Empirical Evaluation of State Event Fault Tree and Fault Tree combined with Markov Chains for the Safety Analysis of Dynamic Embedded Systems

Adrien Mouaffo, Kavyashree Jamboti, Davide Taibi
Software Engineering Research Group
University of Kaiserslautern
Kaiserslautern, Germany
{adrien.mouaffo | jamboti | taibi}@cs.uni-kl.de

Abstract

Most innovation in the automotive industry is driven by embedded systems. They make usage of dynamic adaption to environmental changes or component/subsystem failures for remaining safe. Following this evolution, fault tree analysis techniques have been extended with concept for dynamic adaptation but resulting techniques like state event fault tree analysis, are not widely used in practice.

In this report we present the results of a controlled experiment that analyze these two techniques (State Events Fault Trees and Faul trees combined with markov chains) with regard to their applicability and efficiency in modeling dynamic behavior of dynamic embedded systems.

The experiment was conducted with students of the TU Kaiserslautern to modeli different safety aspects of an ambient assisted living system.

The main results of the experiment show that SEFTs where more easy and effective to use.

Table of Contents

1 Introduction

Most innovation in the automotive industry is driven by embedded systems. They make usage of dynamic adaption to environmental change or component/subsystem failures to maintain a certain level of safety. Following this evolution several safety analysis techniques have been proposed such as fault tree analysis and fault trees combined with Markov Chains, but they are not widely used in practice.

In this report we present the results of a controlled experiment that analyze these two techniques with regard to their applicability and efficiency in modeling dynamic behavior of dynamic embedded systems.

The remainder of this report is structured as follows: Section 2 provides background and related work. In Section 3 design and execution of the experiment are explained, then Section 4 describes performed analysis. In Section 5 we present analysis results and show how we avoided validity threats, finally we conclude in section 6.

2 Background

In this section we give an overall background to understand the performed study. We first describe a dynamic embedded system (Adaptive cruise control). Then we describe safety analysis.

2.1 Adaptive cruise control

Most innovation in the automotive industry is done nowadays through embedded software. They are used for vehicule control systems like anti-lock-bracking systems, electronic stability control, traction control system or adaptive cruise control. These control systems contain an electronic control unit (ECU) with embedded software which, based on value read from dedicated sensors, act on predefined actuator to improve driving comfort and vehicle safety.

Adaptive cruise Control (ACC) is an automotive feature that allows a vehicle's cruise control system to adapt the vehicle's speed to the traffic environment [1]. Sensors placed in front and behind the vehicle are used to detect the speed of cars in front and behind the vehicle. In combinaison with vehicle speed and driving activities (break or throttle) this information is used by the ACC to adapt the vehicle speed and therefore avois collision (see Figure 1).

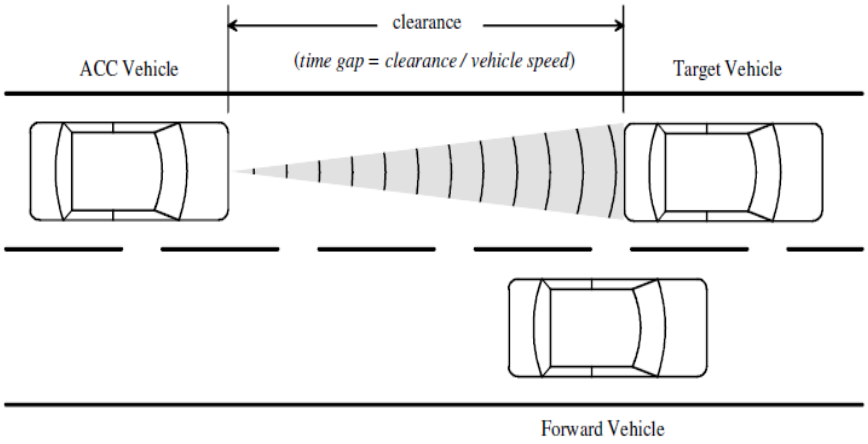

Figure 1 ACC vehicle relation [1]

2.2 Safety Analysis

In this section we introduce some basics on safety of a system. We first introduce some basic vocabulary namely the definition of faults, failures, hazards and accidents [2]. Afterwards we describe safety analysis techniques based on fault tree models. All the descriptions below are only mean to serve as a brief overview and the interested reader is encouraged to refer to the citations for a detailed insight into each of these techniques.

2.2.1 Faults, Failures, Hazards and accidents

The terms failure and fault are the key to any understanding of FT construction. Yet they are often misused. One of them describes the situation(s) to be avoided, while the other describes the problem(s) to be circumvented. In this section we briefly recall the main definitions for performing safety analysis using FT and CFT. For further information we refer to [3, 4, 5].

Failure

Each behavior of a system that differs from the ambient conditions specified behavior although the environmental conditions are specified correctly is called a failure.

Fault/Error

A fault is a static event in a system that may cause a failure. There are many different definitions for a fault. Some of them even differentiate between error and fault. But the main consensus is found in the difference between fault and failure. A fault is a deviation from the specification such as incorrect design or incorrect usage. A failure is a state of a system. A fault may but needn't cause a failure. If a failure is present then it must be cause by one or more faults. Also a fault may be caused by other faults.

Accident, Mishap

An accident is an undesired event that destroys or affects goods such as life or health of humans, economical goods, or the environment.

Hazard

A hazard is a state of the system in scope and its environment in which the occurrence of an accident only depends on uncontrollable influences. A hazard is for example, an open gate in the presence of an

arriving train. Whether an accident occurs depends only on whether the driver of a car near the gate is alerted or not.

2.2.2 Fault Tree Analysis

Fault trees [6, 7, 8] are constructed using a backward searching technique starting with a top event. The causes identified are combined using boolean gates. After its construction, it can provide quantitative results such as the top event probability or qualitative results in the form of Minmal Cut Sets (MCS). A MCS signifies a set of events where the nonoccurrence of even one event prevents the top event from occurring. The MCS can be ranked according to the number of events comprising them and the ones with less number of events need to be ensured that their occurrence probabilities are reduced or eliminated. In cases where the MCS consists of just one event called a single point failure, special attention must be given in order to ensure that it does not occur or its chances are minimized.

