Risk Themes Discovered Through Architecture Evaluations

Len Bass
Robert Nord
William Wood
David Zubrow

September 2006

TECHNICAL REPORT
CMU/SEI-2006-TR-012
ESC-TR-2006-012

Carnegie Mellon
Software Engineering Institute

Pittsburgh, PA 15213-3890

Risk Themes Discovered Through Architecture Evaluations

CMU/SEI-2006-TR-012
ESC-TR-2006-012

Len Bass
Robert Nord
William Wood
David Zubrow

September 2006

Software Architecture Technology Initiative

This report was prepared for the

SEI Administrative Agent
ESC/XPK
5 Eglin Street
Hanscom AFB, MA 01731-2100

The ideas and findings in this report should not be construed as an official DoD position. It is published in the interest of scientific and technical information exchange.

This work is sponsored by the U.S. Department of Defense. The Software Engineering Institute is a federally funded research and development center sponsored by the U.S. Department of Defense.

This work was created in the performance of Federal Government Contract Number FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center. The Government of the United States has a royalty-free government-purpose license to use, duplicate, or disclose the work, in whole or in part and in any manner, and to have or permit others to do so, for government purposes pursuant to the copyright license under the clause at 252.227-7013.

For information about purchasing paper copies of SEI reports, please visit the publications portion of our Web site (http://www.sei.cmu.edu/publications/pubweb.html).

Table of Contents

List of Figures

List of Tables

Acknowledgments

Len Bass visited the National Institute of Computer Technology, Australia (NICTA) while working on parts of this report and would like to thank that organization for hosting him.

Abstract

This technical report analyzes the output of 18 evaluations conducted using the Architecture Tradeoff Analysis Method® (ATAM®) developed by the Carnegie Mellon® Software Engineering Institute. The goal of this analysis was to find patterns in the risk themes identified during those evaluations. The major results are

- a categorization of risk themes
- the observation that twice as many risk themes are risks of "omission" as are risks of "commission"
- a failure to find a relationship between the business/mission goals of a system and the risk themes revealed during an ATAM evaluation of that system
- a failure to find a relationship between the domain of a system being evaluated and the risk themes associated with the development of that system

1 Introduction

The Architecture Tradeoff Analysis Method® (ATAM®) is a method for evaluating software architectures relative to quality attribute goals [Clements 02]. The ATAM, which was developed by the Carnegie Mellon® Software Engineering Institute (SEI), exposes architectural risks that potentially inhibit the achievement of an organization's business and mission goals. The SEI has been doing ATAM evaluations since 1998 and distilling the risks into risk themes since 2000. Risk themes are a summarization and consolidation of the collection of risks found during an evaluation. These themes cover continuously emerging risks that appear repeatedly in the total collection of risks, sensitivities, and tradeoffs, and they have a direct impact on the business drivers and the software architecture. Most evaluations produce an Architecture Evaluation Report as part of their output.

We analyzed 18 final reports dated between 2000 and 2005, and this paper presents the results of that analysis. These ATAM evaluations produced 99 risk themes. Twelve of the systems are for the U.S. Department of Defense, two are for another government agency, and the other four are for commercial organizations. The domains involved range from information systems to embedded systems.

You might assume that there is a connection between articulated quality goals and risk themes. That is, if performance is an explicit goal in the development of the system, there should be either more performance risk themes (because it is more important) or fewer performance risk themes (because more attention is paid to performance). You might also assume that there is a connection between the domain of investigation and the risk themes. For example, systems in the avionics domain might exhibit similar risks. However, the data from our analysis supports neither conclusion: there is no correlation between systems with performance goals and those with performance risk themes, and there is no discernible pattern of risk themes associated with any particular domain.

The data does support the observation that most risk themes discovered during an evaluation cover risks that arise from the lack of an activity rather than the incorrect performance of it.

In summary, the major results of this report are

- a categorization of risk themes

® Architecture Tradeoff Analysis Method, ATAM, and Carnegie Mellon are registered in the U.S. Patent and Trademark Office by Carnegie Mellon University.

- the observation that twice as many risk themes are risks of "omission" as are risks of "commission." That is, the risk themes identify decisions or investigations that were never made rather than those that were made and could lead to undesirable consequences.
- no discernable relationship between the articulated business and mission goals of a system and the risk themes from an ATAM evaluation of that system
- no discernable relationship between the domain of a system being evaluated and the risk themes associated with the development of that system

Our report is organized as follows:

- Section 2 describes the methodology we used to generate categories of risk themes.
- Section 3 lists and describes the final categories we created. It also describes the risk theme categories that were most prevalent in the evaluations we reviewed.
- Section 4 looks at the relationship between risk themes and
  - a system's business and mission goals
  - the system domain
- Section 5 defines risks of omission and risks of commission and provides our results for this categorization.
- Section 6 discusses the implications of our findings on researchers, practitioners, and those who perform architectural evaluations.

2 Methodology for Developing Risk Theme Categories

The *affinity diagram* was originally developed by Kawakita, an anthropologist, to discover meaningful groups of ideas from a raw list [Beyer 98]. Kawakita's idea was to examine the list and let groupings emerge naturally, using the intuition of the analysts, rather than following a preordained categorization. An affinity diagram allows for categories that are not mutually exclusive.

The steps to creating an affinity diagram are as follows:

1. Assemble the team.
   Generating an affinity diagram is typically a team activity that relies on multiple viewpoints and ideas.
2. Write individual statements on note cards or Post-it notes, and give each statement a unique ID number.
   These statements may come from interviews, documents, surveys, brainstorming, or any other source.
3. Group the statements.
   There is no right or wrong way to do this activity, but the groupings should not follow any predetermined categorization. The categories should emerge from the statements and from the team. If you want to put a statement in more than one group, simply write it on multiple note cards.
4. Name each group.
   The name chosen should represent the basic idea shared by all the statements in the group.
5. Cluster the groups.
   Typically, you'll have a large number of groups. All groups have a natural affinity with other groups.
6. Name each cluster.
   The name chosen should represent all the groups in the cluster.

Our three main groups were architecture, process, and organization. Architecture is further refined into runtime and development time qualities. These top-level groups are used only for organizing the other groups and were not used in our analysis. It is worth noting that our top-level categories are three of the four product family concerns discussed by van der Linden [van der Linden 02]: Business, Architecture, Process, and Organization—commonly re-

ferred to as BAPO. Since the ATAM uses business strategy as a criterion to identify risks (and hence risk themes), it is not surprising that no risk themes fall into a business strategy group.

For our set of 99 risk themes, we completed Steps 1 and 2. We then iterated through Steps 3, 4, and 5 three times to confirm that the groupings were stable and meaningful.

