Security and Safety, Vol. 1, 2022001 (2022)
https://doi.org/10.1051/sands/2022001
sands.edpsciences.org

Views OPEN ACCESS

Information Network

Problems and solutions regarding generalized functional safety in cyberspace

JiangXing Wu[1,2,*]

[1] *National Digital Switching System Engineering & Technological R&D Center, Zhengzhou 450002, China*
[2] *Purple Mountain Laboratories, Nanjing 211111, China*

Received: 9 December 2021 / Revised: 17 January 2022 / Accepted: 25 January 2022 / Published online: 22 June 2022

Abstract The common endogenous security problems in cyberspace and related attack threats have posed subversive challenges to conventional theories and methods of functional safety. In the current design of the cyber physical system (CPS), functional safety and cyber security are increasingly intertwined and inseparable, which evolve into the generalized functional safety (S&S) problem. The conventional reliability and cybersecurity technologies are unable to provide security assurance with quantifiable design and verification metrics in response to the cyberattacks in hardware and software with common endogenous security problems, and the functional safety of CPS facilities or device has become a frightening ghost. The dynamic heterogeneity redundancy (DHR) architecture and coding channel theory (CCT) proposed by the cyberspace endogenous security paradigm could handle random failures and uncertain network attacks in an integrated manner, and its generalized robust control mechanism can solve the universal problem of quantitative design for functional safety under probability or improbability perturbation. As a generalized functional safety enabling structure, DHR opens up a new direction to solve the common endogenous security problems in the cross-disciplinary fields of cyberspace.

Keywords Common endogenous security problem, Generalized functional safety (S&S), Dynamic heterogeneity redundancy (DHR) architecture, General robust control mechanism, Coding channel theory, Common cross-disciplinary field problems

Citation Wu JX. Problems and solutions regarding generalized functional safety in cyberspace. Security and Safety 2022; **1**: 2022001. https://doi.org/10.1051/sands/2022001

1 Introduction

According to its classic definition, "functional safety is part of overall safety that depends on a system or device operating correctly in response to its inputs. The safety goal is achieved when each specific safety function is implemented and the required performance level of each safety function is met. [1]". In other words, the safety function of the device or system should be correctly and properly ensured regardless of the presence of random faults, system faults, or common cause failures. However, with the ubiquitous application of digital, intelligent and networking technologies, the connotation, and extension of functional safety have broken through the reliability scope of dealing with function errors or system failures caused by random physical failures in mechanical parts or electronic components. The prevalence of central processing unit (CPU) and software has inevitably introduced cybersecurity issues while improving conventional functional safety. In particular, the threat of "unknown unknown" cyberattacks

* Corresponding author (email: ndscwjx@126.com)

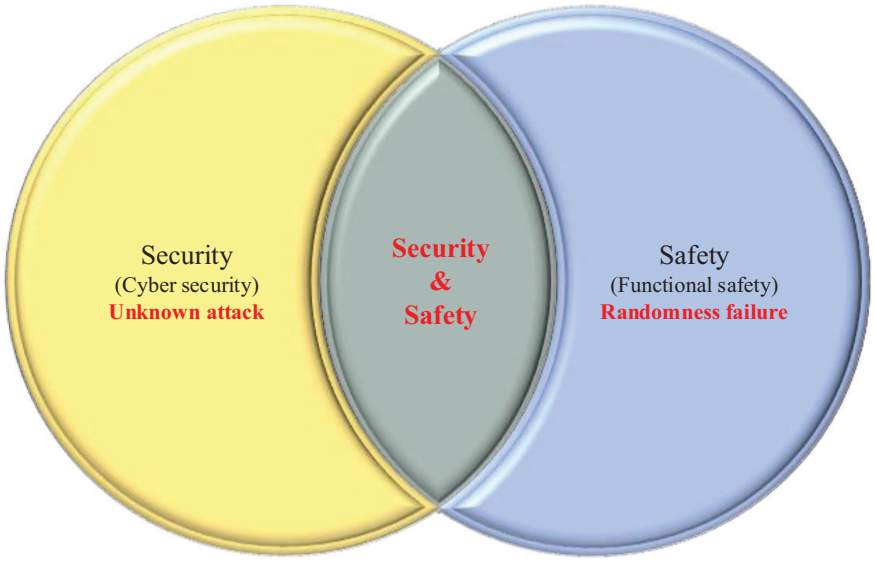

Figure 1. Security & Safety integration problem

[2] targeting the design vulnerabilities or backdoors of cyber physical system (CPS) not only changes the assumptions of randomness and the mathematical properties that can be expressed by probability in the classical reliability theory, but also shakes or even subverts the basis of conventional theories, methods, and practices concerning functional safety. Currently, as the man–machine–thing integration continues, cyber security and functional safety have become an indivisible and integrated security problem, the connotation and denotation of which will continue to expand in depth and width (see Figure 1).

As well known, at present, whether it is a cyber warfare game between countries or a cyberattack initiated by a non-government organization (NGO) or a hacker, the attacker can always take advantage of the offense–defense asymmetry to threaten the deeper integration between Information Technology and physical security infrastructure (PSI) in the era of digital economy. This is largely because, when people enthusiastically embrace intelligent, networked key information infrastructure and industrial control equipment at the technical level, they have not yet been able to get rid of the constraints of classical reliability theories and methods as well as the corresponding practice norms based on issues regarding randomness differential/common-mode failures. Few researches have broken through thinking patterns or original paradigms of scientific development [3] to study the attack theories and methods of software and hardware code design vulnerability or loopholes of critical information infrastructure in the network era, as well as the possible subversive effects on the functional safety design of devices or devices such as CPS. In other words, there are common endogenous security problems in cyberspace that are difficult to solve, so the influence of cyberattacks on target functions can only be avoided or qualitatively described [4, 5] in the current worldwide functional safety design standards of CPS. However, in the context of Internet of everything, the cyberattacks targeting unknown or high-risk vulnerabilities often reduce the reliability or functional safety designs of complex systems like CPS directly into the "nothing-but-eye-candy" dilemma. That is, when the attack fails, the functional safety of the target system may remain at the level of 99.999999 or even higher against the designed indicators, but once the attack succeeds, the system's functional reliability will immediately become a "Logical NOT AND (NAND) gate". Therefore, the era of cyber intelligence urgently calls for new development vitality injected into the design theories and methods as well as testing and evaluation systems of CPS functional safety.

