Wednesday XXX
XX-XX
(Duration: 2 hours)

DEGREES OF MSc, MSci, MEng, BEng, BSc,MA and MA (Social Sciences)

Safety Critical Systems Development (M)

(Answer 3 out of 4 questions)

This examination paper is worth a total of 60 marks

The use of a calculator is not permitted in this examination

INSTRUCTIONS TO INVIGILATORS

Please collect all exam question papers and exam answer scripts and retain for school to collect. Candidates must not remove exam question papers.

1. (a) During the 1990s, the Motor Industry Software Reliability Association (MISRA) introduced a set of guidelines for the use of C in vehicle systems, which became known as MISRA-C. This aimed to provide the benefits of C including easy access to hardware, as well as memory and run-time efficiency while minimizing recognized problems such as limited run-time checking, opaque syntax and areas of the language that are implementation defined.

Briefly comment on the effectiveness of this approach as a means of improving the safety of automotive software.

[5 Marks]

[Seen problem]
In the course, we have looked at a range of language subsets that have been developed to support safety-critical systems. They limit programmers' ability to use 'dangerous' constructs [1 mark]. Ada arguably provides the best example of this – as a language developed to meet the needs of the defense industry [1 mark]. We do not rely on IV&V or good programming practice to avoid these high-risk constructs, the compiler/interpreter prevents them from being used in the first place [1 mark]. On the other hand, it is still possible to write dangerous code in any language [1 mark]. Hence the use of MISRA-C provides some support for the development of safety-critical software but it is not sufficient in itself and must be supplemented by training, sound management and well structured development processes [2 marks]. A range of other solutions are possible.

(b) A new initiative has been created to develop Automotive Grade Linux (AGL). The initial focus is on In-Vehicle-Infotainment but is gradually being expanded to instrument clusters and telematics systems. At present the group developing AGL has not decided on a formal position with respect to the standard ISO 26262.

Why might this create concerns for the future use of AGL?

[5 Marks]

[seen/unseen problem]
ISO 26262 is similar to IEC 61508, both of which have been considered during the course. They are standards that are intended to support functional safety and guide the use of programmable systems; 61508 is a general process standard but 26262 is specifically intended for Automotive applications [1 mark]. Within 26262, ASILs are use to distinguish different levels of integrity [1 mark]. The higher the ASIL then the greater care must be taken in development practices and techniques [1 mark]. One approach might have been to specifically target AGL to support specific integrity levels – however, the original focus on infotainment suggests that safety was not a primary concern [2 marks]. Hence there is a danger as the application of AGL spreads that it may be used at higher levels of integrity for which it was not originally designed [2 marks].

(c) The US National Highway Traffic Safety Administration defines five levels of automation: Level 0 (no automation); Level 1 (function-specific automation); Level 2 (combined-function automation); Level 3 (limited self-driving automation); Level 4 (full self-driving automation).

Explain the increasing problems that arise for the verification and validation of safety critical software as you move from level 0 to level 4 of this automation hierarchy.

[10 Marks]

[Seen/unseen problem]

At level 0, conventional IV&V techniques can be applied [1 mark]. This in itself raises enormous questions – answers might refer to Dijkstra's maxim, which we have looked at in class, characterizing the limits of testing [2 marks]. However, as we move to higher levels of automation a number of additional problems arise. One is associated with testing the range of automated responses that are possible [1 mark]. At level 2 – where automation can integrate different functions, it is important to identify broad interactions between those operations not only in isolation but also in combination; potentially leading to an exponential growth in test cases [2 marks].

Further problems stem from the need to identify interactions between higher levels of automation and environmental factors [1 mark]. In particular, interactions between autonomous functions and systems that operate under conventional modes of control [1 mark]. In other words, testing cannot be in isolation. What happens when a level 4 system must interact with levels 0-3 [1 mark]? This is at the heart of concerns that arise when human drivers violate traffic laws in the presence of autonomous systems [2 marks].

A final set of concerns surround the use of particular implementation techniques that may be needed to realize higher levels of automation. In particular, it is almost impossible to use standard IV&V approached with systems that make extensive use of AI and machine learning [2 marks]. We can prove properties of given code but it can be hard to ensure those properties will always hold when new test cases may force significant revisions in behavior [1 mark].

As with other questions, a range of solutions are possible and marks will be awarded in proportion to the quality and correctness of the argument.

2. (a) Top-down apportionment of risk can be used to set reliability targets for subsystems. For example, an overall reliability target was set for the Airbus A380 and then the risk apportionment was divided amongst individual sub-systems. Verification and validation activities were then used to ensure that those sub-systems could meet the target level of reliability before full manufacturing commenced.

Briefly describe the technical challenges that arise when using risk apportionment to guide the development of complex, safety-critical software.

