Final Project Report WP-E

Document information	
Project Title	ASHiCS
Project Number	E.02.05
Project Manager	[redacted] University of York
Deliverable Name	Final Project Report
Deliverable ID	D0.7
Edition	00.00.02
Template Version	03.00.00
Task contributors	
University of York	

Abstract

The final report of the ASHiCS project provides a publishable summary of the results. In addition it lists all deliverables, dissemination activities, eligible costs, deviations, bills and lessons learned.

Authoring & Approval

Prepared by - *Authors of the document.*		
Name & Company	Position & Title	Date
[redacted]	[redacted]	3 May 2013

Reviewed by - *Reviewers internal to the project.*		
Name & Company	Position & Title	Date
[redacted]	[redacted]	3 May 2013

Approved for submission to the SJU by - *Representatives of the company involved in the project.*		
Name & Company	Position & Title	Date
[redacted]	[redacted]	3 May 2013

Document History

Edition	Date	Status	Author	Justification
00.00.01	28 March 2013	Draft	[redacted]	Submitted to EEC
00.00.02	3 May 2013	Draft		Re-submitted after EEC comments

Intellectual Property Rights (foreground)

This deliverable consists of SJU foreground.

Table of Contents

PUBLISHABLE SUMMARY 4

1 INTRODUCTION 7

1.1 PURPOSE OF THE DOCUMENT 7
1.2 INTENDED READERSHIP 7
1.3 INPUTS FROM OTHER PROJECTS 7
1.4 GLOSSARY OF TERMS 7

2 TECHNICAL PROJECT DELIVERABLES 9

3 DISSEMINATION ACTIVITIES 10

3.1 PRESENTATIONS/PUBLICATIONS AT ATM CONFERENCES/JOURNALS 10
3.2 PRESENTATIONS/PUBLICATIONS AT OTHER CONFERENCES/JOURNALS 10
3.3 DEMONSTRATIONS 10
3.4 EXPLOITATION PLANS 11

4 TOTAL ELIGIBLE COSTS 12

5 PROJECT LESSONS LEARNT 13

6 REFERENCES 14

List of tables

Table 1 – List of Project Deliverables 9
Table 2 – Overview of Billing 12
Table 3 – Overview of Effort and Costs per project participant 12
Table 4 – Project Lessons Learnt 13

List of figures

Figure 1 – Screenshot of the ASHiCS case study scenario in RAMS 4
Figure 2 – Progress of a search over time – horizontal axis is number of simulation runs, vertical axis is fitness (risk) of each run 5
Figure 3 – Near neighbour sampling of original search result 5

Publishable Summary

Safety analysts are starting to worry that large complex systems are becoming too difficult to analyze when part of the system is changed or placed under stress. Traditional safety analysis techniques may miss safety hazards or (more likely) some of the circumstances that can cause them. To help analysts discover hazards in complex systems, ASHiCS has created a proof-of-concept tool that uses evolutionary search and fast-time air traffic control (ATC) simulation to uncover airspace hazards that might otherwise be missed using traditional manual safety analysis.

In the ASHiCS process, simulations sit within what is termed a “search harness”. This describes the evolutionary computation software that “wraps” around the simulation, allowing the search to automatically start, configure, stop and select those simulation runs that are of interest to us. In our case, a simulation that results in a hazard or risk is of interest, and will therefore be judged to have higher “fitness”. The harness ranks the best individuals and creates mutated copies of these for the next generation to see if their fitness can be improved. This next generation of simulations is run, and again each is assessed for fitness. The process is repeated until the levels of evolved fitness in the population either reach a plateau (where no more improvement is likely or possible) or a sufficiently good simulation is found that allows us to stop the search.

As a specific case study, ASHiCS created a fast-time ATC simulation of an en-route sector containing multiple flight paths and aircraft types, and into each run of this simulation we injected a serious incident (cabin pressure loss) that requires one aircraft to make an emergency descent. To create additional complexity and extra workload for the air traffic controller (ATCo), we also introduce a storm moving across the sector. A screenshot of the resulting simulation is shown in Figure 1.