This paper discusses the generalized functional safety problem and gives its definition, and briefly analyzes the "helpless predicament" of the existing cyber security development paradigm and functional safety design specification when facing the generalized functional safety problem. On this basis, this paper introduces the dynamic heterogeneity redundancy (DHR) architecture with generalized functional safety and its working mechanism. Meanwhile, based on the innovative coding channel theory (CCT), this paper proves that the generalized robust control mechanism constructed by DHR can solve the worldwide problem that the functional safety of CPS devices cannot be designed quantitatively and validated against

measurement under the premise that random failures and cyberattacks based on internal vulnerabilities exist simultaneously.

2 Threats of S&S and common endogenous security problems

What is the generalized functional safety (security & safety, S&S for short) problem? We know that the living space of human beings is tangible and based on the physical world, where people's awareness of safety tends to be focused on conventional functional safety, namely the "safety problem". Taking the auto manufacturing sector as an example, a special standard for automotive functional safety ISO26262 was released in 2011 [4]. It is mainly concerned with how to reduce the functional safety risk of automotive electrical and electronic system failure to an acceptable range. However, as we usher in the era of connected intelligent networks featuring the ternary integration of men, machines, and objects, cybersecurity turns out a more prominent problem, which is the "security problem". Also, taking the automotive industry as an example, the cyber security problem in smart cars has quickly penetrated into the traditional functional safety field. Data show that the number of global automotive cyber security incidents has increased nearly eightfold from 2016 to 2020, boosting by about a factor of one fold in 2019 alone [6]. Tens of millions or even hundreds of millions of lines of software and hardware as well as a large number of code design vulnerabilities make enough resources and opportunities available for attackers to cause new security risks, including malicious control of vehicle running state. Different from the risk of privacy or sensitive information disclosure, the former will be directly related to the safety of life and property of passengers.

From this, it is natural to draw a conclusion that, in the deeper integration of information world and the physical world, an extremely contradictory expression is inevitable. On the one hand, all digital infrastructure and intelligent devices have directly turned into typical intelligent and networking CPS systems; on the other hand, conventional reliability or functional safety theories and practices are facing challenges in the era of deep human–machine integration and intelligent interconnection. The boundary between "functional safety" and "cyber security" is "collapsing" at an unprecedented rate, so a new problem of "Integrated Functional Safety" emerged, which is "interwoven and superimposed with Security & Safety". As for such new phenomena caused by a combination of randomness failures with the nature of probability distribution and cyberattack failures difficult to precisely describe with mathematical tools, we might as well name them collectively as generalized functional safety (its definition will be given later). This problem has two distinctive features. First, it has "fusion effects". Cyber security and functional safety problems are not only intertwined, but also existing responses are contradictory, "chaotic, and difficult to sort out". For example, in traditional networks of telecommunication, power, gas, oil transmission, water supply, *etc.*, concerning people's well-being and social security, the destructiveness of functional failures caused by randomness failures is generally controllable, and the recovery is relatively easy. The security risks or hazards arising from cybersecurity issues, however, are extremely uncertain and destructive, where a small leak will sink a great ship, or even lead to domino-like social disasters. Secondly, it has "superimposed effects". Since it is hard for a security problem to die out once it arises, the number and difficulty of the growing problems far outrun humans' problem-solving ability. Therefore, the larger the scale of information technology is, the more serious the cyber security problem would be; and the higher the degree of network connectivity is, the more prevalent the functional safety threat would be. Provided that the following definition is met: "in the target system, if there are both conventional disturbances of some type (*i.e.* randomness failures, errors, malfunctions and other functional safety issues) and cyberattacks based on the target software/hardware vulnerabilities or backdoors, *etc.*," we can say the system suffers from the generalized functional safety problem (S&S problem for short). Accordingly, the definition of "S&S" is as follows: "If there is a robust structure model that not only keeps a given mode's functions within the safety margin of quantifiable designs under the perturbation of some conventional randomness factors, but also safeguards the reliability of the model's functions against the activated cyberattacks based on in-house software/hardware vulnerabilities or backdoors, *etc.*, the very model shall be deemed as generalized functional safety (GFS).".

According to Hegel, a German philosopher Hegel, "Contradiction is the root of all movement and vitality. It is only in so far as something has a contradiction within it that it moves, has an urge and activity.". The double-faceted or even multi-faceted nature of a thing makes its inherent or endogenous

security problems inevitable and ubiquitous. From an engineering perspective, any natural or artificial function has explicit side effects or implicit dark functions, either associated or derived. While we may be able to distinguish between the benign and non-benign attributes of explicit side effects, the nature and potential effects of implicit dark functions remain completely unknown. Therefore, an endogenous safety issue is philosophically an explicit or implicit expression of the endogenous contradictions of an expected function. If we borrow a term of genetics, this scenario is a congenital "genetic defect", which is both individual and common [3]. Common sense tells us that, individual security problems usually do not have a uniform solution, but the common security problems are likely to have universal solutions, which are called the common endogenous security problems. The existence of cyberspace common endogenous security problem [3] is prevalent in cyberspace and supported by at least the following arguments: (1) The phased characteristics and the contradictions between complexity and completeness of human's technology development and cognitive level make it inevitable to completely shake off software/hardware code design defects (or vulnerabilities); (2) In an era of globalization and multi-polarization, it is impossible for any country or enterprise to establish a completely self-sufficient technology chain, supply chain, and industry chain, and thus the backdoor problem cannot be completely eliminated by means of management engineering; (3) At this stage, human's scientific and technological capabilities do not support scrutinizing the software/hardware code backdoors or vulnerabilities in complex systems, whether in terms of theory or engineering technology, which will remain an insurmountable technological challenge at least for the foreseeable future; (4) as information technology, network technology, intelligence technology as well as concomitant deliberate attack factors continue to penetrate into the conventional functional safety field, the cyber security problem makes the classical theories and method of reliability or functional safety based on the premise of physical or logical random failure in the information age no longer stand alone, and the random failure problem inevitably evolves into a generalized functional safety problem with the intertwining of safety & security.