[4 Marks]

[unseen problem]

Before a system is designed in detail it can be very difficult to establish any scientific or engineering basis for the precise apportionment of risk to specific subsystems [2 marks]. Instead, the allocation of the target levels tends to be guided by ad hoc or subjective criteria [1 mark]. Once the system is developed, it can be difficult then to generate enough data to determine whether or not a particular subsystem actually meets very high levels of reliability [1 mark]. If a system fails, this may not violate 1 in 10^6 requirements given a prolonged future period of error free operation [1 mark]. Another concern is that the overall system reliability cannot simply be the composition of individual components because errors might arise from problems in the interaction between subsystems that otherwise function in a reliable manner – these design errors do not obey the stochastic assumptions that are appropriate for more hardware oriented applications [1 mark].

(b) Scaling adjusts the risk apportionment of subsystems to reflect the relative scale of particular components compared to other subsystems. A number of factors can be considered when scaling the apportionment of risk, these include:

- physical size (area of plant or factory);
- size of workforce exposed to a hazard;
- throughput as an indicator of relative importance to company/process;
- inventory of hazardous materials or processes;
- potential impact on surrounding community.

Justify the use of each of these criteria for scaling the risk apportionment of subsystems.

[6 Marks]

[unseen problem]

The physical size (area of plant or factory) can be used as a scaling factor in a subsystem because it provides a crude approximation to the number of physical systems that might be contained within a process. It provides a very crude measure when considering tolerable thresholds for software related applications [up to 2 marks, given maximum of 6].

The size of workforce exposed to a hazard can be used as a scaling factor in a subsystem because it provides a high-level indication of the people who might be affected by the consequences of an adverse event. As with scale, this might be appropriate during the early phases of design and before any more formal risk assessment might be conducted. The workforce exposure obviously does not account for members of the public who could be affected [up to 2 marks, given maximum of 6].

The throughput of a company/process can be used as a scaling factor in a subsystem because again it provides a high-level proxy for the consequences of failure on those higher-level systems. It is arguably less useful as a measure of relative importance to safety [up to 2 marks, given maximum of 6].

The inventory of hazardous materials or processes can be used as a scaling factor in a subsystem because as we have discussed on the course, software itself cannot directly cause harm. This is most often associated with the equipment under control or with other energy sources [up to 2 marks, given maximum of 6].

The potential impact on surrounding community can be used as a scaling factor in a subsystem; as we have seen other measures tend to focus too narrowly on those involved in a process. However, higher levels of compensation are reserved for those who are injured outside the immediate workforce [up to 2 marks, given maximum of 6].

These answers are probably more detailed than I might otherwise expect under exam conditions.

(c) Standards such as IEC61508 associate higher levels of risk with incidents that might affect the general public rather than direct employees of an organization. Higher levels of risk are also associated with incidents that lead to multiple casualties from a single adverse event.

Explain why these distinctions are made and outline the impact that they can have for the development of safety-critical software in complex systems.

[10 Marks]

[seen/unseen problem]

IEC61508 uses risk assessment to derive the Safety Integrity Level of functions that are intended to prevent adverse events [1 mark]. The higher the level of risk then the greter the level of integrity is associated with software and other systems that are intended to either reduce the likelihood or mitigate the consequences of a hazard [2 marks]. This question focuses on two factors that are seen to exacerbate the consequences of an adverse event [1 mark]. Members of the public who may be injured or killed in an incident are treated as a special case because unlike employees they are more likely to be seen as third parties who are unaware of the risks associated with any process nor may they profit directly from it [1 mark]. In contrast, an employee is seen to be more aware of the hazards, have a stake in controlling them and also benefit from the associated processes [1 mark]. Incidents that affect the public and in which multiple injuries/fatalities occur are also widely seen to be more serious than individual incidents both from the standpoint of the public and politicians – in each case regulators are likely to request higher sanctions in the aftermath of an accident or incident [2 marks].

These higher consequence events must be considered during risk assessment and the allocation of SILs to associated mitigations (including software) – in consequence greater resources and attention will be paid to hazards that might affect the public or lead to multiple casualties [2 marks].

Other answers can also be awarded full marks.

3. (a) NASA GB 8719.13 Software Safety Guidebook contains the requirement that "Each NASA project, regardless of its level of safety-criticality, must perform an IV&V evaluation at the beginning of the project, and whenever the project changes significantly".

Briefly explain why it is important to plan for IV&V both at the start of a project and when there are significant project changes.

[4 Marks]

[Unseen problem]
Independent Verification and Validation consists of a broadly defined testing programme that will both establish the value of requirements and also test to determine that those requirements have been met at various stage of implementation and development [2 marks]. It is important to plan IV&V at the start of a project to ensure that there are sufficient resources available to complete associated activities through to deployment [1 mark]. Otherwise, key safety requirements may not be preserved in the final system [1 mark]. It is important also to plan for IV&V during changes after deployment again because safety requirements may be undermined by those changes and increase the risks of adverse events [2 marks].

