# When the Safe State Is the Payoff: A Demand-Origin Framework for Bounding Adversary-Selected Spurious Activation of Safety-Instrumented Functions

Jherrod Thomas
Independent Researcher — The Lion of Functional Safety
jherrodthomas.com · September 2026

Abstract—On September 1, 2026, CISA published ICSA-26-244-03 against the Rockwell Automation Logix platform: CVE-2026-9637, a length-validation defect in Common Industrial Protocol message processing that drives ControlLogix 5580, CompactLogix 5380, GuardLogix 5580 and Compact GuardLogix 5380 controllers into a major nonrecoverable fault clearable only by a physical power cycle. The finding is availability-only, unauthenticated, remotely reachable, and has no workaround. Two of the four affected families are SIL 3 safety controllers, which means the induced fault drives the safety-instrumented function to exactly the de-energized state its safety manual promises. Conventional reliability accounting has no place to record this. The spurious trip rate of a safety-instrumented function is treated as an aggregate of component safe-failure rates, evaluated as an economic quantity, and assumed to sample process states in proportion to their occupancy. None of those three assumptions survives an adversary who can command the trip. We propose a demand-origin partition that separates genuine demand, random spurious activation, and a third class we call commandable spurious activation, and we develop five relations over it. The central results are structural and arithmetic rather than probabilistic. First, the commandable component is bounded by no component reliability parameter and is unaffected by proof testing. Second, its cut order equals the voting threshold K, so every architecture adopted to reduce the probability of failure on demand also reduces the number of channels an adversary must reach. Third, because the adversary selects the moment, the expected consequence of an identical trip is the maximum over observable process states rather than the occupancy-weighted mean, giving an amplification factor that is empirically between two and six for batch chemistry. Fourth — and this is the result that moves the finding out of the availability column — the restoration interval following an induced fault is a period of safety-function unavailability that enters the probability of failure on demand directly. One induced fault per year with an eight-hour physical restoration consumes 91 percent of the entire SIL 3 budget. We instantiate the framework on a GuardLogix-based reactor trip, derive six requirements anchored in IEC 61511-1, IEC 62443-3-3 and IEC 62443-4-1, and argue that "availability-only" is a category error when the asset whose availability is lost is the protection layer.

Index Terms—IEC 61511, IEC 62443, Safety-Instrumented System, Spurious Trip Rate, Denial of Service, Probability of Failure on Demand, Common Industrial Protocol, Operational Technology, Security-Informed Safety, Transient Operations

## I. Introduction

The functional safety community has spent forty years making the safe state reliable. It has spent almost none making the safe state exclusive. Every de-energize-to-trip architecture rests on a premise so deeply assumed that it is rarely written as a premise at all: that the only things able to drive the final elements to their de-energized position are a genuine process demand and the random failure of a component. That premise was true when the path to the final element was a current loop and a relay. It is not true when the path is a routed Ethernet segment and the logic solver accepts unauthenticated messages on a well-known port.

The occasion for this paper is an advisory that most security triage processes would rank as unremarkable. On September 1, 2026, the Cybersecurity and Infrastructure Security Agency published ICSA-26-244-03 against the Rockwell Automation Logix platform, disclosing CVE-2026-9637 [1]. The defect is a length-validation failure in Common Industrial Protocol (CIP) message processing, classified CWE-119, carrying a CVSS v3.1 base score of 7.5 and a CVSS v4.0 base score of 8.7, self-reported by the vendor. A single malformed CIP message places an affected controller into a major nonrecoverable fault (MNRF): the controller halts, outputs assume their configured fault state, and the controller remains there until a person physically power-cycles it. Exploitation requires network reachability to the CIP service on TCP/UDP 44818 and nothing else — no credentials, no engineering-workstation foothold, no operator interaction. There is no workaround; the remedy is firmware [1], [2].

The affected families are ControlLogix 5580, CompactLogix 5380, GuardLogix 5580 and Compact GuardLogix 5380. The last two are SIL 3 / PL e safety controllers. In a plant where a GuardLogix executes a safety-instrumented function (SIF), the consequence of CVE-2026-9637 is that an unauthenticated party with a route to the control network can, at a moment of their choosing, cause the SIF to execute correctly. The controller is not defeated. The safety manual is not violated. The de-energized state is reached exactly as designed, and the plant is down.

In the CVSS vector this is an availability impact with confidentiality and integrity impacts of None, which is the notation's honest summary of what happens to the data. It is a poor summary of what happens to the risk. The reason is that the asset whose availability was lost is the protection layer, and in a protection layer the

transition to the safe state is not free. Lundteigen and Rausand established the taxonomy two decades ago and stated the consequence plainly: spurious activation produces production loss, stress on the affected components and systems, and hazards during system restoration [3]. The last of those three is the one that matters here and the one the availability-only reading discards entirely.

Nor is the exposure confined to process control. The same September fortnight produced ICSA-26-239-03 against Rockwell's OTTO Fleet Manager — a bcrypt work factor set low enough that offline recovery of credentials from a backup is tractable, in the software that holds mission assignment and the site map for fleets of PL d safety-rated autonomous mobile robots [4] — and, three weeks earlier, ICSA-26-225-05 against the ANDRITZ HIPASE-250 hydropower automation platform, whose findings included an undocumented endpoint that reduces audit logging without authenticating the caller [5]. The pattern across all three is identical: a control affecting the safety layer is reachable from a zone whose population was never enumerated in the safety argument.