Note that as is expressed in the theory of contradictions, internal factors need to take effect *via* external factors. As a result, common endogenous security problems usually do not generate a direct negative impact, and only constitute functional safety hazards or threats when disturbed by external factors.

In short, the ubiquitous existence of common endogenous security problems in cyberspace enables attackers to build an effective attack chain, which forms a "generalized functional safety threat" through the accessible attack surface [2] and available software/hardware resources as long as they discover high-risk vulnerabilities in the target system. In other words, any digital, intelligent, networked product technology and application industry cannot avoid the challenges of common cross-disciplinary problems based on generalized functional safety threats.

3 The existing cyber security development paradigm cannot solve S&S problems

The current add-in cyber security development paradigm based on detection analysis and situational awareness is stuck in an "unsolvable" dilemma, since its thinking perspective is to obtain *a priori* knowledge about the attacker and its precise behavioral characteristics. The corresponding methodology is to apply "patch-like" defense based on threat awareness, or active defense based on randomness, diversity, and dynamism (move targets defense (MTD) [7]) to reduce the effectiveness of attack attempts. The problem with the former is that there is no way to guarantee whether the "on-patch" or "post-patch" approach would lead to new common endogenous security problems, regardless of various traditional cyber security technologies like "closing the door and fixing the leak" or "embedding/ internally installing". Even if embedded or implanted with a variety of cyber security protection, like software, hardware, or systems, their inherent design flaws or backdoors are still difficult to prove or testify. The problem with the latter is that there is still a high escape probability even using active protection techniques such as MTD [7]. Because the dynamic scheduling step and the host execution system itself still cannot exclude the possibility of being "bypassed or short-circuited" by high-risk vulnerabilities, the diversity and dynamism cannot change the logical nature of the software and hardware vulnerabilities/backdoors, making it impossible to prevent the collaborative internal and external "raising card" attack [8]. Moreover, the high frequency of dynamic transformations of instructions, data, addresses, and even networks will significantly degrade

the target system service performance or effectiveness. The commonly used add-on cyber security strategies are expected to solve the problem by adding exclusive security components, but ignore the "soul question" of "whether the added security components themselves are secure". In particular, it is difficult for CPS systems concerning life and property security to accept "patch-like" or priori knowledge-based defense strategies and technical architectures, regardless of the network defense approaches including active/passive immunization or dynamic defense technologies with significant performance loss or costly full life-cycle security maintenance. Examples of application areas are industrial control, power systems, fuel supply systems, aviation/aerospace systems, financial systems, the automotive industry, information weaponry, and combat systems. More seriously, the essence of classic reliability technology lies in the quantifiable design and verifiable metrics of functional safety, while the current cyber security measures are either adhering to the concept of "best effort" or facing the native "soul question". For example, the cryptographic algorithms are highly robust in mathematical sense, but the endogenous security problems of the host execution system may make them incompetent. The problem of how to prove the credibility of the trusted root exists from the birth of trusted computing. Other problems are: (1) how to ensure that there is no >51% common-mode vulnerability or backdoor problem in the hardware and software of each node in blockchain; (2) how can the zero-trust architecture guarantee that the functions of those distributed authentication nodes are not "bypassed or short-circuited" by the host system's common endogenous security problems. In other words, it needs a construction with a generalized robust control mechanism that is not only able to counteract the natural factor ingestion, but also can deal with the disturbances caused by cyberattacks [2]. Like the triangle in Euclidean space, the geometric construction determines its endogenous stability. Otherwise, if two issues are dealt with two architectures or processes to "divide and rule", it is still difficult to solve the theoretical and technological conflicts between functional safety and cyber security [9]. Even if the cost and performance are ignored, it is hard to find a quantitative and qualitative solution for the generalized functional security problem.

4 Generalized robust control mechanism and generalized functional safety empowering construction

A well-known axiom goes like this: "Everyone has shortcomings of their own, but when they do the same task independently, it is extremely rare that most people make exactly the same mistake, in the same place, at the same time". This is also the theoretical cornerstone for the democratic system of human society, which explains why the opinions of the majority should be respected. However, this axiom requires at least two constraints: first, for individual members, completing the task is a high-probability event, failing to complete the task is a low-probability event; second, for group members, there is no collective fraud or underground collusion. We refer to this as the relative truth axiom [2] (*i.e.*, consensus mechanisms), which is expressed as a logic diagram, as shown in Figure 2.

In 2013, while studying the endogenous security theory of cyberspace, we found that the True Relatively Axiom is in fact the theoretical and technical bases of the dissimilar redundant structure (DRS) [2, 8, 10] proposed in the 1970s. Studying the DRS in the context of the generalized functional safety (S&S) problem scenario can lead to the following new perceptions:

(1) Under the condition of functional equivalence, any uncertainty (unknown unknowns) existing at the individual level can be converted into "known unknown" probability in the form of differential or common mode at the group level.
(2) Adjusting the number of people, type, task complexity, completion time, and voting strategy is able to control the probability of differential modes, *i.e.*, controlling the impact of unknown problems.
(3) The result of consensus or large number judgments may still be wrong with a small probability, as it is difficult for relativity judgments to distinguish between differential-mode and common-mode problems sometimes.
(4) According to the DRS design approach, taking the expensive "back-to-back" engineering management measures [2, 8, 10] can obtain a global "static heterogeneity" against possible common-mode disturbances. However, the inherent static and deterministic similarity makes it lack of temporal and qualitative robustness under human trial-and-error attacks, *i.e.*, the DRS construction is difficult to sustainably ensure generalized functional safety under cyberattacks [8].

