Using Resilience to Inform Autonomous System Reliability Assessment: A Concept for Autonomous Ships

Bryan M. Gorman and Craig M. Payne

ABSTRACT

This article proposes a resilience engineering approach to augment reliability analysis of complex autonomous systems. This change of approach does not imply a change in reliability metrics per se but requires the addition of resilience considerations to reflect the potentially extreme range of operating environments. In essence, the performance metrics that form the basis for reliability (i.e., preventing anticipated system failures) will have to incorporate measures of a system's resilience, both to rare or unanticipated disturbances and to compromised functioning, as an integrated capability.

INTRODUCTION

Automation in complex systems implies satisfactory performance of a number of linked processes in a changing environment without human intervention. Performance may include processes that control, constrain, or limit the physical state of the system. In contrast, autonomy implies that performance is self-governed.[1] A self-governing process by its definition requires application of rules specified at a level that is more abstract than the physical situation. As such, an autonomous system requires the following types of processes:

- *Observation*—Both sensing and abstraction of concrete events
- *Orientation*—Abstract awareness of operating conditions, both internal and external
- *Decision*—Assessment and selection of alternative actions or plans against one or more abstract objectives, constraints, and limits

For any subsystem of a system, reliability is a measure of probability that the subsystem functions to an expected level when employed.[2] The level of functioning expected is typically defined in relation to assumptions underlying the analysis of peer-level or higher-level functions that are linked to the subsystem under study. In this way, reliability allows subsystem capability analyses to be mutually decoupled into composable units that can be integrated readily into a full-system assessment. This construct fails to capture the mitigating effects of subsystems necessarily integrated into autonomous systems whose function is to sense and respond to events that put its functioning at risk.

Engineering architectures have been defined for an unmanned commercial shipping concept known as Maritime Unmanned Navigation through Intelligence in Networks (MUNIN), which was developed under a 5-year European maritime industry/academic partnership ending in 2016 (see http://www.unmanned-ship.org/munin/). In the United States, the first medium displacement unmanned surface vehicle (MDUSV), named Sea Hunter and pictured in Fig. 1, was launched in April 2016 and is undergoing operational testing administered by the Office of Naval Research. Sea Hunter is the first unmanned ship capable of conducting an extended oceangoing voyage in completely autonomous mode.

Figure 1. Sea Hunter. (Credit: DARPA.)

Reliability, Resilience, and Autonomous Systems

Systems are typically engineered to meet a stated reliability within a range of operating conditions. Systems do not pass acceptance testing if they cannot perform reliably at the boundaries of that range of conditions.[3] Additionally, systems are typically delivered with warnings and instructions regarding operation outside the defined range, particularly if safety is an issue. Ideally, the user of a system ensures that operating conditions are within the appropriate limits before and during use of the system.

In recent years, interest has grown in creating systems that are not only reliable but also resilient (i.e., capable of responding to changes in their environment so that they can continue to operate).[4] Many approaches to resilience include human operators to identify potential failure conditions and to initiate work-around procedures.

In the case of complex autonomous systems, by contrast, a user may not be available to monitor operating conditions and to respond by securing any functions that pose an undue risk to continued operation or taking measures to protect the system physically. Whereas a user may be able to anticipate an imminent disruptive event, an autonomous system operating for an extended period must have resources available to secure its operation and to protect itself under a range of conditions far wider than its nominal operating range. Meaningful reliability measures for autonomous systems, in contrast to traditional probabilistic risk analysis measures, must be

- dynamic, with explicit treatment of disturbance or functional anomaly detection, reaction, and recovery processes;
- interdependent, with explicit treatment of interactions among contributing factors; and
- potentially emergent, with performance factors arising from complex linkages that defy simplistic causal modeling.

Whereas reliability analysis of physical or electronic components is understood to be a complex endeavor requiring a detailed analysis of the physical processes involved in a component's functioning, the type of reliability analysis considered adequate for a system of components is typically a more traditional fault tree analysis.[2] However, this method does not accurately account for mechanisms in place to ensure that faults are contained and functioning is restored. Instead, fault tree analysis relies on an assumption that risk probability and risk severity are analytically separable effects. For an autonomous system, however, they are generally not separable and therefore require a more integrated representative model.