What is missing is not awareness that cyber events can affect safety. IEC 61511-1:2016 §8.2.4 has required a security risk assessment of the SIS since its publication, and the co-engineering literature surveyed by Kriaa et al. [6] and by Lisova, Šljivo and Čaušević [7] has mapped the methodological landscape thoroughly. What is missing is a quantity: a way to write the adversary-commandable share of spurious activation into the same ledger that already holds the spurious trip rate and the probability of failure on demand, so that a safety case can show its effect rather than assert its existence in prose.

The contributions of this paper are as follows.

- A demand-origin partition that splits SIF activations into genuine demand, random spurious activation, and commandable spurious activation, and a tabulation of the properties that distinguish the third class from the second — including three properties under which it behaves in the opposite direction to engineering intuition.
- Five numbered relations giving the partition, the voting-architecture dependence of the commandable cut order, the zone-summed commandable rate, a timing amplification factor that expresses the consequence penalty of adversarial moment-selection, and a restoration-unavailability term that carries an availability-only finding into the probability of failure on demand.
- A structural result: the commandable cut order equals the voting threshold K, so redundancy adopted to lower $\mathrm{PFD}_{\mathrm{avg}}$ simultaneously lowers the number of channels an adversary must reach, inverting the usual relationship between redundancy and robustness.
- A worked application to a GuardLogix-based exothermic batch reactor trip under CVE-2026-9637, with an amplification factor of 5.5 and a restoration term consuming 91 percent of the SIL 3 budget, and six derived requirements traceable to IEC 61511-1 §8.2.4 and §11.2.10, IEC 62443-3-3 SR 5.1, SR 7.1 and SR 7.2, and IEC 62443-4-1 SVV-3.

Section II reviews the background and locates the gap. Section III develops the framework. Section IV applies it. Section V discusses limitations. Section VI concludes.

## II. Background and Related Work

### 0.0.1 A. Spurious Activation as a Reliability Quantity

The reliability treatment of spurious activation is mature and, for its intended purpose, sound. Lundteigen and Rausand separated the three concepts that practice routinely conflates — spurious operation of a component, spurious trip of a subsystem, spurious shutdown of a process — and supplied formulas for the aggregate rate that remain the reference point [3]. Jigar, Liu and Lundteigen revisited the failure classification and propagation assumptions underlying those formulas and produced a corrected set [8]. Innal, Liu and colleagues extended the treatment to redundant architectures, where the spurious and dangerous cut orders move in opposite directions as the voting threshold changes [9]. Lundteigen's survey of open problems in SIS reliability identified spurious activation as underresearched relative to its operational importance [10].

Three assumptions are common to this entire body of work, and all three are appropriate to it. First, the spurious activation rate is an aggregate of component safe-failure rates and is therefore bounded above by component reliability. Second, spurious activations are statistically independent of process state, so the state a trip lands in is sampled according to that state's occupancy fraction. Third, the consequence of a spurious trip is primarily economic, so the design objective is a rate ceiling traded against the probability of failure on demand. IEC 61511-1 §11.2.10 requires that the SIS design consider spurious trips; it does not require the analyst to ask who is causing them.

### 0.0.2 B. Transient Operations and the Cost of a Trip

The second assumption above is where the process safety literature has the most to say, and what it says is unambiguous. Vaughen's analysis of more than fifty incidents attributed in part to transient operations concluded that startups, shutdowns and other non-routine states carry risk out of all proportion to their share of operating time [11]. Santana and colleagues reach the same conclusion from the operational-readiness direction, cataloguing the practices whose absence turns a routine start-up into a loss-of-containment event [12]. The refining industry's own figure — that a facility spends under ten percent of its time in transient operations and incurs roughly half its process safety events there — is repeated widely enough to be folklore, but

the peer-reviewed incident analyses support the direction and rough magnitude [11], [12].

The implication for the present problem is direct and, as far as we can determine, has not been drawn. If the hazard associated with a trip depends strongly on the process state in which the trip occurs, and if the distribution of trip-consequence across states is highly skewed, then who chooses the moment is a first-order parameter of the risk calculation. Random spurious activation samples the occupancy distribution. An adversary does not.

#### 0.0.3 C. Cyber Events in the Process Industry

Empirically, attacks on process facilities are neither hypothetical nor uniformly aimed at the spectacular outcome. Iaiani, Tugnoli, Bonvicini and Cozzani reviewed eighty-two cybersecurity-related incidents in the process industry and adjacent sectors and found that a substantial fraction targeted operational technology directly rather than enterprise systems [13]; their later work develops reference attack scenarios suitable for security vulnerability assessment [14]. The TRITON/TRISIS campaign against Triconex safety controllers remains the canonical case and carries a detail directly relevant here: the plant shut down not because the attacker intended a shutdown but because a defect in the attacker's own code triggered the controllers' internal diagnostics, and it was that shutdown which exposed the intrusion [15], [16]. The safety system's transition to its safe state was, in the only publicly documented attack on an SIS, both an unintended consequence and the detection event.

The security literature has also developed the quantitative vocabulary for the adversary side. McQueen and colleagues' time-to-compromise model expresses attacker progress as a random process parameterized by known vulnerabilities and attacker skill [17], and the attack-tree formalism surveyed by Kordy, Pièvre-Cambacédès and Schweitzer supplies the compositional machinery [18]. Contemporary tutorial and threat-modeling treatments of operational-technology networks [19], [20] and improved vulnerability-scoring work for industrial control systems [21] address the exposure side. What none of these supply is the bridge: a term that a functional safety engineer can add to an existing reliability model.