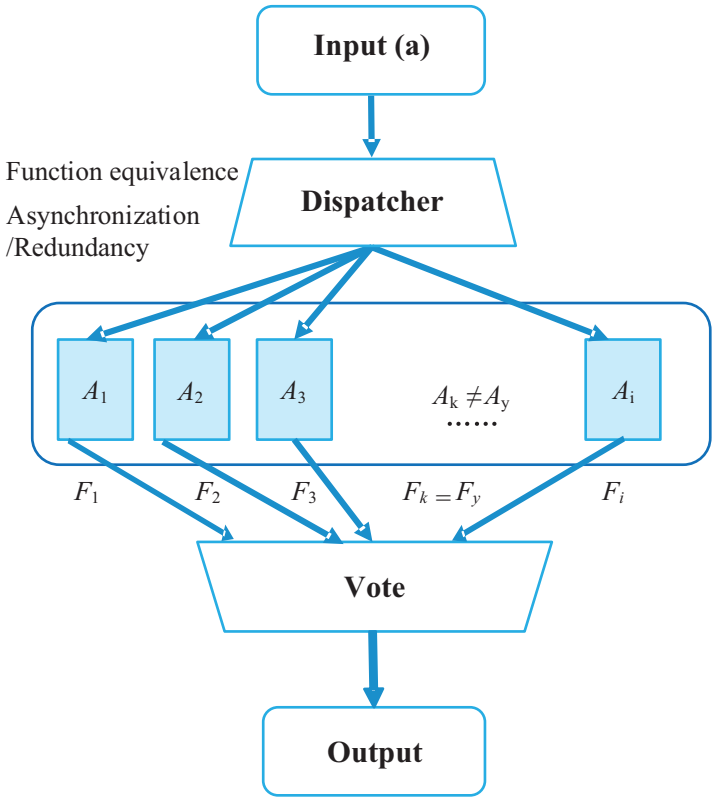

Figure 2. Relative truth logical expression [2]

Based on the above perceptions and the methodological research on endogenous security theory in cyberspace, we proposed a DHR architecture with the generalized robust control mechanism [2, 8, 10], and with its logical expression, as shown in Figure 3.

The red dashed box depicts a heterogeneous redundancy processing scenario set L with M reconfigurable processing scenarios of equivalent function P. For the processing scenario element j ($j = 1, 2, \ldots, M$), the design flaws or vulnerabilities that have different natures (differential mode) from other elements in L are allowed to exist. The workflow is as follows: the sequence of input stimuli (including attacks against the differential-mode vulnerability of a processing scenario) is assigned by the input agency to n elements of the current heterogeneous processing scenario set N (when $n < M$, N can be any of the combinations in C_M^n), and the output vectors of the stimulated processing scenario are transmitted to the multimode iteration verdict module. When there exists an output vector satisfying the requirements of the iterative verdict policy, it is sent to the output agency, which opportunistically generates the output response sequence.

If there is no output vector satisfying the current verdict policy, the verdict module will further change the verdict policy, or drive the feedback controller to select an element j from the set of heterogeneous processing scenario L to replace one of the problem scenarios i in the set of current processing scenarios N, or directly instruct the problem scenario i to perform cleaning or restoration operations until there are multiple same output vectors satisfying the requirements of the verdict policy, leading the DHR loop to the steady state. The feedback controller can also force the replacement of any processing scenario element in the current heterogeneous set of processing scenarios N according to external instructions, and can also set or update the iteration verdict policy according to the verdict parameters. It is straightforward to find out that the DHR has the following generalized robust control characteristics:

(1) As long as a differential-mode output vector appears in the current processing scenario set N, it will be shielded or iteratively corrected by the strategic ruling mechanism, and there is no need to care about the cause of the processing scenario i, to which the differential-mode output vector belongs. In other words, whether it is a randomness failure or a man-made cyberattack, as long

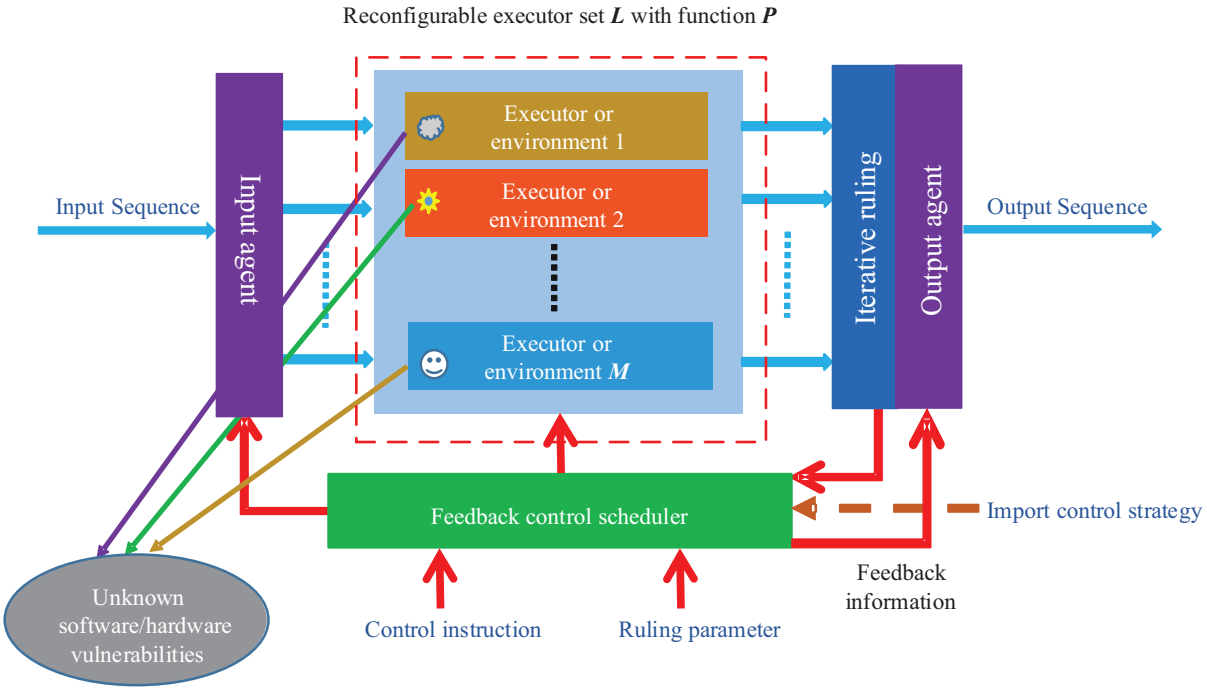

Figure 3. Logical expression of the DHR [3]

as it is a differential-mode problem, it will not affect or continue to affect the correct expression of the function (P) of set N.

