

# Resilient Control System Execution Agent (ReCoSEA)

Craig G. Rieger  
Idaho National Laboratory  
Idaho Falls, Idaho, USA  
Craig.Rieger@inl.gov

Kris Villez  
Purdue University  
West Lafayette, Indiana, USA  
Kris.Villez@gmail.com

**Abstract**—In an increasingly connected world, critical infrastructure systems suffer from two types of vulnerability. The first is the traditionally recognized problem of monitoring the systems for faults and failures, recognizing and analyzing data, and responding with real understanding to the problems of the system. Increasingly complex systems create the opportunity for single points of failure to cascade when inaccurate assessment of system health increases response time or leads to faulty analysis of the problems involved. A second problem involves vulnerability to cyber intrusion, in which malignant actors can mask system degradation or present false data about system status.

A resilient system will protect stability, efficiency, and security. To ensure these three states, the system must react to changing conditions within the system with coordination: no one component of the system can be allowed to react to problems without real consideration of the effects of that action on other components within the system. Systems with multi-agent design typically have three layers of action, a management layer, a coordination layer, and an execution layer. A resilient multi-agent system will emphasize functions of the execution layer, which has the responsibility of initiating actions, monitoring, analyzing, and controlling its own processes, while feeding information back to the higher levels of management and coordination. The design concept of a resilient control system execution agent (ReCoSEA) grows out of these underpinnings, and through the use of computational intelligence techniques, this paper suggests an associated design methodology.

**Keywords**—resilient system, agent, control

## I. INTRODUCTION

### A. Resilient Control System

The design of a resilient control system allows individual agents to act within their spheres of influence so as to correctly react to changes in environment and state without impacting negatively the remainder of the system. The advent of cyber intrusion adds a different problem because the operations of a system have heretofore been functions of monitoring using fallible sensors that provide either continuous or on/off descriptions of system performance. Cyber-security considerations must monitor for intrusions to the system which purposefully create inaccurate data about the system state. Whereas older systems could passively monitor and react to problems within the system, cyber attack presupposes a willful exploitation of vulnerabilities, sometimes of system problems that the malicious actor knows and understands before the system, or even the designer of the system, recognizes.

This paper aims to define an integrated diagnostic and control (IDC) strategy based on an agent design to be resilient against stability, efficiency, and security (SES) performance metrics in networked systems [1]. Prior work has introduced techniques for the embedding of active IDC into an agent for shipboard systems, but has considered primarily the stability aspect of SES performance [2]. The advantages of taking a more comprehensive approach to SES threats [3] as compared to a traditional control system design include:

- Increase the accuracy of estimates for on-line availability of networked systems and their individual components.
- Increase the chance of recognizing potential faults and failures, allowing faults or failures to be identified before they affect system performance [4].
- Decrease the chance of simultaneous appearances of multiple, independent faults and the occurrence of failures which go unnoticed for prolonged times.

Because of the scale of networked systems, a distributed strategy in network awareness is likely to be most successful; the IDC strategy is embedded in an agent-based Resilient Control System Execution Agent (ReCoSEA), in which individual components manipulate their neighboring environments to obtain awareness of potential faults and failures in the components on which they depend and, subsequently, perform corrective actions. Nodes in a network will actively induce disturbances in the networked system to identify potential faults and failures as early as possible. This will occur for the range of SES performance indices and, as described, will include corrective actions that lie outside of control action on industrial processes.

In the paper that follows, the design of a ReCoSEA will be proposed, with specific focus on the fault detection aspects. Section II will provide a background of agent identity. Sections III and IV will provide a perspective on the basis for an integrated non-cyber and cyber fault detection system. Section V will provide the block diagrams of the fault detection and control design framework. This design framework will provide a basis for the ReCoSEA, based upon computational intelligence technologies.