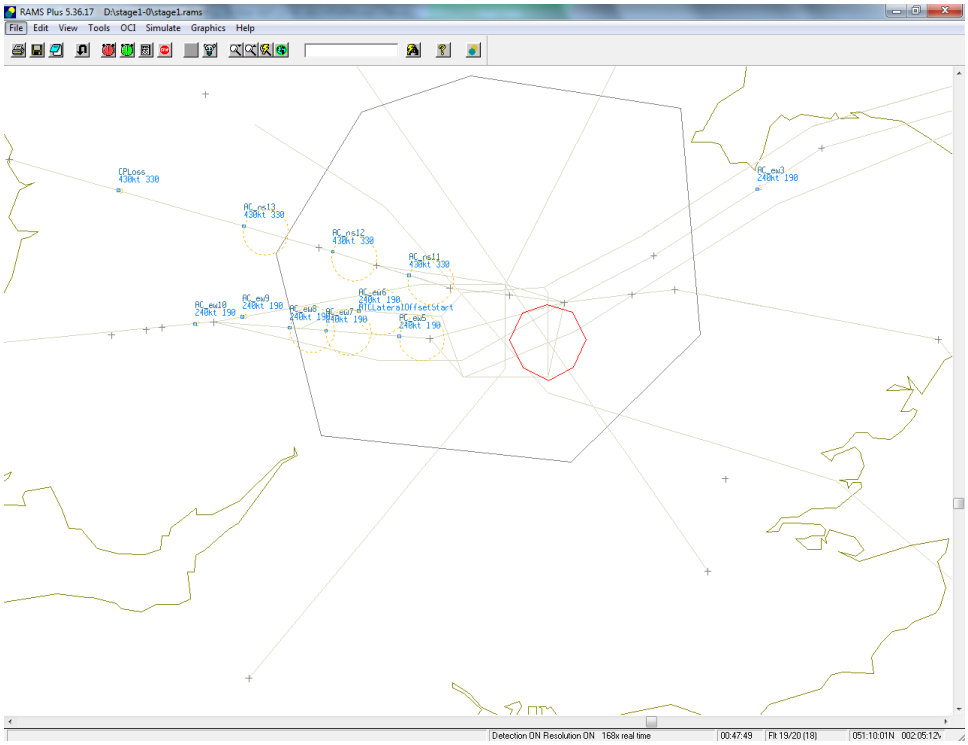

Figure 1 – Screenshot of the ASHiCS case study scenario in RAMS

Given that simulation, we then use a near-neighbor random hill-climber to search for high-risk variants of that situation: we run a wide range of variants, select the subset of variants that caused the most risk, and then mutate the aircraft entry times to create a new set of situation variants that will hopefully have even greater risk. Figure 1 shows a typical progress of a search over time – individual runs vary widely in fitness (i.e. in the level of risk they exhibit), but the trend is for the highest-risk run so far (horizontal bars) to increase over time.

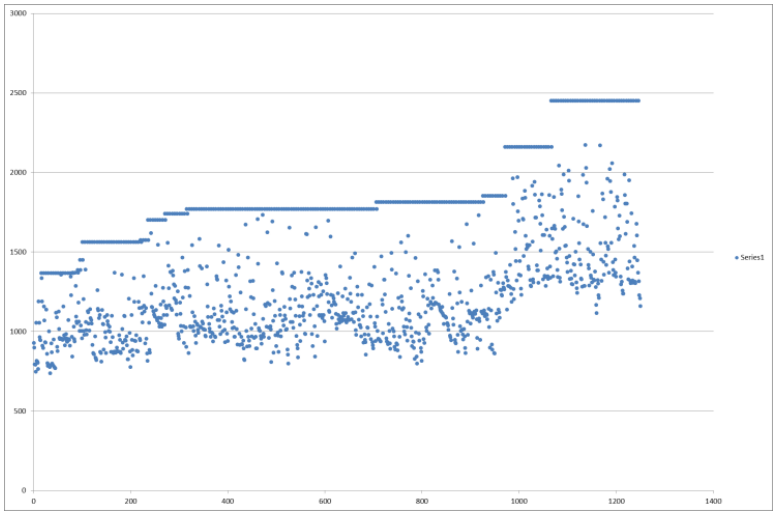

Figure 2 – Progress of a search over time – horizontal axis is number of simulation runs, vertical axis is fitness (risk) of each run

