify a reference element on each metric which corresponds to the abstract level. In the literature, several reference levels have been defined.

Completely Satisfactory the criteria is completely met. It is a saturation level in the sense that one cannot do better than this level in terms of satisfaction.

Budget (target value) This is the expected value in the requirement provided by the customer.

Satisficing This word has been invented by the sociologist H. Simon. The decision maker is happy when this value is reached, even if better elements exist. The user does not basically look for better elements than the satisficing element. The satisficing element usually corresponds to the budget element.

Neutral This level is neither good nor bad. The decision maker is indifferent when he encounters such element. Values better than the neutral element are considered as Good whereas values worse than the neutral element are considered as Bad.

Not satisfied at all The criterion is not met at all for this value. This is also a saturation level as one cannot be worse than this level.

Unacceptable It is similar to "not satisfied at all", except that one means that this is a veto value. An architecture having an unacceptable value on a criterion cannot be selected.

For simplicity in the following case the utility function is described by 3 levels of criteria satisfaction level: **Not satisfying at all value**, **Budget** and **Completely satisfactory** corresponding respectively to 0, 0.5 and 1 satisfaction marks as illustrated in Fig. 2.

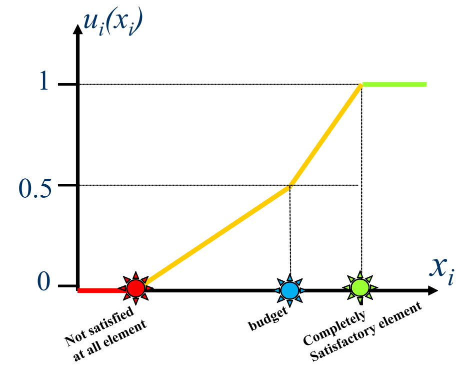

Fig. 2. Shape of a basic utility function

C. MCDA model building

The construction of a MCDA model is composed of the following stages [22]:

*1) **Stage 1**:* It is the structuring phase. The goal is to construct a tree representing a hierarchy of concerns using the basic concepts defined above in which the root represents the overall evaluation, and the leaves are the metrics. All nodes except the leaves to return a numerical evaluation that is a satisfaction degree. The process is decomposed as the following.

Specify concisely the expected issue of the global evaluation. The first step consists in defining in one sentence what the overall evaluation aims at representing. This helps to prevent from taking into account non relevant aspects. This also helps, within future discussions, to retarget the debate by reminding when necessary the objective of evaluation.

Identify the stakeholders. The objective of this step is to itemize the list of stakeholders being involved and in particular to identify all necessary competences. A significant number of stakeholders may intervene in the decision aiding process. Meaning stakeholders are, among others, the **Decision Maker(s)**, responsible of the decision that will be taken. It can be an end user or a customer. Other important stakeholders are the **Expert(s)** which can be any person that may give his opinion in order to help the construction of the evaluation model. Examples of experts are operational and technical experts. The customer may be seen as an expert in some situations. For complex decision problems involving many stakeholders a **Facilitator** may be useful. He is more likely an external consultant rather than a "classic" stakeholder. He helps the stakeholders to structure the debate by bringing a methodology and his external understanding of the problem, by asking the relevant questions and reformulazing what the stakeholders express. He also helps to establish confidence amongst stakeholders and to converge towards a common understanding of the problem shared by all stakeholders.

Build a hierarchy of concerns. The number of relevant criteria is often relatively large and can be larger than several dozen in complex problems.With more than 7 or 8 criteria, psychological studies [19] have shown that human being generally use only simplistic strategies. He decomposes the criteria into two groups: the important ones and the other ones. Only the important criteria are dealt with in a subtle way, taking them separately. The human being makes only a global reasoning on the less important criteria - based for instance on some kind of simple average. The less important criteria are thus not taken into consideration in a subtle way.
For these reasons, one shall proceed at a structuring phase. The aim is to construct a hierarchy of concerns, that is, several nested levels of aggregations. In order to perform such a hierarchy, one shall succeed in grouping the criteria according to a classification that makes sense of the stakeholders. At the end of this step, one shall obtain the relevant criteria together with their organization in a tree. Top-down (objective oriented) and Bottom-up (alternative oriented) approaches are possible even if mixing both approaches is often preferable.