Note that what we finally derived is a categorization, not a taxonomy. It is permissible and even likely that the risk themes from a particular ATAM evaluation could be placed into multiple categories. The important thing is that the risk themes do, in fact, have at least one category in which they can appear. In this aspect, we followed the spirit of the organization structure for quality attribute scenarios in which a particular concrete scenario can be an instance of several different general scenarios, possibly deriving from different quality attributes [Bass 03]. This emphasis on categorization, rather than taxonomy, offers an important benefit: organizations attempting to use our categories while performing their own ATAM evaluations don't have to argue about where to put risk themes that could belong to several categories.

3 Risk Themes Categories

As shown in Figure 1, 15 final categories—the leaf nodes shown—emerged when we applied the Affinity Diagram process to the 99 unique risk themes. Twenty-one risk themes were placed in two categories, and one risk theme was placed in three categories. The categories should be read as "risk themes associated with …" where the ellipsis (…) is the title of the category.

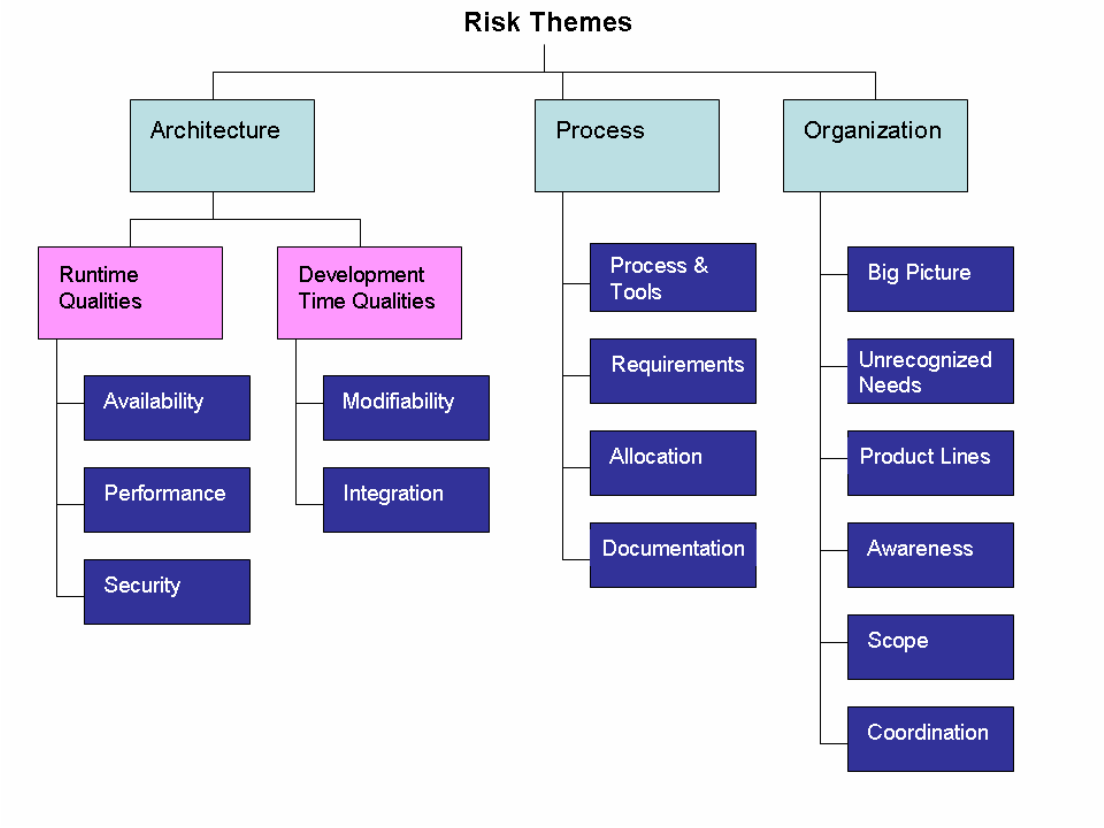

Figure 1: Risk Theme Categories

3.1 Category Descriptions

Below, we describe the 15 final categories and some of the issues they involved. We cannot, for confidentiality reasons, provide the actual risk themes; however, including the issues should give you a sense of the risk themes we placed in each category.

3.1.1 Architecture

Runtime Qualities

Availability: These risk themes mention risks to availability or reliability goals. Issues that arose in this category included

- having a single point of failure
- not including availability mechanisms
- using infrastructure that does not support availability mechanisms

Performance: These risk themes mention problems with achieving performance goals. Issues that arose in this category included

- not knowing performance requirements
- not performing any performance modeling or prototyping
- unfamiliarity with infrastructure choices
- not using known performance mechanisms

Security: These risk themes mention problems with achieving security goals. Issues that arose in this category included

- unknown requirements
- not using known mechanisms to support security goals

Development Time Qualities

Modifiability: These risk themes mention problems with achieving modifiability goals. Issues that arose in this category included

- allocating functionality in a way that jeopardizes portability
- supporting the addition and deletion of different devices
- lack of attention to potential growth paths
- unknown requirements

Integration: These risk themes mention problems associated with integrating various portions of the system. Issues that arose in this category included

- problems with migrating legacy systems
- not using known integration mechanisms
- lack of uniformity in key areas

3.1.2 Process

Process and Tools: These risk themes mention problems with either the development process or the availability of tool support. Issues that arose in this category included

- viewing tools as a solution to particular problems without considering, in the project plan, the resources necessary for tool construction
- relying on untested processes or tools

Requirements: These risk themes refer to problems caused by either uncertainty over requirements or by rapidly changing requirements. Issues that arose in this category included

- lack of attention to important concerns of key stakeholders
- lack of consistent marketing input
- emerging requirements
- disagreement among the stakeholders as to the use of the system
- unclear requirements in certain areas

Allocation: These risk themes refer to problems allocating functionality to system elements. Issues that arose in this category included

- the affect of the allocation on reuse and portability
- management of distribution
- achieving quality of service

Documentation: These risk themes refer to problems resulting from the quality of the documentation. Issues that arose in this category included

- defects in existing architecture diagrams
- lack of documentation for high-priority scenarios
- inconsistency among different views

3.1.3 Organization

Big Picture: These risk themes refer to problems arising from the lack of an overall system perspective. Issues that arose in this category included

- considering applications and infrastructure independently and not paying sufficient attention to their interaction
- lack of any system modeling activities or views to support them
- exclusive focus on functional issues with an associated lack of attention to quality of service issues

Unrecognized Needs: These risk themes refer to problems arising from the failure to consider some important aspect of the architecture necessary for successful system construction. Issues that arose in this category included

- too many uncertainties (which will threaten the project schedule)
- no overall consideration of many issues

- no business goal that speaks to many of the activities the development team must consider
- unknown requirements for quality of service

Product Lines: These risk themes refer to problems associated with the implementation of product lines. Issues that arose in this category included

- the tension between satisfying all the customers' desires and maintaining the integrity of the product line
- the tension between configurability and the management of that configurability
- untested variability mechanisms
- unknown requirements for various market segments
- lack of training and tool support
- lack of explicit definition of commonalities and variabilities