The "search space" (the set of all possible simulation runs) is extremely large and cannot be exhaustively searched for the worst case; this is a problem for safety analysts who need a context to the search results so that they can determine event probabilities. The approach we have taken is to provide a *local* context for the search results – for each high-risk situation found, we explore the space of situations that are very similar. This cannot demonstrate that the worst case scenario has been found, but it can indicate the expected frequency of that result in its near neighborhood. This provides some insight to the nature of the solution space within the near neighborhood of the original result, in terms of the frequency of high risk scenarios and how those scenarios differ from the original.

Figure 3 illustrates the second-stage search for local context by plotting 5000 samples of the near neighbourhood of the best scenario from a single first-stage search. The fitness score of the original is shown as a continuous horizontal line just below the fitness score of 2500 (vertical axis).

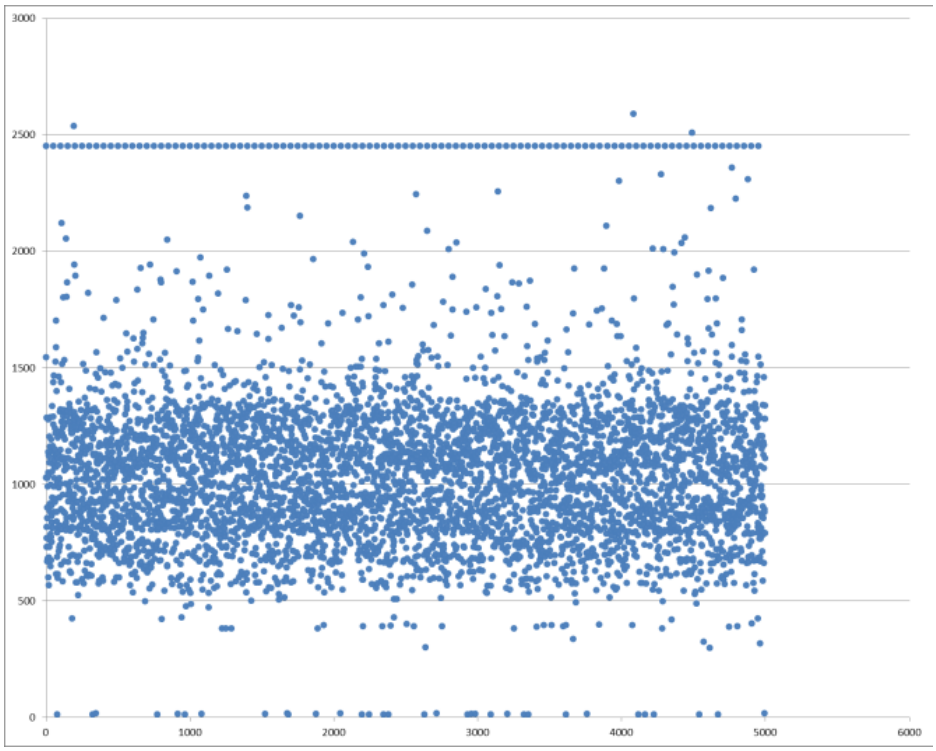

Figure 3 – Near neighbour sampling of original search result

Two things are interesting about this result. First, we can see that the search failed to find the worst case scenario. By extensively sampling the near neighbourhood, we were able to uncover 3 variants who improved on the fitness score of the original scenario (these are the three dots above the horizontal line). However, what is also apparent is the original search result is at (or very close) to the worst cases, something which we have confirmed by careful comparison of the aircraft entry times involved in conflicts. It would appear that although some marginal improvement is possible, the search performed well in terms of finding the worst case scenario.

Secondly, we can see that the vast majority of variants, even within this narrowly defined near neighbourhood of a high ranking scenario, do not come anywhere near the original fitness score. This suggests that there is a relatively narrow parameter band that generated the original high scoring scenario and its close variants. (From our analysis of the variant entry times, it appears that slight variants of the entry times of just 3 out of 20 aircraft are responsible for all the reported conflicts.)