#### 0.0.4 D. Co-Engineering and the Standards Position

IEC 61511-1:2016 §8.2.4 requires a security risk assessment of the SIS specifically, identifying the devices in scope, the threats, the vulnerabilities, the consequences and the countermeasures, and refers the analyst to IEC 62443 for method [22]. IEC 62443-3-3 supplies the system-level requirements, of which three bear on this class of finding: SR 5.1 network segmentation, SR 7.1 denial-of-service protection with a requirement for degraded-mode operation, and SR 7.2 resource management [23]. IEC 62443-4-1 SVV-3 requires robustness and fuzz testing of externally reachable interfaces during development [24]. IEC 61508-2 gives the de-energize-to-trip architecture its credit but contains no clause asking who may command the transition [25].

The co-analysis literature has characterized the general difficulty. Kriaa et al. surveyed the combined approaches available to industrial control systems and classified them by whether safety informs security, security informs safety, or the two are genuinely combined [6]; Lisova and colleagues' systematic review found thirty-three methods, most unvalidated against field data [7]; Kriaa, Bouissou and Laarouchi later proposed a unified framework [26]; and a recent model-driven review documents that the field's output remains predominantly methodological [27]. Sector-specific work on offshore SIS under digitalization reaches a similar conclusion [28].

#### 0.0.5 E. The Gap in One Sentence

No current method expresses adversary-commandable spurious activation as a term in the same reliability ledger that holds the spurious trip rate and the probability of failure on demand, and consequently no safety case shows that an availability-only vulnerability in a logic solver moves both quantities.

### III. Approach: The Demand-Origin Partition

#### 0.0.6 A. Objects and Definitions

Let a SIF comprise a sensor subsystem, a logic solver, and a final-element subsystem, arranged in a K-out-of-N (KooN) voting architecture in which the SIF trips when at least K of N channels demand a trip. Let the SIF be de-energize-to-trip, so the safe state is reached by removal of energy and is therefore also the state reached by any loss of the logic solver's output capability.

Definition 1 (Demand origin). Every activation of the SIF has exactly one origin: a genuine demand, arising from the process condition the SIF exists to detect; a random spurious activation, arising from safe failure of a component; or a commandable spurious activation, arising from an act of an external agent exercising a reachable interface.

Definition 2 (Commandable safe-state path). A commandable safe-state path (CSSP) exists from zone z to a SIF channel if an agent resident in z can, without possessing an authorization credential that the agent is not entitled to, cause that channel to enter its de-energized state. A CSSP is a property of the deployment, not of the product.

Definition 3 (Restoration interval). Following a commandable spurious activation, the restoration interval $\mathrm{MDT}_{\mathrm{r}}$ is the interval from the induced fault until the SIF is again capable of responding to a genuine demand.

Definition 3 is the one practitioners omit, and its omission is what keeps the finding in the availability column. A controller in a major nonrecoverable fault

has not merely stopped controlling; it has stopped protecting. Between the induced fault and the physical power cycle, the SIF's availability is zero. That is a dangerous-side quantity wearing a safe-side costume.

Fig. 1 draws the resulting architecture: the zone population on the left, the conduit and the unauthenticated CIP interface in the middle, and on the right the two distinct consequence paths — the immediate trip, whose severity depends on the process state at that instant, and the restoration window, during which the protection layer is absent.

### 0.0.7 B. The Partition and the Random Component

The total activation rate of the SIF decomposes by origin as

$$_{\text{act}} = _{\text{D}} + _{\text{ST}} + _{\text{CS}} \tag{1}$$

where $_{\text{D}}$ is the genuine demand rate, $_{\text{ST}}$ the random spurious activation rate, and $_{\text{CS}}$ the commandable spurious activation rate. Relation (1) is trivial as arithmetic and consequential as bookkeeping: the third term does not appear in any spurious-trip calculation we have encountered in practice, and the standards that require the spurious trip rate to be evaluated do not name it.

For a KooN architecture with per-channel safe-failure rate $_{\text{S}}$, mean downtime MTTR, and safe-side common-cause factor $_{\text{S}}$, the random component follows the established treatment [3], [8], [9]:

$$_{\text{ST}} \quad \text{C(N, K)} \cdot \text{K} \cdot {}_{\text{S}}^{\text{K}} \cdot \text{MTTR}^{\text{K}-1} + _{\text{S}} \, _{\text{S}} \tag{2}$$

The exponent K in relation (2) is the point of contact with what follows. A spurious trip requires K channels to trip spuriously, whereas a dangerous failure requires $\text{N} - \text{K} + 1$ channels to fail dangerously. The two cut orders move in opposite directions as K varies, which is the familiar trade: 1oo2 halves the dangerous cut order relative to 2oo2 and doubles the spurious exposure, 2oo3 buys back some of each.

### 0.0.8 C. The Commandable Component

The third term obeys none of relation (2)'s structure. Let Z index the zones from which the control network is routable, let $_{\text{z}}$ be the attempt rate originating in zone z, let $\text{n}_\text{z}$ be the number of SIF channels reachable from z over a CSSP, and let $\text{p}_{\text{b,z}}$ be the probability that a deployed control blocks the attempt before it reaches the interface. Then

$$_{\text{CS}} = _{\text{z}\,\text{Z}} \; _{\text{z}} \, (1 - \text{p}_{\text{b,z}}) \cdot 1[\text{n}_\text{z} \quad \text{K}] \tag{3}$$

Three properties follow immediately, and all three run against the direction a reliability engineer expects.