Reliability and Control

A useful way to visualize the difference in approaches to reliability analysis is to consider the difference between open-loop control and closed-loop control.[5] Simple depictions of both are shown in Fig. 2. A system whose operation follows a prescribed procedure using only initial (feedforward) conditions is an open-loop control system. Such a system may have synchronization timing, but its operation is not influenced by conditions it affects. If a disturbance enters the operating environment, the program cannot respond to it. The reliability of such a system can be captured with a static analysis, such as a fault tree, because reliability is related only to the system's ability to perform the set procedure to achieve the desired output under the initial conditions and a prescribed range of disturbances.

In contrast, in a closed-loop control system, a system controller reacts as a function of a desired goal state and of feedback regarding observed conditions under its influence. In this case, if a disturbance enters the operating environment, the controller can sense it and respond appropriately. The reliability of such a system is related to its ability to gain or maintain a state within acceptable tolerance around the goal. Working through such relationships requires a dynamical systems analysis. In autonomous systems, where there is no human intervention even at the highest levels of control, every system necessarily operates as a closed-loop control system.

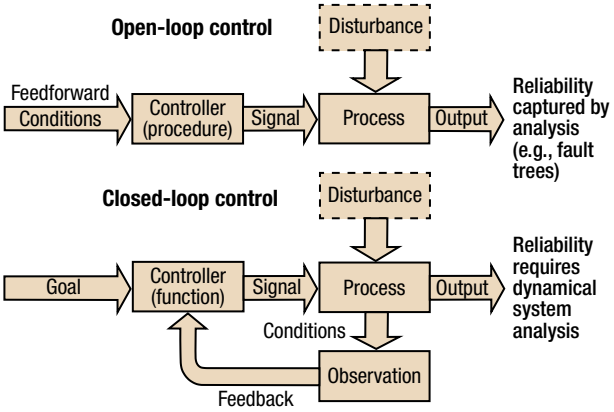

Figure 2. Open-loop and closed-loop control.

Resilience Informs Reliability

This article proposes a resilience engineering approach to assess the reliable performance of complex autonomous systems. This change of approach requires a reassessment of the traditional reliability metrics both at the system and subsystem levels. The resulting resilience-informed metrics are measures of how effectively subsystems can uphold the overall system's productive functioning under adverse conditions.

In resilience engineering, a failure is not simply accepted as a risk with some contained probability. It is instead an event to be assessed for its impact on controlled operation and to be addressed accordingly by an integrated subsystem. In this way resilience focuses attention on the many small, not yet catastrophic, disruptions or failures for which recovery is possible and potentially necessary. Autonomous systems need to be engineered to handle anticipated changes in the operating environment, including extreme variation in conditions or acute disturbances or disruptions. Autonomous systems may also have to respond to changing mission needs while in operation. Such changes may redefine the nature of their controlled operation or the criteria for their success. Even if a system is engineered against every anticipated disruption, unanticipated events may require the system to maintain or regain control robustly on the basis of very limited information.

When resilience is viewed as a process of control, whether it is control of system performance, of key aspects of the operating environment, or of both, measures of resilience for a complex autonomous system are strongly coupled to measures of the degree that the system can maintain or regain control.[6] A system whose performance is in control is operating to the full extent of what its operating environment allows. A system whose performance is out of control is operating outside its capacity to limit risk to its future performance to an acceptable level.

CONTEXT: AUTONOMOUS SHIPS

Setting requirements for resilience demands a thorough search for situations that pose the greatest risk to the system's safe continued operation. Risk analysis scores hazards on their rate or probability of occurrence and the severity of consequences given an occurrence. Often the probability is rated on a four-to-seven-category scale ranging from improbable to frequent,[7] where each successive category either follows a logistic probability scale or traverses a power of 10. Likewise, the severity is also rated on a four-to-seven-category scale ranging from negligible to catastrophic,[7] often following a geometric progression of value lost, such as a power of 10. Hazards with an unacceptable combination of high probability and severity are targeted for engineering development of mitigation systems to reduce the overall risk.