2.2.3 Component Fault Trees

In the previous section we recalled conventional FTs. The modeling of CFTs is a modularization technique to handle FTs for huge systems that consist of more than one component. The components are connected in a functional network via signal ports and the top events of the CFTs correspond to failure modes of the output signals. CFTs are similar to classical FT with some differences. They may contain more than one top event and one basic event may also be connected to more than one logical gate. CFTs have been developed in 2003 by P. Liggesmeyer, O. Mäckel and B. Kaiser, see [4] for further details.

2.2.4 State/Event Fault Trees

State/Event Fault Trees (SEFTs) [9] build on CFTs [4] which are an elegant approach to build fault trees based on failures of the components of a system. A CFT overcomes the drawbacks of the traditional fault tree which only conveys how a failure can occur, but does not specify which components influence each other in a manner that the failure occurs. CFTs can be easily reused as they have clear decomposition semantics based on system architecture. Though CFTs overcome some of the drawbacks of traditional fault trees, they are incapable of handling some other issues of fault trees such as sequence and timing issues of fault tree events. CFTs cannot handle stochastic dependence and cannot be integrated with state-based design models showing the behavior of the system. SEFTs have been designed to overcome the above problems. They allow the modeling of failure of a component showing the internal safety relevant state changes. Unlike traditional FTs or CFTs they make a clear distinction between a state and an event. In the context of SEFTs, a

state is defined as the collectivity of the variable properties of a component that are relevant to its behavior and its reaction to external events and an event is defined as a sudden phenomenon without temporal expansion in the context of discrete event systems. A state or event occurrence in one component can trigger state changes in another component. SEFTs enable the use of a wide range of gates which need not be just boolean operators provides by traditional FTs, gates in SEFTs can be made of boolean operators and state-based models which allow modeling of the order and timing of the occurrence of states and events in an SEFT. Some of the gates used in an SEFT are:

- AND(with n state inputs)
- AND(with n state inputs and one event input)
- OR(with n state inputs)
- OR(with n event inputs)
- History-AND
- Priority-AND

SEFTs are quantitatively analyzed by translation to Petri Nets. The top event probability can be calculated by calculating the probability for the corresponding place in the Petri net.

3 Experiment Design and Execution

In this section we specify the goal of the experiment, describe the design used for the experiment and the procedure followed for its execution.

3.1 Goal

We specify in this section the goal of the controlled experiment. The high level goal is firstly specified using the GQM goal specification template. After that we derive the questions and metrics related to the goal and describe them using a tree structure.

3.1.1 GQM Goal specification

Following the GQM goal specification construct, the goal of the experiment is to:

Analyze state/Event Fault Tree (SEFT) for the purpose of understanding and comparing their applicability and efficiency with and Fault Tree associated to markov chains (FT+MC) with respect to the modeling of safety related aspects of open and dynamic systems.

Figure 2 shows the GQM tree refining the goal related to understanding the applicability and efficiency of SEFT and Figure 3 shows the tree refining the goal related to comparing the applicability of SEFT with FT+MC. Metrics used in both trees are defined as follows:

- **Completeness:** measures the capability of a method to completely model all aspects of the system.
- **Easiness:** measures the effort needed for building the model.
- **Understandability:** measures the effort needed to understand the models built with the technique in relation to the failure logic.
- **Time needed:** measures the time needed for building the models
- **Effort expectancy:** measures the degree of ease associated with the use of the technique.
- **Attitude toward using the methodology:** measures the overall affective reaction to using the technique.
- **Self efficacy:** measures the degree to which the subject believes that they will better perform if they have some help.
- **Performance expectancy:** measures the degree to which the subject believes that using the technique will help him to attain gains in job performance.

– **Metrics for comparison:** measure the difference that the student noticed when applying both techniques.

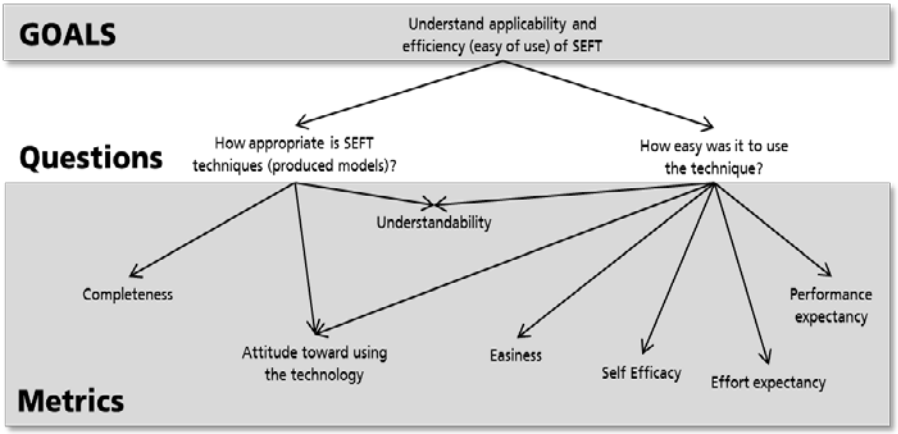

Figure 2 GQM Tree for understanding

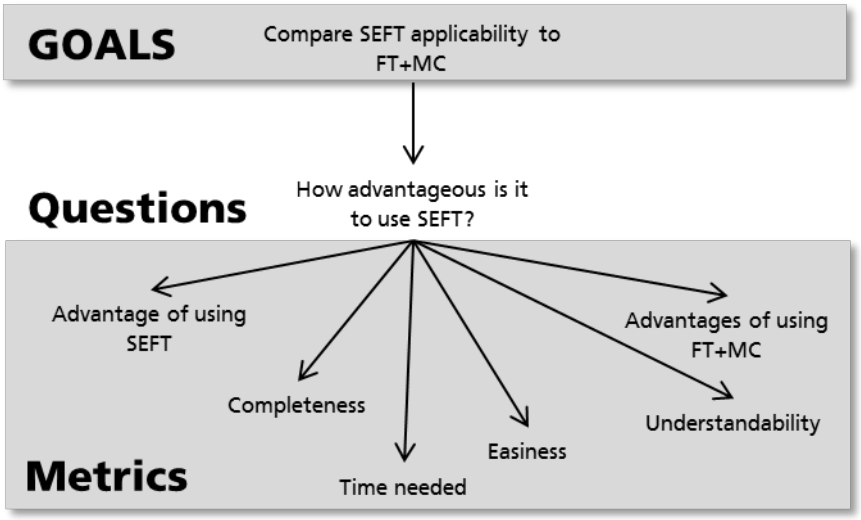

Figure 3 GQM tree for comparison

3.1.2 Hypotheses

In order to compare both techniques regarding the metrics completeness, understandability and easiness we specified following hypotheses:

– **Null hypothesis:**
H0: The score for both techniques are similar.