## II. RESILIENT CONTROL SYSTEM EXECUTION AGENT

The ideas discussed to this point lead to the concept of a resilient agent or ReCoSEA. This agent possesses a mechanism to adjust, within its sphere of influence, to changes within its

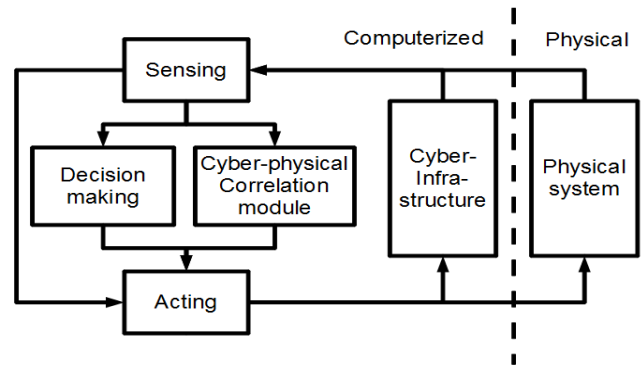
environment. These changes can include the conditions of the control system components or components outside of the control system that nevertheless affect the ability of the control system to fulfill its objectives. These objectives are to fulfill SES performance metrics within the constraints of the operation. For example, in the area of stability, the dynamics of interchange between one agent and another are already implied by existing control-system designs. That is, execution (device) layer elements are associated with unit operations, substations, or an optimally stabilizable entity. This can be seen in chemical plants, where a collection of separate operations make up an integral unit operation [5]. The unit operation, in this case, defines an area of local optimization. Within the operation, many state and input variables may exist. In a plant made up of many unit operations, the process of determining the stabilizable entities normally results in a minimization of the interactions between individual operations. That is, normally only a few state variables will make up the interactions between unit operations. For example, the fluid flow of product from one unit operation to the other must remain within a specified range, as the downstream operation is designed to be stabilized for operation within that range.

Other examples of performance can be taken from the area of security. As cyber attack can affect a control system much like a disturbance, an integrated mechanism is required, not only to distinguish that a fault exists, but to determine also the type of fault to ensure an appropriate control action is taken [6]. From a strictly limited standpoint, a recognized cyber disturbance can be corrected by several means, providing one layer of protection at the control loop level. For a sensor compromise, this could include passive cyber-related actions that include utilizing a known good sensor (or sensor and model, [7][8]) or adjusting the sensory input for the disturbance [9]. However, for an active response, this might include active cyber action that might cut off a communication channel or vary system attributes to attempt a correction that thwarts an attack.

Therefore, the attributes of a ReCoSEA must include a mechanism to detect anomalous events that affect SES performance indices tied specifically to a corrective action. These actions include both feedback-control actions on the associated industrial process as well as other system actions that are specific to human and malicious causes that are not well modeled by traditional means.

In addition, physical and cyber subspaces are commonly considered independently. However, in practice, the assumption that a fault or anomaly detected in physical space can be handled by exclusively physical actions does not necessarily hold. Similarly, disturbances detected in the cyber-space may require actions in physical space. Clearly, there is a need for developing methods that can process data from both spaces and initiate the most appropriate response, whether that is a physical or cyber action, or some combination of the two. Therefore, in addition to the purely cyber and purely physical aspects of fault detection and control, the ReCoSEA will also include a mechanism for correlation between the cyber and physical sides of the agent design, thereby improving the agent’s ability to detect and accurately respond to complex threats that affect the SES performance indices at multiple

levels. Figure 1 depicts a notional breakdown of the elements for ReCoSEA. The agent implements three basic functionalities of sensing, decision making, and acting (vertical layers) in both the cyber and the physical space (horizontal layers). In order to detect, classify and resolve complex anomalies the agent’s decision making contains a cyber-physical correlation module, which connects the anomaly-detection results with the necessary corrective action in the cyber and physical sub-spaces. Computational Intelligence methods can be considered a viable option for implementing the ReCoSEA’s decision making.



**Figure 1. Overall ReCoSEA Cyber-physical Ties**

The end result is to provide a means by which the system can continue to perform as required, maintaining system resilience. The following two sections will describe in further detail the research aspects and objectives in achieving a ReCoSEA design, one that will perform the required fault detection and respond appropriately.