(2) In addition to the perceived presence of a differential-mode output vector, an external control command can also compulsorily activate DHR feedback control loop, leading to a change in the operating environment of individual or multiple elements of the current processing scenario set N, or even directly replacing any or all the elements in set N with other elements in set L.

(3) Once activated, the feedback control circuit will not automatically stop the iterative operations until the emergence of correct output vectors that meet the number required by the iteration ruling strategy. In other words, DHR is designed to be able to withstand trial-and-error blind attacks intrinsically.

(4) If there is a common-mode error vector, there is a probability of common-mode escape, especially when most of the elements in set N have exactly the same error output vector, escape will be inevitable. However, as long as the verdict module can still perceive the differential mode in the verdict result, it will inevitably activate the iterative processing mechanism of the feedback control loop and change the composition, number, or mutual heterogeneity of the elements in the current set N of processing scenarios, as a result of which the sustainable time of the common-mode escape is only related to the iterative convergence algorithm and speed of the DHR loop, *i.e.*, the common-mode escape state is unstable.

(5) Changing the number of elements and the degree of heterogeneity among elements of the heterogeneous processing scenario L within the DHR construction environment, as well as the richness of the iterative verdict strategy and the convergence speed of the feedback control loop, can lead to quantifiable and verifiable security metrics for function P.

In summary, the DHR architecture can not only effectively deal with any random failures inside the structure or attacks toward its common endogenous security problem, but also guarantee that the function P of the target object built on this structure is endowed with the generalized functional safety that can be quantified by design and verification metrics. In other words, as an empowering architecture integrating high reliability, high credibility, and high availability, DHR can provide an integrated solution to the interweaving problems of functional safety and cyber security in the fields or sectors concerned, namely gaining universal S&S features *via* the system architecture technology [8].

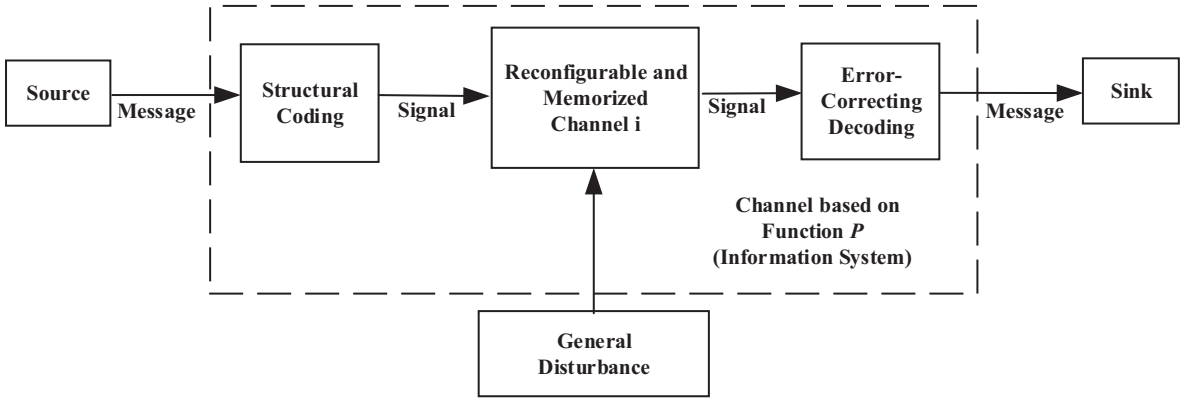

Figure 4. Workflow of a reconfigurable memorized channel to counteract generalized disturbances [12]

It should be noted that, the reliability and trustworthiness of function P based on the DHR architecture cannot be ensured when there are factors aside from the surface of the attack such as password cracking, certificate theft, or high-risk vulnerabilities in the interaction protocol. In other words, although the DHR architecture can effectively prevent the network attacks based on the in-house vulnerability or backdoors, it still needs the support of encryption authentication technology to fight back brute-force attacks at the "front" door.

5 Coding channel theory

5.1 Background

In cyberspace, physical or logical random failures and man-made attacks based on common endogenous security problems can be abstracted into a question of how to correctly process and transmit information in a reconfigurable and memorized channel where random and non-random noises coexist [3]. The errors in information processing and transmission caused by cyberattacks have similar properties with reliability errors and communication noise errors; and the idea of Shannon's second theorem [11] (noisy channel coding theorem) that uses channel coding structure to make correct information transmission in noisy channels has great enlightening significance to the solution of this work.

In a general sense, all kinds of computers and information and communication devices in cyberspace can be regarded as a certain realization of a Turing machine, which can accept, store and execute various algorithms described in the form of program code. Unlike the hypothesis of the memoryless channel in Shannon's communication model [11], in this work, they are abstracted as reconfigurable memorized channels with processing capabilities. Since cyberattacks are distinctly non-random, unlike the Shannon channel random noise precondition, it is necessary to define the generalized disturbance or noise with random and non-random nature in reconfigurable memorized channels, including random communication channel noise, random physical failure, and man-made attack noise, *etc.* Thus, the workflow schematic of generalized disturbance rejection in the reconfigurable memorized channel can be expressed as in Figure 4.

