A comparison of three alternative means for safety critical control

André A. Hauge
PhD student at University of Oslo
Researcher at Institute for Energy Technology, OECD Halden Reactor Project
Halden, Norway
andre.hauge@hrp.no

***Abstract*—This article elaborates upon the strengths and weaknesses tied to three alternative controllers. The context is safety critical control applications. The three kinds of controllers are the human controller, the conventional programmable controller, and the adaptive programmable controller. The safety aspect is particularly emphasised due to the strict requirements put on critical systems for documentable evidence that a certain level of safety is achieved.**

***Keywords*-adaptive; assessment; safety; critical; risk;**

I. Introduction

This article elaborates upon the strengths and weaknesses tied to three alternative controllers, namely:

a) Human controller
b) Conventional programmable controller
c) Adaptive programmable controller (programmable electronic with software using e.g. artificial intelligence techniques)

The three kinds of controllers are compared with respect to:

1) Dependability: Is the controller dependable?
2) Hazard recovery: Does the controller provide ability to mitigate unforeseen hazards?
3) Efficiency: Does the controller provide ability to perform a well defined task with speed and accuracy?
4) Optimization: Does the controller provide ability to optimize operations in dynamic environments?
5) Verifiability: Can the controller behaviour be adequately verified?

This article is structured as follows: Section I introduces the objects to be compared and the properties used for comparison. Section II elaborates upon the chosen properties for each of the alternative controllers. Section III discuss and summarises and section IV concludes.

II. Alternative controllers and evaluation properties

A. Human control

1) Dependable: Cook and Woods state in [1] that incident studies in medicine show that a percentage of 70% to 82% of reported failures can be attributed to human error, and that similar studies within aviation attribute 70% of the incidents to crew error. They also state that incident surveys in a variety of industries attribute similar percentages of critical events to human error. Hollnagel in [2] estimates the human error contribution to accidents typically to be between 70% to 90%. Evidently, humans are not particularly reliable.

2) Hazard recovery: As of today, humans are superior to any technology when it comes to analytical skills, creativity, flexibility, adaptability, and other typical attribute of humans. These skills provide the means needed to cope with changes in environment or operating conditions.

As pointed out by Leveson in [12], human operators are included in complex systems because, unlike computers, they are adaptable and flexible, the human error being an inevitable side effect of this flexibility and adaptability. In addition, Leveson points out that error reports capture the negative events, where the positive effects of human intervention does not shine through. Leveson provides several examples in [12] of events where humans have restored operation when technology failed.

Empirical studies providing failure data for safety critical applications are hard to find; empirical studies on positive variance in such applications are absent. On the issue of providing hazard recovery abilities, it is generally accepted that humans are very strong.

3) Efficiency: Assuming a task is well defined, demanding many computational loops, and a high degree of accuracy, human control is not an efficient means for the job. It does not matter if the task needs to be performed in a short period of time, or a long period of time, computers outperform humans with respect to speed and stamina. Humans are not fast enough to perform many computations in a short period of time, and the likelihood of computation error increases with time.

4) Optimization: If a task is loosely or abstractly defined, calling for interpretation or adapting the implementation for a particular context, humans are very suitable. Finding new ways to reach a goal or improve some process are typical human skills.

5) Verifiability: Human performance is affected by different factors like physical and psychological health, age, emotions and other performance shaping factors. As opposed to computerised controllers it is not possible to inspect and fully determine the successfulness of a human task. Methods

for addressing human performance can be found within the field of Human Reliability Studies (HRA). In [15], three well known HRA methods are evaluated. The three methods all aim at determining how often a human or team of human operators will fail in a task, and showed a general precision of 72% within a factor of 10 of the true human error probability.

B. Conventional programmed electronic control

With a conventional programmed controller, it is meant a controller consisting of well proven programmable hardware implementing software using commonly accepted techniques. What is well proven and what is acceptable is specified in e.g. domain specific safety standards. Conventional controllers will off course vary in dependability based on the design, materials and the development method used.

1) Dependable: Laprie [3], in 1993, describe how traditional systems implementing fault tolerance improve with two orders of magnitude in terms of time to failure compared to non fault tolerant systems; mean time to failure is 21 years as opposed to 6 to 12 weeks for the referenced systems. Assuming our conventional controller is designed with fault tolerance engineering principles, according to Laprie [3] the generally recognised bottleneck for dependability is software, constituting 65% of the failure sources. The reason for this is that computer systems involved in such applications have become increasingly tolerant to physical faults. Littlewood and Strigini [4], in 2000, provides some relevant failure rate data for safety critical systems. They mention that reliability data for critical systems are rarely published. Two representative references are provided though, one on operating experiences of nuclear I&C (Instrumentation & Control) systems [5] and one on avionics systems reliability based on calculation from FAA (Federal Aviation Administration) records [6]. Failure rates for the two were in the range 10E-7 to 10E-8.