Operationalize the concerns. The concerns that have been identified and grouped in a tree in preceding step are abstractions. We need to identify measurable variables representing these abstractions at best. The difficulty is to identify the right measurable datum which synthesizes the different aspects included within the consequence of each studied concern. In practice, there are often many metrics that can be seen as good representative of a given concern. There are two types of metrics: the **natural metric** is typically a statistical indicator whose expression can be easily described. A particular case concerns the proxi-attributes. These are metrics which link on the concern is not at first sight obvious. To illustrate this, let us mention the metric “concentration of pollutant” to measure the consequence on the concern of the effect on health. The **constructed metric** is computed. The stakeholders shall wonder how to specify a given concern into a value that can be computed. This is not always simple, especially when the available information is in limited quantity.
There are two types of **constructed metrics**. The first type is the **aggregated metrics** constructed by aggregating the values of small components such as the “general load of a system combining the load of sub-systems”. The other type of metric is **ex-nihilo metrics** for the purely subjective judgments such as the “workload of an operator operating a system”.

Specification of the preference ordering on the metrics. Each criterion shall depict a different aspect of the overall evaluation. In any case, the larger the utility the better it is. A metric is not any variable having a more or less direct influence on the overall evaluation of the alternatives. There necessarily exists a preference regarding each criterion. This preference indicates the values of the metric that are judged good, fair, bad, ... The preference associated to a metric regarding the consequences on a concern is specified through a preference relation called criterion. A criterion is characterized by a value function which returns a utility to each value of the metric. This is the utility function. This specifies the better and the worse. One shall specify the sense of the preference regarding a metric, all else being equal on the other metrics. The most commonly encountered senses of variation are ”the larger the metric the better” as for a performance metric, ”the smaller the metric the better” as for a cost metric and ”the closer to a given value the better” as for some soft requirements.

Validate a family of criteria. This step validates the hierarchy of criteria. It consists in checking whether the selected criteria and metrics satisfy to some elementary properties. Some important conditions to consider are the set of selected metric form the only variables on which the overall evaluation will be based. All information necessary to assess the alternatives shall be contained in the list of metrics. The set of criteria shall be independent.
Multi-Criteria Decision Aid has formalized the concept of consistent family of criteria. A family of criteria is said to be consistent if the following three properties are fulfilled: **Exhaustiveness** - The family of criteria is sufficient to compare any alternativeness (i.e. there is no missing criterion). If two alternatives are judged identical regarding all the criteria (same preference), then these two alternatives shall be globally judged identically. **Cohesion** - No criterion is useless. For each criterion, there exist at least a couple of alternatives for which one is globally strictly preferred to the second one such that, if one is strictly preferred to the other on this criterion, they are judged identically on the other criteria. **Non-redundancy** - There are no two identical criteria, suppression of one criterion yields the violation of one of the two previous properties.

*2) **Stage 2:*** It consists in quantifying the evaluation tree on the criteria nodes. In other words, we need to construct a judgment for each attribute separately. This amounts to ask “Is this value for this attribute is good or bad?”. This is quantified by a satisfaction function. The construction of this function results from an interview with the domain expert. It is characterized by some thresholds that need to be identified. For the construction of the curve on intermediate values of satisfaction, we use dedicated methods from measurement to quantify the preferences of the expert into numerical utilities. A utility function on a metric is a scale representing the preferences of a stakeholder. There are basically two types of scales: the ordinal scales and the interval scales. An ordinal scale is rather poor, and does not really permit to handle numbers, since usual arithmetic operations are not meaningful here. In an interval scale, the concept of difference makes sense.
Aggregating different criteria requires that they are commensurable. This implies that one shall be able to compare any element of a metric with any element of any other metric. It is not possible to ask to a stakeholder to perform such a comparison since the direct comparison of elements belonging to different metrics does not make sense for human beings. The

MACBETH approach[10] allows solving these difficulties in a way that is completely satisfactory for the stakeholders and is relevant from a mathematical standpoint.

MACBETH approach overview. To explain the MACBETH approach, let us first note that in an interval scale, the notion of comparison of discrepancies makes sense. Moreover, an interval scale is given up to a dilation and a shift. This implies that it is enough to fix two points in an interval scale to entirely fix the scale. This is done for the utility function (which corresponds to an interval scale) on each criterion. As a consequence, the commensurateness assumption will be satisfied if one is able to find on metrics two elements, having the same meaning throughout all criteria. The second ingredient necessary to construct an interval scale - a specification of the difference between pairs of elements of the metrics - in terms of satisfaction degrees.