- **Alternative hypotheses**

H1: The score for both techniques are differents and the score for SEFT is greater than the score FT+MC.

3.2 Participants

Participants are students of the lecture Safety and Reliability Engineering taught at the University of Kaiserslautern during the winter term 2012-2013. They acquired during the lecture background knowledge on fault tree analysis and markov chains.

Participants of each group were trained for applying state event fault trees. Training for applying Fault Trees and Markov Chains was not necessary because they were trained on these techniques during the lecture.

3.3 Material and instruments

We used a questionnaire for getting their feedback on the training. During the experiment participants received material describing the system used for applying SEFT and FT+MC. Each task was described and at the end of the task description a questionnaire was added for getting the impression of each participant after the executuion of the given task. At the end of the experiment each participant then had to fill in a debriefing questionnaire.

3.4 Execution

Here we report details on the design and precedures.

3.4.1 Design and procedure

Subjects were separated in two different groups. Subjects of the first group apply SEFT before applying FT+MC and subjects of the second group apply FT+MC before applying SEFT.

For the experiment we followed the procedure below:

- **Day 1: Training**
  - Introduction into SEFT & Example
  - Review of FTA+MC & Example
  - Application oft he technique on a small example
- **Day 2: Experiment**
  - Recap training

- Answer the training questionnaire
- Introduction to the experiment
- Distribution of material
- For each task
    - Carefully read the task
    - Identify what need to be done
    - Perform changes
    - Review
    - Answer the questionnaire
- Answer the final questionnaire
- Collection of material
- End of the experiment

4 Analysis

During and after the experiment we collected data to be analyzed for testing our hypotheses. In this section we report and analyze those data.

4.1 Data collection and aggregation

Subjects of each group have to apply SEFT and FT + MC on the same systems in 2 tasks: Task 01 and Task 02. After each task they answer a questionnaire of 11 questions. We will reference question Y of task X as TX.Y. E.g. T02.08 represents the 8th question of task 02.

After performing both tasks subjects have to answer a debriefing questionnaire with 16 questions on SEFT and 4 questions on comparing SEFT and FT + MC. We will reference question Y of the debriefing questionnaire as D.Y. E.g. D.06 represents the 6th question of the debriefing questionnaire.

Table 1 shows how these questions are related to the respective metrics. Completeness, easiness, understandability, time needed and quality of produced models are calculated for each task.

Metrics	Questions
Completeness	TX.01: **I am sure that I was able to transfer the description from the system model completely to the SEFT / FT+MC.**
	TX.03: **I was able to identify the locations in the SEFTs / FT+MC that needed to be involved for doing the modifications.**
	TX.04: **I am sure that I was able to identify all the involved locations in the SEFT / FT+MC.**
Easiness	TX.02: **It was easy for me to transfer descriptions of the system model to the SEFT / FT+MC.**
	TX.08: **The SEFTs / FT+MC supported me during the accomplishment of the tasks.**
	TX.09: **It was easy for me to implement the modifications.**
	TX.10: **I was able to make the modifications with minor effort.**
	TX.11: **I was able to re-use a lot from the existing model during the modifications.**
Understandability	TX.05: **Because of the graphical representation of**

	SEFTs / FT+MC it was easy for me to keep the overview of the failure logic.
	TX.06: **The relationship between the SEFTs / FT+MC and system is easy for me to comprehend.**
	TX.07: **The SEFT / FT+MC methodology helped me to keep the overview of the failure logic.**
Time needed	Time nedded for building SEFT
	Time needed for building FT+MC
Performance Expectancy	D.01: **I would find the SEFTs useful in my work.**
	D.02: **Using the SEFTs enables me to accomplish tasks more quickly.**
	D.03: **Using the SEFTs increases my productivity.**
	D.04: **If I use the SEFTs, I will increase my chances of getting a raise. (e.g., by being faster)**
Effort Expectancy	D.05: **The SEFTs methodology is clear and understandable.**
	D.06: **It was easy for me to work with the SEFTs.**
	D.07: **I find the SEFTs easy to use.**
	D.08: **Learning to use the SEFTs was easy for me.**
Attitude toward using the method	D.09: **Using the SEFTs is a good idea.**
	D.10: **The SEFTs make work more interesting.**
	D.11: **Using the SEFTs is fun.**
	D.12: **I like using the SEFTs.**
Self- Efficacy	*I could complete a job or task using the State Event Fault Trees…*
	D.13: **… if there was no one around to tell me what to do as I go.**
	D.14: **… if I could call someone for help if I got stuck.**
	D.15: **… if I had a lot of time to complete the job for which the SEFT was provided.**
	D.16: **… if I had just the built-in help facility for assistance.**
General Assessment and comparison of SEFT with FT+MC	D.20: SEFT are easier to understand than FT+MC
	D.21: The working results will improve when SEFTs are used.
	D.22: FT+MC are more effective with regard to safety analysis than SEFTs.
	D.23: The SEFT is compatible with other methodologies I use.
	D.24: I would prefer working with FT+MC over working with SEFTs.

Table 1 Relation between metrics and questions

4.2 Analysis procedures

The answer for each question was given on a 5 points likert scale ranging from of 1 to 5:

- **1:** The subject strongly disagrees
- **2:** The subject disagrees
- **3:** The subject neither disagrees nor agrees
- **4:** The subject agrees
- **5:** The subject strongly agrees

For aggregating the answers of questions into metrics, we calculate the metric score as followed:

$$\text{Metric score} = \frac{(\sum_{i=1}^{n} \text{question}_i)}{(n * 5)}$$

n is the number ofquestions associated to the metric.