Property 1 (No reliability bound). $_{\text{CS}}$ contains no $_{\text{S}}$. Improving component reliability by an order of magnitude reduces $_{\text{ST}}$ by an order of magnitude and leaves $_{\text{CS}}$ exactly where it was. For the same reason, proof testing — which restores the SIF's dangerous-failure position and detects accumulated safe-side degradation — does nothing to the third term.

Property 2 (Inverted redundancy). The indicator in (3) fires when $\text{n}_\text{z}$ K, so the commandable cut order is K, the same exponent that appears in (2). A 1oo2 architecture adopted to reduce $\text{PFD}_{\text{avg}}$ sets $\text{K} = 1$, which means an adversary who reaches any single channel commands the trip. Redundancy adopted for dangerous-failure tolerance therefore lowers the commandable cut order rather than raising it. This is the opposite of the intuition that redundancy is robustness, and it holds for exactly the same reason that 1oo2 has a higher spurious trip rate than 2oo2 — with the difference that here the additional exposure is not rate-limited by $_{\text{S}}$.

Property 3 (Homogeneous channels are one channel). When the N channels are identical products at identical firmware revisions, a single defect class reaches all of them simultaneously, so the effective $\text{n}_\text{z}$ is N whenever it is one. Diversity, which dependent-failure analysis already recommends against common-cause hardware failure, is the only architectural measure in (3) that does anything at all.

### 0.0.9 D. The Timing Amplification Factor

Let the process occupy states $\text{s} \quad \text{S}$ with occupancy fractions $_{\text{s}}$ summing to unity, and let $\text{C}_\text{s}$ be the consequence of a SIF activation occurring in state s, expressed on any consistent scale. Let $\text{S}_{\text{obs}} \quad \text{S}$ be the subset of states an adversary can distinguish from the information available to them — which, for an agent on the control network, is frequently most of them. A random spurious activation draws its state from . An adversary does not draw; they wait. The ratio of the two expected consequences is the timing amplification factor

$$\text{A} = \max_{\text{s}\,\text{S<sub>obs</sub>}} \text{C}_\text{s} \div _{\text{s}\,\text{S}} \; _{\text{s}} \, \text{C}_\text{s} \tag{4}$$

$\text{A} \quad 1$ always, with equality only when consequence is state-independent or the adversary is blind. Its magnitude is set entirely by the skew of the $\text{C}_\text{s}$ distribution against the distribution, which is to say by exactly the phenomenon the transient-operations literature documents: consequence concentrated in states of low occupancy [11], [12]. Relation (4) is the formal statement of why an adversary need not defeat a safety function to harm a plant. Commanding it at the right moment is sufficient, and the right moment is cheap to wait for.

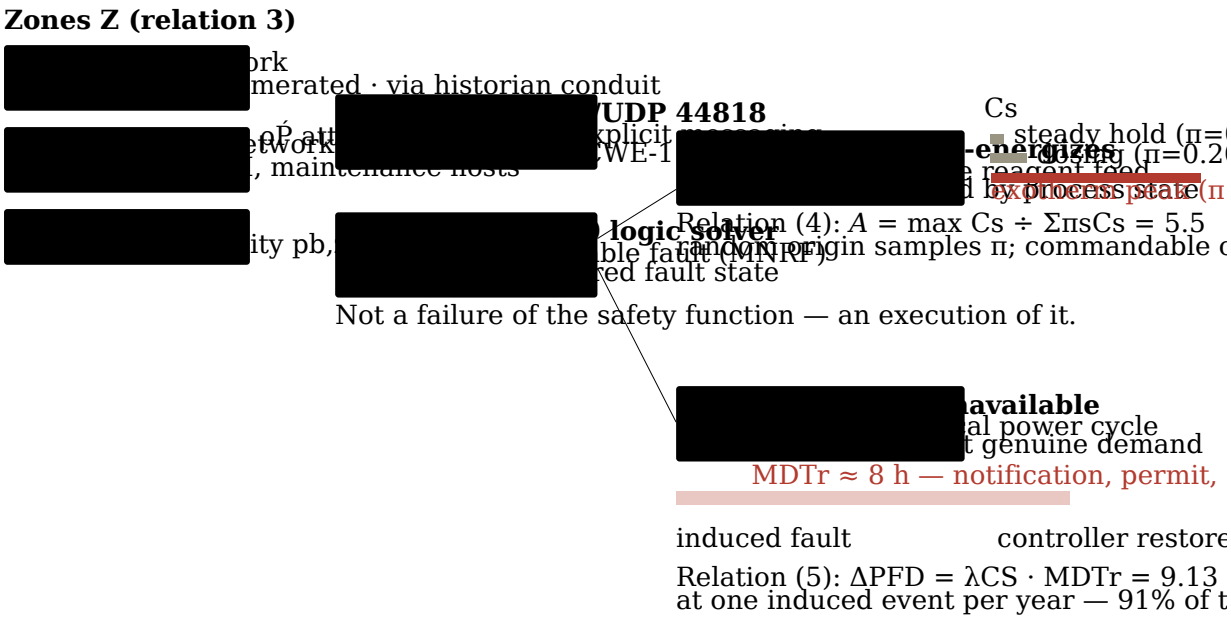
The state-indexed risk contribution of spurious activation is then

$$\text{R}_{\text{sp}} = (\, _{\text{ST}} + \text{A} \cdot \; _{\text{CS}}) \cdot \; _{\text{s}\,\text{S}} \; _{\text{s}} \, \text{C}_\text{s}$$

which is written here without a number because it is an immediate consequence of (3) and (4) rather than an independent relation.