Definition of the reference levels on each attributes. The value of the utility function is interpreted as a degree of satisfaction regarding the preferences of stakeholders. This satisfaction degree corresponds to the same scale on all attributes. Such a degree belongs to the [0,1] unit interval, where 0 corresponds to the total absence of satisfaction and 1 corresponds to the complete satisfaction. The two reference levels are identified to these two levels. The stakeholder is thus asked to identify on each metric two elements :

- The first, named ”U”, is thought by the stakeholder as completely unsatisfactory relatively to his concerns w.r.t. criterion, its satisfaction will be valuated to 0.
- The second, named ”P” is considered as perfectly satisfactory[10], its satisfaction will be valuated to 1.

Identification of referent values on each attribute. If the metric is continuous and is an interval, it is not possible to elicit the utility function for all values of the metric. In practice, we elicit the utility function from the preferences of the stakeholder for only a finite set of elements of the metric. The two reference levels U and P shall be part of the considered subset of the metric in order to normalize the construction of the utility over these elements. The utility function of the metric is constructed by linear interpolation from its value on the identified subset.

Let us discuss on the construction of the elements of the subset when the metric is continuous. Since the [0,1] unit interval is a bounded unipolar scale, and is thus bounded from above and below, the two utilities 0 and 1 correspond to saturation levels. In the example of Figure 3, any element lower than 5 (resp. greater than 13) has a utility equal to 0 (resp. 1). More generally, if the utility function is non-decreasing (i.e. the larger the value of the metric, the better), U is the largest element of the metric for which all the lower values are uniformly judged as unacceptable. Likewise, P is the smallest element of the metric for which all the above values are uniformly perfectly satisfying. The same identification of the relevant levels U and P can be made for non-increasing utility functions and other types of monotonicity.

We identify the two elements U and P as we have just described. These are the first two elements of subset of the metric. Let us explain how to define the other elements of the subset. Firstly, the number of intermediate points between U and P basically corresponds to the complexity of the utility function. In general, it is between 2 (no intermediate point) and 7. For the identification of subset values, the segmentation is not necessarily uniform, depending if the shape is convex, concave or more or less linear. The idea is that the majority of points is placed where variations are the more important. For concave curves, we put more points close to U than to P. This is the case in the example of Figure 3. For convex curves, we put more points close to P than to U. The identification of the subset is easier when the metric is a finite set. In this case, the subset is often equal to the metric definition set.

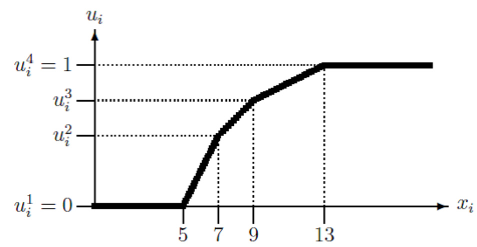

Fig. 3. Piecewise function caracterized by the values {5,7,9,13}

Construction of the utilities on the elements of metric subset. The stakeholder is asked to answer to types of questions regarding the utilities of the elements of the metric subset: **Ordinal information** Given two elements of the metric subset, what is the preference between these two elements? The stakeholder has the choice between three answers: (1) the first is strictly preferred to the second, (2) the first and the second are judged indifferent, (3) the second is strictly preferred to the first. **Cardinal information:** Given two elements for which the second is strictly preferred to the first, what is the satisfaction gain (attractiveness) when going from the first to the second? The answer shall be given within the following finite scale: **Unknown, Very Weak, Weak, Moderate, Strong, Very Strong, and Extreme**. Hesitation between several values can be expressed as an interval of values in the previous finite scale.

In the Figure 3, 13 is strictly preferred to 9, which is strictly preferred to 7, which is strictly preferred to 5. Moreover the cardinal information is gathered in the matrix I.

	13	9	7	5
13	-	Weak	Moderate	Very Strong
9		-	Weak	Strong
7			-	Moderate
5				-

TABLE I
EXAMPLE OF A CARDINAL INFORMATION MATRIX.