Awareness: These risk themes refer to problems associated with a lack of awareness of the activities needed to fully implement and support an architecture. Issues that arose in this category included

- developer training
- the lack of necessary tools
- the lack of guidelines about which mechanisms to use in which contexts
- the lack of planning for interoperability requirements
- the lack of ability to predict properties of the software
- the lack of coordination between architecture teams and implementation teams

Scope: These risk themes refer to problems resulting from unrealistic project goals or the use of immature technology. Issues that arose in this category included

- complexity of the system increasing beyond manageable bounds
- the use of many new technologies
- an unprecedented system scale

Coordination: These risk themes refer to problems associated with a lack of communication between the development team of the system and other important stakeholders. Issues that arose in this category included

- inadequate coordination with external agencies and systems
- inadequate coordination with standards bodies

3.2 Comparison with Other Risk Categorizations

Our categorization was based on an intuitive bottom-up process. In this section, we compare our categories with two others: (1) one based on ATAM evaluations performed by Boeing [O'Connell 06] and (2) one based on a discussion of projects that have failed [Charette 05]. The basis for developing these two categorizations has not been published. As a result, our comparison is subjective in nature.

3.2.1 Categorization from Boeing ATAM Evaluations

O'Connell describes eight ATAM evaluations that Boeing has been involved in [O'Connell 06]. Those evaluations had a collection of risk themes categorized by intuition, although not using the Affinity Diagram process. Table 1 presents the Boeing categories and how they map into the categories we identified.

Table 1: How Boeing's Categories Map into Our Categories

Boeing Categories	**Our Category**
Performance, Scalability	Performance
System Management, Failures	Availability
Security, Information Assurance	Security
Product Line Planning, Lead System Integrator	Product Line
Changing or New Technologies	Modifiability
Legacy, Commercial Off-the-Shelf (COTS) Integration	Integration
New Operational Procedures	Awareness
Unknown or New Requirements	Requirements
Software Architecture Undefined	Big Picture
Architectural Tactics	Unrecognized Needs
Usability	not in our categories—would be classified under Runtime Qualities
Safety	not in our categories—would be classified under Runtime Qualities

The Boeing classification was developed independently from ours, yet the categories in each are very similar. This similarity provides some evidence that the categories (regardless of how they're named) are reasonable ones.

3.2.2 Failure Categories

Charette has examined the literature associated with project failures and enumerated a list of factors that caused them [Charette 05]. This list has a different basis than our list—it examines projects that have already failed (i.e., rather than examining projects under development

to uncover potential problems). Also, his list explicitly includes project aspects, whereas the risk themes that arise from ATAM evaluations are primarily those related to the software architecture. Even with these differences, there is great similarity between Charette's list and the categories we have developed. Table 2 provides the comparison.

Table 2: How Charette's List of Causes of Failure Maps to Our Categories

Charette's Failure Causes	**Our Categories**
unrealistic or unarticulated project goals	Scope
inaccurate estimates of needed resources	Scope
badly defined system requirements	Requirements
poor reporting of the project's status	not in our categories
unmanaged risks	Unrecognized Needs
poor communication among customers, developers, and users	Coordination
use of immature technology	Process and Tools
inability to handle the project's complexity	not in our categories
sloppy development practices	Process and Tools; Documentation
poor project management	Unrecognized Needs
stakeholder politics	not in our categories
commercial pressures	not in our categories

We also identified categories not in Charette's list such as Runtime Qualities, Development Time Qualities, Architectural Perspective, and Product Lines. He developed his list by looking at the reasons why projects fail, and we developed ours by looking at the architectural risks. As a result, the surprise is not that our lists are different but that they overlap in many respects. That is, the ATAM is a valuable tool for identifying some of the common reasons why projects fail. Of course, Charette points out that many of the failed projects he discusses were known to be in trouble and subject to many negative reviews. Still, it is reassuring that ATAM evaluations can discover many of the standard causes of project failure at an early stage of the project life cycle.

Note that the close relation between our categories and Boeing's shows that ours have industrial relevance and value; the comparison to Charette suggests that risk themes discovered in ATAM evaluations have some diagnostic and predictive value for general project performance.

3.3 Counts for the Risk Theme Categories

Next, we present how many ATAM evaluations fell into each of our risk theme categories. Recall that as a result of the Affinity Diagram process, we placed 99 distinct risk themes found in 18 ATAM evaluations into 1 or more categories; 21 risk themes fell into 2 categories, and 1 fell into 3 categories. An ATAM is in a particular category if it had a risk theme

that was placed in that category during the Affinity Diagram process. Figure 2 shows the specific numbers and categories.

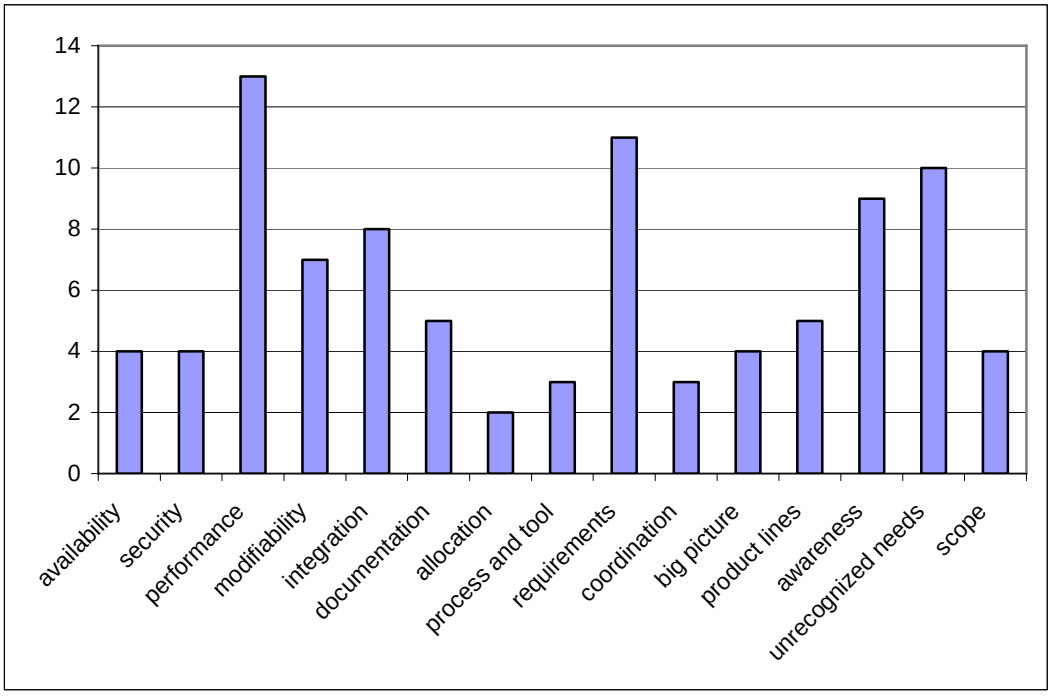

Figure 2: Number of ATAM Evaluations Occurring in Each Risk Theme Category

3.4 Factors that Might Relate to Risk Theme Categories

Figure 3 shows the inputs and activities in an architecture review [Dominick 02]. The inputs are candidate factors that might have a relationship to the risk themes. We use our data set to examine possible relations between these and other factors and the risk themes.