5.2 Basic concept

For the DHR construction, its reconfigurable executor set can be unrolled on the space–time axis as a set of static, functionally equivalent "coding-channels" with heterogeneous processing scenario implementation, and its dynamic expression is to use different coding channels to counteract the effects of generalized disturbances based on the iterative processing mechanism (or intelligent scheduling mechanism). Shannon's channel coding theory focuses more on the "random memoryless channel", while the DHR construction is equivalent to "memorized reconfigurable multi-element redundant channel with random or non-random noise" [12]. Therefore, Shannon's theory and methods cannot be used directly to evaluate the safety or generalized robustness of DHR constructions. It is necessary to develop a "coding channel theory" (CCT)

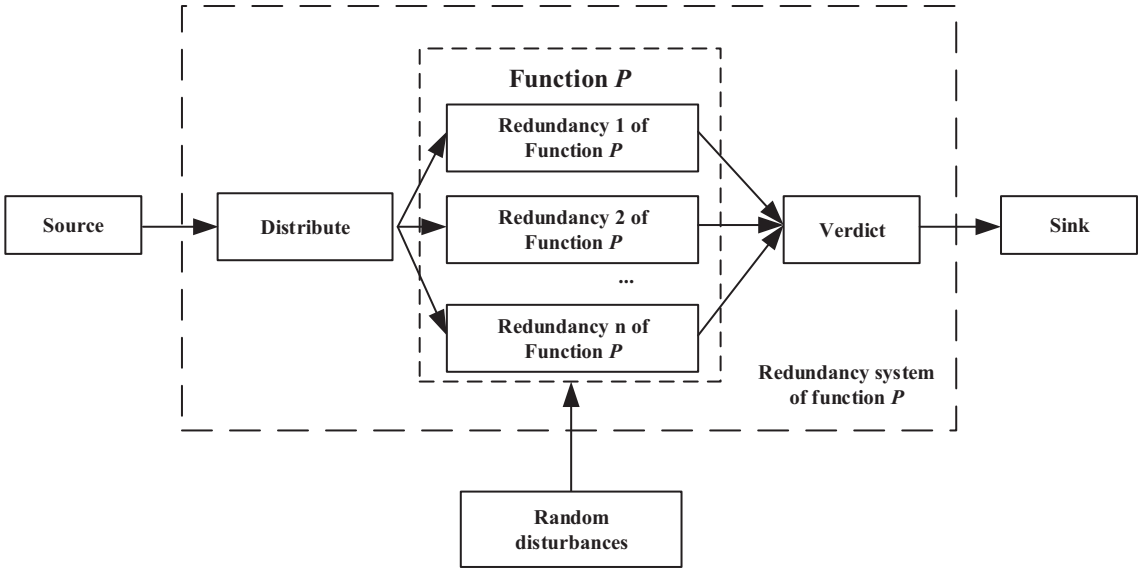

Figure 5. Typical model for reliability assurance [10]

[12] to quantitatively evaluate the performance of the DHR "coding structure" mechanism in suppressing generalized disturbances or noises, providing engineering practice specifications with quantifiable designs and testing metrics for the functional safety of the target object.

In fact, the key to the validity of CCT lies in proving the relevant existence theorem. It is necessary to theoretically clarify how to construct suitable coding channels to provide robust functions targeting a specific discrete memorized channel under the prerequisite of generalized disturbances. The concept of "robustness" refers to the adaptation of appropriate coding/decoding steps and verdict-based dynamic feedback control mechanism, so that even if there is random or artificial additive interference in the system architecture embedded with S&S attributes, the errors in information transmission and processing for a given function can still be controlled within a given threshold. In other words, the CCT is essentially composed of the mathematical model of DHR construction and the two existence theorems together with their related definitions, lemmas, and proofs, which can and should cover Shannon's second theorem.

In summary, the special mechanisms of DHR and CCT allow the problem of cyberattacks against "unknown" design flaws or vulnerabilities in the target hardware and software code to be normalized into the problems that can be effectively addressed by classical functional safety theory and techniques.

5.3 Comparison between S&S and reliability assurance

The expectations of the generalized functional safety (or robust control), the classical reliability assurance and Shannon's reliable communication are all to provide robust service under generalized disturbances; and the corresponding disturbance, fault, error, and failure action chains are described in Chapter 15 of *Cyberspace Endogenous Safety and Security* [10]. The similarities between the above three techniques (generalized functional safety, reliability assurance, and Shannon's reliable communication) are: (1) facing the impact of random or non-random disturbances; (2) facing the threats of faults, errors, and failures; (3) adopting similar solution approaches, *i.e.*, adding redundant "supervisory elements" (in different dimensions) to tolerate differential-mode/common-mode errors, thus interrupting the evolution chain from faults and errors to failures, ultimately reducing the probability of system failure. However, the above three techniques significantly differ from the properties and hypothetical conditions of generalized disturbances and channels (or meta-channel). Their basic models are shown in Figures 5 and 6. The structural encoding, multiple meta-channels, and error correction decoding together form a coding channel with function P. The redundant channels ($i = 1, 2, \ldots, n$) in Figure 5 refer to the sub-channels with equivalent functionality, and the meta-channel ($j = 1, 2, \ldots, n$) in Figure 6 refers to the reconfigurable channel.