*3) **Stage 3:*** It consists in quantifying the evaluation tree on each aggregation node. One needs to aggregate the partial evaluations to obtain higher level evaluations. Considering for instance an aggregation of three criteria, this amounts to know whether the satisfaction attached to an alternative that is for instance good over the first criterion, fair on the second criterion and bad on the last criterion, is considered as rather good, or rather bad. It is likely that the overall satisfaction will equal some value in between. A trade-off or compromise shall be made amongst all the criteria used to compute the aggregation node. This is obtained with the help of compensatory aggregation functions. People use here most often the weighted sum. At the end of this stage, the evaluation model is thoroughly specified.

We are interested here in the construction of the aggregation function at a node. Basically, there are four types of methods to learn the parameters of aggregation methods:

Direct assessment of the parameters. The stakeholders directly assign numerical values to the parameters of the decision model. This approach is not satisfactory. On one hand, the semantics of the weights is not so clearly understood by humans. The concept of weight is used in different methods (weighted sum, weighted ordered average, weighted minimum, weighted majority,...), and the weights have different meaning in each method. On the other hand, for a weighted sum, weights do not make sense in the criteria are not commensurate.

The notion of importance can be rigorously defined once commensurate scales are defined on the metrics. For instance, a criterion is twice as much as important as another criterion if an increase of one unit on the first criterion is equivalent to an increase of two units on the second criterion. This unit or standard corresponds to the commensurateness assumption. In order to identify the precise relative importance ratio between two criteria, one shall ask the stakeholder to identify tradeoffs between these two criteria. Thus, it is delicate to ask directly the weights to the stakeholders. A learning phase by indirect questioning is preferable by large.

Expression of the preferences as a language. In Artificial Intelligence, it is usual to describe all possible models as well-formed formulas defined from a language. This allows a compact representation of the preferences. The idea of these approaches is that the stakeholder can then directly express his preferences in this language. This approach is not possible with the model that we consider.

Elicitation of the parameters from learning examples provided by the stakeholder. A very commonly used approach is to learn the decision model parameters from a set of learning examples provided by the stakeholders. Such examples are typically comparisons of alternatives or assessment of alternatives, which values of all attributes or criteria are known. From the learning data, one can analyze the potential inconsistencies or the incompatibilities with the model that is considered and analyze the completeness of the learning data.The difficulty of this approach arises when the previous set of compatible parameters is large:it is not easy to determine the most relevant learning examples that shall be given in order to reduce the size of the set of compatible parameters as much as possible. This difficulty yields to the last approach.

Elicitation of the parameters from learning examples constructed by the approach. In this approach, the parameters of the decision model are also constructed from a set of learning examples. But instead of asking the stakeholders to provide them, the idea is to construct a set of alternatives from which questions will be asked to the stakeholders. These alternatives are optimally determined so as to maximize the accuracy of the identification of the model parameters. This has some links with some statistical methods such as experiment design, or active machine learning.

The corresponding elicitation process is explained hereafter.

Identification of the relevant alternatives: the binary alternatives. The preferential information that can consider here is a generalization of what is used in the MACBETH approach. The MACBETH approach is dedicated to a weighted sum. For a weighted sum, the weight of a criterion represents its sole importance in the aggregation. We wish to generalize this to a 2-additive Choquet integral. On top of representing the importance of criteria, it can also model interaction between pairs of criteria. Instead of allowing the options to be perfectly satisfactory on one attribute only, one may allow the options to be perfectly satisfactory on two attributes at the same time. These alternatives are called binary as they can take only two values on the different criteria.

Construction of the parameters from the binary alternatives. The stakeholder is asked to provide some ordinal and cardinal information on the binary alternative. The ordinal and cardinal information are of the same nature as presented before. More precisely, the stakeholders are asked to answer questions regarding the utilities of the elements of binary alternative. **Ordinal information:** Given two elements x and y of the binary alternative, what is the preference between these two elements? The stakeholder has the choice between three answers: (1) x is strictly preferred to y, (2) x and y are judged indifferent, (3) y is strictly preferred to x. **Cardinal information:** Given two elements x and y of the binary alternative for which x is strictly preferred to y, what is the satisfaction gain (attractiveness) when going from y to x? The answer shall be given within the following same finite scale than for the utility function with the same pratice when hesitation occurs.

The analysis of the inconsistencies is much more complex than in the utility construction. This is due to the monotony conditions for the 2-additive Choquet integral. The description of the handling of the inconsistency is not the purpose of this document. One can see that the number of elements of binary alternatives increases with the square of the criteria at the level.