Subsystem Functions

To provide a framework to identify hazards to safe operation, system functions are organized by functional groups with limited linkages between groups. In this way, system functions are envisioned as a loosely coupled set of subsystems with qualitatively distinct sets of hazards, most of which can be assessed without invoking a complex monolithic system model. Common functional groups applicable to autonomous ships would likely include the following:

- Voyage planning, execution, and monitoring
- Maneuvering and avoidance
- Observation of environment, ships, and other objects
- Hull integrity and stability, bilge and ballast
- Emergency handling (firefighting, flood control, etc.)
- Propulsion and steering
- Power generation and distribution
- Logging, reporting, and communications
- Security and access control

Architectures based on these functional groups have been published for the MUNIN concept.[8]

Operating Conditions

Disruptions to safe operation are likely sensitive to one or more external and internal operating conditions. To estimate the associated risk accurately, the associated probability and severity measures must be assessed under all applicable conditions, weighted by the likelihood of their occurrence. External conditions are largely random with some degree of predictability, but some internal conditions may persist, such as a component in a state of degraded capability for which restoration, repair, or replacement would require human intervention.

Common operating conditions applicable to autonomous ships would likely include the following[8]:

- Scheduled unrestricted cruise and maneuvering
- Cruise and maneuvering under limitations due to mission requirements
- Cruise under limitations due to degraded capability
- Traffic detection and collision avoidance
- Ice or object detection and avoidance
- Legally restricted navigation
- Weather routing
- Low visibility
- Loss of communication or GPS

- Propeller fouling
- Emergencies (collision, flooding, fire)

Combinations of these conditions provide the context to assess a system's ability to maintain its essential functioning.

ELEMENTS OF RESILIENCE

In resilience engineering, it is necessary to understand that events will put the control of system performance at risk. It is equally necessary to understand the causes of those events in terms that will lead to effective mitigation strategies. A model for system performance and control must serve as a frame of reference to understand issues posed by a range of untoward events, enabling identification of the critical measures that determine that the system is not in control.

When devising measures, it is preferable to strive for simplicity but also to recognize the implications of complex linkages in the system. Simple measures generally lead to simpler designs for mitigation, but overly simple designs may lack the robustness required to handle the expected range of untoward events. The desired model for controlled system performance not only informs the engineering of mitigation systems but also becomes the basis for the system's own awareness and control of state—anticipating events, securing functions, and restoring control under stressing or compromised operating conditions.

In any unplanned and uncontrollable event that can be anticipated, there is potential for operational or functional failures. These failures can be categorized as untoward events, or disturbances, which can lead to associated subsequent losses of capability or operational capacity, or disruptions. Disturbances will always occur; however, disruptions can be avoided.

Three Resilience Strategies

Resilience can be separated into three largely distinct strategies: prevention, response, and recovery. Their effects are depicted in Fig. 3 (adapted from Ref. 4). With adequate preparation and resources, disruptions can be prevented. The measure of prevention is the degree of reduction in the frequency of a given disruption resulting in a loss with similar magnitude and recovery time.

If a disruption does occur, actions may be taken to reduce the probability and/or severity of further losses. These actions make up a response and are measured by mitigation of the fractional loss of one or more operating capacities for disruptions with similar frequency and recovery time. The response is optimal if the lowest point of the operating capacity is maximized.

After the response, damage can be assessed and efforts to regain full functionality can begin. This recovery process largely begins after the loss of operating capacity has halted, but it can overlap with the response period. Recovery is measured by both the amount of operating capacity recovered and the rate of recovery for disruptions with similar risk. All three aspects of resilience play key roles in determining the ability to maintain and restore operating capacity in the face of potential disruptions.