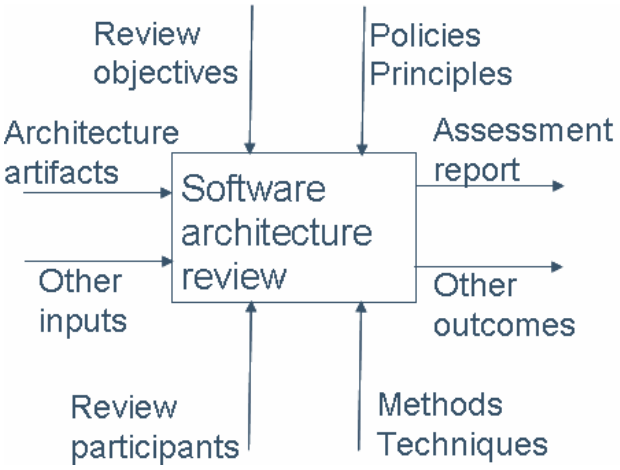

Figure 3: Activity View (in IDEF) of Software Architecture Review

Next, we discuss some of the possible types of relationships shown in this figure that might affect the risk themes.

- **architecture artifacts.** One input to an ATAM evaluation consists of a system description. Systems can be characterized in terms of the domain they fall into; their style and complexity; and their criticality, size, and dynamism [Boehm 03]. In this report, we provide one analysis in terms of domain. The underlying style might be a factor that enters into the risk themes that were discovered. To date, we have not investigated this idea further.
- **other inputs.** Inputs to an ATAM evaluation also include the business and mission goals. A conjecture is that the product-related risk themes (the ones we characterized under the "architecture" branch of our risk themes) are related to the business and mission goals for a system. We explore one such conjecture in Section 4.1.
- **review objectives.** The review objectives of an ATAM evaluation are expressed in terms of the business goals.
- **review participants.** Two types of review participants might be related to the risk themes:
  1. development team. There are, potentially, a wide variety of factors that affect the development team including
     - team members' skill set
     - the organization's level of process maturity
     - the organization's culture
     - whether team members are organic to the customer or work for a subcontractor
     - whether team members are local or distributed and, if distributed, whether they are distributed across one continent or across multiple continents

     We did not investigate the relation between the development team and risk themes in the work reported here.
  2. evaluation team. Different members of the evaluation team have different areas of expertise and, hence, are likely to look for and find risks in those areas. Each evaluation team consisted of at least four individuals. Analyzing the relation between the evaluation team and the risk themes discovered during ATAM evaluations would require more data than we have collected thus far.
- **methods and techniques.** The methods and techniques that were input for the ATAM evaluations were part of the ATAM itself and were not customized. Hence, their use cannot explain the variation in the risk themes.
- **policies and procedures.** An ATAM evaluation receives no specific inputs in terms of policies and procedures. As we will see, many of the risk themes are expressed in terms of organizational activities that might be a subject of policies or procedures. However, since an ATAM evaluation does not collect specific inputs in this area, we could not perform any analysis with a data set derived from the final reports of the ATAM evaluations.

Finally, the systems that the SEI is asked to evaluate are a small subset of the systems constructed each year and are not a random subset of all systems by any means. At least one of the stakeholders has to engage the SEI in order to perform the evaluation, and the system must be large or complex enough to justify its cost. It is possible that the systems we evaluate are those that are thought to have severe risks or are pushing the state of the art with respect to complexity and technology. We have no way of exploring this factor further because doing so requires data from ATAM evaluations performed by other organizations—data that we currently lack.

3.5 Most Prevalent Risk Theme Categories

The following four risk theme categories are exhibited by over 50% of the ATAM evaluations: (1) Performance, (2) Requirements, (3) Awareness, and (4) Unrecognized Needs. We discuss each one below briefly.

3.5.1 Performance

Possible reasons for the large number of ATAM evaluations that experienced risks with performance are

1. Performance is a pervasive property in every system whether it involves real-time deadlines, user response, or the number of transactions processed per minute. As such, it is a property that can be examined in every ATAM evaluation. Security and high availability, on the other hand, are only properties of interest in specific systems, so the number of ATAM evaluations in which they are of interest is smaller.
2. Much is known about performance, so it is relatively easy for evaluators to focus on it.
3. The evaluation team frequently included experts in performance who tended to look for problems in their own areas of expertise.

In a subsequent section, we explore the connection between performance as an articulated business goal and performance as a risk theme.

3.5.2 Requirements

“Badly defined system requirements” is one cause of system failure that Charette mentions [Charette 05]. It is also a cause of frustration for developers. Since the ATAM works by gathering input from developers, any frustrations they have regarding system requirements would be readily available to the evaluators. On the other hand, changing and volatile requirements are a fact of life in virtually every development effort. Many architectural techniques [Bass 03, Ch. 5] exist to allow for easy system modification which, in turn, helps manage volatile requirements—as long as those requirements are foreseeable during the design phase of the development. One area of deeper investigation is to determine whether the changing and volatile requirements that lead to risk themes in ATAM evaluations could be accommodated by standard architectural techniques.

3.5.3 Unrecognized Needs

The common aspect of these risk themes is that the developers failed to consider something that was important. Either they were falling behind in the schedule and cutting corners in terms of modeling or analysis, or they just overlooked something. All the development teams for the systems we evaluated were under time pressure. Unfortunately, the desire to omit some aspect of sound development practices under such circumstances is going to grow. Charette identifies “poor project management” and “sloppy development practices” as factors in software failures—both of which are exemplified by the failure to consider important factors. Another area of deeper investigation is to determine which development processes provide the best mechanisms for insuring that unrecognized needs are, in fact, recognized.

3.5.4 Organizational Awareness

The fact that a system lives within a business and organizational context is often overlooked by those responsible for developing budgets and schedules. The business and organizational context generates requirements for training, for coordination with stakeholders, and for responsible planning. The large number of ATAM evaluations that fell into this risk theme category suggests that these problems are widespread.

4 Predictors of Risk Themes

4.1 Business and Mission Goals as a Predictor

In this section, we describe the results of examining the articulated business and mission goals of a system with respect to the risk themes uncovered during an ATAM evaluation of that system. If they were related in some way, projects could begin risk mitigation strategies for particular types of risks at the project's inception rather than waiting until the risks emerge. In short, after focusing on the relation between performance as a business goal and performance as a risk theme, we found no such relation in this one case. A side effect of looking for relations statistically is that the chances of finding a relation increases with the number of relations examined. For example, if a relation is deemed significant at the .05 level (meaning that there is a 5% probability of being incorrect in an assessment), 20 such assessments will almost certainly result in a relation being found. That assumption led us to focus on the most likely source of such a relation—performance as an articulated business goal and performance as a discovered risk theme.