The overall ASHiCS process produces a set of high-risk variant situations, which can then be studied in depth. This study can start in the original simulation, and then progress to higher-fidelity models and complementary analysis approaches. The contribution of ASHiCS is to identify the situation types that that generate the worst cases; analysts and can then investigate how to prevent that configuration of inputs leading to a hazard in the air sector being modeled.

The aim of ASHiCS was to develop a proof of concept approach to the automatic identification of hazards in complex systems. Our two-stage search process not only demonstrates the identification of hazardous scenarios, it helps analysts to understand the context in which these hazards occur and thus their place in the risk landscape of the whole system. With further work the approach would provide valuable practical benefits.

1 Introduction

1.1 Purpose of the document

The purpose of this document is to:

- Summarise the technical results and conclusions of the project (Publishable Summary);
- Provide a complete overview of all deliverables;
- Provide a complete overview of all dissemination activities (past and in progress). Where appropriate, provide feedback from presentations. Describe exploitation plans.
- Provide a complete overview of the billing status, eligible costs, planned and actual effort (incl. an explanation of the discrepancies).
- Analyse the lessons learnt at project level.

1.2 Intended readership

This document's intended readership are the personnel at EUROCONTROL and SESAR JU who are concerned with final acceptance of the ASHiCS project results. As a summary of the ASHiCS project, it may also be of interest to ATM planners, modellers and safety analysts interested in automated searches for hazards using fast time ATM simulation software such as RAMS Plus.

1.3 Inputs from other projects

We have had no input from other projects.

1.4 Glossary of terms

Evolutionary search

Form of search algorithm that uses selective pressure and mutation to improve a population of candidate solutions over many generations.

Evolutionary strategy

Pragmatics of evolutionary search relating to rate, range and restrictions of mutation, crossover, combination or other means of furthering good genes, population size, fitness selection policy, number of generations, etc.

Fitness function

Process used to select individuals from the population of candidate solutions by a ranking score assigning to each solution.

Search heuristics

Means of effectively guiding the search algorithm through the search space.

Search Landscape

Imaginary visualisation of a search space in which the fitness of each individual in a set's population is shown as a measure of vertical height with individuals of similar fitness being placed close together. By plotting a curve between the heights of individuals a landscape can be drawn with peaks representing areas in the solution space that contain the fittest individuals. This visualisation is extremely pervasive within the search literature, however it has many theoretical problems: i) there are no horizontal axis which can place the individuals geometrically within a set so the notion of similar solutions lying close to one another is hard to justify; ii) the visualisation breaks down completely in high dimensionality (i.e. where many factors may affect fitness levels), as there are likely to be areas of "impossible" gene combinations that cannot be realised in a solution.

Weighted fitness function

In a multi-objective fitness function, it is possible to assign greater “weight” to certain factors within the fitness evaluation so that the search favours solutions presenting those characteristics over others.

2 Technical Project Deliverables

Number	Title	Short Description	Approval status
D1.1	Scenario Description	*Identified our choice of scenarios, the simulation software that was chosen and how we intended to develop the scenarios over time.*	Approved
D1.2	Model Description	*Described our baseline scenario constructed to test that the application interfaces work correctly with RAMS Plus and that the search algorithms can identify the worst case scenario for our safety incident to occur. Also described how we intend to extend the baseline scenario to more complex traffic patterns and controller workloads.*	Approved
D2.1	Risk Measure Description	*Described the instruments we planned to use to measure risk (or perceived risk) in the ASHiCS RAMS Plus scenarios.*	Approved
D2.2	Method Description Technical Report	*Described the search process for ASHiCS in detail, including our then-current characterisation of the search landscape*	Approved
D3.1	Algorithm Evaluation Technical Report	*Gave an assessment of the current state and future direction of the ASHiCS search harness algorithms, alongside details of new additions to the Stage 2 scenario (simulation of severe weather impact) that allowed us to increase the complexity of the air sector and workload of controllers. Initial results from this revised model are included.*	Approved
D3.2	Algorithm Evaluation Technical Report	*This final deliverable for ASHiCS discussed the implementation and results of the two-stage search process proposed in the previous deliverable.*	Submitted