IV. Industrial case-study

The design of complex systems such as radio communication products requires taking into account various and sometimes contradictory concerns such as security and performance. Indeed, radio communication equipment exhibits strong requirements in terms of size, weight, power consumption, security and real-time performance. One of the most challenging aspects in system engineering is to analyze the combination of numerous concerns.

A. *Secure Radio Architecture.*

A secure radio platform is basically divided into three parts: The **Red security domain** receives sensitive information from the user point of view (data plan) such as plain text data that need to be ciphered; The **Black security domain** deals with nonsensitive information that are ciphered for data information and may be ciphered or not for control information; An **Information security domain** (InfoSec) handles communications between Red and Black domains. It ciphers data information from Red to Black domain and deciphers them from Black to Red domain using cryptographic channels. Control information may go between Red and Black domains without ciphering using bypass channels. For strong security and safety needs, a physical separation is enforced for the Red, Black and InfoSec domains. Each domain is implemented by a dedicated board in the radio equipment and has its own independent processor. The introduction of multi-core processors, hypervisor and separation kernel technologies in embedded systems allows a new security/safety architecture with a logical separation between the Red, Black and InfoSec domains. Basically, each domain may be implemented on a single multi-core processor. Multiple processors may be replaced by a single multi-core processor at lower frequency. This reduces power consumption as it roughly grows linearly with the processor frequency and the number of processors.

B. *Hardware/Software Architecture.[6]*

A radio platform is the set of software and hardware layers that provide the services required by the Software Radio Protocol (SRP) application layer through Application Programming Interfaces (APIs). A radio platform includes system components: Radio Devices (RD) (e.g. Ethernet Device, Audio device) and others Services (e.g. management service, IP and routing service). The SRP application and Software-Defined Radio (SDR) platform components may be designed for different security/safety levels (e.g. Common Criteria (CC) for security and/or DO178 for safety). Figure 4 presents the SRP application high level architecture.

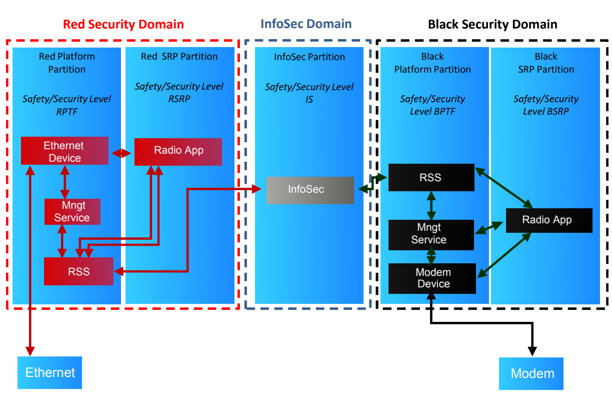

Fig. 4. SRP application high-level architecture

In addition to the SRP application components (Red and Black Radio App), the use case architecture consists of the following SDR platform components: **the Ethernet Device** abstracts an Ethernet network interface of the target SDR platform, **the Management Service** checks and dispatches control and management requests to SRP and platform components. For instance, it allows the configuration of component properties such as the MAC address or the transmission power of the radio equipment, **the Radio Security Service** (RSS) provides security channels to cipher/decipher user information, and forward control information without encryption (bypass), **the Modem Device** abstracts the Physical layer implemented on DSP, FPGA and the Radio Frequency (RF) front-end(s).

C. *The experiment.*

This experiment focus evaluations on the SRP sub-system of the SDR. Based on the same logical model, different deployments of the component-based radio application may be compared according to various criteria. Each security/safety partition may be physically isolated in boards or logically isolated in virtual machines (VMs). As described above there are three security domains: Red, InfoSec and Black domains, and five security/safety partitions for Red Application components, Red Platform components, InfoSec components, Black Application components and Black Platform components. Each domain and partition may have different Saftey/security level depending on the final product target application. Considering the number of boards and VMs reliable variants have been automatically derived from a variability model connected to business architecture patterns described in [11]. These reliable variants are presented in the Table II.