We begin by presenting a histogram of the business and mission goals articulated by our 18 ATAM evaluations (see Figure 4). The categories we use are those of Kazman and Bass [Kazman 05].

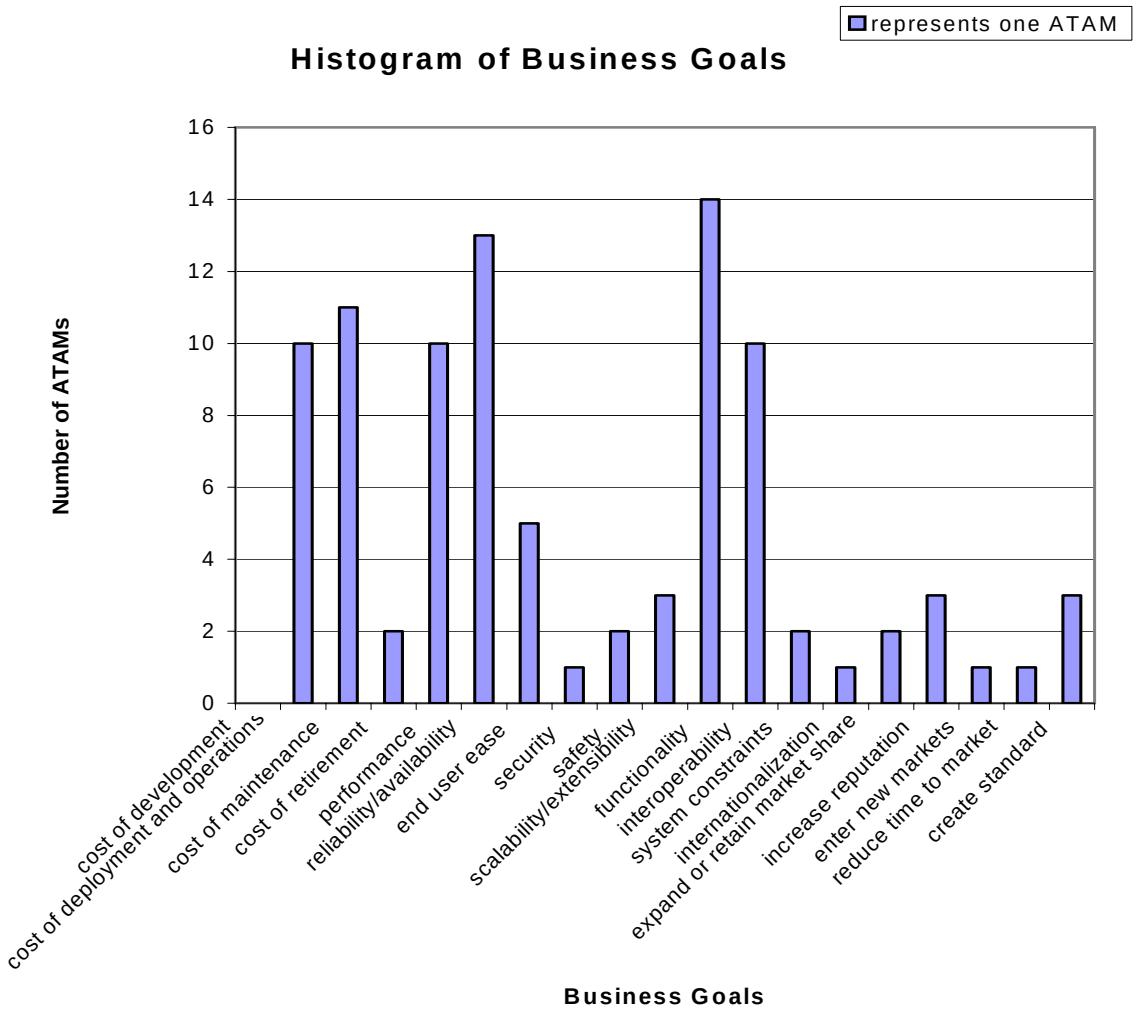

Figure 4: Business and Mission Goals Articulated in the ATAM Evaluation Data

As shown above, more than half of the ATAM evaluations expressed performance as an important business goal. Since more than half also had performance as a risk theme category, we wonder whether there is a relationship between the ATAM evaluations that have articulated performance as a business goal and those that have risk themes under the Performance category. Intuitively, an express business goal of performance might indicate more pressure on performance and, hence, result in more performance risk themes. Alternatively, having performance as an express business goal might mean that developers pay more attention to performance, resulting in fewer performance risk themes. The data support neither interpretation.

Table 3 shows a cross-tabulation of the ATAM evaluations based on the existence of a performance business goal and a performance risk theme. The interpretation of a cross-tabulation table is that the diagonals are the indicators of the correlation. If all the items are

on the diagonal downward to the right, there is a 1.0 correlation. If they are all on the diagonal upward to the right, there is a -1.0 correlation. In our case, the correlation of the ATAM evaluations based on the categories of Performance Business Goals and Performance Risk is .194. Thus, there is no correlation—either negative or positive—between the performance business goal and the performance risk themes for ATAM evaluations.

Table 3: Cross-Tabulation of ATAM Evaluation Data

		Business and Mission Goals	
		Not Performance	Performance
Risk Themes	Not performance	3	2
	Performance	5	8

4.2 Domain as a Predictor

Another possible predictor of risk themes is the domain of the system being evaluated. Table 4 lists the domains of the systems being evaluated and the number of evaluations in each domain. To determine the domain of the system being evaluated, we asked SEI staff members who were not involved in writing this report and who were familiar with the system to categorize it.

Table 4: Tally of ATAM Evaluations in Each Domain

Domain	Number of ATAM Evaluations
Avionics	3
Command, Control, Communications, Computer, Intelligence, Surveillance, and Reconnaissance (C4ISR)	1
Command and control	4
Command and intelligence	1
Distributed infrastructure	1
Embedded information systems	2
Embedded control systems	2
Information systems	1
Information, surveillance, reconnaissance	1
Mission computing	1
Modeling and simulation	1

Our goal in this analysis was to look for similarities in the risk theme patterns of ATAM evaluations from similar domains. That is, we asked ourselves whether the three evaluations in the avionics domain have risk themes in the same set of risk theme categories (or a similar set). We excluded all the domains that had only one ATAM evaluation and were left with four domains: (1) avionics, (2) command and control, (3) embedded information systems, and (4) embedded control systems. We first discuss how we visually represent the similarities among ATAM evaluations within a domain, and then we discuss a statistical test we applied to the domains. In both cases, the evidence is that there is no similarity among the risk themes of ATAM evaluations in any of the four domains.