(b) NASA guidance GB 8719.13 also advocates a change management process that includes the use of cross-indexing to link software changes, requirements, code, component versions and testing.

Why is it important to support each of these forms of traceability? Explain how safety cases can be used to assess the impact that software changes have upon the overall safety of a complex system.

[6 Marks]

[seen/unseen problem]

It is important to support traceability so that regulators, customers and other developers [1 mark] can see how high level requirements are satisfied within the code and then how testing is used to demonstrate that the requirements are met as intended [1 mark]. When software is updated or changed, these relationships may have to be revised to show how the amended software components continue to meet safety objectives [1 mark]. Similarly, if requirements change the previous code may not meet the new constraints [1 mark]. New tests may identify flaws in the relationships between requirements and code [1 mark].

Safety cases, especially those developed in GSN, provide a high level argument using the evidence from testing and other forms of IV&V to show that an implementation meets particular requirements [2 marks]. If test data changes then you can trace the links in the diagram to show what areas of a safety argument might be undermined by any change [1 mark]s. Conversely, if a requirement changes these same links can be used in the opposite direction to determine whether any tests now need to be repeated or revised [2 marks].

(c)c) The US Federal Aviation Administration (FAA) recently issued a circular on 'tool qualification guidance' that includes the following requirement:

"If your legacy system software was previously approved using ED-12 / DO-178 or ED-12A / DO-178A, and you intend to use a new or modified tool for modifications to the legacy system software, use the criteria of ED-12C / DO-178C, section 12.2, to determine if tool qualification is needed. If you need to qualify the tool, use the software level assigned by the system safety assessment for determining the required Tool Qualification Level (TQL), and use ED-215 / DO-330 for the applicable objectives, activities, guidance, and life cycle data. You may declare your qualified tool as having satisfied ED-215 / DO-330 and not the legacy system software as having satisfied ED-12C / DO-178C".

Write a brief technical report for a project manager explaining what this means for future software development projects involving legacy code.

[Unseen problem]

There are many different solutions to this question – at its heart is the growing idea in standards such as DO-178C that you have to approve not just the code you develop but also the tools you use to develop that code [2 marks] – such as model based development environments [up to 2 marks for examples of the tools mentioned here]. This raises complex questions when legacy code might have been developed using tools that themselves did not meet the standards set out in more recent requirements [2 marks]. The FAA approach states that when you use new tools to modify legacy code then those tools should be assessed using the more recent standards to a level appropriate to the criticality of the legacy code [2 marks[even though that legacy code itself might not meet all of the requirements for those standards (eg 178C) [2 marks]. It is important to note that these requirements apply to the use of tools throughout the development lifecycle [1 mark].

[10 Marks]

4. Write a technical report for the senior management of a safety-critical software company explaining the problems that can arise when attempting to improve safety culture across the software supply chain.

Your answer should explain how some of the risk assessment techniques that have been presented in the course might be used to focus resources on those software suppliers, which have the greatest impact on **system safety**.

[20 Marks]

[Unseen essay]
Many different solutions can be provided for this question. Safety culture describes the everyday working practices that companies use to ensure that they mitigate risks so that they are as low as reasonably practicable (ALARP within the UK approach) [2 marks]. Solutions might refer to the use of safety management systems – using the insights from previous adverse events to inform the risk assessments that then direct appropriate mitigations [2 marks]. They might also look at resilience engineering with its focus more on reinforcing successful interactions than on learning from a small number of adverse or near miss events [2 marks]. In either case, good answers must address the problems of encouraging strong safety culture beyond the boundaries of a single organization [1 mark]. Supply chain requirements are increasingly important given the rise of outsourcing within safety-related companies – this creates concerns because you outsource the service and not the risk [2 marks]. Hence, it may be necessary for safety related organisations to ask more detailed question about implementation mechanisms and any associated safety/reliability measures than in more general industries [2 marks].

The reference to risk assessment in the second part of the question suggests that companies can use measures of frequency and consequence to direct attention to the safety culture of key elements within the supply chain [2 marks]. One approach would be to use a functional decomposition of the service provided by external companies [1 mark]. Another would be to look at key components or artifacts within their own production processes [1 mark]. Either way, the expectation would be that more critical suppliers provide additional evidence of higher levels of safety maturity than those with less important roles [2 marks]. Evidence might refer to SMS documentation, training records, contingency and business continuity plans etc [3 marks]. At higher levels of criticality, failure to provide such evidence might lead companies to seek new suppliers or to take functions in house [2 marks]. At the very least, additional incident reporting mechanisms or audit procedures might be used [2 marks].