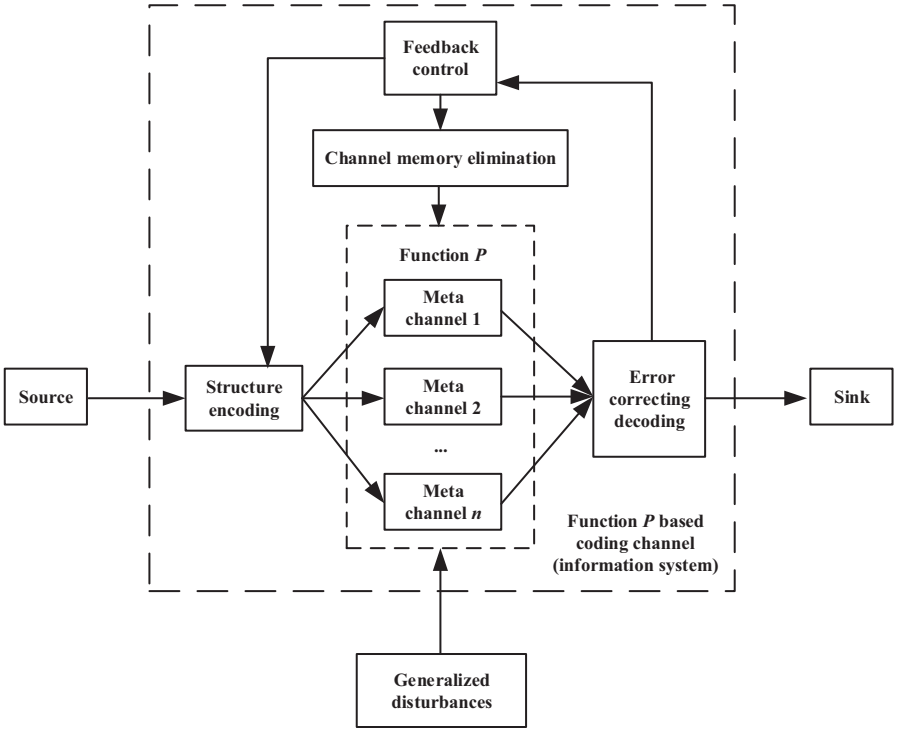

Figure 6. Model of DHR architecture [10]

5.4 Coding channel existence theorem

In a random noise memoryless channel, the meta-channel refers to the sub-channel within the coding channel for message delivery and processing. For any random noise, the probability of an error in the meta-channel output $P_e < 1$. Due to the memoryless nature of the reconfigurable meta-channel, for the random noise at any moment t, the probability of an error in the meta-channel output $P_e(t) < 1$. Therefore, in the memoryless channel with random noise, the n meta-channels have the same mathematical expressions with the Shannon channel in Shannon's second theorem. In chapter 15 of reference [10], the coding channel existence first theorem, the randomness lemma of random noise with memorized channel, the randomness lemma of random noise with memorized channel, and the coding channel existence second theorem are explained in detail, and the relevant proofs are given.

The mathematical model for information transmission in the DHR with the feedback memory-elimination construction channel is also given in reference, which is shown in Figure 7.

It has been proved that the DHR with the feedback memory-elimination construction channel scheme, ensures that the cyberattack model in system failure evaluation can be normalized to the random noise model and be processed by the coding channel [12]. This means that in the DHR construction, the proven reliability theories, methods, testing, and evaluation measures can be used to deal with function failure problems with seemingly two different mathematical properties, thereby providing a universal solution for the intertwined problems of functional safety and cyber security (generalized functional safety) in an integrated manner.

6 Conclusion

Due to the prevalence of common endogenous security problem in cyberspace, it is difficult for existing cyber security technologies to give quantitative security indicators under the "unknown unknown" cyberattacks. In addition, the prerequisites and basic assumptions of the various current cyber security technologies have "unclosed" logical problems or even paradoxes, so the traditional security stereotype that is based on the priori knowledge dependent "anti-virus anti-Trojan" cannot solve the integrated S&S

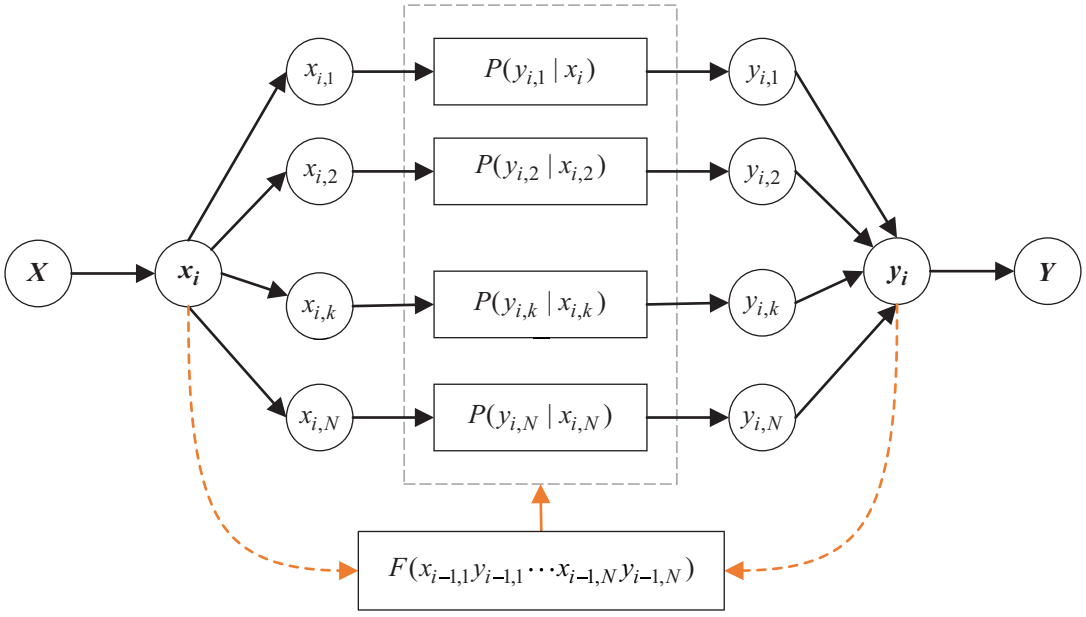

Figure 7. Information transmission in the DHR with the feedback memory-elimination construction channel [12]

problem, nor can the "divide-and-conquer" mode be expected to obtain the generalized functional safety with quantifiable design and test metrics.

Obviously, the generalized functional safety problem not only overturns the hypothetical prerequisites of the classical reliability theory, but also introduces unprecedented crisis to the classical functional safety theories, methods, and practice norms. Therefore, many disciplines and related technical fields will inevitably face this common cross-cutting problem, and there is an urgent need to create new theories, methods, and reliability empowering architectures that can solve the S&S problem in an integrated manner.

6.1 Claims required to be emphasized

(1) The DHR architecture based on iterative control and policy verdict mechanism, has features similar to the inaccuracy measurement effect in quantum physics and mimetic defense fog in biology [2, 8, 10], which can be an integrated solution to the S&S problem triggered by random failures or cyberattacks based on software and hardware design flaws, without relying on (but not excluding) external priori knowledge and additional defense measures. In other words, DHR is an empowering architecture to obtain generalized functional safety, which not only can meet the application requirements of quantifiable design and metrics, but also has generalized robust control properties that are more reasonable than any current structures [4].