Consider what it means for two ATAM evaluations to have their risk themes in the same set of risk theme categories. We can divide the 15 risk theme categories into three groups with respect to those two ATAM evaluations:

1. those that came up in zero ATAM evaluations
2. those that came up in one ATAM evaluation
3. those that came up in two ATAM evaluations

If two ATAM evaluations had identical categories of risk themes, we would expect most of the themes to fall into groups 1 and 3 and group 2 to be empty. This expectation generalizes to looking for similarities in *N* ATAM evaluations—that is, the majority of the themes would be associated with either zero or *N* of the evaluations, and none would be associated with only one. Even if the ATAM evaluations did not have identical categories of risk themes, we would expect spikes building toward *N* with another spike at zero.

Figure 5 contains four graphs that depict the risk theme category distribution in each domain mentioned above. Note that in each graph, the Y axis shows the number of risk themes, and

the X axis shows the number of ATAM evaluations. None of the graphs shows any tendency to bunch toward the right (i.e., the higher number of ATAM evaluations that exhibit common risk themes). From this visualization, we see no evidence that risk theme categories are similar for systems in a particular domain.

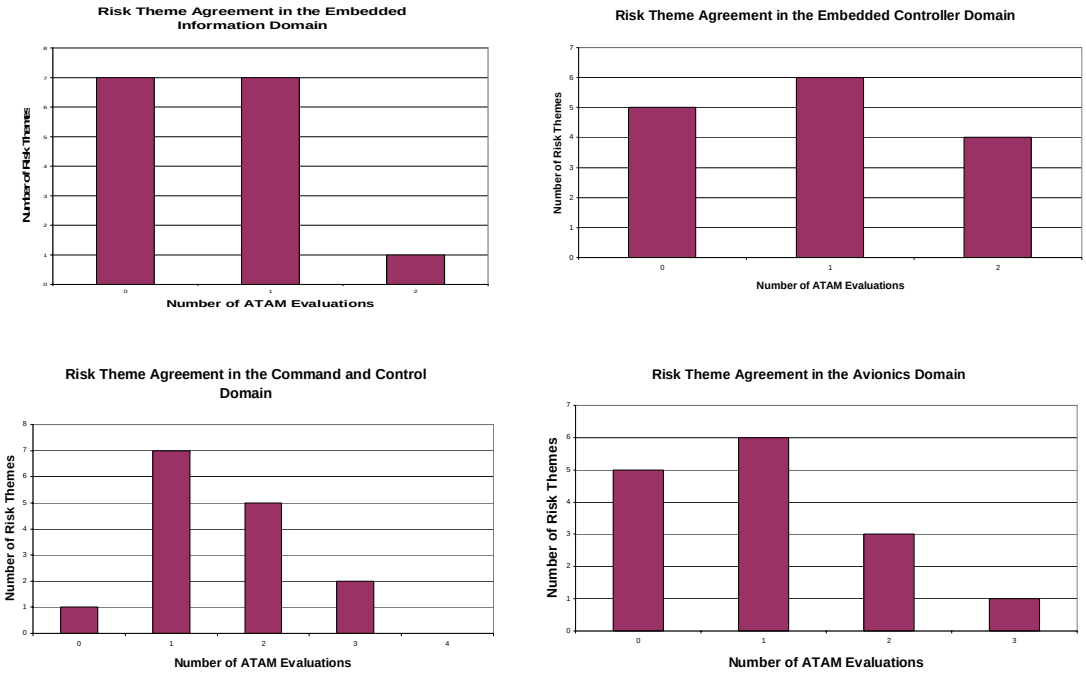

Figure 5: Visualization of Similarity of ATAM Evaluation Risk Themes

We also have a more formal, statistically oriented measure of similarity that yields similar results. We analyzed the data in terms of the percentage of ATAM evaluations within each domain that had at least one risk within the risk theme category across the 15 risk theme categories. For instance, if there were three ATAM evaluations in a domain and only one had a risk theme in the Availability category, the score for Availability would be .33. We interpret this value as percent agreement on the presence of the risk theme category within the domain. This computation was done across all 15 risk theme categories for each domain. Note that this formulation does not give "credit" for agreement in terms of the absence of a risk theme. This assumption was made because of the explicit focus on the presence of risk themes in this investigation.

Summarizing the percent agreement values across the risk categories yields an average percentage agreement that can be used to characterize the degree of similarity in the risk theme patterns for the ATAM evaluations within the domain.

For the four domains where we had two or more ATAM evaluations, these values ranged from .131 to .415. From a statistical viewpoint, some amount of agreement is likely to occur by chance. Therefore, various adjustments were made. One of the best-known adjustments is the Kappa statistic that is computed as follows:

$$K = \frac{P(a) - P(e)}{1 - P(e)}$$

where P(a) is the observed percent of agreement and P(e) is an estimate of agreement that would occur by chance.

Two alternatives were considered to estimate the chance agreement. First, in keeping with the traditional formulation of Kappa, we computed the average presence of a risk theme within the domain. This calculation is simply the average of the number of risk themes present considering all possible opportunities for a risk theme to be present (i.e., number of risk themes * number of ATAM evaluations). A second formulation of chance was to compute the percent agreement as described above for all the ATAM evaluations that were not in the focus domain and use this value as our estimate of the chance agreement.

We used various combinations of the above in our analysis. Some formulations are more conservative and others more liberal with regard to characterizing agreement—hence, the similarity of the risk theme patterns for ATAM evaluations within a domain. By using a combination of formulary methods, we believe the real value of agreement will be bounded by the agreement values resulting from the conservative and liberal formulations.

In the instances where a domain had only two ATAM evaluations, the traditional notions of agreement were used. However, we also computed the domain agreement by excluding the risk themes where neither ATAM evaluation had a risk in that category.

The following observations are based on the preceding analysis:

- None of the domains exhibit strong similarity in their risk theme profiles. This finding is true for both the unadjusted percent agreement (excluding instances where there is agreement on the absence of a risk theme) and the chance-corrected Kappa statistic.
- Focusing only on the total agreement for the presence of a risk theme, we found (for the 15 risk themes):
  - avionics (three ATAM evaluations): one risk theme
  - command and control (four ATAM evaluations): zero risk themes
  - embedded information system (two ATAM evaluations): one risk theme
  - embedded controllers (two ATAM evaluations): four risk themes

Based on this analysis, there is no evidence supporting an assertion that domains would exhibit a common risk theme category profile.

One of our conjectures is that the organizational context plays a part in the generation of risk themes. However, because we lack the necessary data on ATAM evaluations conducted for the same organization, we can’t pursue that supposition with our current data set.