### III. FAULT DETECTION/CONTROL ON NON-CYBER PERFORMANCE INDICES

Today’s digital world is networked. Indeed, many needs are fulfilled by means of connected individual elements that depend on each other, yet seek to fulfill individual goals. In the case of anomalous events, such dependencies may have severe consequences as a fault or failure of an individual element propagates through a network. It serves well, therefore, to investigate whether networked systems can be improved in their design and operation so to achieve more resilience with respect to failing modes of individual, or groups of, elements. However, the nature of the data suggests that SES performance of control systems should be divided into two categories: 1) data that are primarily of a continuous control-system design, using sensors that provide continuous or on/off indications of status and, specifically, process efficiency, stability, and physical security, and 2) data associated with cyber security that have often been event based and heuristic in nature, not easily aligning with traditional process data. This section discusses the former, the process indicators, with cyber-security indices covered in the section that follows.

One element of resilience is an increased robustness in design and operation. This is usually conceived of as an optimization

problem for the worst-case scenario. Once designed or implemented, resilience becomes a passive approach, aimed at handling all situations within assumed bounds. However, robustness comes at a cost because optimization for the worst case often requires that one obtain suboptimal solutions for normal situations [10]. In contrast, fault-tolerant control assumes proper fault detection and identification (FDI) mechanisms as well as feedback controls [11]. An active approach to resilience is therefore proposed. To this end, the following elements are necessary:

- Proper detection of anomalous events
- Proper isolation and identification (diagnosis) of anomalous events
- Proper accommodation in response to anomalous events.

These requirements are usually studied in the context of active IDC. In such an approach, a system detects and diagnoses faults and/or failures on-line and takes action following identification (feedback loop). It is this action-taking paradigm that makes this approach active. However, the detection and diagnosis mechanisms are usually passive; i.e., the system or process under surveillance is passively monitored without triggering any response from the system in any way. This poses the following problems:

- Failures go unnoticed for prolonged times [12]. In many instances, individual elements of a networked system are only active under particular circumstances. This means that failures or faults can go unnoticed for a long time since the individual elements are never requested to execute a task [13]. This is especially true for safety or backup systems. Even if fault-tolerant control is available for such elements, one should expect significant delay in accommodating actions following the need for their use.
- System design anticipates that failures will occur one at a time [14]. The introduction of an anomalous event in a particular element usually triggers events in a sequence of connected elements [15]. It is possible that some of these are not functioning or fail completely, and it is for this reason that robustness and fault-tolerance are built-in. However, these measures are often based on single fault or failure assumptions. Given that connected elements may not have been triggered for longer times, multiple failures may become apparent suddenly and simultaneously. As a consequence, the robustness and fault-tolerant design or operation may not suffice anymore to properly accommodate for a new event. Unfortunately, this is not uncommon, hence the saying “Trouble never comes alone.”

In particular, current state-of-the-art techniques lack the ability to actively probe and check performance of individual elements in view of early fault and failure detection and diagnosis. In addition, they lack the means to systematically avoid the appearance of multiple problems at one time. In

designing a ReCoSEA, these objectives are proposed as part of an integrated solution.

#### IV. FAULT DETECTION/CONTROL ON CYBER PERFORMANCE INDICES

Next-generation cyber research for control systems must find its basis in resilience, meaning it will be fundamentally of a proactive nature. Current methods of understanding a cyber threat only after its release are typical of a clinical response to normal human diseases, but not indicative of a healthy situation. Zero-day vulnerabilities defeat this philosophy, and as nation-state players hone their skills, new threats will be presented against existing defenses. This will make the threat even more potent, not unlike an antibiotic-resistant strain of bacteria. With the mathematical description of a complex network defined, the ability to utilize many linear and nonlinear techniques, even those applied now to control theory designs, may be hypothetically applied to cyber-security design. New methodologies can be applied in a proactive or a feedback fashion, or in combination, to provide an active response to cyber threats [16].