Solution	Description	Board(s)	VM(s)
S1	1 board per safety partition	5	0
S2	1 board per security domain	3	0
S3	1 board per sec. dom. and 1 VM per saf. part.	3	5
S4	1 board and 1 VM per security domain	1	3
S5	1 board and 1 VM per safety partition	1	5

TABLE II
CANDIDATE ARCHITECTURE VARIANTS.

D. *Evaluation Criteria.*

A solution may be applicable to a specific usage in the SDR productline. The objective is selecting the best design for a hand-held SDR (low power, safe and secure, high availability). The identified criteria are listed in the Table III.

Id	Criteria	Not satisfying at all value	Budget value	Completely satisfactory value
1	Software part. for Security	0	3	3
2	Hardware part. for Security	1	3	3
3	Software part. for Safety	0	3	3
4	Hardware part. for Safety	1	3	3
5	Used CPU Resource ratio	0.55	0.50	0.25
6	SoC Lifetime (h)	50000	60000	90000
7	Communication overhead (us)	1500	1000	200
8	Maintenance period (h)	90000	70000	60000
9	Cost per equipment (Euro)	1000	600	200
10	Power consumption (mWh)	2000	1700	1000

TABLE III
EVALUATION CRITERIA.

The Criteria are aggregated into several Aggregations mapping the principal concerns of the architecture description. Such mapping is often questionable and requires experts consensus. The resulting Aggregations are the following:

- **Security** aggregates Criteria 1 and 2;
- **Safety** aggregates Criteria 3 and 4;
- **Availability** aggregates Criteria 5 and 6;
- **RoI** aggregated Criteria 8 and 9;
- Criteria 7 and 10 remain untouched.

Aggregations and Criteria are aggregated under a single Aggregation, the model root, representing the overall assessment. The resulting model overview is represented in Figure 5.

Experts expressed the constraint that a solution presenting a top-level Aggregation or Criterion evaluated as "Not satisfying at all" must be evaluated under "Budget". This constraint has been translated by automated learning into a nearly global complementarity between these Aggregations and Criteria weights; the complementarity between criteria is interpreted as the min of evaluation of theses criteria by the 2-additive Choquet integral. The resulting relative weight schema is represented in the Figure 6.

V. Results and conclusion

For concrete understanding of the proposed approach value we compare results of the proposed approach with the usual weighted sum approach, both are using the same aggregation model and utility functions. The expressiveness of weighted sum make us adapt manually the relative weight of top-level Aggregations or Criteria as described in the Figure 7.

Results are compiled in the Table IV[2]. The comparison criteria is the overall score interpreted as a utility function, the higher is the better: a 0 score is interpreted as "Not satisfactory at all", a 0.5 score is interpreted as "Budget" and a 1 score is interpreted a "Completely satisfactory". The objective is then selecting solutions evaluated above 0.5 and, in case of multiple selection, choose the one with the highest mark.

Solution	Weighted sum @1000Mhz	Myriad @1000Mhz	Weighted sum @800Mhz	Myriad @800Mhz
S1	0.41	0.24	0.43	0.29
S2	0.54	0.36	0.56	0.43
S3	**0.62**	0.32	0.64	0.39
S4	0.59	**0.42**	**0.74**	**0.78**
S5	0.56	0.39	0.72	0.76

TABLE IV
OVERALL EVALUATION RESULTS.

A. Results

At 1000 Mhz Weighted sum proposes four satisfying solutions meanwhile Myriad-based evaluation cannot find any satisfying. Looking into details, Power consumption is "not satisfying at all" in all cases. Good scores on other criteria are compensating Power. The Myriad-based evaluation make us conclude there is **no satisfying solution**. This conclusion is operationally valid for the target usage. Because Power is outside acceptable range while CPU Resource exceeds the expectations, lowering SoC frequency may improve Power evaluation while keeping CPU Resource satisfying. At 800 Mhz Weighted sum evaluation proposes four satisfying solutions. Other criteria scores balance the unacceptable score of Power for S2 and S3 meanwhile S4 and S5 are identified as satisfying solutions. Myriad-based evaluation exclude all solutions other than S4 and S5. These solutions are operationally acceptable, both methods agree on this.

[2]Full evaluation models and results are available on demand.

B. Comparison justification

When deciding, one has to justify the choice. Myriad generates an argumentation report justifying evaluation and comparison towards the evaluation model. As example, the following is the raw result generated from the Myriad evaluations comparison at 800 Mhz, focusing on S4.