5 Categorization into Risks of Commission and Risks of Omission

In addition to the categorization based on affinity diagrams, we divided the risk themes into categories based on the type of risk theme:

1. risks of commission: risk themes that result from problematical decisions within the architecture; for example

   > "The chosen operating system does not support the multi-processing and partitioning of memory that would prevent forbidden accesses. This lack of support requires the developers to create a concurrency system on top of the operating system. The higher those kinds of services are in the layered structure of a software system, the more complex they are to implement and the higher the performance penalty is. This situation exposes the platform to the risk of performance overruns and fault propagation."

   That is, the architects decided to use a particular operating system, and doing so generated a risk.

2. risks of omission: risk themes that result from not performing certain activities; for example

   > "Risks arise from the lack of an overarching architectural point of view that includes applications, application commonality, and the application framework."

   That is, the architects did not define an overall architecture, and that led to the risks enumerated.

3. neither commission nor omission: Some risk themes are neither risks of commission nor risks of omission, and some risk themes are ambiguous; for example

   > "The system under review was already a complex system when it 'only' had to meet one set of needs, with an expert staff responsible for developing, maintaining, and operating the system. However, the system must now, or in the near future, meet the demands of new, diverse user communities including test, integration, analysis, and training. To meet these demands, the system is growing in complexity."

   This risk theme is worded in a way that does not identify the particular risks being identified, so it is not clear whether it is a risk of omission or commission.

For our 99 risk themes, 25 were risks of commission, 57 were risks of omission, and 17 were neither. In general, it's unclear whether a particular categorization of items is repeatable. One measure widely used to check for repeatability is Cohen's Kappa—a method in which two

people place items into categories based on a documented set of rules. If the Kappa measure is greater than .7, the categorization is considered to be repeatable. Two independent categorizations of our 99 risk themes into the three categories above yielded a Cohen's Kappa of .82.

Using the same set of rules, Boeing categorized the 28 risk themes for its set of ATAM evaluations: 16 were risks of omission, 11 were risks of commission, and 1 was neither.

The ratios of risks of omission to risks of commission from the Boeing and SEI data sets were 1.5:1 (for Boeing) and 2.3:1 (for the SEI). In either case, the risks of omission are more numerous than the risks of commission.

Knowing that the majority of the risk themes are risks of omission is important to

- evaluators—since it indicates that they need to be vigilant in detecting those risks
- developers—since it indicates that they need to make sure all aspects of a project are considered during development

Figure 6 displays the SEI risks of commission and omission as a percentage of the categories defined by the affinity diagram.

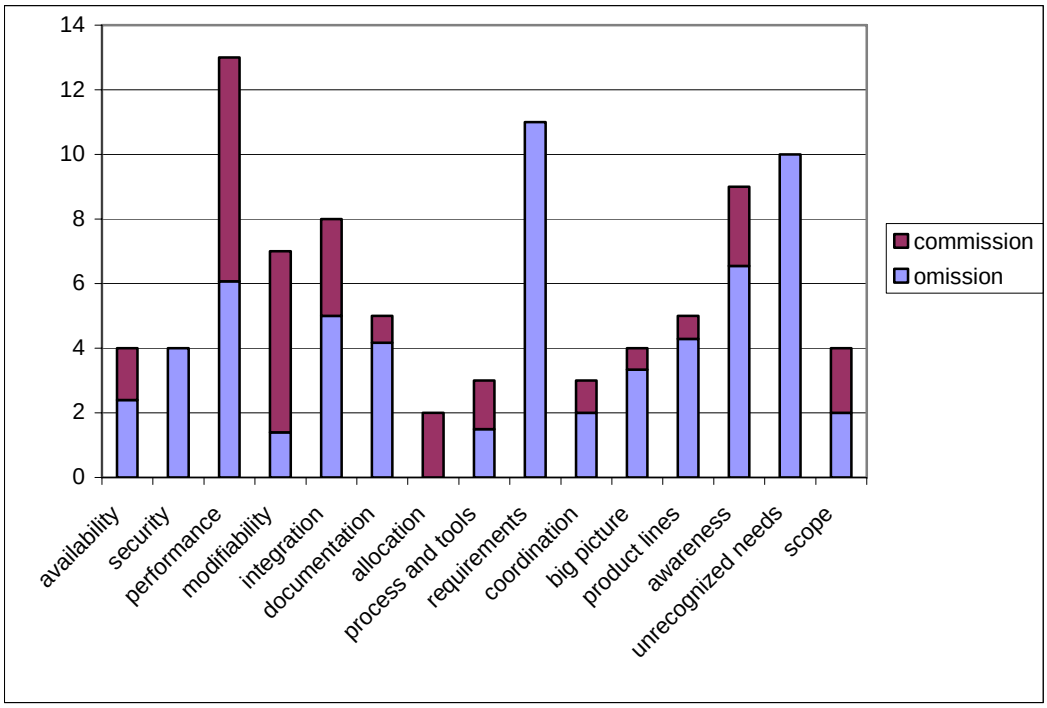

Figure 6: Risks of Omission and Commission Overlaid on an Affinity-Diagram-Based Categorization (as a Percentage)

Figure 6 shows that four risk theme categories (Security, Requirements, Allocation, and Unrecognized Needs) contained either all omission risks or all commission risks. Of these four, two contained all omission risks. Risk themes having to do with requirements uncertainty are,

in our data set, risk themes of omission. An examination of the data reveals that these risk themes are not the result of architectural decisions but instead are the result of having a high level of uncertainty in the requirements. Risk themes associated with the lack of considering unrecognized needs are, by definition, risk themes of omission. Risk themes associated with the allocation of functionality are, by definition, risk themes of commission, since they are concerned with architectural decisions that allocate functionality to processors or components. The security risk themes are also all risks of omission, but that is not definitional. One interpretation of this phenomenon is that security requirements are hard to meet and the architectural teams ignored some important aspect of meeting them. We have no evidence to support this conjecture other than the observation that all risk themes having to do with security in our data set were risk themes of omission.

In the other categories, there is no discernible pattern to explain the distribution of risk themes of omission and commission.

6 Applications of These Results and Conclusions

Our analysis of the ATAM evaluation reports leads us to two main conclusions:

1. Risks of omission are much more numerous than risks of commission. We've observed this trait in the ATAM evaluations performed by both the SEI and an independent organization.
2. There is no evidence of any relationship between the articulated business/mission goals and the risk themes discovered during an ATAM evaluation or between the domain and the risk themes.

This analysis has applications for various communities. Next, we discuss specifically how practitioners, researchers, and those who perform ATAM evaluations can apply these results.