The quantity and diversity of a control system’s vulnerabilities are related to the system’s security. However, we currently have few effective ways of modeling how these vulnerability, device, and system attributes affect an adversary; neither can we easily determine or predict the degree to which the system is immune to and resilient to an attack. Cyber security design, assessment, and sensing for critical infrastructure must take into account that some vulnerabilities are inherently less severe than others, that not all devices in the system have the same value to the attacker or the same value to the defender, and that not all vulnerabilities are equally accessible to an adversary. Being able to anticipate the attacker’s likely attack objectives, strategy, processes, and decisions is clearly valuable. Measures and models to simulate and predict these elements, coupled with methods to evaluate the trade-offs among defense options, would enable organizations to improve their security resource allocations and to balance security with other needs and constraints in the critical infrastructure. In particular, to more effectively design and assess the security posture of a control system, it would be useful to have relatively inexpensive tools and techniques, based on sound scientific and engineering principles, for the design and implementation of quantifiable aspects of security.

In thinking about active cyber security for control systems, an analogy to control theory can be developed. By its nature, the primary reason for having a control system is to operate a process with stability, whether the process is performed in an oil refinery, chemical plant, or electric transmission system. The controller design, based on control theory, provides changes to the control elements to regulate the process, based upon feedback on the state. Those control elements may be valves, switches, or any number of devices. In looking at passive control of cyber security, similar processes can be conceptually envisioned. Even within current communications security technology, such as intrusion-detection and prevention devices (IDS/IPS), certain characteristics or threat signatures are recognized and, in the case of IPS, reacted to by restricting traffic. However, the approach taken in IDS/IPS design has

several limitations [17]. First these systems look at historical patterns, whether signature-based or anomaly-based, which are not necessarily predictive of what may be seen in the future in an attack from an intelligent adversary. Second, these detection methods are not foolproof and invariably require unfortunate tradeoffs between false positives and false negatives. Lastly, applying restrictions on traffic flow as the result of a detected threat may end up limiting functionality of a control system’s communications for no valid reason, and may even be used by an attacker in a denial of service attack. Even if there is good reason to restrict the traffic, it is not possible to be comprehensive while preventing false positives.

What might provide a more synergistic option to cyber detection is a mechanism to integrate cyber sensing information, including that from IDS/IPS, with physical data. Following the same parallel to control theory, active cyber security feedback loops would involve a mechanism of representative and reproducible sensing, a mathematically based mechanism to model the information streams in the system, and associated theory for control of the system and information streams. Each of these three elements comes with its own difficulty in finding a solution and its own merit for research, depicted in Figure 2 as Normalize, Rationalize and Prioritize, followed by response actions. In what we will introduce in the paper, a cyber feedback loop approach will be taken to integrate physical and security sensors to understand anomalies in behavior. These new methods will allow the design and implementation of more secure systems in the presence of known and unknown software and device vulnerabilities and degradations.

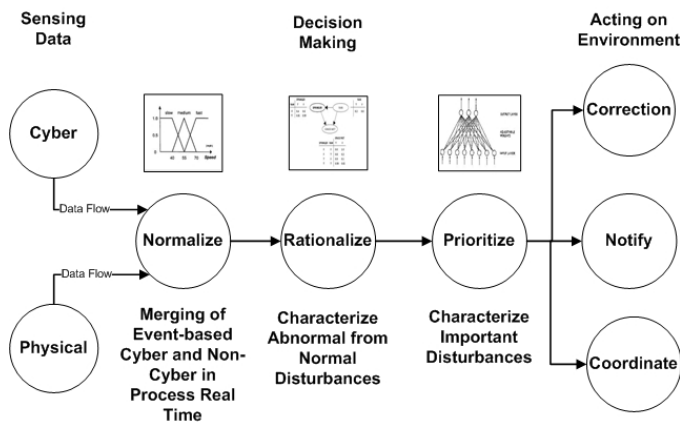


Figure 2. ReCoSEA Attributes

## V. ACTIVE FAULT DETECTION DESIGN

Referring to Fig. 2, a mechanism for integrated fault detection design can be based upon a combination of intelligent techniques that have been utilized for both cyber and physical (industrial process) anomaly detection. This normalization aspect will be addressed in the combination of fuzzy membership functions based upon the two types of data, physical and cyber, correlated to a common sensor. While the degradation impacts to the sensor can be indicated by anomaly analysis of the physical data provided by the sensor, it may also be characterized by cyber detection anomalies that

indicate modification of the sensor data. The task of the “Normalize” aspect will be to associate the incoming data with an individual sensor or type of information. The “Rationalize” aspect will characterize anomalies. Through the comparisons of both the cyber and physical aspects, a judgment will be characterized and a statistic provided that will be proportional to the belief that the sensor has not been impacted. The “Prioritize” aspect will determine the level of importance in mixed initiative (human + automation) response to correlate to the different anomalies detected. The “Correction,” “Notify,” and “Coordinate” aspects will be provided by a mixed initiative response—which provides corrective responses that can be at a local control loop level as well as at a supervisory level—and Notify, which implies a human in the loop for appropriate response.