Equation 1 Metric score calculation

The **metric score** is calculated per metric for each subject. It is a percentage value which expresses how close the score is to the ideal answer that is the subject strongly agrees about all questions which are related to the metric. Values for the metric score range from 0,2 to 1,0. 0,2 if the subject strongly disagrees for all questions and 1,0 if the subject strongly agrees for all questions (Figure 4). If a score is less than or equal to 0,6 then the result is not considered positif.

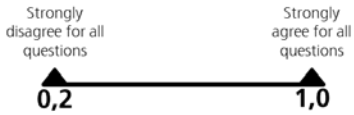

Figure 4 Metric score range

For each metric a descriptive statistic analysis is performed for giving quantitative statistical information abouth the metric. Then a hypothesis test is performed to gain confidence on the results (median) from the descriptive statistics.

For comparing the metric scores between both techniques we first test the data for normality with a Shapiro-Wilk's test. When scores for both techniques are normally distributed we perform an Independent T-Test for comparing the means. Else we perform a median test for comparing the medians.

4.3 Analysis results

In this section we report analysis results for each metric define in the GQM tree (Figure 2 and Figure 3).

4.3.1 Completeness

Completeness measures the capability of a method to completely model all aspects of the system.

As shown in Table 2 the completeness score for SEFT has a median of 0,80 and for FT+MC a median of 0,70. A hypothesis test confirms the results of median calculation for both techniques with an acceptable confidence level (0,009 for SEFT and 0,022 for FT+MC)

As shown in Table 3 the completeness score for SEFT has a median of 0,80 and for FT+MC a median of 0,73 in the group of subjects who performs SEFT before FT+MC. The completeness score is similar for both techniques (median = 0,66) in the group of subjects who performs FT+MC before SEFT.

Descriptives			
		Statistic	**Std. Error**
Completeness_Task1_SEFT	Mean	,7267	,04350
	Median	,8000	
	Variance	,038	
	Std. Deviation	,19455	
	Minimum	,33	
	Maximum	1,00	
Completeness_Task2_FTMC	Mean	,6967	,03667
	Median	,7000	
	Variance	,027	
	Std. Deviation	,16398	
	Minimum	,47	
	Maximum	1,00	

Table 2 General completeness score

Descriptives				
	Groups		**Statistic**	**Std. Error**
Completeness_Task1_SEFT	Group1 (SEFT before FT+MC)	Mean	,7818	,05400
		Median	,8000	
		Variance	,032	
		Std. Deviation	,17911	
		Minimum	,33	

		Maximum	1,00	
	Group2 (FT+MC before SEFT)	Mean	,6593	,06708
		Median	,6667	
		Variance	,040	
		Std. Deviation	,20123	
		Minimum	,40	
		Maximum	1,00	
Completeness_Task2_FTMC	Group1 (SEFT before FT+MC)	Mean	,7091	,04608
		Median	,7333	
		Variance	,023	
		Std. Deviation	,15282	
		Minimum	,47	
		Maximum	1,00	
	Group2 (FT+MC before SEFT)	Mean	,6815	,06164
		Median	,6667	
		Variance	,034	
		Std. Deviation	,18493	
		Minimum	,47	
		Maximum	1,00	

Table 3 Completeness score per group

Completeness scores for both techniques normaly distributed as assessed by the Shapiro-Wilk's test (Sig.= 0,142 for SEFT and Sig. = 0,210 for FT+MC). Therefore we perform a paired-samples t-test to compare the mean obtained for both techniques (Figure 5). Completeness score for SEFT (0,72 ± 0,19) was better than the completeness score for FT+MC (0,69 ± 0,16). The mean difference (0,03) was not statistical significant (p=0,415)

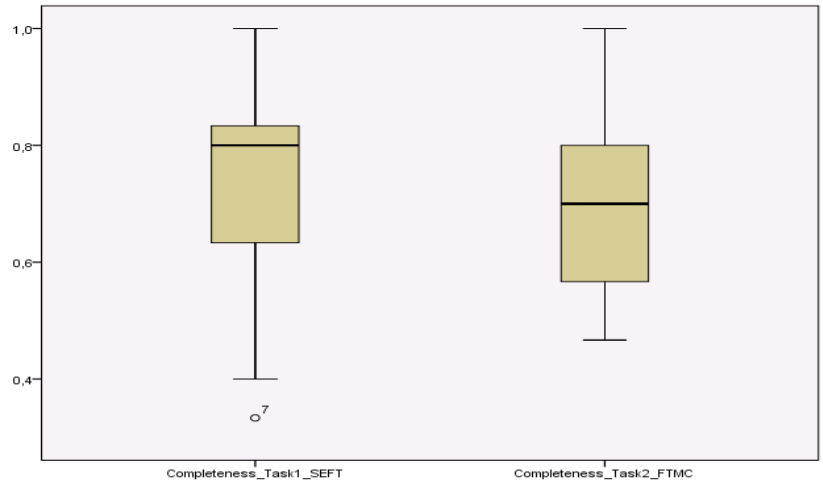

Figure 5 Boxplot: Completeness score

4.3.2 Easiness

Easiness measures the effort needed for building the model.

As shown in Table 4 the easiness score for SEFT has a median of 0,76 and for FT+MC a median of 0,78. A hypothesis test confirms the results of median calculation for both techniques with an acceptable confidence level (0,019 for SEFT and 0,002 for FT+MC)

As shown in Table 5 the easiness score for SEFT has a median of 0,76 and for FT+MC a median of 0,68 in the group of subjects who performs SEFT before FT+MC. The easiness score for SEFT has a median of 0,76 and for FT+MC a median of 0,80 in the group of subjects who performs FT+MC before SEFT.