Table 1 – List of Project Deliverables

3 Dissemination Activities

3.1 Presentations/publications at ATM conferences/journals

ASHiCS: Automating the Search for Hazards in Complex Systems *[1]*

1st SESAR Innovation Days, December 2011, Toulouse, France

With increasingly complex systems to manage, safety analysts are starting to express concern that large complex systems are becoming too difficult to predict or guarantee safety when part of the system is changed or placed under stress. In order to help analysts discover hazards within complex systems, we propose a new generation of tools that make use of automated search heuristics and simulation to uncover hazards that might otherwise be missed using traditional (manual) safety analysis.

Searching air sectors for risk *[2]*

2nd SESAR Innovation Days, November 2012, Braunschweig, Germany

ASHiCS permits the automatic discovery of high risk air traffic scenarios. In this paper we describe the evolutionary search used in ASHiCS and present an analysis of the project's Stage 2 solution landscape. We suggest that random or exhaustive search is infeasible given the size of the solution space presented by simple air traffic scenarios, and that standard linear regression modelling is unlikely to find traffic input patterns that indicate the presence of high risk. While ASHiCS successfully targets high risk scenarios, we remain faced with very large search spaces in future Systems of Systems (SoS) models, and hope that our investigations into linear regression modelling will lead to a generic technique of practical value in the dimension reduction of these large search spaces.

3.2 Presentations/publications at other conferences/journals

Searching for Risk in Large Complex Spaces [3]

EvoStar, April 2013, Vienna, Austria

ASHiCS (Automating the Search for Hazards in Complex Systems) uses evolutionary search on air traffic control simulations to find scenario configurations that generate high risk for a given air sector. Weighted heuristics are able to focus on specific events, flight paths or aircraft so that the search can effectively target incidents of interest. We describe how work on the characterization of our solution space suggests that destructive mutation operators perform badly in sensitive, high dimensional spaces. Finally, our work raises some issues about using collective risk assessment to discover significant safety events and whether the results are useful to safety analysts.

The Discovery and Quantification of Risk in High Dimensional Search Spaces [4]

GECCO, July 2013, Amsterdam (***confirmed – will definitely occur***)

We describe a technique used by the ASHiCS project to discover high risk air traffic control (ATC) scenarios. The search space is extremely large and cannot be exhaustively searched for the worst case, creating a problem for safety analysts who require a context to search results so that event probabilities can be determined. While providing context cannot demonstrate that the worst case scenario has been found over all input permutations, it can indicate the expected frequency of that result in its near neighbourhood, allowing analysts to focus on a much reduced parameter range when investigating those aircraft in conflict.

3.3 Demonstrations

None.

3.4 Exploitation plans

In terms of similar research, we have just applied for EPSRC funding for a project in a similar domain. This will also use heuristic search over a simulation to look for a diversity of “bad things”, and thus will be able to build on the ASHiCS work. It represents a slight change of direction, however: instead of hazard analysis, it will be explicitly about software testing. The system studied will be a mobile robot controller (initially, a ground vehicle because the ground vehicle is both richer in complications than the air, and easier to understand for non-experts).

We are also looking at avenues to get funding for a PhD student to address a similar topic in a longer-term, more academic way. One current funding target is DSTL in the UK, who have recently issued a call for PhD bids that covers this area.

The shift from hazard analysis to software testing and from ATM SoS to individual vehicles is largely motivated by the ease with which faults can be seeded (and their effects understood). It has become clear from the ASHiCS project that large-scale hazard analysis is a very difficult thing to study; single-vehicle (controller) testing is much more tractable. Of course, if we have strong success with this more tractable problem, we will look at returning to the ASHiCS problem using the techniques we have developed.

4 Total Eligible Costs

Note: the costs in Table 2 below do not include the final invoice as it has not yet been raised.