S4 is clearly preferred to S1 on the criterion "SDR Overall Assessment". **"S4" is preferred to "S1"** since the intensity of preference of "S4" over "S1" on the criteria "Communication overhead", "RoI" and "Power" is **MUCH LARGER** than the intensity of preference of "S1" over "S4" on criterion "Availability".

S4 is a bit preferred to S2 on the criterion "SDR Overall Assessment". **"S4" is preferred to "S2"** since **the large importance of the criteria "RoI" and "Power" reinforces** the relative strength of "S4" compared to "S2" on these criteria, **the small importance of criterion "Availability" minimizes** the relative strength of "S4" compared to "S2" on this criterion.

S4 is preferred to S3 on the criterion "SDR Overall Assessment". **"S4" is preferred to "S3"** since the intensity of preference of "S4" over "S3" on the criteria "Communication overhead", "RoI" and "Power" is **MUCH LARGER** than the intensity of preference of "S3" over "S4" on criterion "Availability".

S4 is almost similar to S5 on the criterion "SDR Overall Assessment". **"S4" is preferred to "S5"** since the intensity of preference of "S4" over "S5" on the criteria "Communication overhead" and "RoI" is **MUCH LARGER** than the intensity of preference of "S5" over "S4" on nothing.

Despite its automatic syntax, the generated argumentation helps in producing justification report, producing a complete argumentation of the evaluation for each Aggregation.

C. Conclusion

In this paper we illustrated the use of a tooled method for comparing evaluations of different solutions to a given problem in an objective way. Evaluating is always a irrational activity. The preference model synthetizes experts know-how. It is built by a method attempting to bring maximum rationality: utility and weights are computed by automated learning based on experts decisions. An often used approach consists in changing the criteria weights during the decision process in order to reach the expected alternative. This is of course very debatable. The methodology proposed represents better the decision maker preferences than when fixing directly utilities and weights. A preference model addresses a mean of answering to a question, not the definition of truth. If the radio was designed for being embedded in a vehicle, the preference model, although having the same aggregation structure, would have different utilities and weight for criteria such as Power consumption or CPU resource usage. The results would have been different.

The presented tooled method gives an evaluation thanks to a preference model, with evaluation result justifications. Deciding implies responsibility and for this reason remains

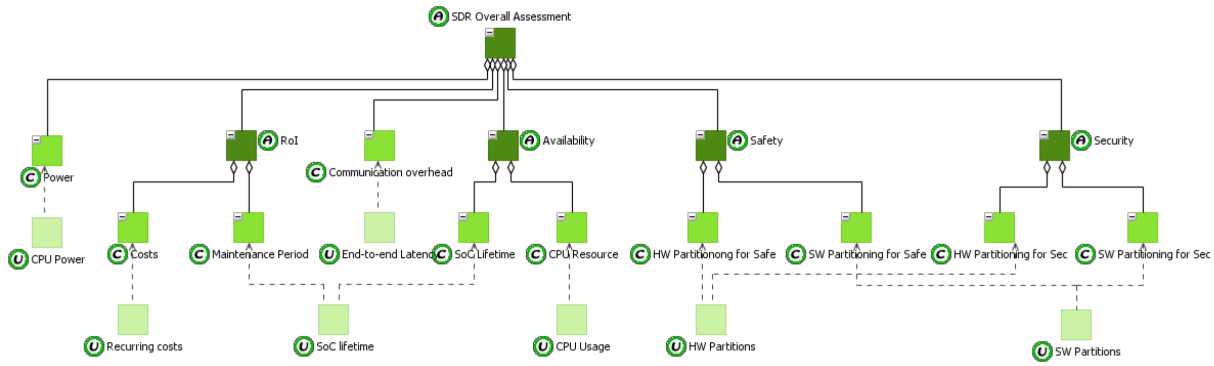

Fig. 5. MCDA model

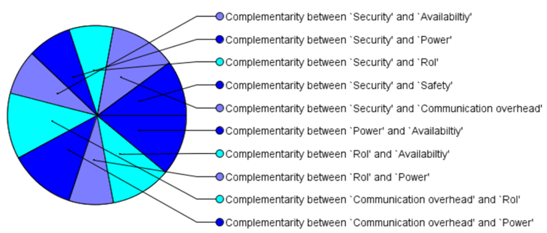

Fig. 6. Aggregation and Criteria relative weights, MYRIAD method.