Descriptives			
		Statistic	**Std. Error**
Easiness_Task1_SEFT	Mean	,7060	,04609
	Median	,7600	
	Variance	,042	
	Std. Deviation	,20613	
	Minimum	,24	
	Maximum	1,00	
Easiness_Task2_FTMC	Mean	,7440	,03035
	Median	,7800	
	Variance	,018	
	Std. Deviation	,13574	

	Minimum	,56	
	Maximum	1,00	

Table 4 General easiness score

Descriptives				
	Groups		Statistic	Std. Error
Easiness_Task1_SEFT	Group1 (SEFT before FT+MC)	Mean	,7455	,05689
		Median	,7600	
		Variance	,036	
		Std. Deviation	,18870	
		Minimum	,36	
		Maximum	1,00	
	Group2 (FT+MC before SEFT)	Mean	,6578	,07575
		Median	,7600	
		Variance	,052	
		Std. Deviation	,22725	
		Minimum	,24	
		Maximum	1,00	
Easiness_Task2_FTMC	Group1 (SEFT before FT+MC)	Mean	,7345	,04424
		Median	,6800	
		Variance	,022	
		Std. Deviation	,14672	
		Minimum	,56	
		Maximum	1,00	
	Group2 (FT+MC before SEFT)	Mean	,7556	,04292
		Median	,8000	
		Variance	,017	
		Std. Deviation	,12875	
		Minimum	,56	
		Maximum	1,00	

Table 5 Easiness score per group

Easiness scores for both techniques are normaly distributed as assessed by the Shapiro-Wilk's test (Sig.= 0,167 for SEFT and Sig. = 0,119 for FT+MC). Therefore we perform a paired-samples t-test to compare the mean obtained for both techniques (Figure 6). Easiness score for SEFT (0,70 ± 0,20) was less than the easiness score for FT+MC (0,74 ± 0,13). The mean difference (0,038) was not statistical significant (p=0,399)

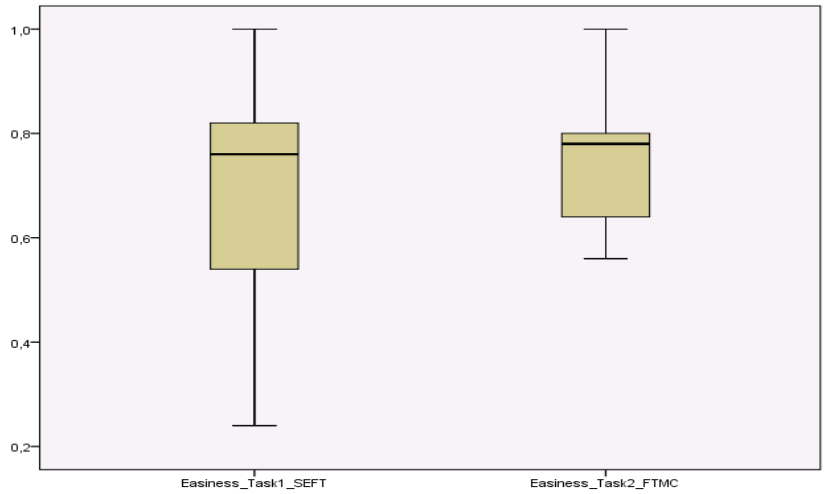

Figure 6 Boxplot: Easiness score

4.3.3 Understandability

Understandability measures the effort needed to understand the models built with the technique in relation to the failure logic.

As shown in Table 6 the understandability score for SEFT has a median of 0,80 and for FT+MC a median of 0,73. A hypothesis test confirms the results of median calculation for both techniques with an acceptable confidence level (0,002 for SEFT and 0,009 for FT+MC)

As shown in Table 7 the understandability score for SEFT has a median of 0,80 and for FT+MC a median of 0,73 in the group of subjects who performs SEFT before FT+MC. The understandability score for SEFT has a median of 0,66 and for FT+MC a median of 0,80 in the group of subjects who performs FT+MC before SEFT.

Descriptives			
		Statistic	**Std. Error**
Understandability_Task1_SEFT	Mean	,7467	,03541
	Median	,8000	
	Variance	,025	
	Std. Deviation	,15835	
	Minimum	,53	
	Maximum	1,00	
Understandability_Task2_FTMC	Mean	,7367	,03635
	Median	,7333	
	Variance	,026	

	Std. Deviation	,16255	
	Minimum	,33	
	Maximum	1,00	

Table 6 General understandability score

Descriptives				
	Groups		Statistic	Std. Error
Understandability_Task1_SEFT	Group1 (SEFT before FT+MC)	Mean	,7879	,04102
		Median	,8000	
		Variance	,019	
		Std. Deviation	,13603	
		Minimum	,53	
		Maximum	1,00	
	Group2 (FT+MC before SEFT)	Mean	,6963	,05891
		Median	,6667	
		Variance	,031	
		Std. Deviation	,17673	
		Minimum	,53	
		Maximum	1,00	
Understandability_Task2_FTMC	Group1 (SEFT before FT+MC)	Mean	,7030	,05640
		Median	,7333	
		Variance	,035	
		Std. Deviation	,18706	
		Minimum	,33	
		Maximum	1,00	
	Group2 (FT+MC before SEFT)	Mean	,7778	,04157
		Median	,8000	
		Variance	,016	
		Std. Deviation	,12472	
		Minimum	,60	
		Maximum	1,00	

Table 7 Understandability score per group

Understandability scores for SEFT is not normaly distributed as assessed by the Shapiro-Wilk's test (Sig.= 0,046). Therefore we perform a median test to compare the median obtained previously. The median test retain the null hypothesis: the median of difference equals 0 (Figure 7).