2) Hazard recovery: Conventional controllers, developed according to good engineering principles, can certainly possess robustness with respect to changes in environment, but they will never be adaptive. In order for conventional controllers to provide high performance in changing environments, the main strategy is to increase the controller robustness. The successfulness of increasing the robustness depends on the designer's ability to foresee events that might occur during operation, the effect being that the controller will never get more robust than accounted for during design.

3) Efficiency: One of the strengths of computerised controllers is computational power. Its speed, accuracy and ability to provide continuous operations in computing tasks are unprecedented. At least this is valid as long as the computing task and operating environment is well defined and known.

4) Optimization: In uncertain or slowly changing operating environments conventional control have clear short comings. Schumann and Gupta state in [9] that conventional systems have proven ineffective to deal with catastrophic changes or slow degradation of complex, highly non-linear systems like aircrafts, spacecrafts, robots or flexible manufacturing systems.

5) Verifiability: Assuming an error free software implementation, the software would not be a contributor to system failures. Hardware on the other hand, at one point in time error free, will experience failures. One reason is the inevitable effect of material degradation in hardware, which at some point in time may result in arbitrary behaviour. Issues related to hardware failure are commonly mitigated by redundancy and other means for achieving fault tolerance.

Focusing on software aspects, software faults are systematic. A deterministic software component failing at some specific input will fail again if fed with identical input. That being said, the intention of a well defined software engineering process is to avoid faults being introduced, remove latent faults or by design build in detection and tolerance to faults. Once experiencing a failure, one expects that the triggering fault will be removed or in any other way handled. The remaining faults, those not experienced will manifest as failures random. Assuming we have found a failure rate for some software component, this rate describes the rate at which unresolved latent faults are triggered and become failures. The problem however is to provide assurance before commissioning that such a level has been achieved. Domain specific standards provide requirements and/or guidance on the development and verification of critical systems. There are two prevailing styles for arguing safety, either process assurance based (focusing on the development process) or product evidence based (focusing on the operational behaviour). Differing in the philosophy on how safety shall be argued, the intention is however to show through various claims, arguments and evidence that a system is adequately safe for its purpose. The system failure rates documented by [5] and [6] provide confidence that the engineering principles advocated by laws, regulations and standards have had an effect in producing highly reliable and safe systems.

C. Adaptive programmed electronic control

1) Dependable: Adaptive programmed controllers do not necessarily deviate with respect to hardware or software platforms with what used in traditional programmed systems. The difference is in the techniques applied to fulfil some function. In that manner, the high level of dependability achieved in traditional critical controllers should be achievable for adaptable controllers too. So an adaptive controller has basically the same failure characteristics as conventional controller with some additional risks. In conventional controllers, simplicity and transparency are valuable properties in order to provide verifiable and certifiable software. Adaptive controllers, developed with techniques which produce

controllers that are not easily analysable, adds an uncertainty with respect to if there are any new hazards introduced or any increase in the likelihood of a hazard occurring as opposed to use a traditional system.

An adaptive system is dynamic of nature; it is a feature that the system will change during operation. This non-determinism do not imply that the system is not safe. The challenge with non-determinism in general is to handle its potential negative effects, providing a need for proper assessment and verification methods. The strict requirements tied to critical systems demands that a change proposed to a system is not effectuated before it has been assessed that there are no negative impact on safety of this change. Utilising on-line learning adaptive components, one must assure that any change during its lifetime will not have an adverse effect on the system.

2) Hazard recovery: A representative example is described by Schumann and Gupta in [9]. [9] refers to a NASA project called IFCS (Intelligent Flight Control System). The IFCS project utilised an on-line adaptive neural network in order to optimize aircraft performance during normal and adverse conditions. The neuro-controller was designed to enable a pilot to maintain control and safely land an aircraft that had suffered a major systems failure or combat damage. Control surface failures may conflict with the design assumptions of an aircraft flight control system, with the effect that it is unable to handle the situation. The IFCS neuro-controller compensate for discrepancies between a reference model of the flight dynamics to any ”new” flight dynamics model in order to maintain the best possible flight performance. The adaptive neural network software ”learns” the new flight characteristics, on-board and in real time, thereby helping the pilot to maintain or regain control and prevent a potentially catastrophic aircraft accident.

In adverse and unpredictable situations, adaptive systems offer abilities that are difficult to implement with conventional techniques.

3) Efficiency: With regard to computational power, adaptive programmed systems possess the same abilities as conventional programmed systems. The difference is not the platform which provides these abilities; it is the techniques used to implement some control function. In addition many of the commonly applied techniques within this field, e.g. neural networks, fuzzy logic or genetic algorithms, are parallel computing friendly.