### 0.0.10 E. The Restoration Term

The final relation is the one that moves the finding out of the availability column. During the restoration




Branch B is the branch an availability-only classification discards. It is a dangerous-side qua

Figure 1: The commandable safe-state path and its two consequence branches. The left column is the zone population able to reach the logic solver's CIP service; relation (3) sums over it. The centre shows the induced major nonrecoverable fault, which is not a failure of the safety function but an execution of it. The upper right branch is the trip itself, whose consequence is indexed by the process state at the moment of activation and is therefore subject to the amplification of relation (4). The lower right branch is the restoration window of relation (5), during which the SIF is unavailable to a genuine demand — the branch that an availability-only classification discards, and the one that enters $\mathrm{PFD}_{\mathrm{avg}}$.

interval of Definition 3, the SIF cannot respond to a genuine demand. That is unavailability in the strict IEC 61508 sense, and it enters the average probability of failure on demand additively:

$$\mathrm{PFD}_{\mathrm{avg}} = \mathrm{PFD}_{\mathrm{rand}} + \ _{\mathrm{CS}} \cdot \mathrm{MDT}_{\mathrm{r}} \qquad (5)$$

Relation (5) has no security content whatever. It is the standard treatment of known downtime, applied to a downtime source that the standard treatment does not enumerate. Its force is arithmetic. For a SIL 3 SIF the ceiling on $\mathrm{PFD}_{\mathrm{avg}}$ is $10^{-3}$ [25]. A restoration requiring a person to reach the cabinet and power-cycle the controller is optimistically eight hours in a staffed plant and considerably longer at a remote or unmanned facility. Eight hours is $9.13 \times 10^{-4}$ of a year. One induced fault per year therefore consumes 91.3 percent of the entire SIL 3 budget before a single component failure has been counted. Four per year — a plausible figure for an adversary probing rather than committing — yields $3.65 \times 10^{-3}$, which is outside SIL 3 altogether and sits in the SIL 2 band.

Table I sets out the properties that distinguish the three origin classes.

| Property | Genuine demand | Random spurious | Commandable spurious |
|---|---|---|---|
| Rate bounded by component reliability | No (process property) | Yes ( _S) | No |
| Reduced by improved component quality | No | Yes | No |
| Reduced by proof testing | n/a | Partially | No |
| Cut order in KooN | N−K+1 (dangerous) | K | K |
| Effect of adding redundancy for PFD | Improves | Worsens | Worsens |
| Mitigated by channel diversity | n/a | Partially ( _S) | Yes |
| Process state at activation | Correlated with hazard | Sampled from | Selected: max C_s |
| Appears in conventional STR calculation | No | Yes | No |
| Appears in conventional PFD_avg | n/a | No | |

No — but should, via (5)
Detected by the SIS diagnostics
n/a
Usually
As a fault, not as an attack

TABLE 1. Properties of the three demand-origin classes. The right-hand column is the one that carries the argument: every row on which the commandable class differs from the random class is a row on which existing reliability practice has no place to record the difference.

## IV. Worked Example: A GuardLogix Reactor Trip Under CVE-2026-9637

### 0.0.11 A. The Item and Its Architecture

Consider a semi-batch exothermic reactor with a high-temperature/high-pressure SIF assessed at SIL 3 by layer-of-protection analysis. The SIF comprises three transmitters in 2oo3 voting, a GuardLogix 5580 safety controller, and two de-energize-to-trip block valves on the reagent feed. Trip action isolates reagent feed and opens the quench path. This is an ordinary configuration; nothing about it is chosen to make the argument work.

The controller is on a Level 2 control network that is routed, for engineering and historian access, to a Level 3 site network. Firmware is V35.012 — within the affected set of ICSA-26-244-03 [1]. The CIP service is reachable on TCP/UDP 44818 from any host on Level 2, and, through the historian conduit, from a set of Level 3 hosts whose population nobody has enumerated since commissioning.

### 0.0.12 B. Applying the Partition

Step 1 — Origins. $_{D}$ is taken as 0.1/yr from the LOPA. $_{ST}$ is the vendor-supported figure for the assembled loop, on the order of 0.05/yr.

Step 2 — CSSP enumeration. The logic solver is a single channel in the sense of relation (3): whatever the sensor voting, one malformed CIP message to the controller drives the whole SIF to its de-energized state. K is therefore effectively 1 at the logic-solver layer regardless of the 2oo3 transmitter arrangement — an instance of Property 2 in its starkest form, since the 2oo3 arrangement was adopted precisely to avoid single-channel trips. Two zones reach the interface: Level 2 with an enumerated population, and Level 3 through the historian conduit with an unenumerated one. No authentication is required on the CIP explicit-messaging path in the deployed configuration [1], [2], so $p_{b,z}$ is whatever the conduit firewall rules supply and nothing more.

Step 3 — State indexing. Table II gives the occupancy and consequence profile. Consequence is normalized so that a trip from steady hold, which costs a batch and a restart, is 1.

Process state s
Occupancy _s
Consequence C_s
_sC_s
In S_obs?
Steady hold
0.55
1
0.55
Yes
Reagent charging / dosing
0.20
4
0.80
Yes
Exotherm peak
0.08
25
2.00
Yes
Start-up / heat-up
0.10
8
0.80
Yes
Planned shutdown / quench
0.07
6
0.42
Yes
Totals
1.00
—
4.57
—

TABLE 2. State-indexed consequence profile for the semi-batch reactor. Consequence is normalized to a trip from steady hold. The exotherm-peak entry dominates because a feed isolation during peak heat release removes the cooling duty's matching reagent control while the accumulated reagent inventory continues to react, which is the standard runaway precursor.