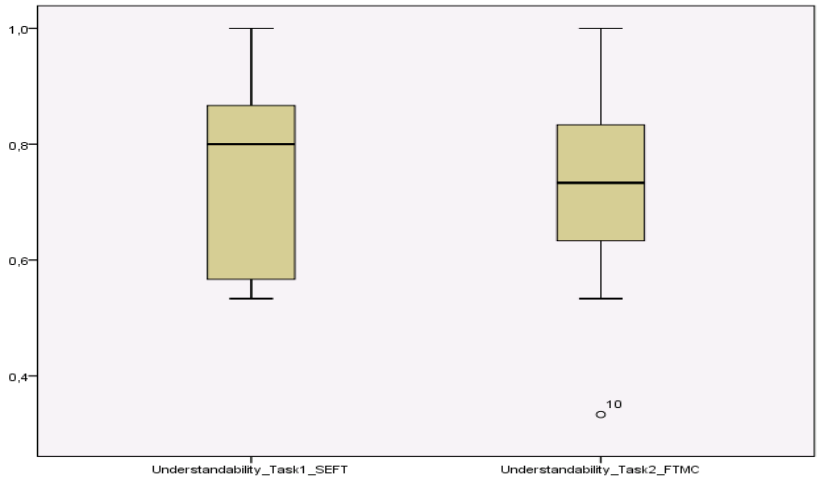

Figure 7 Boxplot: Understandability

4.3.4 Time needed

Time needed: measures the time needed for building the models.

As shown in Table 8 mean time needed for SEFT is 22,36 and 16,27 for FT+MC.

As shown in Table 9 mean time needed for SEFT is 28,67 and 16,17 for FT+MC in the group of subjects who performs SEFT before FT+MC. Mean time needed for SEFT is 14,80 and 16,40 for FT+MC in the group of subjects who performs FT+MC before SEFT.

Descriptives			
		Statistic	**Std. Error**
Time_Task01_SEFT	Mean	22,36	2,694
	Median	22,00	
	Variance	79,855	
	Std. Deviation	8,936	
	Minimum	10	
	Maximum	40	
Time_Task02_FTMC	Mean	16,27	1,342
	Median	15,00	
	Variance	19,818	
	Std. Deviation	4,452	
	Minimum	7	
	Maximum	23	

Table 8 General time needed

Descriptives				
	Groups		Statistic	Std. Error
Time_Task01_SEFT	Group1 (SEFT before FT+MC)	Mean	28,67	2,603
		Median	27,50	
		Variance	40,667	
		Std. Deviation	6,377	
		Minimum	22	
		Maximum	40	
	Group2 (FT+MC before SEFT)	Mean	14,80	1,881
		Median	15,00	
		Variance	17,700	
		Std. Deviation	4,207	
		Minimum	10	
		Maximum	21	
Time_Task02_FTMC	Group1 (SEFT before FT+MC)	Mean	16,17	,833
		Median	15,00	
		Variance	4,167	
		Std. Deviation	2,041	
		Minimum	15	
		Maximum	20	
	Group2 (FT+MC before SEFT)	Mean	16,40	2,977
		Median	20,00	
		Variance	44,300	
		Std. Deviation	6,656	
		Minimum	7	
		Maximum	23	

Table 9 Time needed per group

Time needed for both techniques are normaly distributed as assessed by the Shapiro-Wilk's test (Sig.= 0,805 for SEFT and Sig. = 0,443 for FT+MC). Therefore we perform a paired-samples t-test to compare the mean obtained for both techniques (Figure 8Figure 6). Time needed for SEFT (22,36 ± 8,93) was higher than the time needed for FT+MC (16,27 ± 4,45). The mean difference (6,091) was not statistical significant (p=0,069)

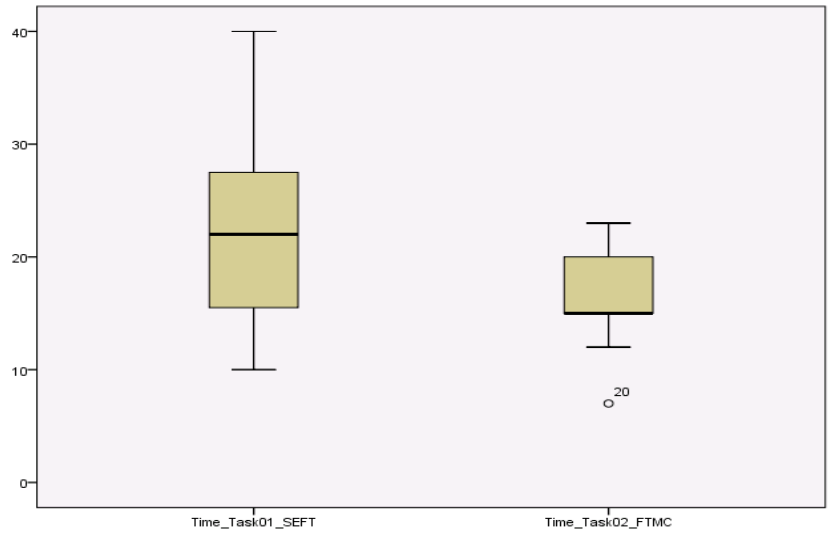

Figure 8 Boxplot: Time needed

4.3.5 Performance expectancy

Performance expectancy measures the degree to which the subject believes that using the technique will help him to attain gains in job performance.

As shown in Table 10 the performance expectancy score for SEFT has a median of 0,72. A hypothesis test confirms the results of median calculation with an acceptable confidence level (0,015).

As shown in Table 11 the performance expectancy score for SEFT has a median of 0,75 in the group of subjects who performs SEFT before FT+MC and a median of 0,60 in the group of subjects who performs FT+MC before SEFT.

Descriptives			
		Statistic	**Std. Error**
Performance_Expectancy_SEFT	Mean	,6795	,03333
	Median	,7250	
	Variance	,024	
	Std. Deviation	,15634	
	Minimum	,20	
	Maximum	,90	

Table 10 General Performance expectancyscore

Descriptives				
	Groups		Statistic	Std. Error
Performance_Expectancy_SEFT	Group1 (SEFT before FT+MC)	Mean	,7273	,02967
		Median	,7500	
		Variance	,010	
		Std. Deviation	,09840	
		Minimum	,55	
		Maximum	,90	
	Group2 (FT+MC before SEFT)	Mean	,6318	,05771
		Median	,6000	
		Variance	,037	
		Std. Deviation	,19141	
		Minimum	,20	
		Maximum	,85	

Table 11 Performance expectancy per group

4.3.6 Effort expectancy

Effort expectancy measures the degree of ease associated with the use of the technique.

As shown in Table 12 the effort expectancy score for SEFT has a median of 0,75. A hypothesis test confirms the results of median calculation with an acceptable confidence level (0,005).