4) Optimization: Schumann and Gupta in [9] refer to several successful studies with respect to performance using on-line learning adaptive control, by the use of neural networks. They state that traditional control has proven ineffective to deal with catastrophic changes or slow degradation of complex, highly non-linear systems like aircrafts or spacecrafts, robots or flexible manufacturing systems. The need for adaptable control technology is addressed by several authors, particularly by avionics and aerospace researchers as a means to provide increased performance, see e.g. [7]–[9].

5) Verifiability: Schumann and Gupta in [9] state that although the neuro-adaptive controllers offer many advantages, they have not been used in mission- or safety-critical applications, because performance and safety guarantees cannot be provided at development time.

The increased system complexity which the adaptivity feature adds is problematic on the issues of assessment. Assurance must be provided before commissioning that any change due to the adaptability feature does not lead to any new hazards or in any way increase likelihood of existing hazards during its operational lifetime.

Dijkstra addressed the problem of non-determinism with a formal approach to derivation of programs with guarded commands back in 1975 [14]. By the guarded commands construct, statements are executed only if the guard is true, if the guard is false, the statement will not be executed. So, although efforts on the topic of handling non-determinism has been studied decades ago, current papers on verification of adaptive and reasoning systems, see [7]–[11], does not provide statements about consensus in methods or some prevailing method solving the problem. In fact, Jacklin et al. in [8] and Schumann and Gupta in [9], state that the non-determinism is still a challenge. Jacklin et al. further state that the real verification and validation problem faced by learning systems is proving that the learning process is convergent and repeatable, that the convergence rate is acceptably fast and that the learning process is stable.

III. Discussion and Summary

Comparing different controllers by their relative risk is biased unless we also compare their respective positive variance. In order to assure that accidents do not occur, the *Resilience Engineering* concept, [13], both focus on how to handle anticipated errors and how the positive side of being flexible and adaptive provides ability to handle unforeseen events. In Hollnagels book on the concept of Resilience Engineering, [13], it described how improvements of safety have been dominated by hindsight, reactive rather than proactive. The book advocates principles strengthening the ability of systems to anticipate and adapt to the potential for surprise and failure. The properties used to compare the three controllers, described in the introduction, reflects this philosophy in that the *Hazard recovery* and *Optimization* criteria focus on the different controllers ability to provide positive variance given a fault situation or optimal control given a normal control situation.

In comparing three different kinds of implementation of a critical control function there certainly are differences in risk. The human controller offers high degree of flexibility and adaptability to changes in environment and ability to handle unforeseen events. Although there is a lack of empirical studies on the positive effect of human intervention

in emergency and accident situations, authors like Leveson [12] argue that this is the case. The downside of human performance is low reliability, human error causing over 70% of the failures according to incident studies.

Traditional programmed systems offer high degree of dependability. Empirical studies show a very low system failure rate, in the range 10E-7 to 10E-8. Individual system component failures, being a problem in the past in that component failure propagated to system failure is largely mitigated by fault tolerance techniques. These systems are highly analysable in the sense that a white box assessment style is possible. Although traditional programmable systems offer much in terms of dependability, they do not offer much with respect to flexibility and adaptability.

Adaptive systems provide the strengths of computerised systems like computational power and accuracy. They also offers flexibility and adaptability, a valued strength in many application areas, see [9]. Although it is possible to assess an adaptive system by looking at its internals, there is a lack of commonly acceptable methods to do so. The challenge in assessing and verifying the absence of adverse effects make certification problematic, the effect of which slows down utilisation.

Table I summarise what we addressed in section II. A grading colour and numbering scheme is used to qualitativly compare the tree controllers with respect to the different properties described in section I. The scheme should be interpreted as follows:

- Blue colour, VH: Very high degree
- Green colour, H: High degree
- Yellow colour, G: Good degree
- Orange colour, L: Low degree
- Red colour, VL: Very low degree

We justify the colouring in Table I as follows:

- Dependability: Studies show that humans are error prone and that conventional systems obtain high dependability. Adaptive systems add complexity, unresolved with regard safety assessment, but there are no indications that adaptive systems should be drastically more hazardous than conventional controllers.
- Hazard recovery: The ability to handle unforeseen events is clearly different of the different controllers. Humans are superior. Robust design in conventional control provides some tolerance to unforeseen events but clearly limits itself to what can be accounted for during design phase. Adaptive systems provide better hazard recovery ability as they have some degree of flexibility; on the other hand this flexibility has some bound established during design.
- Efficiency: Ability to handle huge amount of data fast and accurate is a characteristic of both adaptive and conventional systems.
- Optimization: Assuming the degree of change on some plant and its environment is moderate/within-design, one might expect that the difference in performance between a human and an adaptive controller would be smaller than in the events covered by the *Hazard recovery* property which includes failure events.
- Verifiability: Conventional systems can be assessed by different commonly accepted means, e.g. testing or formal reasoning. Adaptive systems can possibly be assessed with similar methods as conventional systems; they offer the possibility to look inside. For humans a white box style assessment is not applicable, verification of behaviour typically involve probabilistic reasoning. Human reliability assessment methods do not provide an exact science as such although probabilistic reasoning methods on human performance provide very useful information.