Date	Deliverables on Bill	Contribution for Effort	Contribution for Other Costs (specify)	Status
30/12/2011	D0.0; D1.1	€34,581	Consumables: €78.05	Paid
31/12/2012	D0.1; D0.2; D1.2; D4.1	€69,362		Paid
31/12/2012	D0.3; D2.1; D0.4; D2.2; D0.5	€95,730		Paid
01/04/2013	D0.6; D4.2; D3.1; D0.7; D3.2; D4.3	€80,899	Consumables: €523.60; Travel: €7,006.38	Pending
GRAND TOTAL		€280,572	Consumables: €601.65; Travel €7,006.38[1]	

Table 2 – Overview of Billing

Company	Planned man-days	Actual man-days	Total Cost	Total Contribution	Reason for Deviation
The University of York	359	359	€280,572	€280,572	
GRAND TOTAL	359	359	€280,572	€280,572	

Table 3 – Overview of Effort and Costs per project participant

[1] This is an estimate of travel costs – final costs may deviate slightly.

5 Project Lessons Learnt

What worked well?
The project built on many years of ideas and research at York (in both simulation for hazard analysis and heuristic search). This meant that the project could start on the ATM domain straight away without a lot of generic preparatory work.
Choosing RAMS as the simulation platform (rather than a hand-coded simulator using generic simulation tools) gave some credibility to the results, and meant that we were forced to understand some ATM issues.
Scoping the project early on by agreeing the scenarios that we would use helped us meet deadlines (it was clear what needed to be done) and gave us clear examples to discuss with ATM experts.
The decision, late in the project, to focus on sensitivity analysis of the results, and to exploring the context (neighbourhood) of each interesting run, gave us some interesting results (which have proved publishable).
Agreement that interim deliverables (Dx.1) would be small work-in-progress reports allowed the project to proceed roughly as originally planned while still smoothing the payments.
What should be improved?
Working with RAMS was difficult throughout the project, and decidedly shaped what it was possible to do. In a future project of this nature we would aim to have a simulation tool developer/vendor as project partner. As a fallback option, using an open-source simulator would acceptable, as long as that simulator had some track record and an established user base (at present, we are not aware of any such simulator).
In ASHiCS we spent much effort spent on simulating operational airspace situations. This was a necessary prerequisite to our work, but it meant that we could not explore networked computer systems issues that might have impinged on those operational situations.
We entered the ASHiCS project planning and expecting to find qualitatively novel hazards in ATM situations, and to “know them when we saw them”. In a future project, we would take a narrower approach, at least in this university-based research work: we would couch the problem in terms of finding known problems in simulations (e.g. the effect of manually seeded faults). We would only aim to search for novel, unexpected problems if we were deploying the tool as part of a real systems development i.e. in an operational environment.
York’s lack of ATM experience entailed a long learning curve for the project staff, and although this was recognised by EEC/SJU at the time of award we were never able to get adequate subject-matter-expert support to fully compensate for this. EEC gave us access to experts, but they invariably had little incentive to work with us; doing so meant sacrificing either their own time or their day-to-day work, at apparently no remuneration for themselves or their employer.
Producing a deliverable every four months had some value as a pacemaker, but the resulting reports served little purpose other than a record of work done. We are not aware of anyone other than the project officer actually reading these reports, and the time spent on them did reduce the time we could spend on peer-reviewed conference and journal papers. (As alluded to above, our original bid had one deliverable every 8 months; this was increased post-award (at EEC’s request) to every 4 months in order to smooth the profile of payments to the university)

Table 4 – Project Lessons Learnt

6 References

[1] Clegg K, Alexander R. ASHiCS: Automating the Search for Hazards in Complex Systems. *Proceedings of the 1st SESAR Innovation Days*, Toulouse, France, 2011.

[2] Clegg K, Alexander R. Searching air sectors for risk. *Proceedings of the 2nd SESAR Innovation Days*, Braunschweig, Germany, 2012.

[3] Clegg K, Alexander R. Searching for Risk in Large Complex Spaces. *Proceedings of EvoStar*, Vienna, Austria, 2013.

[4] Clegg K, Alexander R. The Discovery and Quantification of Risk in High Dimensional Search Spaces. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, Amsterdam, 2013