Resources for Resilience

Maintaining or regaining controlled operation typically requires sufficient resources of at least five mutually dependent types[9]: (*i*) time, (*ii*) knowledge, (*iii*) readiness, (*iv*) material, and (*v*) energy. Figure 4 illustrates the role of these resources in the context of maintaining closed-loop control of operations. The figure's inner cycle is an observe–orient–decide–act loop[10] depicting common identifiable processes that together lead to controlled progress toward an intended goal. Each process leads to a product, such as the events composing the state of the operational environment, including internal states, such as the state of the system's readiness to execute one or more alternative plans of action. Observation leads to information feedback, which is used to build a construct, or model, of the situation as needed for effective control. This construct is informed by higher-level knowledge of the situation, which allows generation of useful assumptions about

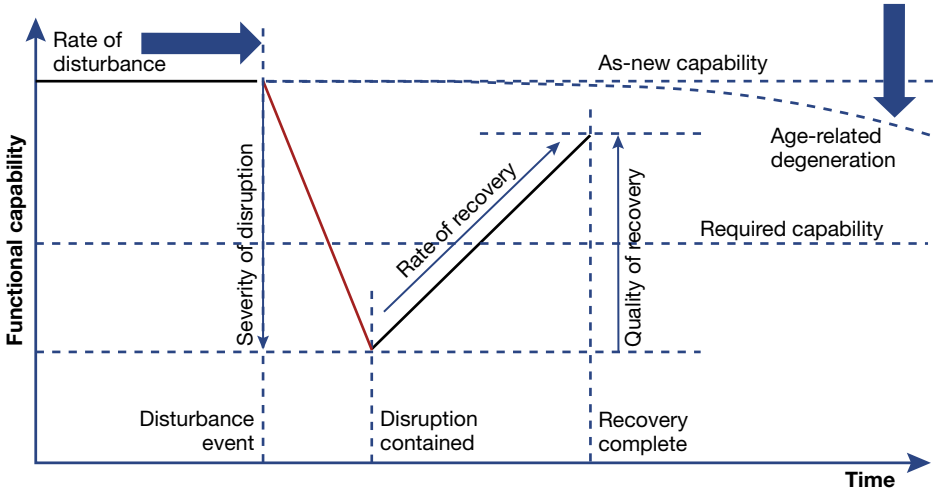

Figure 3. Elements of resilience. (Adapted from Ref. 4.)

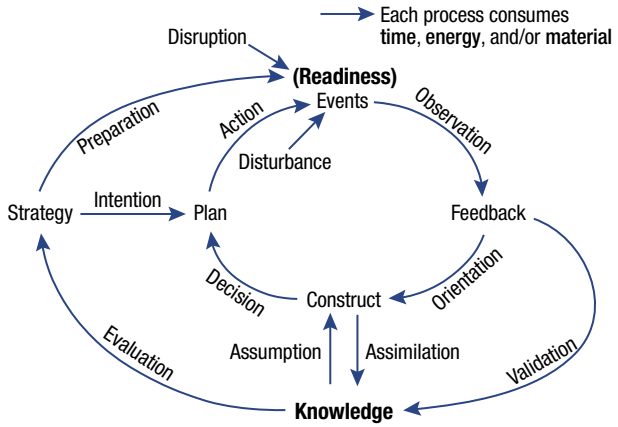

Figure 4. Resources for resilience in a control context.

the past, present, and predicted future situation, to fill gaps in the construct. A decision process is used to evaluate alternative plans of action for feasibility and predicted effectiveness, leading to the selection of a plan to execute. Both the construct and new feedback allow refinement of knowledge though processes of assimilation and validation. Evaluation of the situation can lead to changes in higher-level strategy and subsequently changes in intention governing the plan of action. Strategy also drives preparation of the system to achieve a state of readiness for effective control.

In the context of this control process, the time available to perform a critical activity is a constraint that establishes the sufficiency of other types of resources. Accurate and relevant knowledge of the system's state and its environment, informed by observations, leads to improved selection among alternative methods of maintaining or restoring control. Those alternative methods are enabled by the existence of subsystems engineered to restore control; they compose the system's readiness, or preparedness. To restore control to a physical state, subsystems require available stores of material, energy, or both.

Example: Redundancy in Critical Systems

An autonomous ship's machinery must be both physically and functionally reliable. The propulsion machinery, in particular, is prerequisite to safe navigation. Increased reliability in critical systems can be achieved by a number of measures. Redundancy, for example, is a simple strategy and can be achieved physically by redundant systems that perform the same function in the same way. It can also be achieved functionally by redundant processes that perform the same essential function or meet the same need by perhaps a more robust, if less efficient, means. Redundancy is one particular strategy that lends itself to traditional probabilistic risk analysis methods but also is understood easily as a resilience strategy.