The occupancy-weighted mean consequence is 4.57. The maximum over observable states is 25. Relation (4) gives A = 25 / 4.57 = 5.47. A random spurious trip costs 4.57 normalized units in expectation; a commanded one costs 25. The trip is identical. The safety function is identical. Only the clock differs.

The observability assumption deserves a note in both directions. An adversary who has reached the CIP service can generally read process tags, so $S_{obs} = S$ is the realistic case. An adversary restricted to network-timing inference might distinguish only start-up from running, giving $S_{obs}$ = {steady hold, start-up} and A = 8 / 4.57 = 1.75. Even the blind-adversary bound is a 75 percent consequence penalty over the random case, which is to say that relation (4) does not depend on an unusually capable attacker to produce a non-trivial number.

Step 4 — The restoration term. The MNRF requires a physical power cycle [1]. The reactor hall is staffed on day shift only; the realistic restoration interval spans notification, permit, access and restart of the controller, and eight hours is not pessimistic. By relation (5), with $_{CS}$ conservatively set to one event per year, the restoration contribution to $PFD_{avg}$ is $9.13 \times 10^{-4}$. The SIL 3 ceiling is $10^{-3}$ [25]. The randomly-derived $PFD_{rand}$ for this loop, before any of this, is perhaps $4 \times 10^{-4}$. Their sum is $1.31 \times 10^{-3}$, and the SIF is no longer SIL 3.

This is the whole argument in one line of arithmetic. An advisory whose CVSS vector reads availability-only, whose integrity impact is None, and whose triage in most organizations is "patch at next window," has removed the SIF's integrity level. Not degraded it — removed it, in the sense that the verification claim on which the LOPA's risk reduction was predicated no longer holds.

Step 5 — The bypass term, unquantified. IEC 61511-1 §11.2.10's concern with spurious trips is not only economic. An adversary able to induce trips repeatedly is running a campaign against the operators' confidence in the safety layer, and the documented endpoint of that campaign is the operational bypass. We decline to put a number on this. We note only that it is the mechanism by which a sequence of availability-only events becomes a dangerous-side outcome with no further technical action by the attacker, and that it is invisible to every relation in Section III.

### 0.0.13 C. A Second Instance: Fleet Manager as a Safety-Relevant Configuration Store

ICSA-26-239-03 presents the same structure in a different domain [4]. OTTO Fleet Manager is not a safety controller; the autonomous mobile robots it dispatches carry their own PL d protective stop. But the Fleet Manager holds the site map and the mission assignment, which is to say it holds the assumptions under which the on-robot safety functions were validated. A credential recovered offline from a backup is a CSSP to the configuration rather than to the de-energized state, and relation (3) applies unchanged with $n_z$ counting the robots whose mission parameters the compromised account can rewrite. The consequence-indexing of relation (4) also applies, because the

states in which a fleet-wide mission change is most harmful — shift change, mixed human-robot traffic in a staging area — have low occupancy and high $C_s$. We do not develop this case further here; we note it because it shows that the framework's objects are the reachable interface and the state-indexed consequence, not the safety controller specifically.

### 0.0.14 D. Derived Requirements

ID
Requirement
From
Anchor
R1
The SIS security risk assessment shall enumerate, for each SIF, every zone from which a commandable safe-state path exists to any channel, and shall record the actor population of each such zone.
(3)
IEC 61511-1 §8.2.4
R2
The spurious trip analysis shall report the commandable component separately from the random component, and shall state explicitly that the commandable component is not bounded by component reliability and is not reduced by proof testing.
(1), (2), (3)
IEC 61511-1 §11.2.10
R3
The SIF's consequence analysis shall be indexed by process operating state, and the residual risk statement shall report both the occupancy-weighted mean and the maximum over adversary-observable states.
(4)
IEC 61511-1 §8.2.2, §9
R4
Where a logic solver fault requires physical intervention to clear, the restoration interval shall be estimated from the staffing and access model of the actual installation and shall be entered into PFD_avg as a known unavailability.
(5)
IEC 61508-2 §7.4; IEC 61511-1 §11.9
R5
Conduits terminating on a safety logic solver's protocol service shall enforce allow-listed source addressing and shall be verified against SR 5.1; degraded-mode operation under protocol flooding shall be demonstrated against SR 7.1 rather than assumed.
(3)
IEC 62443-3-3 SR 5.1, SR 7.1, SR 7.2
R6
Protocol-parsing interfaces reachable without authentication shall be subject to documented robustness and fuzz testing, and the resulting evidence shall be named in the safety case's supporting-tool and component-justification record.
Property 1
IEC 62443-4-1 SVV-3; IEC 61508-2 §7.4.2.12

TABLE 3. Requirements derived from the worked example, with the relation that motivates each and the standards clause under which it would be written.

R4 is the cheapest of the six and the most consequential. It requires no new analysis technique, no security expertise, and no vendor cooperation — only that an existing arithmetic operation be performed on a number the plant already knows. It is also the requirement whose absence allowed CVE-2026-9637 to be classified as an availability finding by everyone who looked at it, including, on the evidence of the advisory's own scoring, the vendor who reported it.

## V. Discussion

The commandable rate is not a frequency and should not be reported as one. Relation (3) sums attempt rates, and there is no defensible base rate for $_z$. CISA reports no known public exploitation of CVE-2026-9637 [1]. We use $_{CS} = 1/\text{yr}$ in Section IV not as a prediction but as a unit-scaled conditional: the arithmetic of relation (5) is best read as "what one event per year costs," and the reader who believes the true rate is a tenth of that should divide the restoration term by ten and observe that it still consumes nine percent of the SIL 3 budget. The framework's defensible use is comparative and conditional, not predictive.