As shown in Table 13 the effort expectancy score for SEFT has a median of 0,80 in the group of subjects who performs SEFT before FT+MC and a median of 0,62 in the group of subjects who performs FT+MC before SEFT.

Descriptives			
		Statistic	Std. Error
Effort_Expectancy_SEFT	Mean	,7286	,03725
	Median	,7500	
	Variance	,029	
	Std. Deviation	,17071	
	Minimum	,40	
	Maximum	1,00	

Table 12 General effort expectancy score

Descriptives				
	Groups		Statistic	Std. Error
Effort_Expectancy_SEFT	Group1 (SEFT before FT+MC)	Mean	,7773	,04121
		Median	,8000	
		Variance	,019	
		Std. Deviation	,13668	
		Minimum	,55	
		Maximum	1,00	
	Group2 (FT+MC before SEFT)	Mean	,6750	,06158
		Median	,6250	
		Variance	,038	
		Std. Deviation	,19472	
		Minimum	,40	
		Maximum	1,00	

Table 13 Effort expectancy per group

4.3.7 Attitude with using SEFT

Attitude with using the methodology measures the overall affective reaction to using the technique.

As shown in Table 14 the attitude withSEFT score has a median of 0,70. A hypothesis test confirms the results of median calculation with an acceptable confidence level (0,011).

As shown in Table 15 the attitude withSEFT score has a median of 0,72 in the group of subjects who performs SEFT before FT+MC and a median of 0,70 in the group of subjects who performs FT+MC before SEFT.

Descriptives			
		Statistic	Std. Error
Attitude_with_Technology_SEFT	Mean	,7000	,03795
	Median	,7000	
	Variance	,030	
	Std. Deviation	,17393	
	Minimum	,20	
	Maximum	1,00	

Table 14 General attitude with SEFT score

Descriptives				
	Groups		Statistic	Std. Error
Attitude_with_Technology_SEFT	Group1 (SEFT	Mean	,7450	,03202
		Median	,7250	

	before FT+MC)	Variance	,010	
		Std. Deviation	,1012 4	
		Minimum	,65	
		Maximum	1,00	
	Group2 (FT+MC before SEFT)	Mean	,6591	,06565
		Median	,7000	
		Variance	,047	
		Std. Deviation	,2177 4	
		Minimum	,20	
		Maximum	,90	

Table 15 Attitude with SEFT score per group

4.3.8 Self efficacy

Self efficacy measures the degree to which the subject believes that they will better perform if they have some help.

As shown in Table 16 the self efficacy score for SEFT has a median of 0,70. A hypothesis test confirms the results of median calculation with an acceptable confidence level (0,029).

As shown in Table 17 the self efficacy score for SEFT has a median of 0,77 in the group of subjects who performs SEFT before FT+MC and a median of 0,60 in the group of subjects who performs FT+MC before SEFT.

Descriptives			
		Statistic	**Std. Error**
Self_Efficacy_SEFT	Mean	,6881	,03235
	Median	,7000	
	Variance	,022	
	Std. Deviation	,14824	
	Minimum	,35	
	Maximum	,90	

Table 16 General delf efficacy score

Descriptives				
	Groups		**Statistic**	**Std. Error**
Self_Efficacy_SEFT	Group1 (SEFT before FT+MC)	Mean	,7650	,02986
		Median	,7750	
		Variance	,009	
		Std. Deviation	,0944 3	

		Minimum	,60	
		Maximum	,90	
	Group2 (FT+MC before SEFT)	Mean	,6182	,04733
		Median	,6000	
		Variance	,025	
		Std. Deviation	,15696	
		Minimum	,35	
		Maximum	,80	

Table 17 Self efficacy score per group

4.3.9 Comparison

Metrics for comparison: measure the difference that the student noticed when applying both techniques.

SEFT_Better represents the answer on if subjects think that SEFT are better than FT+MC and FTMC_Better represents answers on if subjects think that FT+MC are better than SEFT. As shown in Table 18 the SEFT_Better score has a median of 0,66 and the FTMC_Better score a median of 0,70. A hypothesis test confirms the results of median calculation with an acceptable confidence level only for FTMC_Better (0,029).

As shown in Table 19 the SEFT_Better score has a median of 0,73 in the group of subjects who performs SEFT before FT+MC and a median of 0,60 in the group of subjects who performs FT+MC before SEFT. And the FTMC_Better score has a median of 0,70 in the group of subjects who performs SEFT before FT+MC and a median of 0,80 in the group of subjects who performs FT+MC before SEFT.

Descriptives			
		Statistic	**Std. Error**
SEFT_Better	Mean	,6455	,03205
	Median	,6667	
	Variance	,023	
	Std. Deviation	,15032	
	Minimum	,20	
	Maximum	,87	
FTMC_Better	Mean	,7000	,03543
	Median	,7000	
	Variance	,028	
	Std. Deviation	,16619	
	Minimum	,40	

	Maximum	1,00	

Table 18 General comparison score

Descriptives				
	Groups		Statistic	Std. Error
SEFT_Better	Group1 (SEFT before FT+MC)	Mean	,7091	,02737
		Median	,7333	
		Variance	,008	
		Std. Deviation	,09079	
		Minimum	,60	
		Maximum	,87	
	Group2 (FT+MC before SEFT)	Mean	,5818	,05249
		Median	,6000	
		Variance	,030	
		Std. Deviation	,17408	
		Minimum	,20	
		Maximum	,80	
FTMC_Better	Group1 (SEFT before FT+MC)	Mean	,6818	,05012
		Median	,7000	
		Variance	,028	
		Std. Deviation	,16624	
		Minimum	,40	
		Maximum	1,00	
	Group2 (FT+MC before SEFT)	Mean	,7182	,05191
		Median	,8000	
		Variance	,030	
		Std. Deviation	,17215	
		Minimum	,50	
		Maximum	1,00	

Table 19 Comparison score per group

4.3.10 Comments from subjects

A coding analysis [coding analysis] was performed to analyze comments made by subjects and results are shown in Table 20.