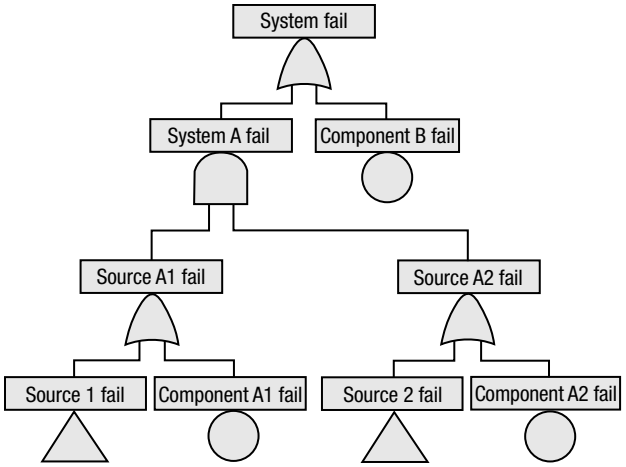

Figure 5. Fault tree.

RETHINKING THE FAULT TREE

Deciding whether to use, adapt, or create alternative models or methods for reliability analysis requires assessing their suitability to capture the essential system linkages and relationships among system states and configurations that impact reliability. In traditional probabilistic reliability analysis, these linkages and relationships are summarized in a tree-like diagram called a fault tree.[2] An example is depicted in Fig. 5. Leaves of the tree represent nominally independent logical conditions whose states (true or false) aggregate via logical linkages (and, or, etc.) to higher-level logical states, commonly called house events. At the top of the tree is a failure condition that is visually linked to all its related elemental conditions in such a way that computing the probability of failure from the probability of elemental conditions is a rule-based process of interpreting the linkages and aggregations. Fault trees are often an adequate description of failure probabilities for reliability analysis. The reliability is considered an inclusive aggregation of the many failures that would be interpreted as a case of a system not behaving reliably.

Dynamic Fault Trees

Although fault tree aggregations are often expressed as rates of occurrence, these rates are still only a static view of relationships among conditions related to failure. Failure rates cannot be tracked over time without some adaptation. One such adaptation is the dynamic fault tree.[11] It was created to enable the evaluation of failure risk, and hence reliability, as a function of time. Time dependence is added to the probabilistic models for elemental conditions, and time requirements are added to a house events table, which relates the relevance of house events to failure as the system evolves through discrete changes of state over time. Reaction and recovery time are reflected in the evolution, and time requirements can be established accordingly.

Dynamic Probabilistic Risk Analysis

An alternative approach to fault trees is to apply dynamical systems analysis to the computation of risk. This approach is common in component reliability analysis, where reliability is a function of measures on physical processes described by a set of simultaneous differential equations. This approach, however, is much less common for systems of systems. Some of the earliest uses of this approach in the commercial sector have involved continuous event trees, and simplifications of them to equivalent discrete event trees, for commercial nuclear reactor safety analysis.[12] Other methods of this type include Monte Carlo simulation and Markov cell-to-cell mapping techniques. Recent developments in autonomous ground and airborne vehicles have motivated the use of this alternative approach in pursuit of better reliability assessment of systems with autonomous control.[13]

CONCLUSION

The concept of reliability as static property of a complex autonomous system is limited as a guide to its engineering. Reliable performance is instead a quality of how the system works to maintain its essential functioning—the broader concept of resilience. The system's properties relevant to its resilience can be characterized as measures of its potential to continue to function in the face of irregular variations, disruptions, or degraded operating conditions. Engineering for resilience in autonomous systems is not simply a matter of adjusting procedures and tolerances. Rather, it requires continuous active monitoring and control of system and subsystem performance. These control functions make an autonomous system cognitive in the sense that it works to identify and minimize influences that would disrupt its essential functioning. If the methods we choose to assess reliable system performance do not account for these resilience strategies, they are at best incomplete. Methods that treat the maintenance of reliable performance as the dynamical process it is show promise of accurately accounting for resilience in the wave of autonomous systems we expect to see in the near future.