Consequence indexing imports a known difficulty. The $C_s$ values in Table II are structured engineering judgment, and the amplification factor inherits all of their uncertainty. Two mitigations apply. First, A is a ratio, so multiplicative errors common to all states cancel; only the skew matters. Second, the direction of the result is robust: A  1 by construction, and A is large exactly when the transient-operations literature says consequence is concentrated [11], [12]. An analyst who distrusts our numbers can still perform step 3 and obtain a defensible lower bound by setting the consequence of every state except the worst to the steady-hold value.

Property 2 is uncomfortable and should be. The claim that redundancy adopted for dangerous-failure tolerance worsens the commandable exposure is not an argument against redundancy. Dangerous-failure tolerance is the reason the architecture exists and the term it improves is the term the LOPA credits. The claim is narrower: the trade that voting architecture makes between dangerous and spurious cut orders has always been understood as a trade against random spurious activation, and the commandable component rides along on the spurious side of that trade without being priced. Where diversity is available it prices the term downward; where it is not — and the worked example's homogeneous firmware revision is the common case — it does not.

The restoration term generalizes past this advisory. Relation (5) applies to any fault mode requiring physical intervention, whether induced or not. That includes latched faults, fail-secure lockouts, and the increasingly common pattern in which a security control's response to an anomaly is to halt the component. A control that converts a network event into a manual site visit is, in a protection layer, converting an availability event into a PFD contribution. We suspect this pattern is widespread and unmeasured.

Threats to validity. The worked example is constructed rather than drawn from an incident, because no public incident of this class exists — the only documented SIS attack shut the plant down by accident rather than design [15], [16]. The consequence profile is representative of batch chemistry and would be much flatter for a continuous process with a benign trip state, in which case A approaches 1 and most of Section III-D's force evaporates. The framework is therefore sharpest exactly where the transient-operations risk is highest, which is a feature of the argument but also a limit on its generality.

Where the standards lens stops. IEC 61511-1 §8.2.4 requires the security risk assessment; it does not require its output to be expressed in the units the rest of the safety lifecycle uses, and so in practice the assessment produces a separate document with a separate vocabulary that the SIL verification calculation never reads. That gap is organizational, not technical. Relations (1) through (5) are arithmetic that a functional safety engineer can perform without a security specialist. The obstacle to performing them is that no clause names the quantity, and nothing that is not named gets a column in the spreadsheet.

## VI. Conclusion and Future Work

An availability-only vulnerability in a safety logic solver is not an availability-only finding. Because the asset is a protection layer, the induced transition to the safe state carries the consequence of a trip at a moment the adversary chose, and the interval before restoration is a period in which the protection layer is simply absent. This paper has expressed both effects as terms in the reliability ledger that safety engineers already maintain: a timing amplification factor that replaces the occupancy-weighted mean consequence with the maximum over adversary-observable states, and a restoration-unavailability term that enters the average probability of failure on demand additively. Applied to a GuardLogix-based reactor trip under CVE-2026-9637, the first gives an amplification of 5.5 and the second consumes 91 percent of the SIL 3 budget at one induced event per year, together

taking the SIF out of its assessed integrity level without any component having failed.

Three directions follow. First, the consequence-indexing step wants empirical support: incident databases record the process state at the time of a trip often enough that the $C_s$ distribution could be estimated for common unit operations rather than judged per project. Second, the bypass mechanism named in Section IV-B is the most consequential unquantified path in the framework and is amenable to the same treatment that human-reliability analysis gives to other procedural violations. Third, and most practically, relation (5) should be proposed as an addition to the SIL verification calculation itself rather than to the security risk assessment, on the argument that a known unavailability with a known duration has always belonged in $PFD_{avg}$ and that the only novelty here is the cause.