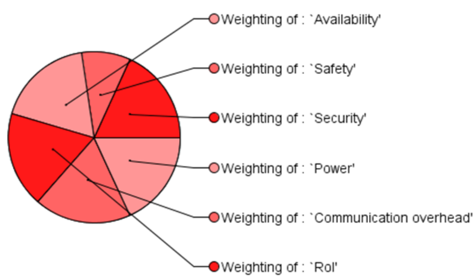

Fig. 7. Aggregation and Criteria relative weights, Weighted sum method.

an expert activity. The role of tooling is only to help decision maker in his task.

According to the experiment both evaluation methods highlight the best solutions. The Weighted sum proposes false positives because it requires independent variables. This hypothesis is not true in our case: we need to model "if one of Criteria is Not Satisfying At All then Evaluation is under Budget". Giving artificially strong weights to such criteria does not work here because of compensation. The Weighted sum is not adapted to decision making problems such as choosing a design for a given usage because one of application hypothesis is not satisfied. Myriad uses the Choquet integral for aggregating criteria satisfaction. It acts as a Weighted sum when variables are independent and manages variables interaction. It is an adapted tool to the problem class illustrated.

The presented tooled method requires the capability of sorting criteria, not possible when the criteria number is high and/or preference sorting is not possible. In this case one needs a new preference model and aggregation function class such as Generalized Additive (GAI) model[18].

References

[1] http://www.esteco.com
[2] http://www.simul8.com
[3] http://visadecisions.com/
[4] http://www-05.ibm.com/de/events/embedded-world/WP-Smarter-system-development.pdf
[5] Defense acquisition guidebook. https://dag.dau.mil/
[6] International Radio Security Services API Specification. WINNF-09-S-0011. V2.0.0
[7] Manual for the operation of the joint capabilities integration and development system
[8] Biltgen, P., Ender, T., Cole, B., Mavris, D., Biltgen, P., Mavris, D., Molter, G.: Development of a collaborative capability-based tradeoff environment for complex system architectures. (2006)
[9] Biltgen, P., Mavris, D., Molter, G.: A methodology for technology evaluation and capability tradeoff for complex system architectures
[10] Bana e Costa, C., Vansnick, J.: A theoretical framework for Measuring Attractiveness by a Categorical Based Evaluation TecHnique (MACBETH). (1994)
[11] Degueule, T., Filho, J.F., Barais, O., Acher, M., Noir, J.L., Madelénat, S., Gailliard, G., Burlot, G., Constant, O.: Tooling support for variability and architectural patterns in systems engineering (2015)
[12] Gerd Gigerenzer, W.G.: Heuristic decision making. Annual Review of Psychology 62, 451–482 (2011)
[13] Kazman, R.; Abowd, G.B.L.W.M.: Saam: A method for analyzing the properties of software architectures
[14] Kazman, R.; Abowd, G.B.L.W.M.: Software Reliability Theory. Encyclopedia of Software Engineering (1994)
[15] Kazman, R.; Klein, M.C.P.: Atam: Method for architecture software engineering institute (2000), http://www.sei.cmu.edu/publications/documents/00.reports/00tr004.html
[16] Klein, M., Ralya, T., Pollak, B., Obenza, R., Gonzales Harbour, M.: A Practitioner's Handbook for Real-Time Analysis (1993)
[17] Labreuche, C., Le Huédé, F.: Myriad: a tool suite for MCDA pp. 204–209 (September 7-9 2005)
[18] Labreuche, C., Grabisch, M.: A comparison of the GAI model and the choquet integral w.r.t. a k-ary capacity pp. 54–65 (2015), http://dx.doi.org/10.1007/978-3-319-23240-9_5
[19] Miller, G.: The magical number seven, plus or minus two: Some limits on our capacity for processing information. The Psychological Review 63, 81–97 (1956)
[20] Montmain, J., Labreuche, C., Imoussaten, A., Trousset, F.: Multi-criteria improvement of complex systems. Information Sciences 291, 61–84 (2015)
[21] Moody, K.: Issues and recommendations for instituting capability-based planning in the canadian forces (2005), http://www.cfc.forces.gc.ca/papers/csc/csc31/mds/moody.doc
[22] Roy, B.: Decision aiding today: what should we expect? (1999)