REFERENCES

[1]Vamvoudakis, K. G., Antsaklis, P. J., Dixon, W. E., Hespanha, J. P., Lewis, F. L., et al., "Autonomy and Machine Intelligence in Complex Systems: A Tutorial," in *Proc. 2015 American Control Conf.*, Chicago, IL, pp. 5062–5079 (2015).

[2]Kuo, W., and Zuo, M. J., *Optimal Reliability Modeling: Principles and Applications*, John Wiley & Sons, Hoboken, NJ (2003).

[3]Kossiakoff, A., and Sweet, W. N., *Systems Engineering: Principles and Practices*, John Wiley & Sons, Hoboken, NJ (2003).

[4]Ayyub, B. M., *Risk Analysis in Engineering and Economics*, 2nd Ed., CRC Press, Boca Raton, FL (2014).

[5]Dorf, R. C., and Bishop, R. H., *Modern Control Systems*, 12th Ed., Pearson Education, London (2011).

[6]Hollnagel, E., Woods, D. D., and Leveson, N. G. (eds.), *Resilience Engineering: Concepts and Precepts*, CRC Press, Boca Raton, FL (2006).

[7]U.S. Department of Defense, *Department of Defense Standard Practice: System Safety*, MIL-STD-882E (11 May 2012).

[8]Rødseth, Ø. J., and Tjora, Å., "A System Architecture for an Unmanned Ship," in *Proc. 13th International Conf. on Computer and IT Applications in the Maritime Industries* (COMPIT 2014), Redworth, UK (2014).

[9]Hollnagel, E., and Woods, D. D., *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*, CRC Press, Boca Raton, FL (2005).

[10]Ford, D., "The Essence of Winning and Losing," 1995 lecture viewgraphs attributed to J. R. Boyd, http://danford.net/boyd/essence.htm.

[11]Čepin, M., and Mavko, B., "A Dynamic Fault Tree," *Reliab. Eng. Syst. Safe.* **75**(1), 83–91 (2002).

[12]Smidts, C., "Probabilistic Dynamics: A Comparison Between Continuous Event Trees and a Discrete Event Tree Model," *Reliab. Eng. Syst. Safe.* **44**(2), 189–206 (1994).

[13]Hejase, M., Kurt, A., Aldemir, T., Ozguner, U., Guarro, S., et al., "Dynamic Probabilistic Risk Assessment of Unmanned Aircraft Adaptive Flight Control Systems," in *Proc. 2018 AIAA Information Systems-AIAA Infotech @ Aerospace*, Kissimmee, FL, pp. 1–11 (2018).

Craig M. Payne, Force Projection Sector, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

Craig Payne is an operational assessments leader and a Principal Professional Staff member in APL's Force Projection Sector. He received a B.S. in chemical oceanography from the University of Washington and an M.S. in applied physics (ASW curriculum) from the Naval Postgraduate School. He is a retired U.S. naval officer who spent his career serving on surface combatants and aircraft carriers and teaching at the U.S. Naval Academy where he wrote the textbook *Principles of Naval Weapon Systems*. While at APL Craig has been highly involved with the development of unmanned surface vessel capabilities and testing while acting as the Office of Naval Research medium displacement unmanned surface vehicle technical organization lead and the APL project manager. His e-mail address is craig.payne@jhuapl.edu.

Bryan M. Gorman, Force Projection Sector, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

Bryan M. Gorman is a statistical physicist in APL's Force Projection Sector. He received a Ph.D. in 1994 in physics from the Florida State University. At APL he serves as chief scientist for submarine warfare programs and as supervisor of the Analytics and Modeling Section of the Operational and Threat Assessment Group. His research interests have included analytical and high-performance computational modeling of complex systems, advancing model-based systems engineering, decision analysis, and game theory under uncertainty, and improving accuracy and reducing complexity in performance, reliability, and resilience modeling. He is a member of the American Physical Society, the Society for Industrial and Applied Mathematics, the Association for Computing Machinery, the Military Operations Research Society, and the National Defense Industrial Association. His e-mail address is bryan.gorman@jhuapl.edu.