## References


1. Cybersecurity and Infrastructure Security Agency, "Rockwell Automation Logix platform," ICS Advisory ICSA-26-244-03, Sep. 1, 2026. [Online]. Available: https://www.cisa.gov/news-events/ics-advisories/icsa-26-244-03
2. 1898 & Co., "Denial of service vulnerability in Rockwell Automation Logix controllers," Advisory brief, Sep. 3, 2026. [Online]. Available: https://1898advisories.burnsmcd.com/denial-of-service-vulnerability-in-rockwell-automation-logix-controllers
3. M. A. Lundteigen and M. Rausand, "Spurious activation of safety instrumented systems in the oil and gas industry: Basic concepts and formulas," Rel. Eng. Syst. Saf., vol. 93, no. 8, pp. 1208–1217, Aug. 2008, doi: 10.1016/j.ress.2007.07.004.
4. Cybersecurity and Infrastructure Security Agency, "Rockwell Automation OTTO Fleet Manager," ICS Advisory ICSA-26-239-03, Aug. 27, 2026. [Online]. Available: https://www.cisa.gov/news-events/ics-advisories/icsa-26-239-03
5. Cybersecurity and Infrastructure Security Agency, "ANDRITZ HIPASE-250 and HIPASE-250 SCALA," ICS Advisory ICSA-26-225-05, Aug. 13, 2026. [Online]. Available: https://www.cisa.gov/news-events/ics-advisories/icsa-26-225-05
6. S. Kriaa, L. Piètre-Cambacédès, M. Bouissou, and Y. Halgand, "A survey of approaches combining safety and security for industrial control systems," Rel. Eng. Syst. Saf., vol. 139, pp. 156–178, Jul. 2015, doi: 10.1016/j.ress.2015.02.008.
7. E. Lisova, I. Šljivo, and A. Čaušević, "Safety and security co-analyses: A systematic literature review," IEEE Syst. J., vol. 13, no. 3, pp. 2189–2200, Sep. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8556001
8. A. A. Jigar, Y. Liu, and M. A. Lundteigen, "Spurious activation analysis of safety-instrumented systems," Rel. Eng. Syst. Saf., vol. 156, pp. 15–23, Dec. 2016, doi: 10.1016/j.ress.2016.06.015.
9. F. Innal, Y. Liu, M. A. Lundteigen, and Y. Barros, "Spurious activation and operational integrity evaluation of redundant safety instrumented systems," Rel. Eng. Syst. Saf., vol. 196, art. 106738, Apr. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S095183201930540X
10. M. A. Lundteigen, M. Rausand, and I. B. Utne, "Reliability of safety instrumented systems: Where to direct future research?" Process Saf. Progress, vol. 29, no. 4, pp. 372–378, Dec. 2010, doi: 10.1002/prs.10390.
11. B. K. Vaughen, "Understanding and managing the risk during transient operations," Process Saf. Progress, vol. 39, no. 4, art. e12146, Dec. 2020, doi: 10.1002/prs.12146.
12. H. M. Santana et al., "Operational readiness: Good practices for process plant start-up to avoid process safety accidents," Process Saf. Progress, vol. 42, no. 3, art. e12470, Sep. 2023, doi: 10.1002/prs.12470.
13. M. Iaiani, A. Tugnoli, S. Bonvicini, and V. Cozzani, "Analysis of cybersecurity-related incidents in the process industry," Rel. Eng. Syst. Saf., vol. 209, art. 107485, May 2021, doi: 10.1016/j.ress.2021.107485.
14. M. Iaiani, A. Tugnoli, and V. Cozzani, "Identification of reference scenarios for security attacks to the process industry," Process Saf. Environ. Protect., vol. 161, pp. 334–349, May 2022. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0957582022002
15. A. Carcano and M. Lostri, "TRITON: The first ICS cyber attack on safety instrument systems," Nozomi Networks / Black Hat USA 2018 white paper, Aug. 2018. [Online]. Available: https://i.blackhat.com/us-18/Wed-August-8/us-18-Carcano-TRITON-How-It-Disrupted-Safety-Systems-And-Changed-The-Threat-Landscape-Of-Industrial-Control-Systems-Forever-wp.pdf
16. MITRE Corporation, "Cyber risk to mission case study: Triton," Tech. Rep. AD1183008, Defense Technical Information Center, 2022. [Online]. Available: https://apps.dtic.mil/sti/trecms/pdf/AD1183008.pdf
17. M. A. McQueen, W. F. Boyer, M. A. Flynn, and G. A. Beitel, "Time-to-compromise model for cyber risk reduction estimation," in Quality of Protection: Security Measurements and Metrics, Boston, MA, USA: Springer, 2006, pp. 49–64, doi: 10.1007/978-0-387-36584-8_5.
18. B. Kordy, L. Piètre-Cambacédès, and P. Schweitzer, "DAG-based attack and defense modeling: Don't miss the forest for the attack trees," Computer Science Review, vols. 13–14, pp. 1–38, Nov. 2014. [Online]. Available: https://arxiv.org/abs/1303.7397
19. "Cyber security of OT networks: A tutorial and overview," arXiv:2502.14017, Feb. 2025. [Online]. Available: https://arxiv.org/abs/2502.14017
20. "Threat-based security controls to protect industrial control systems," arXiv:2501.13268, Jan. 2025. [Online]. Available: https://arxiv.org/abs/2501.13268
21. "Vulnerability assessment of industrial control system with an improved CVSS," arXiv:2306.08631, Jun. 2023. [Online]. Available: https://arxiv.org/abs/2306.08631
22. International Electrotechnical Commission, IEC 61511-1:2016 — Functional Safety: Safety Instrumented Systems for the Process Industry Sector, Part 1: Framework, Definitions, System, Hardware and Application Programming Requirements, Geneva, Switzerland, 2016.
23. International Electrotechnical Commission, IEC 62443-3-3:2013 — Industrial Communication Networks: Network and System Security, Part 3-3: System Security Requirements and Security Levels, Geneva, Switzerland, 2013.
24. International Electrotechnical Commission, IEC 62443-4-1:2018 — Security for Industrial Automation and Control Systems, Part 4-1: Secure Product Development Lifecycle Requirements, Geneva, Switzerland, 2018.
25. International Electrotechnical Commission, IEC 61508-1:2010 and IEC 61508-2:2010 — Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems, Geneva, Switzerland, 2010.
26. S. Kriaa, M. Bouissou, and Y. Laarouchi, "A new safety and security risk analysis framework for industrial control systems," Proc. Inst. Mech. Eng. Part O: J. Risk Rel., vol. 233, no. 2, pp. 151–174, Apr. 2019, doi: 10.1177/1748006X18765885.
27. "Model-driven safety and security co-analysis: A systematic literature review," J. Syst. Softw., vol. 220, art. 112251, Feb. 2025, doi: 10.1016/j.jss.2024.112251.
28. B. Gjerde and M. A. Lundteigen, "Cybersecurity of offshore oil and gas production assets under trending asset digitalization contexts: A specific review of issues and challenges in safety instrumented systems," Eur. J. Security Res., vol. 6, pp. 43–63, 2021, doi: 10.1007/s41125-021-00076-2.