(2) The economic cost of the DHR structure is strongly related to the number of functions (or facilities) that need to be protected in the target system and the deployment scheme of "security barriers". Taking the market transaction price of the same type of COTS-level high-reliability products as a reference, the cost increase is less than 20% [8].

(3) The dissimilarity redundancy structure, regardless of its high design and maintenance costs, lacks in temporal robustness and quality robustness under cyberattack conditions [8], and thus is not a candidate solution to the S&S problem.

6.2 Things required to pay attention

(1) If the heterogeneity between elements in set K is large enough, there is no theoretical possibility of common-mode escape [8]. However, in engineering practice to ensure the heterogeneity to be

sufficiently large, currently it can only be achieved with high product design costs and whole-life maintenance costs, like the case of DRS. Fortunately, the DHR architecture can eliminate the differential- or common-mode expressions caused by current cyberattacks, by simply recovering, reconfiguring, or replacing elements in set N of current processing scenarios through its own feedback control-based dynamic iterative processing mechanism. In other words, the dynamic heterogeneity formed by the feedback iteration mechanism is used to reduce the static heterogeneity requirement among processing scenarios in K. This means that, the engineering implementation allows a certain degree of combinatorial common-mode vulnerability among functionally equivalent heterogeneous elements in the processing scenario set K. In addition, as long as the elements in the current processing scenario set N can be selected or reconstructed appropriately, the persistent attack against a combinatorial common-mode vulnerability can be defeated.

(2) If the successful attack over common-mode vulnerability is strongly related to the current processing scenario set N, then the dynamics, diversity, and uncertainty of set N can bring too much uncertainty to the building process and stable maintenance of the attack chain [8]. For example, some processing scenarios have honeypots installed, some have sandboxes set, and some have firewalls deployed; then the same vulnerability in different processing scenarios requires different stealthy attack methods. In addition, a coordinated common-mode escape has to be achieved within the DHR construction that does not have precise synchronization mechanisms, which exponentially increases the attack complexity and the difficulty of maintaining the stability of the attack chain.

(3) If there exist processing scenario-independent common-mode vulnerabilities (including backdoors or Trojan-Horse with cross-platform capabilities) and the potential impact caused by attacks is limited to the action scope of the application software where the common-mode vulnerability exists, the DHR construction theoretically has no desired efficacy in that case [8]. Therefore, in engineering practice, the methods to detect on time and block such common-mode vulnerabilities in application software must be studied.

At present, the generalized functional safety empowered by the DHR architecture is being demonstratively applied in China's domestic industry fields of Information Technology (IT)/Information Communication Technology (ICT)/Industrial Control System (ICS)/CPS and smart cars, and the industrial applications show that "the results remarkably meet theoretical expectations". Meanwhile, the purple mountain lab (PML) has established a global, round-the-clock cyberspace endogenous security testbed – NEST, which can support prototype testing for various types of COTS-level endogenous security or generalized functional safety products. In addition, the Zhejiang Lab (ZJLab) is also establishing the world's largest ICS endogenous security testbed, which is expected to be completed in 2023.

As an important derivative result, the DHR architecture can solve the pain point problem of the insurance industry that the generalized functional safety "cannot be quantified accurately", so that the insurance-type services related to cyber security and functional safety can get rid of the embarrassment of small markets compared to the large scale of digital economy. Therefore, the DHR architecture has the potential to contribute indispensable influence on financial industry for the development of a new generation of information technology and product markets with generalized functional safety attributes, and can also give play to the "leverage" role of the capital market in technology and industrial development to completely reverse the imbalance of the cyberspace attack and defense pattern.

Conflict of Interest

The author declares no conflict of interest.

Data Availability

No data are associated with this article.

Acknowledgements

Thanks Caixia Liu, Yuanyuan Liu and the anonymous reviewers for their helpful comments and suggestions.

Funding

This work was supported by the National Natural Science Foundation Innovation Group Project (61521003).

References

[1] Part0 IEC 61508. Functional Safety. Geneva: International Electrotechnical Commission, 2005.
[2] Wu JX. An Introduction to Cyberspace Mimic Defense. China: Science Press, 2017 (Chinese).
[3] Wu JX. Development paradigms of cyberspace endogenous safety and security. Sci China Inf Sci 2022; 65: 156301. https://doi.org/10.1007/s11432-021-3379-2.
[4] ISO – International Organization for Standardization. ISO 26262 Road Vehicles Functional Safety. 2011. https://quality-one.com/iso-26262/.
[5] Parts 1 – 7 IEC 61508. Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems. Geneva: International Electrotechnical Commission, 2005.
[6] Upstream Security's 2020 Global Automotive Cybersecurity Report. https://upstream.auto/upstream-security-global-automotive-cybersecurity-repot-2020/.
[7] Jajodia S, Ghosh AK and Swarup V. Moving Target Defense. China: Springer, August 2011.
[8] Wu JX. Cyberspace Mimic Defense: General Robust Control and Endogenous Safety & Security. China: Springer, December 2019.
[9] Kavallieratos G, Katsikas S and Gkioulos V. Cybersecurity and safety co-engineering of cyber physical systems – A comprehensive survey. Future Internet 2020; **12**: 65.
[10] Wu JX. Cyberspace Endogenous Safety and Security: Mimic Defense and General Robust Control. China: Science Press, 2020 (Chinese).
[11] Shannon CE and Weaver W. The Mathematical Theory of Communication. Urbana: University of Illinois Press, 1949.
[12] Wu JX. Cyberspace endogenous safety and security. Engineering 2021, in press. https://doi.org/10.1016/j.eng.2021.05.015.

JiangXing Wu is currently a professor and the Director of China National Digital Switching System Engineering and Technological R&D Center (NDSC). He was selected as an academician of China Academy of Engineering in 2003. He is an active scientist in the communication and information system as well as in the computer and network technologies. He created the mimic computing and mimic security theory, and developed the first mimic computer in the world. His general research interests include the communication and information system, computer architecture, and cyber security.