Pairing the appropriate computational intelligence technique to the individual aspect is important relative to the type of data that must be analyzed. For the Normalize aspect, the use of fuzzy logic provides a mechanism to perform comparisons across various data sources [18], [19]. For cyber, this can be the clustering of the encrypted data from known secure sensors, where inconsistencies in the continuous data are already associated with individual sensor communications [20]. Both the cyber and physical variables can have their own membership functions that are based on a time-based extraction of the data. Through the use of the rule base, the comparison of the encrypted and unencrypted channels provides residuals that can then be used as a means for assertion of undesired anomalies in the Rationalize aspect. Although a much more extensive set of fuzzy comparisons can be used, Figure 3 provides one example where variations in value due to latencies or other issues of direct comparison can be simply performed.

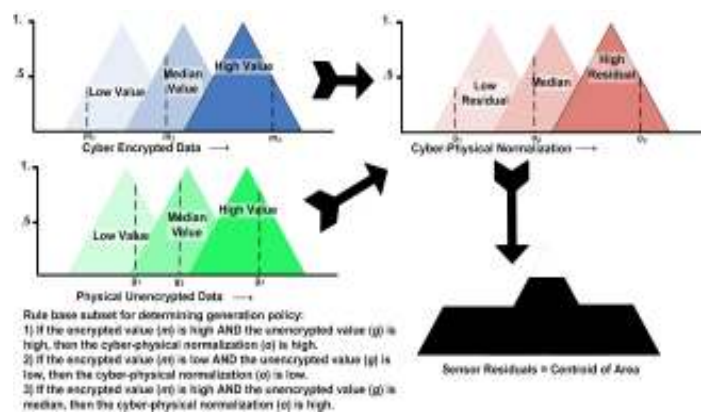


Figure 3. Fuzzy Logic Normalize Aspect

For the “Rationalize” aspect, the application of a Bayesian Belief Network (BBN) provides a means to analyze statistically the incoming data [21], [22]. The BBN can take the 0-1 statistical output of the Normalize defuzzification aspect and provide a conclusion as to where relevant anomalies exist based upon history. Given a BBN, such as that in Figure 4, a next step would be to characterize the necessary Bayesian probabilities associated with the BBN. Bayes’ Formula and the manipulated probabilities for the BBN can be

more simply expressed in terms of a hypothesis ( $H$ ) and data ( $D$ ):

$$P(H | D) \propto P(D | H)P(H)$$

where:

- $P(H)$  is the prior probability of  $H$ ; the probability that the hypothesis is correct before comparing to the data.
- $P(D|H)$  is the conditional probability; the likelihood of seeing the data  $D$  given that the hypothesis  $H$  is true.
- $P(H|D)$  is the posterior probability; the probability that the hypothesis is true, given the data  $D$ .

Considering the information provided by the management layer, the output of the Normalize defuzzification can itself be considered a probability that must be linked, in this case becoming the prior probability of the hypothesis  $H$ . That is, the input represents the preference of accepting a particular judgment for realignment based upon the management policy for realigning assets. The conditional probability or likelihood that reflects what is desired is borne out in maintaining a state awareness of the anomalies. The posterior probability reflects the confirmation that the hypothesis is true, driving the continued realignment of the power system to reflect the policy. This implies that the hypothesis provides what might be considered a selection of potential control actions or notifications to be parsed by the Prioritize aspect.