Advatanges	Disadvantages
Clear process Clear mapping with system strucure Diagrams are more informative More gates Easy to decompose	Time consuming More complicated than FT+MC Too many information
Difficulties	**Improvement suggestion**
Checking every components Keep trace of consistency	Tool support Better graphical representation of

Too many gates Manual effort	elements (gates, events, …) Enhanced traceability Notation

Table 20 Comments from subjects

4.4 Analysis Summary

Table 21 shows a summary of our analysis. In the column hypothesis check, a + (resp. -) shows the satisfaction (resp. non satisfaction) of the main hypothesis:

- SEFT obtains better results than FT+MC for completeness, easiness, understandability, time needed
- SEFT obtains results better than 0.6 for Effort Expectancy, Attitude toward using the method, Self- Efficacy and Performance Expectancy.

Metric	Mean / Median		Hypothesis Check
	SEFT	FT + MC	
Completeness	,8000	,7000	+
Easiness	,7600	,7800	-
Understandability	,8000	,7333	+
Time needed (min)	22,36	16,27	-
Effort Expectancy	,7500		+
Attitude toward using the method	,7000		+
Self- Efficacy	,7000		+
Performance Expectancy	,7250		+

Table 21 Analysis summary

5 Discussion

FTs with Markov chains are a popular tried and tested technique to model dynamic systems in the industry. This technique has high degree of appropriateness and easiness as they are easily understandable.

The appropriateness of SEFT models for modeling dynamic systems is also high as users are able to completely represent safety scenarios using SEFTs. Although it was also found that SEFTs were not so difficult to understand and it was possible for users to understand the semantics of the modeling elements and successfully use them to create failure models, the measures obtained indicate that understanding of SEFTs is not as high as we had assumed it would be.

Our studies also showed that they could use SEFT with more confidence if they were provided with assistance. Users expect their performance to increase if they used SEFTs to model safety scenarios. The efficiency of building SEFTs is also influenced by tool availability as SEFTs have a large number of modeling elements with constraints imposed on the connections between them. This is also reflected by the fact that users required more time to perform some tasks on SEFTs as compared to FTs with Markov Chains.

6 Validity threats

In this section we discussed how validity threats [10] were avoided.

6.1.1 Conclusion validity

Due to the number of subjects we were able to obtain necessary power for performed statistical tests.

Before performing test preconditions (normality, independence of variables …) were checked to make sure that they are satisfied.

To get reliable measures questionnaires were checked by an expert on empirical studies.

Subjects have similar background and knowledge about safety analysis which part of the lecture Safety and Reliability for Embedded Systems.

6.1.2 Internal validity

Subjects were trained with techniques before experimentation. A more comprehensive training was performed only for SEFT because they already practice with fault trees and markov chains during the lecture.

Subjects were graduate students of the lecture safety and reliability engineering and were sufficiently motivated because the experiment was also use as practical exercise for the lecture.

Forms used for collecting data were checked by an expert on experimentation.

To avoid the effect of learning the whole group was divided into two subgroups: one applying SEFT before FT+Markov chains and the other one applying FT+Markov chains before SEFT.

6.1.3 Construct validity

The experiment's goal was refined into clearly defined metrics and measures to avoid misunderstandings.

Due to the time constraints we only use one system. But the operation to be performed was design to avaoid threats du to mono-operation bias.

To limit the effect of learning the whole group was divided into two subgroups: one applying SEFT before FT+Markov chains and the other one applying FT+Markov chains before SEFT.

Subjects were not aware of the hypothesis to be tested or measures to be taken.

6.1.4 External validity

We tried to have subjects closed enough to industrial setting by selecting graduate students in a class of safety and reliability for embedded systems.

The proposed system was derived from a concept car developed at Fraunhofer IESE (close enough to real setting).

7 Conclusion

In order to analyze the applicability and efficiency of SEFTs and Faul trees combined with markov chains on modeling safety aspects of dynamic and adaptive systems, we performed a controlled experiment where subjects have to apply these techniques on an AAL system and provide their feedback on using these techniques. The experiment was conducted with students of the TU Kaiserslautern and consisted of two parts: a training session where subjects were trained on using the techniques and the main experimental session where they used the techniques for modeling different safety aspects of an ambient assisted living system. Results of the experiment show that SEFTs where more easy and effective to use. Since we could not obtain enough data to statistically support our results and therefore are planning replicating the experiment with more subjects.

8 References

[1] "Adaptive cruise control system overview," 5th Meeting of the U.S. Software System Safety Working Group Anaheim, California USA, April 12th-14th 2005.

[2] S. Spang, "Scrutinizing the impact of security on safety on an communicating vehicle pla-toon." VIERForES Meilensteinbericht zum Arbeitspaket 5.5.2.

[3] R. Schwarz, "Modulare security-modelle zur komponentenbasierten modellierung komplexer eingebetteter systeme," Fraunhofer IESE, Milestone report to Workpackage 6.1.3, 2009.

[4] B. Kaiser, P. Liggesmeyer, and O. Mïckel, "A new component concept for fault trees," in *Proceedings of the 8th Australian Workshop on Safety Critical Systems and Software (SCSS03)*, 2003, pp. 37–46.

[5] A. Avizienis, J.-C. Laprie, B. Randell, and C. Landwehr, "Basic concepts and taxonomy of dependable and secure computing," *IEEE Transactions on Dependable and Secure Computing*, vol. 1, no. 1, pp. 11 – 33, jan.-march 2004.

[6] W. E. Vesely, F. F. Goldberg, N. H. Roberts, and D. F. Haasl, *Fault Tree Handbook*. U.S. Nuclear Regulatory Commission, 1981.

[7] P. Liggesmeyer, *Qualitätssicherung softwareintensiver technischer Systeme*, ser. Forschung in der Softwaretechnik. Spektrum Akademischer Verlag, 2000.

[8] N. G. Leveson, *Safeware - system safety and computers: a guide to preventing accidents and losses caused by technology*. Addison-Wesley, 1995.

[9] B. Kaiser, "State event fault trees: a safety and reliability analysis technique for software controlled systems," Ph.D. dissertation, Technical University of Kaiserslautern, 2006.

[10] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in software engineering: an introduction*. Norwell, MA, USA: Kluwer Academic Publishers, 2000.