Controller	**Dep.**	**Haz. rec.**	**Eff.**	**Opt.**	**Ver.**
Human	VL	VH	VL	VH	L
Conventional	VH	L	VH	L	H
Adaptive	H	G	VH	H	L

Table I
COMPARISON SUMMARY

IV. CONCLUSION

There is a need for adaptive controllers. Adaptive programmed controllers provide an opportunity to yield reliable as well as flexible behaviour, studies on the subject has proven so. High reliability should be obtainable with means similar to what used in conventional systems, there is no basic difference in the hardware and software platform used to develop and implement such systems. The resilience towards unanticipated events or system variability can be obtained by the inherent flexibility offered by adaptive programming techniques. Whether the reason for utilising adaptive systems is increased safety or increased performance, there are unresolved issues related to assessment and verification of no adverse effects. The lack of commonly accepted methods on the assessment of adaptable systems may hinder the certification which again slows down utilisation. This calls for further research on the subject.

ACKNOWLEDGMENT

This paper is a preliminary result of a PhD study on the use of adaptive control technology in a safety critical context. The PhD study is scheduled from September 2008 to August 2012. Focus is on safety issues and the ability to assess and prove the absence of adverse effects. The author would like to thank supervisor Dr. Ketil Stølen at University of Oslo, Norway and Dr. Bjørn Axel Gran at Institute for Energy Technology, Halden, Norway for their guidance and helpful comments.

REFERENCES

[1] R. I. Cook and D. Woods, *Operating at the Sharp End: The Complexity of Human Error* In Bogner MS (Ed.), Human Error in Medicine, 1994

[2] E. Hollnagel, *Human reliability analysis: Context and control*, Academic Press, 1993

[3] J.C. Laprie, *Dependability of Computer Systems: from Concepts to Limits*, 12th International Conference on Computer Safety, Reliability and Security, 1993

[4] B. Littlewood and L. Strigini, *Software reliability and dependability: a roadmap*, Proceedings of the Conference on The Future of Software Engineering, 2000

[5] A. Laryd,*Operating experience of software in programmable equipment used in ABB Atom nuclear I&C application*, Advanced Control and Instrumentation Systems in Nuclear Power Plants, 1994

[6] M. Shooman, *Avionics Software Problem Occurrence Rates*, in Proceedings of International Symposium on Software Reliability Engineering, 1996

[7] G.S. Tallant and P. Bose and J.M Buffington and V.W. Crum and R.A. Hull and T. Johnson and B. Krogh and R. Prasanth, *Validation & verification of intelligent and adaptive control systems*, IEEE Aerospace Conference, 2006

[8] S.A. Jacklin and J.M. Schumann and P.P. Gupta and M. Richard and K. Guenther and F. Soares, *Development of Advanced Verification and Validation Procedures and Tools for the Certification of Learning Systems in Aerospace Applications*, Infotech at Aerospace conference, American Institute of Aeronautics and Astronautics, 2005

[9] J. Schumann and P. Gupta, *Monitoring the Performance of a neruo-adaptive Controller*, Proceedings of the 24th international workshop on Bayesian inference and maximum entropy methods in Science and engineering, 2004

[10] S.A. Jacklin and M.R Lowry and J.M. Schumann and P.P. Gupta and J.J. Bosworth and E. Zavala and J.W. Kelly and K.J. Hayhurst and C.M. Belcastro, *Verification, Validation, and Certification Challenges for Adaptive Flight-Critical Control Systems Software*, Guidance, Navigation, and Control Conference and Exhibit, American Institute of Aeronautics and Astronautics, 2004

[11] B. Taylor and M. Darrah and C.Moats, *Verification and validation of neural networks: a sampling of research in progress*, Proceedings of SPIE, 2003

[12] N.G. Leveson, *Safeware: system safety and computers*, Addison-Wesley Publishing Company, 1995

[13] E. Hollnagel, *Resilience engineering: concepts and precepts*, Ashgate Publishing, 2006

[14] E.W. Dijkstra, *Guarded commands, nondeterminacy and formal derivation of program*, Communications of the ACM, 1975

[15] B. Kirwan, *The validation of three Human Reliability Quantification techniques, THERP, HEART, and JHEDI: Part III - Practical aspects of the usage of the techniques*, Applied Ergonomics, 1997