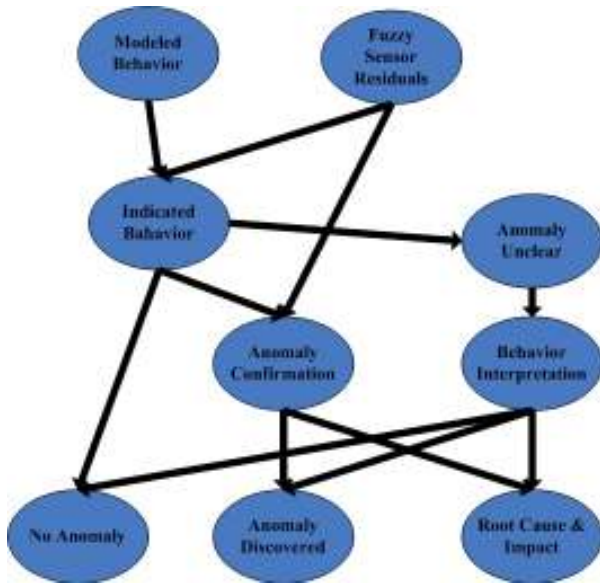


Figure 4. BBN Rationalize Aspect

Of additional note in the BBN framework is a two pronged approach in which (1) the output of the Normalize aspect provides the data for the first cut anomaly analysis and (2) additional fault detection techniques can be brought to bear for final disposition when conclusions are indeterminate. This inclusion generalizes the design better, and provides a means to represent introduction of other techniques where more appropriate to the data. Root cause and impact interpretations are maintained from the BBN evaluation.

The Prioritize aspect will perform the function of prioritizing actions or responses that involve both the human and direct digital action to valves, breakers or other field devices in an industrial control system. In order to provide the appropriate response, the immediacy of the impact and type of response is necessary. A neural network can then be used to selectively align the response based upon this information that accompanies the anomaly statistic from the Rationalize aspect to perform a selection of this appropriate response [22][23][24][25]. In what is rationalized in Figure 5, some sources of anomaly and the level of impact are mentioned. While the cyber and physical aspects were normalized earlier in the agent, the responses from the automation and/or human can vary depending on the source. One reason for this is suggested in the need to secure cyber channels from a malicious actor, in addition to correcting the affects within the automation, such as a sensor selection methodology [10] [26].

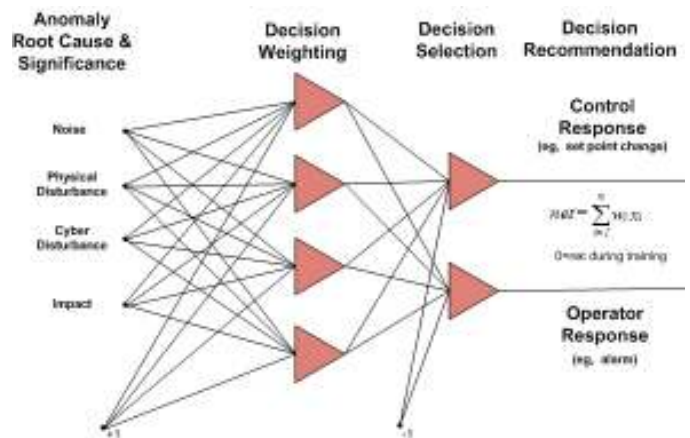


Figure 5. Neural Network Prioritize Aspect

Depending on the association of the ReCoSEA aspects with individual industrial process infrastructure, including sensors and field devices, the complexity of the decision recommendation and the post-processing of the decisions will also vary. In the Correction, Coordinate and Notify aspects, the local and supervisory control will be performed in automation, and where a human is required, an indication of the plant status and interaction necessary. If the ReCoSEA is desired to correlate to an area of local optimization, a term mentioned earlier for a chemical plant, the resulting agent could be simply described by Figure 6. The Prioritize neural network provides the final Correction of the industrial process variable. The Coordinate response provides an adjustment of the set point, in the case where the industrial process needs to move to a new operating point. Finally, the Notify response provides the operator a warning that indicates whether something should be closely monitored, or perhaps a manual corrective action is needed.

## VI. SUMMARY

A generalized design methodology has been suggested in this paper for analyzing and acting upon anomalies in cyber physical systems, such as industrial control systems. The