6.1 Applications for Practitioners

Practitioners should have three takeaways from our analysis:

1. The large number of risks of omission and the large number of risk themes related to a the lack of coordination suggest that architects should, at the inception of a project, identify the activities they should be performing and the entities with which they should be coordinating. They can treat these things as risks to be mitigated or as items on a process checklist or some other means of managing these unknowns. However, given the prevalence of these types of risk themes in our ATAM evaluations, architects should consider them early in a project.
2. The large number of risks in the Awareness category suggests that the architect needs to inform the organization about the implications of the architecture and the system being constructed. The organization should also be aware of the effects of constraints such as resource and schedule constraints. The risk themes in this category ranged from schedule problems to understanding what it means to institutionalize architecture-centric practices to coordination activities. In any case, these risks are all outward-looking from the project and suggest activities that must be undertaken to ensure the project's success.
3. The field of software design encompasses many techniques for dealing with the problems of late-arriving or underspecified requirements. Interfaces can be made more abstract to allow a broader range of a module's use, intermediates can be inserted between the producer and the consumer of data or services, and binding can be deferred through a variety of techniques. All of these techniques are well-known and available to architects

[Bass 03, Ch. 5]. The high number of risk themes in the category of Requirements Uncertainty and Volatility suggests that these techniques are not used as widely as they should be. Architects should identify areas of likely unknown or changing requirements, determine how they impact a design, and then use these mechanisms to reduce the impact of late-arriving or changing requirements.

6.2 Applications for Researchers

This work identifies or supports a number of questions that researchers can ask:

1. Are there predictors for the risk themes of a project? We presented negative evidence with respect to business/mission goals and the domain. What are the other possible predictors? One conjecture is that risks derive from organizational characteristics. If this is true, researchers need to identify those characteristics and the risks that derive from them and then determine how to reduce those risks. The data set we analyzed for this report does not include a sufficient number of ATAM evaluations from any single organization to enable a thorough analysis of this issue.
2. Are there any architectural techniques that can be used to reduce the impact of risks arising from unmet organizational coordination needs or organizational awareness requirements?
3. Are there any other software engineering techniques that can be used to reduce the impact of risks arising from unmet coordination needs or organizational awareness requirements? Process enactment tools, for example, would seem to have a role in ensuring that the required coordination is achieved.

6.3 Recommendations for ATAM Evaluators

One recommendation for those performing ATAM evaluations is to express their risk themes as risks. That is, a risk theme should express explicitly the consequences that could occur if it is not addressed. The SEI report on categorizing risk [Carr 93] describes the elements of a good risk expression. A second recommendation for those performing ATAM evaluations is to use the risk theme categories presented in this report as a guide to the production of risk themes. Templates for risk themes could be developed to help architects carry out both of these recommendations.

6.4 Conclusion

In this report, we presented some analyses based on a data set of 18 ATAM evaluations. As with any such data set, the analyses we performed are just a subset of those possible. The results we have achieved thus far, however, contain applications for practitioners, researchers, and those performing ATAM evaluations and point the way for further analysis.

Bibliography

URLs are valid as of the publication date of this document.

[Bass 03] Bass, L.; Clements, P.; & Kazman, R. *Software Architecture in Practice*, Third Edition. Boston, MA: Addison-Wesley, 2003 (ISBN 0-321-15495-9).

[Beyer 98] Beyer, H. & Holtzblatt, K. *Contextual Design: Defining Customer-Centered Systems*. New York, NY: ACM Press, 1998 (ISBN 1-558-60411-1).

[Boehm 03] Boehm, B. & Turner, R. *Balancing Agility and Discipline: A Guide for the Perplexed*. Boston, MA: Addison-Wesley, 2003 (ISBN 0-321-18612-5).

[Carr 93] Carr, M.; Konda, S.; Monarch, I.; Ulrich, F.; & Walker, C. *Taxonomy-Based Risk Identification* (CMU/SEI-93-TR-006, ADA266992). Pittsburgh, PA: Software Engineering Institute, 1993. http://www.sei.cmu.edu/publications/documents/93.reports/93.tr.006.html.

[Charette 05] Charette, R. N. "Why Software Fails." *IEEE Spectrum 42*, 9 (September 2005): 42 - 49. http://ieeexplore.ieee.org/xpls/abs_all.jsp?isnumber=32236&arnumber=1502528&count=9&index=5.

[Clements 02] Clements, P.; Kazman, R.; & Klein, M. *Evaluating Software Architectures: Methods and Case Studies*. Boston, MA: Addison-Wesley, 2002 (ISBN 0-201-70482-X).

[Dominick 02] Dominick, L.; Hilliard, R.; Kahane, E.; Kazman, R.; Kruchten. P.; Kozaczynski, W.; Obbink, H.; Postema, H.; Ran, A.; & Tracz, W. *Software Architecture Review and Assessment (SARA) Report, Version 1.0*. http://philippe.kruchten.com/architecture/SARAv1.pdf (2002).

[Kazman 05] Kazman, R. & Bass, L. *Categorizing Business Goals for Software Architectures* (CMU/SEI-2005-TR-021, ADA444917). Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University, 2005. http://www.sei.cmu.edu/publications/documents/05.reports/05tr021.html.

[O'Connell 06] O'Connell, D. *Boeing's Experiences Using the SEI ATAM and QAW Processes*. http://www.sei.cmu.edu/architecture/saturn/2006 /OConnell.pdf (2006).

[van der Linden 02] van der Linden, F. "Software Product Families in Europe: The Esaps & Cafe Projects." *IEEE Software 19*, 4 (July/August 2002): 41 - 49. http://ieeexplore.ieee.org/xpls/abs_all.jsp?isnumber=21951& arnumber=1020286&count=26&index=8.

[Wikipedia 06] Wikipedia. *Cohen's Kappa*. http://en.wikipedia.org/wiki/Cohen's_kappa (August 2006).

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
(Leave Blank)	September 2006	Final

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
Risk Themes Discovered Through Architecture Evaluations	FA8721-05-C-0003

6. AUTHOR(S)

Len Bass, Robert Nord, William Wood, & David Zubrow

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Software Engineering Institute Carnegie Mellon University Pittsburgh, PA 15213	CMU/SEI-2006-TR-012

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
HQ ESC/XPK 5 Eglin Street Hanscom AFB, MA 01731-2116	ESC-TR-2006-012

11. SUPPLEMENTARY NOTES

12A DISTRIBUTION/AVAILABILITY STATEMENT	12B DISTRIBUTION CODE
Unclassified/Unlimited, DTIC, NTIS	

13. ABSTRACT (MAXIMUM 200 WORDS)

This technical report analyzes the output of 18 evaluations conducted using the Architecture Tradeoff Analysis Method® (ATAM®) developed by the Carnegie Mellon® Software Engineering Institute. The goal of this analysis was to find patterns in the risk themes identified during those evaluations. The major results are

- a categorization of risk themes
- the observation that twice as many risk themes are risks of "omission" as are risks of "commission"
- a failure to find a relationship between the business/mission goals of a system and the risk themes revealed during an ATAM evaluation of that system
- a failure to find a relationship between the domain of a system being evaluated and the risk themes associated with the development of that system

14. SUBJECT TERMS	15. NUMBER OF PAGES
software architecture evaluations, ATAM evaluations, risk themes	42

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL