

XPhysICS: Cross-Physical-Domain Threat Grounding for Industrial Control Systems Security

SANGSHIN PARK, University of Utah, USA

JAINTA PAUL, University of Utah, USA

LAWRENCE PONCE, University of Utah, USA

MD RAIHAN AHMED, University of Utah, USA

MU ZHANG, University of Utah, USA

LUIS GARCIA, Kahlert School of Computing, University of Utah, USA

Industrial control system attacks are usually documented in terms of the plant where they occurred: its sensors, actuators, process stages, and control logic. Yet many attacks express a more general physical pattern—such as suppressing flow, corrupting chemical dosing, or driving a vessel toward overflow—that may also matter in a different plant. The challenge is deciding when such a threat remains meaningful on a new system rather than relying on similar component names or broad semantic labels.

We present XPhysICS, a methodology for grounding documented cyber-physical threats onto a specific target system. XPhysICS first converts source evidence into a provenance-linked description of what is manipulated, what physical consequence is expected, and what observations the source evidence calls for. Once this analyst-guided abstraction, its vocabulary and schema version, and a machine-readable target contract are fixed, XPhysICS applies deterministic grounding checks. An accepted result can be represented as a *validation slice* that records the mapped roles, signals, dependencies, and context intended to support later evaluation. Grounding is kept separate from questions such as whether the slice is adequate, whether the effect is dynamically reachable, or whether a particular analysis can operate on it.

We study 83 threat abstractions across continuous-process and manufacturing sources, using separate denominators for the corresponding evaluations. The continuous-process study evaluates 78 abstractions against target contracts spanning water treatment, water distribution, hydropower, and chemical processes. Selected cases are then exercised through controlled perturbations of simulator-role signals. We also test compatibility with several analysis styles, including the released upstream GeCo implementation, and conduct a three-objective, one-target realizability study using a paper-derived search reproduction under specified plant models and command surfaces. Across these evaluated settings, the results support treating explicit target checks and traceable evidence as separate from semantic similarity alone.

Additional Key Words and Phrases: industrial control systems, cyber-physical systems, cross-physical-domain threat grounding, threat grounding, security validation

1 Introduction

Industrial control systems (ICS) operate physical processes such as water treatment, power generation, oil and gas handling, and manufacturing, where a cyber action can change physical behavior. Well-known incidents span logic and setpoint modification [15], protocol abuse for equipment reconfiguration [23, 34], and safety-system subversion [14]. Water-sector incidents likewise show that chemical-dosing manipulations can recur across distinct facilities [25, 31]. Although the affected plants differ, the underlying pattern may again involve familiar physical roles—pumps, valves, tanks, interlocks, and their control relationships.

Authors' Contact Information: Sangshin Park, u1418114@utah.edu, University of Utah, Salt Lake City, Utah, USA; Jainta Paul, u1471999@utah.edu, University of Utah, Salt Lake City, Utah, USA; Lawrence Ponce, u1384059@utah.edu, University of Utah, Salt Lake City, Utah, USA; Md Raihan Ahmed, u1374605@utah.edu, University of Utah, Salt Lake City, Utah, USA; Mu Zhang, muzhang@cs.utah.edu, University of Utah, Salt Lake City, Utah, USA; Luis Garcia, la.garcia@utah.edu, Kahlert School of Computing, University of Utah, Salt Lake City, Utah, USA.

Sharing incident reports helps operators learn from these patterns. For example, NERC’s Electricity Information Sharing and Analysis Center published a technical account of the 2015 Ukrainian power-grid attack for broader defensive use [24]. Yet such reports are written in the language of the source plant: its tags, vendor-specific components, control logic, and topology. A dosing-corruption threat does not automatically fit another plant simply because both contain chemical tanks. The target must provide enough documented role, process-stage, dependency, and observation evidence to apply explicit checks. Our central question is therefore: given threat evidence from one ICS and a description of another, does the source abstraction satisfy the implemented checks for that target?

Existing representations solve related but different problems. MITRE ATT&CK for ICS [7] organizes tactics and techniques, while knowledge graphs and ontologies organize assets, threats, and relationships [6, 18, 27, 28, 33]. These systems support classification and semantic correspondence, but similar labels do not show that a particular target can support or evaluate the same physical effect. At the other end of the spectrum, controller-level analyses can provide detailed target evidence when the necessary code and toolchains are available. What is missing is a method that connects documented source threats to the evidence available for a specific system under test (SUT).

We present XPHYSICS, a framework for this *target-conditioned grounding* problem. XPHYSICS represents source evidence by recording what is manipulated, the expected physical consequence, what observations the source evidence calls for, and where each claim came from. It represents the target through a contract that describes available roles, signals, dependencies, and evaluation rules. With those inputs fixed, the grounder applies five deterministic implementation checks: component coverage, coarse role/type compatibility, source-stage coherence, minimum-tag slice viability, and rule-surface intersection. An accepted result can then be materialized as a *validation slice* that records the target roles, signals, dependencies, timing assumptions, and provenance selected for later analysis. Source abstraction may require analyst judgment; the grounding decision is deterministic only after the abstraction, vocabulary, schema version, and target contract have been fixed (Section 3.2).

The core procedure does not require target controller code. When such code is available, *CrossPLC* can contribute routines, tags, read/write relationships, and state information for role recovery and slice construction. Its original translation and consolidation path was developed as part of SCADMAN [2]; we augment it for XPHYSICS. This enrichment, along with temporal guardrails, remains optional.

The validation slice also separates grounding from what happens afterward. Predictive, state-aware, phase-aware, control-integrity, and provenance systems [1, 2, 5, 19, 37] can consume different views of the target context without becoming part of the grounding definition. We test this interface using XPHYSICS-native consumer lanes, prototype adapters, and the unmodified upstream GeCo implementation; these are compatibility studies, not comparative detector evaluations. We likewise study dynamic realizability separately by applying a paper-derived reproduction of the ICSFlux search method [32] to three frozen groundings. That experiment does not execute the published ICSFlux artifact or constitute an implemented XPHYSICS–ICSFlux workflow.

Our evaluation uses source and target settings spanning water treatment, water distribution, hydropower, chemical processes, and a smaller manufacturing corpus. It reports broad grounding coverage, detailed water-domain validation slices, nine additional Hydro/GRFICS executions, downstream compatibility studies, and the three-objective realizability experiment. Because these studies answer different questions, their denominators and conclusions are reported separately.

These studies are organized around four research questions:

RQ1: Structural grounding. Under the evaluated target contracts, which source abstractions satisfy the implemented grounding predicates across the target process families?

RQ2: Slice evidence. For selected accepted groundings, do the constructed validation slices provide the declared signal, dependency, and timing evidence, and what target-side consumer outcomes arise under the evaluated perturbations?

RQ3: Interoperability. Can the same validation-slice representation supply the inputs required by distinct downstream analysis styles, including the unmodified upstream GeCo implementation?

RQ4: Dynamic realizability. For a bounded set of frozen groundings, does realizability differ across the evaluated plant models and command surfaces while structural grounding remains fixed?

Contributions. The primary contribution of this work is a provenance-aware, target-conditioned methodology for grounding documented ICS threats into target-specific validation slices. Supporting contributions provide the representations, mechanisms, and empirical evidence required to evaluate this methodology.

- **Target-conditioned threat grounding.** We introduce a framework that separates analyst-guided abstraction of source-side threat evidence from deterministic target grounding. Once the source abstraction, vocabulary and schema version, and target contract are fixed, candidate mappings are evaluated using five explicit eligibility criteria covering component coverage, implemented role/type compatibility, source-stage coherence, minimum-tag slice viability, and rule-surface intersection (Sections 3.1–3.3). The resulting grounding record distinguishes candidate-generation outcomes from rejection under individual eligibility criteria.
- **Validation slices for target-side evaluation.** We introduce the validation slice as a target-specific representation that records mapped manipulation and consequence paths, observable signals, dependencies, timing assumptions, and consumer-relevant context (Section 3.4). This representation enables grounding acceptance, slice adequacy, consumer applicability, and consumer outcome to be evaluated separately rather than represented by a single cross-system applicability decision.
- **Machine-validated evidence and target-contract representations.** We provide schema-validated representations for source abstractions and target contracts (Table 1 and Table 2), together with versioned vocabularies and an evidence ledger that preserves provenance across source evidence, abstraction fields, and grounding decisions. Optional CrossPLC-based controller enrichment, temporal guardrails, and bounded downstream consumers evaluate additional uses of the resulting validation-slice representation.
- **Multi-domain evaluation with explicit evidence levels.** We evaluate XPhysICS across water-treatment, water-distribution, hydro/water-energy, and chemical-process settings. The evaluation reports continuous-process grounding coverage, detailed executable water-domain validation slices, nine bounded-support Hydro/GRFICS cases, downstream consumer compatibility, and a bounded dynamic-realizability study. These results characterize distinct evidence levels and identify the scope within which cross-domain threat semantics remain structurally groundable and target-side evaluable.

2 Operational Context and Problem Statement

Why plant-specific grounding matters. Figure 1 illustrates the distinction between recurring cyber-physical roles and target-specific applicability. Components such as pumps, valves, tanks, and interlocks may appear across multiple plants without providing equivalent physical roles, process context, or observability. For example, an overflow-related manipulation documented for one water-treatment stage may have candidate counterparts on another target, but

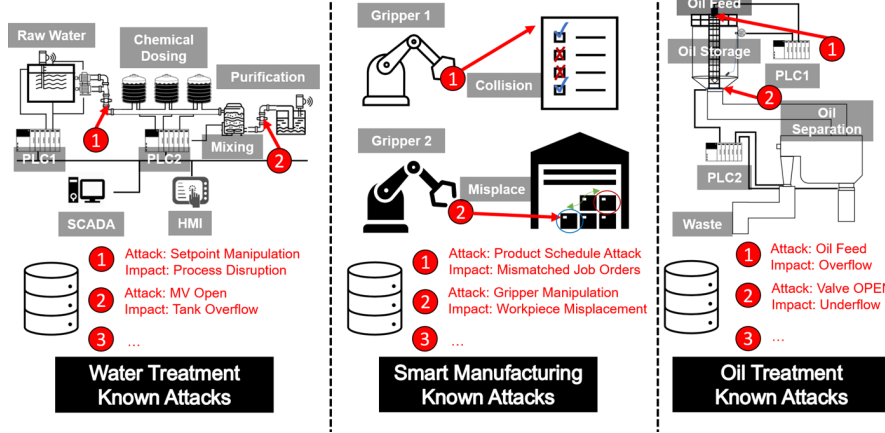


Fig. 1. Recurring cyber-physical threat patterns across ICS domains. Water-treatment, manufacturing, and oil/chemical-process systems may expose similar physical roles, including pumps, valves, tanks, interlocks, and control interactions, despite differences in plant topology, process structure, and physical medium.

component-name or type similarity alone does not establish whether the target can evaluate the corresponding effect. The implemented grounding filters check component coverage, role/type, source-stage, minimum-tag slice viability, and rule intersection; complete source-role coverage and slice adequacy require separate evidence.

2.1 Problem Statement

Given source artifacts describing an ICS threat and target artifacts describing a selected SUT, XPHYSICS determines whether the source threat abstraction is structurally admissible under the target contract. For an accepted grounding, XPHYSICS constructs a target-specific *validation slice* containing the manipulated and consequence paths, observable signals, dependencies, timing assumptions, and consumer-relevant context required for subsequent target-side evaluation. A *rule surface* denotes the target variables, dependencies, and operating context over which the target contract defines applicable evaluation rules.

The problem therefore comprises three stages: constructing a reusable threat abstraction from heterogeneous source evidence, evaluating candidate mappings against explicit target-conditioned eligibility criteria, and constructing a validation slice whose adequacy can be assessed independently of the grounding decision. Downstream consumer applicability and outcome are evaluated subsequently and do not determine grounding acceptance. Section 3 formalizes these stages.

2.2 Use by Downstream Analyses

Validation slices provide a common target-conditioned representation for downstream analyses with different input requirements. Invariant-based analyses may consume ranges, correlations, mass-balance relationships, or rate-of-change constraints; predictive analyses may require time-series windows and expected trajectories; state- or phase-aware analyses additionally require operating context; and provenance-oriented analyses may use links between target-side observations, grounding decisions, and their supporting evidence. The validation slice exposes the target roles, signals, dependencies, contextual information, and provenance needed to construct these consumer-specific views.

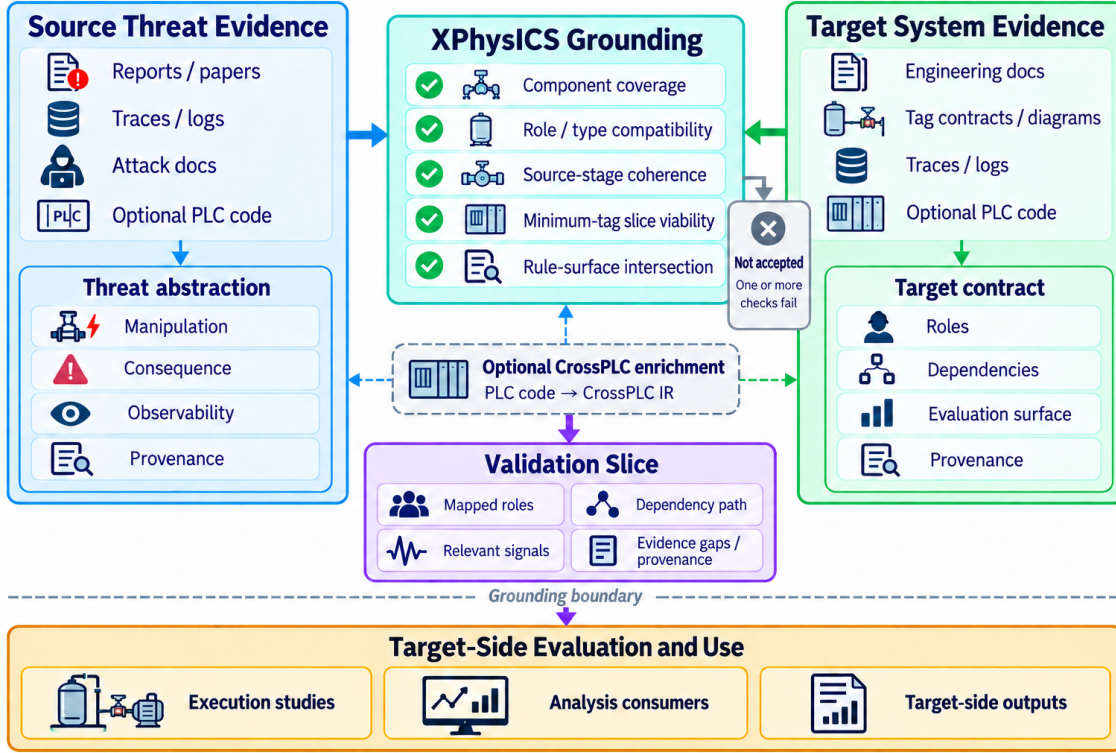


Fig. 2. Overview of XPhysICS. Evidence from a documented source threat and a target system is converted into a threat abstraction and target contract. The grounder applies the implemented eligibility checks and either does not accept a candidate when one or more checks fail or returns a target-specific validation slice. Execution and downstream analysis occur beyond the grounding boundary. CrossPLC provides optional controller-code enrichment.

Consumer applicability remains separate from grounding and slice adequacy. A validation slice may be structurally grounded and adequate for target-side evaluation while falling outside the modeled surface of a particular consumer. Similarly, the outcome produced by an applicable consumer does not alter the underlying grounding decision. This separation allows the same grounded representation to support heterogeneous downstream analyses without making any individual consumer part of the definition of grounding.

2.3 Design Requirements and Scope

The operational setting motivates four design requirements. **Target specificity** requires each grounding decision to be conditioned on a specified SUT and its available evidence rather than represented as an abstract semantic correspondence. **Explicit eligibility** requires candidate mappings to satisfy the declared implementation-level filters before grounding acceptance, while recording limitations those filters do not establish. **Slice adequacy** requires the resulting validation slice to expose the signal coverage, dependency closure, and timing evaluability needed for the intended target-side evaluation, with adequacy reported separately from grounding and consumer outcome. **Auditable**

provenance requires source evidence, target evidence, mapping decisions, and eligibility outcomes to remain traceable throughout the grounding process. Section 3.3 formalizes the grounding criteria, while Section 3.5 defines the subsequent adequacy assessment.

XPHYSICS operates over heterogeneous ICS artifacts, including attack descriptions, dataset documentation, engineering diagrams, manuals, target contracts, time-series traces, and controller artifacts when available. Controller artifacts are not required by the core grounding procedure; when present, they provide optional evidence for role recovery and validation-slice construction through program-side tags, routines, read/write relationships, and available state information.

The primary evaluation focuses on continuous-process settings spanning water treatment, water distribution, hydro/water-energy, and chemical processes. Additional studies examine manufacturing/program-artifact grounding, controller-code enrichment, temporal guardrails, and protocol evidence under separate evidence surfaces and denominators. These extension studies are reported independently of the continuous-process grounding matrix and the primary validation-slice studies.

3 XPHYSICS Framework

Figure 2 summarizes the XPHYSICS methodology. The framework connects source-side threat evidence to target-conditioned security evaluation through four stages: source threat abstraction, target-conditioned grounding, validation-slice construction, and slice adequacy with downstream interfaces. The objective is not to establish equivalence between source and target plants, but to determine whether documented source threat semantics satisfy explicit structural and evaluability requirements for a specified target SUT.

We represent a source evidence bundle as

$$\mathcal{S} = (\mathcal{D}_s, \mathcal{X}_s, \mathcal{P}_s),$$

where $\mathcal{D}_s$ denotes source documents and engineering artifacts, $\mathcal{X}_s$ denotes traces or benchmark data when available, and $\mathcal{P}_s$ denotes optional controller or program artifacts. A target SUT is represented as

$$\mathcal{T} = (\mathcal{D}_t, \mathcal{X}_t, \mathcal{P}_t, \mathcal{R}_t),$$

where $\mathcal{R}_t$ denotes target-side rule surfaces or consumer-relevant variables. Rather than representing a comprehensive plant model, XPHYSICS produces a grounded, target-specific validation object that provides structured input for subsequent target-side evaluation.

Running example. Figure 3 illustrates the principal transformations performed by XPHYSICS. Source evidence describing a cyber-physical manipulation is first represented in terms of its effect family, manipulated roles, consequence roles, observability obligations, and provenance. Grounding then attempts to map these source-side requirements onto target entities and signals and evaluates the resulting candidate against the target contract. When the grounding criteria are satisfied, the mapped entities and signals form the basis of a target-specific validation slice.

This example illustrates the distinction between semantic similarity and target-conditioned grounding. A semantically related tank is insufficient to establish an admissible grounding. Grounding is accepted only when the candidate satisfies the implemented component-coverage, role/type, source-stage, minimum-tag, and rule-intersection filters. The example is a recorded structural mapping, not an executed attack or a measured consumer outcome.

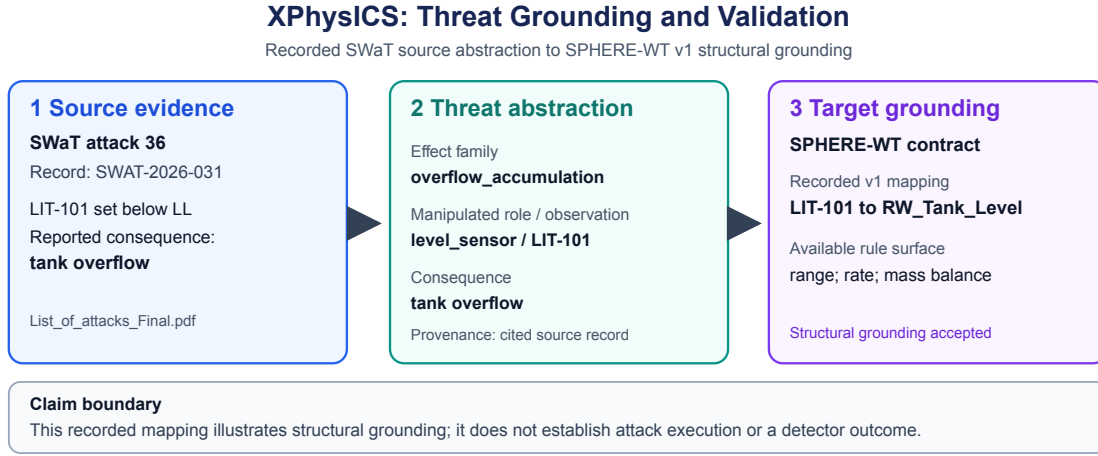


Fig. 3. Source-backed worked grounding example. SWaT attack 36 is recorded as SWAT-2026-031, with a LIT-101 low-reading manipulation and a reported tank-overflow consequence. The recorded v1 structural grounding maps its level-sensor role to the SPHERE-WT RW_Tank_Level signal. This illustration does not claim that the source attack was executed on WT or that a downstream consumer detected it.

3.1 Source Threat Abstraction

The first transformation maps heterogeneous source evidence into a structured threat abstraction. Source evidence may include engineering documentation, manuals, incident reports, benchmark traces, attack annotations, and, where available, controller or program artifacts. XPhysICS represents these heterogeneous evidence sources within a common abstraction schema, providing the structured source semantics required for subsequent target-conditioned grounding.

For each candidate source-side threat pattern, XPhysICS constructs

$$\tau = (f, R_m, R_c, O, \Pi),$$

where f is the *effect family*, R_m is the set of *manipulated roles*, R_c is the set of *consequence roles*, O is the set of *observability obligations*, and Π is the *provenance bundle* linking the abstraction to its supporting source evidence. This representation is intentionally narrower than a complete executable model of the source plant. It preserves the manipulated roles, expected physical consequence, observations required to assess that consequence, and supporting evidence required for subsequent grounding.

Table 1 provides an operational definition of each field, including its representation and data type, the vocabulary constraining its admissible values, the supporting source evidence, and the mechanism by which the field is populated, whether through analyst interpretation, source-specific parsing, or inheritance from prior work. The provenance bundle Π maintains traceability between each populated field value and its supporting source evidence.

Worked example. For the source record in Figure 3, the analyst assigns $f = \text{overflow_accumulation}$ based on the explicitly reported tank-overflow consequence. R_m contains LIT-101 (role `level_sensor`), the named manipulated signal in SWaT attack 36. The source-evidenced observability set is $O = \{\text{LIT-101}\}$; no pump, valve, or flow tag is added to this source record merely because it might be useful downstream. The provenance bundle Π records List_of_attacks_Final.pdf, attack 36, the structured_parse extraction method, and high confidence. The v1

Table 1. Operational definitions of the source threat abstraction $\tau = (f, R_m, R_c, O, \Pi)$. Origin distinguishes OTThreat-informed constructs from XPhysICS-specific constructs.

Construct	Operational definition	Origin / assignment
Effect family f	Normalized physical-consequence category drawn from a closed, versioned extraction vocabulary. Target grounding operates over the narrower grounding vocabulary declared in scope by the target contract (Section 3.3).	OTThreat-informed vocabulary; analyst- or parser-assigned.
Manipulated roles R_m	Source-side component roles manipulated by the threat, represented by an identifier and functional role with optional process-stage context.	XPhysICS-specific; analyst- or parser-assigned.
Consequence roles R_c	Components or consequence records associated with the realized physical effect f following manipulation of R_m .	XPhysICS-specific; assigned from supporting source evidence.
Observability obligations O	Signals or observations indicated by the source evidence as relevant to assessing the stated effect; these obligations seed target-side slice construction (Section 3.4).	XPhysICS-specific; may be empty when the source evidence does not identify corresponding observations.
Provenance Π	Evidence-lineage metadata linking populated abstraction fields to the source document, supporting excerpt, extraction method, and confidence assignment.	OTThreat-informed; provenance metadata are recorded by the evidence workflow, while confidence may involve analyst or reviewer judgment.

grounder records an accepted structural mapping of this abstraction to the SPHERE-WT RW_Tank_Level target signal. The target contract offers range, rate, and mass-balance rule surfaces, but their availability does not itself establish a detector outcome for this source record.

Construction of τ may require analyst interpretation or source-specific parsing when the available evidence is ambiguous. For example, an observation such as “LIT-101 reads erratically” without an explicit consequence does not uniquely determine f ; such evidence requires additional interpretation or an appropriately qualified confidence assignment. This semantic-abstraction step is distinct from target grounding. Once τ , the vocabulary and schema version, and the target contract are fixed, the grounding procedure in Section 3.3 applies the specified eligibility criteria deterministically without additional semantic interpretation.

When PLC or program artifacts are available, *CrossPLC* provides optional controller-aware enrichment of the source or target evidence layer. We represent this enrichment as

$$\mathcal{I} = (Q, T, E^{rw}, E^{call}, S),$$

where Q denotes routines or function blocks, T denotes controller tags, E^{rw} represents read/write dependencies, E^{call} represents call relations, and S denotes recovered controller-state structure when available. This information can provide additional evidence for role recovery and validation-slice construction but is not required by the core grounding procedure.

3.2 Analyst Workflow

The XPhysICS workflow separates analyst-guided source abstraction from deterministic target grounding and subsequent target-side evaluation:

- (1) **Source evidence** (input). The workflow begins with source artifacts, such as engineering documentation, dataset descriptions, incident reports, or structured workbooks, that describe a source-side threat.
- (2) **Evidence selection** (manual/parser-assisted). An analyst or source-specific parser (Section 4.1) identifies evidence describing the manipulated components, physical consequence, and relevant observations.
- (3) **Semantic abstraction** (manual/parser-assisted). The selected evidence is represented as $\tau = (f, R_m, R_c, O, \Pi)$ using the vocabulary defined in Table 1. Unstructured evidence may require analyst interpretation, whereas compatible structured artifacts may be processed by source-specific parsers.
- (4) **Normalization** (automated). Component identifiers and role labels are canonicalized according to the schema-defined vocabulary. Values outside the permitted enumerations are rejected.
- (5) **Review and adjudication** (manual, partial coverage). A subset of extracted abstractions is compared with the corresponding source evidence for component, effect-family, and observability consistency. Appendix A reports the internal spot-check over 20 sampled records. This procedure provides a bounded extraction check rather than an independent multi-analyst agreement study, as discussed in Section 6.1.
- (6) **Target-contract construction** (manual, per target). An engineer specifies the available target signals, dependency relationships, effect-family scope, rule surfaces, and relevant study context in a target contract (Table 2). The contract is defined at the target-SUT level rather than independently for each source threat.
- (7) **Candidate mapping** (automated). The grounder constructs candidate mappings from $R_m \cup R_c \cup O$ to entities and signals declared in the target contract (Algorithm 1).
- (8) **Eligibility evaluation** (automated). The predicates χ_{map} , χ_{type} , χ_{stage} , χ_{slice} , χ_{rule} are evaluated for the fixed source abstraction and target contract without additional semantic interpretation.
- (9) **Validation-slice construction** (automated). An accepted grounding is materialized as $\mathcal{V}(\tau, \mathcal{T}, \mu)$.
- (10) **Downstream evaluation** (optional). Slice adequacy and consumer applicability are assessed independently of grounding acceptance, after which applicable consumers may be evaluated over the resulting slice.

Open-ended semantic interpretation is concentrated in source evidence selection and abstraction, whereas normalization, candidate mapping, eligibility evaluation, validation-slice construction, and downstream consumer execution operate according to predefined procedures once their inputs are fixed. Target-contract construction and bounded review additionally require manual input. The workflow therefore distinguishes analyst-guided semantic abstraction from deterministic target grounding without characterizing the complete process as either fully automated or fully manual.

3.3 Target-Conditioned Grounding

Given a source abstraction τ , XPHYSICS attempts to instantiate it on a specified target SUT. The target provides a grounding surface

$$\Gamma_t = (V_t, E_t, U_t, \Sigma_t, \mathcal{R}_t),$$

where V_t denotes target entities or process roles, E_t denotes dependency or flow relationships, U_t denotes unit and engineering metadata, Σ_t denotes observable target signals, and $\mathcal{R}_t$ denotes the target rule surface. When controller-aware enrichment is available, the grounding surface may additionally incorporate $\bar{I}_t$.

A candidate grounding is a partial mapping

$$\mu : R_m \cup R_c \cup O \rightarrow V_t \cup \Sigma_t$$

from source-side roles and observability obligations to target-side entities and signals. The v1 implementation permits individual source components or observation obligations to remain unresolved; its component-coverage predicate requires at least one mapped component. XPHYSICS accepts the resulting candidate only when

$$G(\tau, \mathcal{T}, \mu) = \chi_{\text{map}}(\mu) \wedge \chi_{\text{type}}(\mu) \wedge \chi_{\text{stage}}(\mu) \wedge \chi_{\text{slice}}(\mu) \wedge \chi_{\text{rule}}(\mu).$$

Here, χ_{map} records whether at least one source component has a candidate target mapping. χ_{type} applies the implemented coarse source-role/target-tag-type check to each mapped component; it is not a physical-unit checker. χ_{stage} checks whether numeric process-stage labels declared on source components span at most adjacent stages. χ_{slice} expands mapped target tags through declared target relationships and requires the resulting tag set to meet a target-specific minimum size. Finally, χ_{rule} requires at least one target rule in $\mathcal{R}_t$ to reference a tag in that expanded set (Section 4.2). These are implementation-level admissibility filters. Grounding acceptance does not establish complete role coverage, a target-side manipulation-to-consequence path, or validation-slice adequacy; those stronger properties require separate evidence.

The extraction and grounding vocabularies have distinct scopes. Source abstractions may use the seven effect families defined in Table 1, whereas the reported grounding evaluation is restricted to a five-family subset. Abstractions assigned to either of the two extraction-only families therefore fall outside the reported grounding surface rather than constituting rejected source threats. This distinction preserves the broader source representation while delimiting the effect families covered by the current grounding results.

Given a fixed source abstraction, vocabulary and schema version, and target contract, the grounding procedure evaluates these eligibility criteria deterministically. The resulting decision establishes whether the candidate satisfies the implemented component-coverage, role/type, source-stage, minimum-tag, and rule-intersection filters. It does not establish that every source role was resolved or that a target dependency path connects the manipulation and consequence roles. Controller-aware evidence may supplement subsequent slice construction when program artifacts are available but is not required by the core grounding procedure.

Grounding and downstream evaluation are distinct. The χ_{rule} predicate establishes that an accepted grounding intersects at least one evaluation rule declared by the target contract. Grounding acceptance does not depend on the applicability or outcome of the bounded predictive, state-aware, or phase-aware consumers, nor on that of the upstream GeCo implementation evaluated in Section 5.6. Consumer applicability is assessed only after grounding acceptance and validation-slice construction.

Accordingly, XPhysICS distinguishes multiple evidence layers. *Grounding acceptance* establishes acceptance under the implemented structural-admissibility filters for the target contract. *Slice adequacy* (Section 3.5) assesses signal coverage, dependency closure, and timing evaluability. *Dynamic realizability*, when evaluated, asks whether the grounded effect can be produced under a specified plant model and command surface (Section 5.7). *Consumer applicability* determines whether a particular downstream analysis can operate on the resulting slice. Finally, *consumer outcome* records the result produced by an applicable consumer. These layers answer distinct evidentiary questions; a result at a downstream layer does not retroactively alter an earlier grounding decision.

3.4 Validation Slice Construction

For an accepted grounding, XPhysICS instantiates a target-specific validation slice

$$\mathcal{V}(\tau, \mathcal{T}, \mu) = (P_m, P_c, \Sigma_v, \Omega_v),$$

where P_m is the *manipulated path*, P_c is the *consequence path*, $\Sigma_v \subseteq \Sigma_t$ is the set of slice-observable signals, and Ω_v is the *consumer obligation set*.

The validation slice is the core methodological object of XPhysICS. It defines the bounded, target-specific structure that preserves the manipulated control surface, expected effect path, observable signals required to assess the effect, and obligations required by subsequent target-side analyses. This representation provides an intermediate abstraction between source-side semantic correspondence and whole-plant equivalence. Semantic correspondence alone does not establish that a target exposes sufficient structure for evaluation, whereas requiring equivalence between complete plants would impose unnecessary constraints on target-specific analysis. The validation slice therefore captures the target-side structure required to evaluate a grounded threat under the available evidence.

3.5 Slice Adequacy and Downstream Interfaces

The adequacy of a grounded slice is assessed with respect to the requirements of the intended target-side evaluation. XPhysICS characterizes slice adequacy along three dimensions: signal coverage, dependency closure, and timing evaluability. These criteria assess whether the slice provides the observations, process relationships, and temporal resolution required to evaluate the grounded effect; they do not establish fidelity of the target as a complete plant model or dynamic realizability of the effect.

Slice adequacy is classified as *strong*, *partial*, or *insufficient*. A *strong* slice satisfies the declared adequacy requirements for the intended evaluation. A *partial* slice retains an accepted grounding but has identified limitations in signal coverage, dependency closure, or timing evaluability. An *insufficient* slice does not satisfy the minimum adequacy requirements for the intended target-side evaluation. For strong or partial slices, XPhysICS may derive consumer-specific views from the grounded slice and its associated provenance. These views support invariant-family analyses, bounded predictive consumers, bounded state- or phase-aware consumers, provenance-oriented analyses, and optional temporal guardrails.

Adequacy labels are distinct from consumer applicability and outcome labels. A strong slice may, for example, be applicable to a particular consumer yet produce a *nominal-confounded* or *near-threshold* outcome. Conversely, a grounded slice may fall outside the modeled surface of a particular consumer without changing its grounding or adequacy classification. These distinctions preserve the evidentiary scope of each stage throughout the evaluation in Section 5.

4 Implementation

We implement XPHYSICS as a target-conditioned pipeline that transforms source-side threat evidence and target-side system evidence into structured validation artifacts for a selected SUT. Following Figure 2, source artifacts are represented as threat abstractions, target evidence is specified through machine-validated contracts, deterministic grounding evaluates candidate mappings, and accepted groundings are materialized as validation-slice bundles for subsequent target-side evaluation.

4.1 Evidence Records and Threat Abstractions

XPHYSICS begins by registering source artifacts that describe reusable ICS threat patterns, including incident reports, academic attack papers, benchmark annotations, engineering documentation, structured workbooks, and selected controller artifacts. Each artifact is represented by a source-document record containing its domain, document identifier, provenance tier, admissibility metadata, source location, and extraction metadata. These records preserve the provenance required to relate structured threat abstractions to their supporting source evidence.

The extractor converts source records into ThreatAbstraction objects implementing the representation defined in Section 3.1. Each object records the effect family, manipulated roles, consequence roles, observability obligations, provenance links, mechanism metadata, timing information when available, confidence, supporting source evidence, and extraction metadata. The primary evaluation uses structured or curated extraction paths. LLM-assisted extraction is retained as an exploratory support mechanism and is not used as the basis for the reported grounding results.

For sources that provide structured local artifacts, XPHYSICS employs source-specific parsers to construct the common threat abstraction. The WADI parser, for example, extracts attack points, actuator and sensor state changes, physical impacts, and labeled attack intervals from workbook and trace artifacts, supplementing evidence available in document-based sources. CISS and other supporting sources are processed analogously when compatible structured artifacts are available. Regardless of the source-specific extraction procedure, each parser produces the common abstraction schema required by the subsequent grounding procedure.

4.2 Target Contracts and Study Bundles

Grounding is conditioned on a selected target SUT. We represent the target grounding surface Γ_t (Section 3.3) using machine-validated target contracts that describe available target signals and roles, dependency relationships, effect-family scope, and the rules and operating context available for target-side evaluation. Study bundles associate these contracts with nominal traces, controlled perturbation cases, and timing information used to exercise grounded validation slices.

The primary target studies use contracts for the SPHERE [17] water-treatment (WT) and water-distribution (WD) systems. Additional hydropower, chemical-process, CISS [20], and manufacturing studies use the same representation where the required target evidence is available. Their results are reported under the corresponding evidence scope and denominator in Section 5.

Target contracts are schema-validated before grounding or execution. Contracts that do not satisfy the required schema are rejected before grounding rather than interpreted with missing fields. Table 2 summarizes the components most relevant to grounding and validation-slice construction. The reported experiments use versioned schemas and executable study contracts.

Table 2. Target-contract components used by XPhysICS for grounding and validation-slice construction.

Contract component	Role in XPhysICS
Effect-family scope	Declares the source-threat effect families within the target’s current grounding scope.
Target signals and roles	Defines the target signal-and-role surface over which candidate mappings are evaluated.
Dependency relationships	Provides target relationships used during validation-slice construction and dependency/connectivity evaluation.
Threat-surface scope	Declares the boundary directions and threat classes represented by the target study.
Target rules and context	Defines the target rules and operating context used to determine rule-surface applicability and support subsequent slice evaluation.

The target-contract grounding vocabulary is a five-family subset of the seven-family extraction vocabulary introduced in Table 1. Consequently, a source abstraction may be valid within the extraction representation while falling outside the effect-family scope supported by the current target contracts. Such abstractions are reported outside the current grounding surface rather than as rejected source threats.

4.3 Deterministic Grounding and Slice Generation

The deterministic grounder evaluates each source abstraction τ against a target grounding surface $\Gamma_t = (V_t, E_t, U_t, \Sigma_t, \mathcal{R}_t)$, where V_t denotes target entities or roles, E_t denotes dependency or flow relationships, U_t denotes unit and engineering metadata, Σ_t denotes observable target signals, and $\mathcal{R}_t$ denotes the target rule surface. Candidate mappings from source roles and observability obligations to target entities and signals are evaluated using the five eligibility predicates defined in Section 3.3.

The resulting grounding record contains the source threat identifier, target SUT, mapped tags or components, applicable target rules, slice requirements, per-check diagnostics, and provenance links. Grounding acceptance establishes acceptance under the implemented structural-admissibility filters for the target contract. It does not establish slice adequacy, dynamic realizability, consumer applicability, or consumer outcome.

Algorithm 1 specifies the grounding procedure at the granularity implemented by XPhysICS. Candidate construction applies an ordered matching procedure consisting of exact or normalized tag-name matching, configured aliases, role-preference lookup, and a role-keyword search over target-tag descriptions. The first compatible candidate in the declared contract order is selected, and the selected mapping method and associated metadata are retained in the grounding record. Determinism here follows from fixed inputs, fixed precedence, and fixed contract order; it is not a claim that all alternative candidate orders converge to the same mapping.

The implemented role/type check performs a coarse compatibility test between the source role (e.g., `level_sensor`) and the declared target-tag type (e.g., `real` or `bool`); it does not implement general physical-unit reasoning. Stage coherence uses numeric process-stage labels declared on source components and excludes candidates spanning non-adjacent source stages; it does not inspect the stages of mapped target tags. Observability obligations are resolved separately by exact, normalized, or configured-alias lookup, and unresolved obligations are omitted from the initial expansion set. Slice viability expands the mapped target-tag set through the correlation, causality, mass-balance, and trip-response relationships declared by the target contract until no additional tags are introduced and evaluates the resulting set against the target-specific minimum slice size. Rule applicability requires at least one declared target rule to intersect the resulting tag set.

The current component-coverage predicate requires at least one source component mapping, rather than total coverage of every component or observation obligation. Grounding acceptance requires all five eligibility predicates to hold. For accepted and rejected candidates, the grounding record retains the outcome of each predicate, providing an explicit basis for the resulting decision.

Accepted groundings are materialized as validation-slice bundles implementing $\mathcal{V}(\tau, \mathcal{T}, \mu) = (P_m, P_c, \Sigma_v, \Omega_\theta)$ as defined in Section 3.4. Each bundle records the manipulated and consequence paths, required signals and dependencies, perturbation specifications when applicable, nominal and perturbed traces when available, timing assumptions, operating context, adequacy metadata, and consumer-facing outputs. Rather than representing a complete digital twin or exhaustive model of the target, the bundle encodes the bounded target-specific structure required for subsequent evaluation of the grounded effect. Slice adequacy is assessed independently in terms of signal coverage, dependency closure, and timing evaluability, yielding *strong*, *partial*, or *insufficient* classifications.

4.4 Bounded Validation Consumers

Downstream consumers operate on validation slices rather than directly on source evidence. The implementation includes four bounded consumer families. The invariant-family consumer evaluates target-declared range, rate-of-change, correlation, and mass-balance rules over slice signals and dependencies and serves as the primary native consumer for the WT/WD validation studies. A GeCo-style predictive lane [37] trains a predictive model, calibrates thresholds using target nominal traces, and evaluates residual or cumulative-sum deviations over perturbation cases. A SAIN-style state-aware lane [1] infers coarse target states from available tags and applies state-conditioned thresholds to residual invariants. A SCAPHY-style phase-aware lane [19] infers coarse operating phases from target state and actuator tags and applies phase-conditioned value/delta scoring with persistence-gated alerts.

The GeCo-, SAIN-, and SCAPHY-style lanes are bounded XPHYSICS-native implementations of selected analysis patterns motivated by the corresponding published systems. They are used to evaluate whether the validation-slice representation provides the information required by these downstream analysis styles. Each lane produces consumer-specific measurements, including detected windows, alert timing, summary metrics, nominal false-positive rates, and case-level outcomes. These measurements characterize the bounded implementations rather than comparative performance against the corresponding published systems. Compatibility with the upstream GeCo implementation is evaluated separately in Section 5.6.

4.5 Optional Enrichment Paths

XPHYSICS supports two optional enrichment mechanisms: controller-code evidence and temporal guardrails. These mechanisms supplement particular validation slices when the required artifacts are available but are not prerequisites for the core grounding procedure.

When source or target PLC/program artifacts are available, XPHYSICS can invoke CrossPLC. Its original controller-code translation and consolidation path was developed as part of SCADMAN [2]; for XPHYSICS, we augment it to recover program-side evidence including routines or function blocks, controller tags, read/write dependencies, call relationships, and available state structure. In the current evaluation, this evidence supports controller-aware studies in code-rich water and manufacturing cases. CrossPLC is not required for grounding, and its extension results are reported separately from the continuous-process grounding matrix.

When a validation slice has a suitable temporal specification, XPHYSICS can attach an optional Signal Temporal Logic (STL) guardrail evaluated with RTAMT [29]. Given a temporal specification φ_τ and target trace x , the guardrail

Algorithm 1 Deterministic grounding of one source abstraction τ against one target contract Γ_t .

Require: source abstraction $\tau = (f, R_m, R_c, O, \Pi)$; target contract $\Gamma_t = (V_t, E_t, U_t, \Sigma_t, \mathcal{R}_t)$
Ensure: grounding record with success flag, tag mappings, applicable rules, slice requirements, and per-check diagnostics

```

1:  $M \leftarrow \emptyset$  ▷ candidate tag mappings
2: for each component  $c \in R_m \cup R_c$  do
3:    $m \leftarrow \text{MAPCOMPONENTTOTAG}(c, \Gamma_t)$  ▷ exact  $\rightarrow$  normalized  $\rightarrow$  alias  $\rightarrow$  role-preference  $\rightarrow$  lexical
4:   if  $m \neq \perp$  then  $M \leftarrow M \cup \{m\}$ 
5:   end if
6: end for
7:  $\chi_{\text{map}} \leftarrow (M \neq \emptyset)$ 
8:  $\chi_{\text{type}} \leftarrow \chi_{\text{map}} \wedge \bigwedge_{m \in M} \text{ROLETYPEDCOMPATIBLE}(m)$ 
9:  $S \leftarrow \{\text{process\_stage}(c) : c \in R_m \cup R_c\}$ 
10:  $\chi_{\text{stage}} \leftarrow (\text{stages in } S \text{ are pairwise adjacent, or } |S| \leq 1)$ 
11:  $O_t \leftarrow \emptyset$ 
12: for each observation obligation  $o \in O$  do
13:    $s \leftarrow \text{FINDTARGETTAG}(o, \Gamma_t)$  ▷ exact  $\rightarrow$  normalized  $\rightarrow$  configured alias
14:   if  $s \neq \perp$  then  $O_t \leftarrow O_t \cup \{s\}$ 
15:   end if
16: end for
17:  $T \leftarrow \text{EXPANDTOFIXPOINT}(M \cup O_t, \mathcal{R}_t)$  ▷ correlation/causality/mass-balance/trip-response relationships
18:  $\chi_{\text{slice}} \leftarrow (|T| \geq \text{minimum\_slice\_tags})$ 
19:  $A \leftarrow \{r \in \mathcal{R}_t : \text{tags}(r) \cap T \neq \emptyset\}$ 
20:  $\chi_{\text{rule}} \leftarrow (A \neq \emptyset)$ 
21:  $\text{success} \leftarrow \chi_{\text{map}} \wedge \chi_{\text{type}} \wedge \chi_{\text{stage}} \wedge \chi_{\text{slice}} \wedge \chi_{\text{rule}}$ 
22: return ( $\text{success}, M, A, T, (\chi_{\text{map}}, \chi_{\text{type}}, \chi_{\text{stage}}, \chi_{\text{slice}}, \chi_{\text{rule}})$ )

```

computes robustness

$$\rho(\varphi_r, x).$$

The sign and magnitude of this robustness value provide a slice-level assessment of whether the evaluated trace satisfies the specified temporal property. In the current evaluation, this mechanism is exercised as bounded extension evidence for WADI $\rightarrow$ WD boundary cases rather than as a grounding criterion or full-system verification mechanism.

4.6 Assessment Outputs and Provenance

The final XPhysICS assessment package records the grounding decision, validation-slice metadata, adequacy classification, available consumer outputs, contextual limitations, and provenance links for the selected SUT. Depending on the analyses applied to a slice, the package may additionally contain typed alerts, classification results, provenance records, or an indication that the available target evidence is insufficient for the requested evaluation.

We additionally exercise the validation-slice output structure through prototype downstream adapters. A SCADMAN-motivated classification adapter [2] applies post-analysis classification to detector windows, while an ICSTracker-motivated provenance adapter [5] expands rule templates over eligible mappings. These prototypes evaluate additional uses of the XPhysICS output representation and are reported separately in Appendix E.

XPhysICS maintains an evidence ledger informed by the provenance representation used in OTThreat [30]. The ledger records source documents, extraction runs, canonical threat objects, field-level evidence spans, provenance tiers, admissibility metadata, consumer-readiness metadata, and grounding records. It is implemented as a relational SQLite

database with JSON, CSV, and Cypher exports for inspection and analysis. This representation preserves traceability between source evidence, structured threat abstractions, and subsequent grounding results.

5 Evaluation

We evaluate XPHYSICS across a sequence of distinct evidentiary questions. Starting from heterogeneous source evidence, XPHYSICS constructs structured threat abstractions, evaluates them against implemented structural-admissibility filters on target systems, materializes accepted groundings as validation slices, and examines subsequent target-side evaluation where the required execution substrate is available. The evaluation therefore reports broad continuous-process grounding, detailed validation-slice studies, downstream consumer compatibility, a bounded dynamic-realizability study, and extension studies under separate evidence surfaces and denominators.

The reported measurements correspond to different stages of the methodology. *Grounding acceptance* indicates that a source threat abstraction satisfies the deterministic eligibility criteria defined by the target contract. *Slice adequacy* assesses whether the resulting validation slice provides the signal coverage, dependency closure, and timing evaluability required for the intended target-side evaluation. *Dynamic realizability*, where evaluated, asks whether the grounded effect can be produced under a specified plant model and command surface. *Consumer applicability* determines whether a particular downstream analysis can operate on an eligible slice, while *consumer outcome* records the result produced by an applicable consumer. Extension studies are reported separately when they exercise additional representation, implementation, or evidence surfaces rather than the primary continuous-process validation path.

5.1 Source Semantics and Provenance

Target-conditioned grounding requires source evidence to be represented as structured, traceable threat semantics. The evaluation corpus contains 83 canonical source-threat abstractions drawn from 12 source documents spanning Secure Water Treatment (SWaT), Water Distribution (WADI), OilTreatment, and Fischertechnik manufacturing artifacts. Of these, 78 form the continuous-process grounding corpus: 51 from SWaT, 17 from WADI, and 10 from OilTreatment. The remaining five Fischertechnik abstractions exercise manufacturing/program-artifact grounding and are reported separately.

Table 3 summarizes this source-evidence layer. The 83 abstractions comprise 55 high-, 18 medium-, and 10 low-confidence records. For WADI, structured extraction refines coarse attack-table entries into concrete attack points, effect families, attack windows, and observability context. These abstractions and their associated provenance constitute the source-side inputs to target-conditioned grounding.

XPHYSICS uses OTThreat [30] as part of the lineage informing its evidence representation while maintaining a separate evidence ledger for the current framework. The broader ledger contains 333 threat objects and 1,947 field-level evidence spans across 16 source documents. It records source excerpts, extraction metadata, confidence, and other provenance required to trace structured abstractions to their supporting evidence. The 333-object ledger therefore represents a broader evidence and provenance population, not 333 independently validated, grounded, or executable attacks. The evaluation denominators used below are the 78 continuous-process abstractions and five manufacturing/program-artifact abstractions, with executable claims restricted further to the cases actually instantiated and exercised. Appendix A reports the additional ledger composition, vocabulary, selection, and spot-check details.

Table 3. Source semantics and provenance. XPhysICS extracts typed source-threat abstractions and preserves field-level evidence before target grounding. Counts are evaluation artifacts produced by XPhysICS, not external benchmark labels.

Evaluation object	Result	Interpretation
Structured threat corpus	83 threats from 12 documents; 55 high-confidence, 18 medium-confidence, 10 low-confidence. Continuous-process matrix sources: SWaT 51, WADI 17, OilTreatment 10 ($n = 78$). Manufacturing extension: Fischertechnik 5.	Provides typed source semantics across water, oil/chemical, and manufacturing evidence for deterministic target grounding.
WADI source deepening	Naive PDF parsing: 2 rows; XPhysICS: 15 A1 threats + 14 A2 attack windows	Source-specific extraction recovers concrete attack points, effect families, and timing/observability context missed by table scraping.
XPhysICS evidence ledger	333 curated threats; 1,947 field-level evidence spans; 100% core-field coverage	Our OTThreat-aligned audit layer links paper-facing threat objects to source evidence; it is provenance support, not the grounding algorithm.
Manual extraction sample	20 sampled items: 14 full pass, 6 partial, 0 fail	A bounded spot-check found no outright abstraction failures among sampled paper-facing records.

5.2 Continuous-Process Grounding Coverage

RQ1: Structural grounding. We first evaluate the extent to which continuous-process threat abstractions are structurally groundable across the evaluated target process families. This analysis considers the 78 continuous-process abstractions derived from SWaT, WADI, and OilTreatment and applies the deterministic grounding procedure to water-treatment, water-distribution, hydro/water-energy, and chemical-process target surfaces. The five Fischertechnik abstractions are excluded from this matrix because they exercise manufacturing/workcell and program-artifact semantics and are evaluated separately in Section 5.8.

For each source abstraction τ , the grounder constructs candidate target mappings μ and evaluates them using the eligibility predicates defined in Section 3.3: component coverage, implemented role/type compatibility, source-stage coherence, minimum-tag slice viability, and rule-surface intersection. Figure 4 reports the resulting grounding coverage. Each cell represents the proportion of source abstractions satisfying these criteria for the corresponding target process surface. Grounding coverage therefore measures acceptance under the implemented structural-admissibility filters for the target contract and does not measure slice adequacy, dynamic realizability, or downstream consumer performance.

Grounding coverage is highest for water-domain sources evaluated against water-domain targets and lower for several adjacent process families. Differences in coverage reflect the eligibility conditions imposed by the corresponding target contracts, including component coverage, implemented role/type compatibility, source-stage coherence, minimum-tag slice viability, and intersection with the declared rule surface. The matrix therefore characterizes the observed scope and boundaries of structural grounding across the evaluated source–target combinations.

5.3 Testing Grounded Water-Domain Cases

RQ2: Slice evidence. Grounding acceptance does not by itself establish that the resulting target-side representation provides sufficient execution evidence for downstream evaluation. We therefore examine two deeper water-domain studies, SWaT→WT and WADI→WD, by materializing selected accepted groundings as executable validation slices. The corresponding study bundles provide target signals, dependency relationships, nominal traces, controlled perturbation cases, timing information, and the metadata required to assess slice adequacy and downstream outcomes.

Case selection and execution. The five detailed cases are not a random sample of the 78 continuous-process source abstractions. Of these abstractions, 61 ground to WT and 59 to WD (Figure 4). We construct an executable validation case only when the mapped manipulated and consequence signals and the required dependencies are represented in

Target-Conditioned Grounding Coverage

How many source threats can be grounded to each SPHERE target family?

Source Corpora	 Water Treatment flagship	 Water Distribution flagship	 Hydro Power support	 Chemical (GRFICS/TE) support
 SWAT n=51	38/51 74.5%	37/51 72.5%	30/51 58.8%	20/51 39.2%
 WADI n=17	15/17 88.2%	14/17 82.4%	12/17 70.6%	11/17 64.7%
 OilTreatment n=10	8/10 80.0%	8/10 80.0%	7/10 70.0%	5/10 50.0%

Legend

Cell = grounded / total threats
Green intensity = grounding coverage %

Grounded = passed target-conditioned mapping
and all 5 deterministic eligibility checks

Fig. 4. Continuous-process grounding coverage across target SUT families. Rows denote source threat corpora and columns denote target process families. Each cell reports accepted groundings relative to the source-threat denominator for that row. The matrix comprises 78 continuous-process abstractions from SWaT, WADI, and OilTreatment; the five Fischertechnik manufacturing abstractions are evaluated separately under a manufacturing/program-artifact denominator.

the target study bundle and when both nominal evidence and a viable controlled perturbation surface are available. From this eligible set, we report five cases selected to span multiple effect families and to include a boundary case in addition to clearer downstream responses (Table 4). These experiments therefore examine a selected subset for which the required target-side execution substrate was available and do not imply execution of all 78 continuous-process threats. All five reported slices satisfy the declared adequacy requirements for signal coverage, dependency closure, and timing evaluability.

Source grounding versus WT probe. Figure 3’s SWaT attack 36 record lists only LIT-101 as manipulated and reports tank overflow. Its v1 structural grounding maps that role to WT’s RW_Tank_Level; the WT slice includes RW_Tank_PR_Valve as target context, not as a mapping of MV-101 from this source record. The type, stage, structure, and rule-surface eligibility checks are satisfied.

The separate WT high-spoof test applies a positive offset to RW_Tank_Level. Attack 36 instead sets LIT-101 below LL, so this test probes the mapped WT signal but does not replay the source attack’s direction. The native invariant consumer yields three nominal-subtracted violations: two rate-of-change and one mass-balance. This demonstrates a response to the controlled WT perturbation, not adversary access, exploit feasibility, or equivalent behavior under other plant dynamics.

Table 4. Water-domain validation-slice outcomes. Controlled target-side perturbations probe slices supported by accepted groundings; they need not be direction-matched replays of source attacks. Outcomes are labeled as clean, nominal-confounded, or near-threshold.

Grounding and target case	Target-side result and interpretation
SWaT→WT: Tank-level high spoof. Positive level offset on the WT tank slice; a target-side probe, not a direction-matched replay of SWaT attack 36.	Nominal-subtracted analysis yields three invariant violations: two rate-of-change and one mass-balance. This is the cleanest WT invariant case.
SWaT→WT: Pump-flow low spoof. Pump/flow suppression instantiated on the WT slice.	Raw analysis yields nine correlation violations, but nominal-subtracted analysis yields zero. We treat this as nominal-confounded and do not count it as a clean invariant success.
WADI→WD: Supply-flow low spoof. Low-flow effect instantiated on the WD supply slice.	No baseline invariant violations occur; the first firing appears only at the -64 perturbation boundary with shallow range/correlation evidence. We report this as a near-threshold boundary case.
WADI→WD: Supply-flow high idle. High-flow/pressure effect instantiated during an idle supply condition.	The slice produces 11 range violations, giving a clean WD flow/pressure case.
WADI→WD: NaOCl level step. Dosing/storage-level perturbation instantiated on the WD chemical slice.	The slice produces three mass-balance/rate-of-change violations. This is a clean WD dosing case, although it is outside the GeCo predictive surface.

The SWaT pump-flow case further illustrates the separation between grounding, adequacy, and consumer outcome. The slice satisfies both grounding and adequacy criteria, but all nine raw correlation violations also occur in the nominal trace and disappear after nominal subtraction. The consumer outcome is therefore classified as *nominal-confounded*, while the grounding and slice-adequacy classifications remain unchanged. The WADI→WD cases illustrate the same distinction: an adequately grounded slice may produce a response that survives the nominal baseline or may remain near the selected consumer’s decision boundary without changing the underlying grounding.

These studies characterize validation slices as an intermediate target-conditioned representation between source-derived threat semantics and downstream evaluation. The slices preserve the signals, dependencies, and context required for the corresponding analyses while maintaining separate grounding, adequacy, and consumer-outcome classifications. The perturbations are controlled modifications of simulator-role signals applied to otherwise nominal SPHERE executions; they are not captured or independently executed attacks. Appendix B reports the per-case adequacy metadata and raw and nominal-subtracted outcomes, while Section 6.1 discusses the resulting evidentiary limitations.

5.4 Consumer Compatibility over Shared Slices

RQ3: Interoperability. We next evaluate whether a common validation-slice representation can provide the information required by multiple downstream analysis styles. Table 5 summarizes four bounded consumer lanes evaluated over the WT/WD slices: the native invariant-family consumer, a bounded GeCo-style predictive lane [37], a bounded SAIN-style state-aware lane [1], and a bounded SCAPHY-style phase-aware lane [19]. The latter three are XPhysICS-native implementations of selected analysis patterns motivated by the corresponding systems; they do not execute the published implementations.

The invariant-family lane provides the native rule-based evaluation of the WT/WD validation slices, including the nominal-confounded WT pump-flow case. The bounded GeCo-style lane identifies both WT cases and two of the three WD cases at a target nominal FPR of 0.0; the NaOCl dosing case lies outside its modeled surface. The bounded SAIN-style lane removes the nominal false-positive behavior observed for WT while retaining the reported WT and WD coverage. The bounded SCAPHY-style lane additionally identifies the WD low-flow boundary case, increasing the observed WD coverage from 2/3 to 3/3 within that bounded implementation.

Table 5. Consumer utility from shared validation slices. Rows consume the same WT/WD slice artifacts using different bounded analysis styles.

Consumer family	WT slice results	WD slice results	Role in evaluation
Invariant checks	Raw 2/2; 1 clean case remains after nominal subtraction.	2/3; high-flow and NaOCl are strong, low-flow is partial.	Native slice consumer; preserves WT nominal caveat.
GeCo-style predictive [37]	2/2 hits; nominal FPR 0.0; mean F1 0.4167.	2/3 hits; nominal FPR 0.0; mean F1 0.1025; NaOCl outside modeled surface.	Bounded predictive compatibility lane.
SAIN-style state-aware [1]	2/2 hits; nominal FPR improves from 0.0167 to 0.0; mean F1 0.66.	3/3 hits; nominal FPR 0.0; mean F1 1.0.	Bounded state-aware compatibility lane.
SCAPHY-style phase-aware [19]	2/2 hits; nominal FPR 0.0; mean F1 0.5661.	3/3 hits; recovers low-flow case; nominal FPR 0.0; mean F1 0.57.	Bounded phase/context-aware compatibility lane.

Across these lanes, the validation-slice representation supplies signals and context used by several downstream analysis styles, including invariant evaluation, predictive residual analysis, state-conditioned thresholds, and phase-conditioned scoring. These results provide evidence of consumer compatibility with the shared slice representation under the evaluated implementations. They are not comparative measurements against the corresponding published systems and do not establish fidelity to those systems. Implementation scope and consumer-specific details are reported in Appendix D.

5.5 Bounded-Support Validation Slices: Hydro and Chemical-Process Targets

We next examine nine bounded-support cases spanning SWaT→Hydro, WADI→GRFICS, and EPIC→Hydro. These cases use the same target-contract and controlled-perturbation methodology as the primary validation studies (Section 4.2). Because corresponding nominal-subtracted baselines are unavailable, the resulting evidence is limited to validation-slice executability and observed bounded-consumer response rather than the stronger outcome interpretation available for the WT/WD flagship cases. Appendix C reports the full adequacy and timing metadata.

Across the nine cases, pressure- and flow-related perturbations produce less consistent activation of dependency-based rules than of range rules. A similar pattern is observed for the WD low-flow case (Section 5.3). Interpretation of these cases is constrained by two aspects of the experimental design: several Hydro cases share a target execution bundle, and nominal-subtracted baselines are unavailable. Consequently, variation in observed downstream rule response does not modify the structural grounding or slice-adequacy classifications. These cases provide evidence of validation-slice executability and bounded consumer compatibility across the evaluated target families without establishing target-independent behavior or validation against independently executed attacks.

5.6 Upstream GeCo as a Consumer

The GeCo-style lane in Section 5.4 is a bounded XPHYSICS-native implementation of a predictive analysis pattern motivated by GeCo. We separately evaluate compatibility with the published upstream GeCo implementation by executing the unmodified artifact, including its released per-case retraining and Industrial Protocol Abstraction Layer (IPAL) evaluation pipeline, over nine XPHYSICS-grounded slices spanning three studies. For each case, the evaluation uses 60 target rows, 21 of which correspond to the perturbation window. The SPHERE-derived bundles are represented in IPAL state format, after which the upstream GeCo pipeline performs its released training and scoring procedure without modification.

The upstream evaluation additionally reports eTaF1, the event-aware metric used by GeCo. As an artifact-level consistency check, we also executed the released evaluation on GeCo’s native benchmark material. The WADI result

Table 6. Bounded-support validation slices on hydro and chemical-process targets. All nine satisfy the reported slice-adequacy requirements but lack nominal-subtracted baselines; the observed responses therefore characterize slice executability and bounded consumer compatibility rather than clean consumer outcomes. [†]Shared Hydro executions provide convergent source-side grounding evidence while sharing the same target-side execution evidence.

Case	Effect / perturbation	Observed response	Interpretation
SWaT→Hydro: reservoir level	overflow, HY_Res_Level +20	range, rate-of-change fire (2/2)	Both declared rule types respond
SWaT→Hydro: turbine speed	speed excursion, HY_Speed_Pct +70	range, rate-of-change fire (2/2)	Both declared rule types respond
SWaT→Hydro: flow	pressure/flow, HY_Flow -90	range fires only (1/3)	Limited observed dependency-rule response
WADI→GRFICS: tank pressure	pressure/flow, TE_Tank_Pressure +1500	rate-of-change fires only (1/2)	Limited observed rule response
WADI→GRFICS: tank level	overflow, TE_Tank_Level +40	rate-of-change and causality responses	Mixed observed rule response
WADI→GRFICS: feed-1 flow	pressure/flow, TE_Feed1_Flow -250	causality, range fire (2/3)	Broadest observed GRFICS rule response
EPIC→Hydro: turbine speed	speed excursion, HY_Speed_Pct +70	range, rate-of-change fire (2/2), identical to SWaT-sourced case	Shared target execution [†]
EPIC→Hydro: generated power	speed excursion, HY_Power_MW -40	range, correlation fire (2/2)	Both declared rule types respond
EPIC→Hydro: flow	pressure/flow, HY_Flow -90	range fires only (1/3), identical to SWaT-sourced case	Limited observed dependency-rule response; shared execution [†]

matches the retained comparison baseline available with the evaluated artifact. The corresponding SWaT comparison baseline was not retained and therefore could not be independently rechecked.

The nine slice executions do not correspond to nine independent target-side observations. Two cases share the same Hydro target execution and nominal evidence as corresponding cases derived from different source-side provenance, and consequently produce identical GeCo measurements. These cases therefore provide convergent grounding evidence from distinct source-side provenance while sharing target-side execution evidence.

Across the nine executions, GeCo obtains recall 1.0 with $tp = 21$ and $fn = 0$ for each case, while false positives range from 2 to 3. Only two distinct eTaF1 values occur across the nine cases. These measurements therefore provide little case-level differentiation among the finer interpretations reported in Table 6. The concentration of observed outcomes is consistent with the large controlled perturbations and small per-case evaluation setting used here, although this experiment does not isolate the cause of the limited variation. The controlled-perturbation limitations described in Section 6.1 continue to apply.

This experiment establishes that the evaluated XPhysICS validation slices can serve as inputs to the unmodified upstream GeCo pipeline and can be scored using its released event-aware evaluation procedure. The evidence therefore supports downstream consumer compatibility for the evaluated cases. It does not establish transfer quality, case-level discrimination, independent target-side replication, comparative superiority, or validation against independently executed attacks.

5.7 Dynamic Realizability: A Bounded Example

RQ4: Dynamic realizability. Structural grounding does not establish that a grounded effect can be produced through a particular plant model and legitimate command surface. We examine this distinction for three frozen XPhysICS groundings on the GRFICS/Tennessee-Eastman target: one pressure-family objective (pressure_flow_perturbation, $P > 3100$) and two level-family objectives (underflow_depletion, $L \leq 5$; overflow_accumulation, $L \geq 95$). For these

cases, we use a paper-derived reproduction of the physics-guided search method described by ICSFlux [32] as a separate check of dynamic realizability.

Each search objective is compiled from a frozen grounded threat and specifies the corresponding violation condition over the evaluated observable and command surface. The search procedure is separate from the controlled additive perturbation used to exercise the validation slices in the preceding studies. This experiment therefore asks a different question: whether an already-grounded objective can be reached under a specified model and command surface. It does not evaluate the published ICSFlux artifact and does not by itself remove the limitations associated with the controlled perturbations used elsewhere in the evaluation.

Against the plant model and command surface specified in the ICSFlux appendix, the pressure objective is reached, whereas the two level objectives are not reached under the evaluated search and extremal-command campaign. In the additional extremal campaign, every level-relevant command is held at its corresponding extreme for 400 simulated hours while the controller retains the level within the non-violating band. Under the live OpenPLC-driven plant configuration used for this bounded study, all three objectives are reached. These observations are limited to three objectives on a single target and provide evidence that structural grounding and model-specific dynamic realizability are distinct properties. The empirical negative result is bounded by the evaluated model, command alphabet, and campaign; it is not a general impossibility proof. The difference between the two execution settings should not be interpreted as a property of the grounding decision or of the target independently of those conditions.

5.8 Extension Evidence and Boundaries

Finally, we examine whether the XPHYSICS evidence representation can support additional artifact types and analysis surfaces beyond the primary continuous-process studies. These experiments exercise different denominators and evidence forms and are therefore reported separately rather than combined with the continuous-process grounding percentages or the primary validation- slice results.

CISS [20] contributes additional water-treatment attack semantics. The Fischertechnik TXT→Siemens S7 study exercises manufacturing/program-artifact grounding using program, tag, and state evidence. CrossPLC exercises optional controller-aware enrichment by exposing controller-side tags, routines or function blocks, read/write dependencies, call relations, and available state structure in code-rich water and manufacturing cases. These results do not alter the denominator of the continuous-process grounding matrix. STL/RTAMT guardrails [29] provide bounded temporal robustness evaluation for WADI→WD boundary cases. EPIC/Hydro [3] and PowerDuck/GOOSE [38] contribute power- and protocol-oriented evidence, including GOOSE flooding, insertion, replay, and suppression traces, while VetPLC-style artifacts [39] contribute event and timing obligations for manufacturing-oriented settings.

Taken together, these studies show that the XPHYSICS representation can be instantiated over additional evidence types and mechanisms. They do not provide the same target-side evidence as the primary WT/WD studies and are therefore interpreted under their own experimental scope. Appendix F reports the corresponding extension-evidence ledger.

6 Discussion

Grounding and downstream evaluation. The grounding matrix answers a narrow question: which source abstractions pass the implemented component-coverage, role/type, source-stage, minimum-tag, and rule-intersection filters for each target contract? A pass does not establish detector performance, attack severity, exploitability, slice adequacy,

or dynamic realizability. A rejected candidate fails one or more grounding checks; a partial or insufficient slice instead identifies a limitation in the evidence available after grounding.

The validation slice records the mapped paths, signals, dependencies, timing assumptions, and context associated with an accepted result. Later studies ask different questions about that record. For example, a strong slice may still produce a *nominal-confounded* consumer outcome because the same violation appears in the nominal trace, or it may not expose the inputs expected by a particular consumer. Neither case changes the grounding decision.

Complementarity with downstream analyses. XPhysICS produces target-conditioned validation artifacts rather than a new detector or provenance-analysis system. The bounded GeCo-, SAIN-, and SCAPHY-style lanes (Section 4.4) are XPhysICS-native implementations of selected analysis patterns motivated by the corresponding systems and are used to evaluate whether the validation-slice representation provides the required consumer inputs. The upstream GeCo study (Section 5.6) separately demonstrates that the evaluated XPhysICS-grounded slices can be processed by the released GeCo pipeline. Its evidentiary scope is consumer compatibility rather than transfer quality, case-level discrimination, or comparative detector performance.

The classification- and provenance-oriented adapters in Appendix E exercise additional output interfaces motivated by SCADMAN and ICSTracker without executing those published systems. Together, these studies exercise several output interfaces; only the GeCo study establishes compatibility with a released upstream implementation.

Evidence quality and grounding boundaries. The quality of XPhysICS’s output depends on the completeness and accuracy of the available source and target evidence. Missing process diagrams, undocumented bypasses, stale tag names, unit or type mismatches, and vendor-specific aliases may prevent otherwise plausible candidate mappings from satisfying the grounding criteria or may reduce the adequacy of the resulting slice. XPhysICS therefore retains field-level provenance and explicit eligibility outcomes rather than filling missing evidence through unconstrained inference. This design makes the evidentiary basis and residual uncertainty of each result inspectable, which is particularly important for legacy systems with incomplete or inconsistent documentation.

Optional enrichment paths. CrossPLC and STL/RTAMT provide optional evidence for particular studies without changing the requirements of the core grounding procedure. In the water-treatment controller study, CrossPLC contributes controller-side evidence for role and dependency recovery while leaving the continuous-process grounding denominator unchanged. The STL/RTAMT study evaluates temporal robustness around the WADI→WD boundary case and therefore supplements slice-level evaluation rather than serving as an alternative grounding mechanism. These studies illustrate how additional controller-code and temporal-logic evidence was used in the evaluated cases without making either source a prerequisite for grounding.

Scope of extension evidence. The extension studies in Section 5.8 exercise evidence surfaces and denominators distinct from the continuous-process grounding matrix. These studies include manufacturing and program-artifact grounding, controller-code enrichment, protocol traces, and temporal obligations. Accordingly, their results are reported separately and are not included in the continuous-process grounding percentages or interpreted as equivalent to the primary WT/WD executable studies.

Security and deployment considerations. Grounding over physical roles, process context, dependencies, observable signals, and declared rule surfaces reduces dependence on exact source–target tag-name correspondence. The procedure nevertheless remains sensitive to the integrity and completeness of its input evidence. Incorrect or outdated manuals, misleading engineering diagrams, stale configuration information, or adversarially manipulated evidence can affect source abstraction or target contract construction. Provenance tracking, schema validation, and cross-source

corroboration provide partial safeguards, while stronger mechanisms for evidence integrity and attestation remain deployment concerns.

Practical deployment also requires engineering effort beyond the grounding algorithm itself, including alias reconciliation, unit and type normalization, timestamp alignment, target-contract construction, and maintenance of target-side dependency information. These activities affect the coverage and adequacy of the resulting validation slices and should therefore be treated as part of the deployment boundary rather than as hidden assumptions of the grounding procedure.

Scope of deterministic grounding. The deterministic claim of XP_{HYSICS} applies only after the source abstraction, vocabulary and schema version, and target contract have been fixed. Source evidence selection, semantic abstraction, confidence assignment, and target-contract construction may involve analyst interpretation or source-specific parsing and are outside this deterministic claim. In addition, the current grounding vocabulary is narrower than the extraction vocabulary (Table 1), so some valid source abstractions fall outside the effect-family scope currently supported by the target contracts.

Interpretation of controlled-perturbation evidence. The five flagship SWaT→WT/WADI→WD cases and the nine bounded-support Hydro/GRFICS cases (Section 5.5) use controlled additive perturbations applied to simulator-role signals during otherwise nominal executions. These perturbations provide evidence that the corresponding validation slices can be executed under the evaluated target configurations and that the relevant target-side signals and dependencies exhibit observable responses. They are not captured or independently executed attacks and therefore do not establish adversary access, exploit feasibility, realistic manipulation magnitude, or equivalent consumer behavior under subtler or differently structured perturbations. These limitations motivate the separate reporting of grounding, slice adequacy, consumer applicability, and consumer outcome.

Structural grounding and dynamic realizability. Structural grounding establishes that a source threat abstraction satisfies the implemented component-coverage, role/type, source-stage, minimum-tag, and rule-intersection filters. Producing the corresponding effect through a particular plant model and command surface is a separate question.

The bounded study in Section 5.7 illustrates this distinction for three frozen groundings. Their realizability differs between the analytical plant model used by the paper-derived search reproduction and the evaluated OpenPLC-driven execution configuration. This result provides a bounded example in which structural grounding remains unchanged while model-specific realizability differs. The observation is limited to three objectives on one target and should not be generalized to other groundings, models, command surfaces, or target systems.

6.1 Threats to Validity

Construct validity. The central construct evaluated in this work is target-conditioned grounding, defined as acceptance under five eligibility predicates against a fixed source abstraction, vocabulary and schema version, and target contract. The validity of this construct therefore depends on both the representational vocabulary and the selected grounding criteria. The effect-family, role, and rule-surface representation was developed from the source and target corpora considered in this study and may not capture threat semantics outside those domains. The reported grounding evaluation covers only five of the seven effect families available during source abstraction; abstractions assigned to the two remaining families fall outside the reported grounding surface.

The five eligibility predicates (χ_{map} , χ_{type} , χ_{stage} , χ_{slice} , χ_{rule}) operationalize structural admissibility for the evaluated target contracts but are not derived from an external standard. Alternative vocabularies, mapping procedures, or eligibility criteria could therefore produce different grounding decisions. The implemented role/type check is coarse rather

than general physical-unit reasoning. Component coverage is existence-based, stage coherence uses source-declared stage labels, slice viability uses a minimum expanded-tag count, and rule relevance uses tag intersection. The reported grounding matrix therefore measures acceptance under these implemented filters, not independently adjudicated mapping correctness. Finally, grounding and validation-slice execution do not establish that an equivalent exploit would succeed against deployed target control logic.

Internal validity. Source-abstraction construction involves analyst judgment. The evaluation includes a bounded manual spot-check over 20 records (Appendix A) rather than an independent multi-analyst study. Consequently, the evaluation does not provide an inter-annotator agreement measure or establish that independent analysts would construct identical abstractions from the same source evidence.

The deep validation studies are also subject to selection effects. The five flagship cases and nine bounded-support cases were selected from cases for which the required target contract, observable signals, dependency structure, and viable perturbation surface were available. They therefore do not form a random or exhaustive sample of the 78 continuous-process abstractions. Effect-family diversity and the inclusion of a boundary case also influenced case selection.

The target-side perturbations are controlled synthetic modifications rather than independently executed attacks. Large additive offsets may be easier for threshold- or residual-based consumers to identify than subtler, gradual, or dynamically generated manipulations. This limitation also affects the interpretation of the upstream GeCo results (Section 5.6), which provide limited differentiation across the evaluated cases. The separate realizability study reduces reliance on the same additive perturbation mechanism for three selected objectives but does not resolve this limitation for the broader validation corpus.

External validity. The deepest executable evidence remains concentrated in water-domain groundings: SWaT into a water-treatment target and WADI into a water-distribution target within the SPHERE experimental environment. The Hydro and chemical-process studies in Section 5.5 extend validation-slice construction and controlled execution to additional process families, but they remain research-environment evaluations and do not constitute independently captured cross-domain attacks.

The evaluation therefore does not establish generalization to ICS families outside the evaluated water, energy, chemical-process, and bounded manufacturing settings; to substantially different control architectures; or to targets for which the required role, signal, dependency, or nominal evidence cannot be constructed. Incomplete target evidence is reflected in failed grounding criteria or partial/insufficient slice adequacy rather than treated as evidence of transferability.

7 Related Work

Model-Based CPS Safety and Security Coengineering. System Aware and Mission Aware cybersecurity use systems-engineering artifacts to connect mission objectives, system structure, attack vectors, unsafe actions, hazards, resilience modes, and sentinel-based monitoring [11, 12, 16, 21, 22]. Bakirtzis et al. extend this perspective through an ontological metamodel that relates model-based systems-engineering entities, including components, links, functions, control actions, feedback, and operational context, to safety-, security-, and resilience-oriented concepts such as attack vectors, loss scenarios, hazards, unsafe actions, resilient modes, and sentinels [9]. Related approaches further connect system models with attack-vector or vulnerability knowledge to support design-time security analysis [8, 10].

This body of work emphasizes design-time traceability among architecture, functions, mission objectives, attack vectors, losses, and resilience decisions. XPhysICS addresses a complementary problem: target-conditioned grounding

of threat evidence documented for one ICS onto a selected target SUT. Rather than requiring a complete mission-engineering or system model, XPhysICS evaluates whether a structured source threat abstraction satisfies the target contract and, when it does, constructs a validation slice containing the manipulated and consequence paths, observability requirements, and target-side evaluation context required for subsequent analysis.

ICS Threat Intelligence, Ontologies, and Evidence Representations. Security ontologies and knowledge graphs provide structured representations of cyber-physical threat intelligence. Prior work includes industrial-IoT ontologies [27], CTI knowledge-graph construction using distant supervision [33], LLM-assisted CTI extraction systems [13], and OT-oriented representations such as OTThreat [30]. These approaches provide vocabularies, extraction mechanisms, and structured representations for organizing threat, asset, and evidence information.

XPhysICS builds on this representation-oriented foundation by connecting structured source threat abstractions to explicit target contracts. Its evidence ledger preserves provenance across source evidence, abstraction fields, candidate mappings, grounding decisions, and validation-slice construction. Grounding further requires candidate mappings to satisfy the implemented component-coverage, role/type, source-stage, minimum-tag, and rule-intersection criteria. The distinction is therefore between representing source-side threat semantics and evaluating their structural admissibility on a particular target.

Target-Side Detection, Monitoring, and Provenance Consumers. A broad class of ICS security systems performs analysis after the relevant target signals, controller behavior, or execution context have been characterized. GeCo learns discrete state-space models from benign traces and detects departures from learned behavior [37]. SAIN derives state-aware invariants from PLC program states and transitions [1]. SCAPHY relates SCADA-host execution phases to physical behavior [19]. SCADMAN evaluates control-loop consistency across distributed controllers [2], while ICSTracker reconstructs provenance-oriented causal relationships for ICS forensics [5].

These systems focus on downstream analysis over an instrumented or otherwise characterized target. XPhysICS instead addresses the preceding problem of constructing target-conditioned context from externally documented threat semantics. Its validation slices identify manipulated and consequence paths, observable signals, dependencies, and consumer-relevant context that can then be exposed to downstream analyses.

Sections 5.4 and 5.6 examine this relationship through bounded XPhysICS-native consumer lanes and the released upstream GeCo implementation, respectively. Appendix E additionally evaluates prototype output interfaces motivated by SCADMAN and ICSTracker. These studies have different evidentiary scopes and are used to assess downstream compatibility rather than comparative detector performance or fidelity across systems.

PLC Analysis, Code-Level Testing, and Formal Guardrails. PLC-oriented analysis and testing techniques provide assurance at the controller-program and scan-cycle levels. Attack-pattern mining, static or dynamic program analysis, and PLC fuzzing can expose implementation-level weaknesses or execution behaviors [35, 36]. ICSQuartz, for example, provides a vendor-agnostic analysis path for IEC 61131-3 Structured Text through a RuSTy-to-LLVM toolchain [36].

CrossPLC serves a different role within XPhysICS. When controller artifacts are available, it lifts heterogeneous PLC project formats into a common representation of tags, routines or function blocks, read/write relationships, call relations, and available state information. This representation provides optional controller-level evidence for role recovery, dependency recovery, and validation-slice construction, but is not required by the core target-grounding procedure.

Temporal-logic monitoring provides another complementary analysis surface. RTAMT supports robustness-based evaluation of cyber-physical temporal specifications [29]. XPhysICS uses STL/RTAMT as an optional slice-level mechanism when a grounded validation slice exposes the signals and thresholds required by a temporal specification. These checks supplement target-side evaluation and remain separate from grounding acceptance.

Physics-Guided Target Search. ICSFlux [32] searches for legitimate command sequences that drive a target toward a specified violation condition given a black-box controller, a physical model, and an explicit safety/security constraint. XPhysICS addresses a different stage: it determines whether externally documented threat semantics are structurally admissible on a selected target. In relation to the three ICSFlux inputs, XPhysICS provides information relevant to formulating the violation constraint; it does not derive the physical model or black-box controller. The two approaches therefore address complementary questions of target-conditioned threat grounding and model-specific dynamic realizability.

Section 5.7 examines this distinction through a bounded study that applies a paper-derived reproduction of the ICSFlux search method to three frozen XPhysICS groundings. The study treats structural grounding and dynamic realizability as separate evidentiary questions. It does not evaluate the published ICSFlux artifact and does not constitute an implemented XPhysICS–ICSFlux workflow.

8 Conclusion

This paper asked when threat knowledge documented for one industrial control system can be meaningfully evaluated on another. XPhysICS addresses that question by turning source evidence into a provenance-linked threat abstraction and checking it against a target contract. Once the abstraction, vocabulary, schema version, and contract are fixed, the grounding checks are deterministic. An accepted result can be materialized as a validation slice that records the mapped roles, signals, dependencies, timing assumptions, and provenance. Whether that slice is adequate, whether the effect is dynamically reachable, and what a downstream analysis reports remain separate questions.

The grounding matrix evaluates 78 continuous-process abstractions under the implemented target-contract filters. The deepest executable evidence remains the water-domain SWaT→WT and WADI→WD studies. Nine Hydro and chemical-process cases extend controlled slice execution to additional target families, but they are research-environment perturbations, not independently captured cross-domain attacks. The released upstream GeCo study demonstrates compatibility for the evaluated slices. A separate three-objective, one-target study shows that a grounding decision can remain fixed while realizability changes with the evaluated plant model and command surface.

Within these evaluated contracts and cases, semantic similarity alone was not enough to characterize target applicability. The source interpretation had to remain traceable to evidence, and candidate mappings had to pass explicit target checks. Grounding results and downstream experimental results should therefore be read as answers to different questions, within the limitations in Section 6.1. Future work will broaden the target-contract vocabulary, examine individual grounding criteria and simpler mapping baselines, and extend deep executable evaluation to additional target families and control architectures.

Acknowledgments

This material is based upon work supported in part by the U.S. National Science Foundation under Award Nos. 2425711 (FMitF) and 2330066 (SPHERE). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the U.S. National Science Foundation.

A Source Evidence and Extraction Details

Table 7 provides additional source-evidence and extraction details complementing the summary in Table 3. It reports source- and domain-level evidence, manual spot-check results, and provenance metadata associated with the source abstractions used in the evaluation. The evidence representation is informed by OTThreat [30], while the evidence ledger and its associated records are implemented as part of XPHYSICS.

Table 7. Source-evidence and extraction ledger used in the evaluation, including provenance metadata and manual spot-check results.

Evidence object	Appendix detail
Structured extraction	Result: 83 threats from 12 documents; 55 high-confidence, 18 medium-confidence, and 10 low-confidence. The source set includes SWaT [26] (51), WADI [4] (17), OilTreatment (10), and Fischertechnik (5). Claim boundary: these are coverage and confidence counts for the paper-facing corpus, not a full extraction precision/recall evaluation.
WADI [4] deepening	Result: naive PDF parsing yields 2 rows; XPHYSICS recovers 15 A1 threats and 14 A2 attack windows, a 7.5× gain in attack-level threat records. Claim boundary: A1 seeds extraction, A2 contributes timing/observability windows, and A3 provides clean context; this is source-specific extraction, not a general PDF-table benchmark.
CISS [20] semantics	Result: 187 parsed rows, 73 process-relevant rows, and 20 high-confidence derived threats from CISS [20]. Claim boundary: CISS contributes source semantics and observability support; we do not treat it as perturbation replay.
Manual sample	Result: 20 sampled items: 14 full pass, 6 partial, and 0 fail. Claim boundary: this bounded spot-check found no outright abstraction failures, but it is not a full precision/recall measurement.
XPHYSICS evidence ledger	Result: 333 curated threats, 1,947 field-level evidence spans, and 100% core-field coverage. Claim boundary: the ledger is an artifact of XPHYSICS, aligned with OTThreat [30]; it supports provenance and auditability, not the grounding algorithm itself.

B Water-Domain Validation-Slice Details

Table 8 expands Table 4 with adequacy, perturbation, and consumer-outcome metadata for the water-domain validation slices. Raw consumer responses are reported separately from nominal-subtracted outcomes because some violations observed during the perturbation window are also present in the corresponding nominal traces. Nominal subtraction is therefore used to distinguish responses that remain after accounting for the nominal baseline from nominal-confounded outcomes and to identify cases near the evaluated consumer’s decision boundary.

Table 8. Detailed water-domain validation-slice evidence, including slice adequacy, controlled-perturbation metadata, raw consumer response, nominal-subtracted outcome, and interpretation. Target signals are defined by the corresponding SPHERE target contracts.

Grounding case	Detailed outcome
SWaT→WT: tank-level high spoof	Adequacy: target signals declared in the WT contract; 100% source/target/dependency coverage; timing evaluable. Outcome: nominal-subtracted delta yields 3 violations: 2 rate-of-change and 1 mass-balance. Interpretation: clean WT invariant case.
SWaT→WT: pump-flow low spoof	Adequacy: target signals declared in the WT contract; 100% source/target/dependency coverage; timing evaluable. Outcome: raw analysis yields 9 correlation violations, but nominal-subtracted delta yields 0. Interpretation: nominal-confounded; not counted as a clean invariant success.
WADI→WD: supply-flow low spoof	Adequacy: target signals declared in the WD contract; 100% source/target/dependency coverage; partial verdict. Outcome: no baseline invariant violations; first firing occurs at −64 with shallow range/correlation evidence. Interpretation: near-threshold boundary case.
WADI→WD: supply-flow high idle	Adequacy: target signals declared in the WD contract; 100% source/target/dependency coverage; strong verdict. Outcome: 11 range violations. Interpretation: clean WD flow/pressure case.
WADI→WD: NaOCl level step	Adequacy: target signals declared in the WD contract; 100% source/target/dependency coverage; strong verdict. Outcome: 3 mass-balance/rate-of-change violations. Interpretation: clean WD dosing case; outside the GeCo-style predictive surface.

C Bounded-Support Validation-Slice Details (Hydro/GRFICS)

Table 9 provides the detailed adequacy and consumer-response metadata for the nine bounded-support cases summarized in Table 6. All nine satisfy the declared adequacy requirements for their respective studies. Nominal-subtracted baselines are unavailable for these cases, so the observed rule responses are not assigned the *clean* consumer-outcome classification used for flagship cases with corresponding nominal evidence. The earliest recorded rule violation occurs at the perturbation onset (sample 20) in all nine cases. This timing is reported as execution metadata supporting slice executability and should not be interpreted as a substitute for nominal-subtracted evidence or as evidence of independently executed attacks.

Table 9. Adequacy and consumer-response metadata for the nine Hydro/GRFICS bounded-support cases. “Earliest” denotes the sample index of the first recorded rule violation; the controlled perturbation begins at sample 20 for every case. Observed rule responses characterize bounded consumer behavior without a nominal-subtracted baseline.

Case	Expected types	rule	Viol.	Earl.	Observed rule types (expected hit / total)
swat_hydro_uc1 / dro_reservoir_level_high_spoof	hy- range; rate_of_change		23	20	range; rate_of_change (2/2)
swat_hydro_uc1 / dro_speed_pct_high_spoof	hy- range; rate_of_change		23	20	range; rate_of_change (2/2)
swat_hydro_uc1 / dro_flow_low_spoof	hy- range; correlation; causality		23	20	range; rate_of_change (1/3 expected; additional rate_of_change)
wadi_grfics_uc1 / grfics_tank_pressure_high_spoof	range; rate_of_change		2	20	rate_of_change (1/2)
wadi_grfics_uc1 / grfics_tank_level_high_spoof	range; rate_of_change		3	20	causality; rate_of_change (1/2 expected; additional causality)
wadi_grfics_uc1 / grfics_feed1_flow_low_spoof	range; correlation; causality		25	20	causality; range; rate_of_change (2/3 expected; additional rate_of_change)
epic_hydro_uc0 / dro_speed_pct_high_spoof	hy- range; rate_of_change		23	20	range; rate_of_change (2/2)
epic_hydro_uc0 / dro_power_mw_low_spoof	hy- range; correlation		22	20	correlation; range (2/2)
epic_hydro_uc0 / dro_flow_low_spoof	hy- range; correlation; causality		23	20	range; rate_of_change (1/3 expected; additional rate_of_change)

D Bounded Consumer Implementations

Table 10 documents the implementation scope of each bounded consumer lane over the common XPHYSICS validation-slice representation. The GeCo-, SAIN-, and SCAPHY-style lanes are XPHYSICS-native implementations of selected predictive, state-aware, and phase-aware analysis patterns motivated by the corresponding published systems. They are used to evaluate whether the validation-slice representation provides the signals and contextual information required by these analysis styles. They are not evaluations of the corresponding published implementations. Compatibility with the released upstream GeCo implementation is evaluated separately in Section 5.6.

Table 10. Bounded consumer implementations over the common validation-slice representation. The “-style” lanes are XPHYSICS-native implementations of selected analysis patterns motivated by the corresponding published systems and are evaluated as bounded consumers rather than reproductions of those systems.

Consumer family	Bounded implementation details
Invariant checks	Slice inputs: signal ranges, correlations, mass-balance relationships, and rate-of-change constraints. Calibration: thresholds from nominal traces. Scoring: violation count and nominal-subtracted delta. Boundary: native XPHYSICS consumer and primary validation lane.
GeCo-style predictive [37]	Slice inputs: time-series windows and signal correlations. Calibration: nominal-trace training and threshold calibration. Scoring: hit/miss, nominal FPR, and F1. Boundary: bounded predictive lane inspired by GeCo; not upstream code reuse or faithful reproduction.
SAIN-style state-aware [1]	Slice inputs: target signals and coarse discrete process state. Calibration: state-conditioned thresholds from nominal traces. Scoring: hit/miss, nominal FPR, and F1. Boundary: bounded state-aware lane; not full SAIN reproduction.
SCAPHY-style phase-aware [19]	Slice inputs: target signals, phase labels, and operating context. Calibration: phase-segmented nominal calibration. Scoring: hit/miss, nominal FPR, and F1. Boundary: bounded phase/context-aware lane; not full SCAPHY reproduction.

E Additional Downstream Adapter Evidence

Table 11 summarizes XPhysICS-authored prototype adapters for post-analysis classification and provenance-oriented workflows. The adapters use analysis patterns and output structures motivated by SCADMAN and ICSTracker but do not execute the published systems. Their measurements therefore characterize the XPhysICS-authored prototypes rather than the corresponding published systems. This evidence is distinct from the upstream GeCo study in Section 5.6, which executes the released GeCo pipeline over XPhysICS-grounded validation slices.

The reviewed SCADMAN-Light artifact was not treated as an implementation of the published SCADMAN analysis in this evaluation. Within the reviewed artifact, the available implementation applies a weighted attack-signature classifier with a fixed decision threshold rather than exposing the control-flow-attestation mechanism considered in our SCADMAN- motivated adapter. This observation is limited to the reviewed SCADMAN-Light artifact and is not a claim about other SCADMAN implementations. No corresponding implementation assessment was performed for semantic-icstracker.

Post-analysis classification adapter. The prototype consumes detector windows, evaluates target-grounded temporal predicates, and emits typed events with robustness and provenance metadata. Under the evaluated 1 FA/hr operating point, adding the XPhysICS-derived semantic classification stage changes precision from 0.375 to 1.000 and F1 from 0.545 to 0.800, while recall changes from 1.000 to 0.667. These measurements characterize the evaluated XPhysICS prototype and operating point; they are not measurements of the published SCADMAN system or a comparison against it.

Provenance-workflow adapter. The prototype consumes grounded mappings, eligibility outcomes, projected edges, impact chains, and lineage records and instantiates rule templates over eligible mappings. In the evaluated workflow, all 17 templates are instantiable and produce 2,074 typed mappings. The corresponding rerun agreement measurement is 0.95 under the evaluated configuration. These measurements characterize the XPhysICS prototype workflow and do not constitute an execution or evaluation of the published ICSTracker system.

Table 11. Prototype downstream-adapter evidence. Measurements characterize the XPhysICS-authored adapters under the evaluated configurations; SCADMAN and ICSTracker motivate the corresponding analysis patterns but are not executed as baselines.

Prototype adapter	Verified prototype evidence and claim boundary
Post-analysis classification/typing	Result: at 1 FA/hr, baseline P/R/F1 was 0.375/1.000/0.545; adding XPhysICS semantics yielded 1.000/0.667/0.800, corresponding to a precision gain of +0.625 and F1 gain of +0.255. What it exercises: post-analysis typing over detector windows using grounded temporal predicates and provenance. Boundary: these measurements characterize an earlier XPhysICS prototype adapter only. SCADMAN is cited as motivating literature and was not independently reproduced here; the evaluated prototype does not implement SCADMAN’s control-flow-attestation mechanism and therefore does not measure SCADMAN performance or mechanism fidelity.
Provenance-workflow adapter	Result: 17/17 rule templates instantiated, 2,074 typed physics-consistent mappings, and 0.95 reproducibility across reruns. What it exercises: rule expansion, provenance-pack generation, eligibility checks, and typed path export. Boundary: these measurements characterize the prototype provenance workflow only. ICSTracker is cited as motivating literature; because implementation fidelity has not been independently established, these results do not constitute an ICSTracker reproduction or performance evaluation.

F Extension Evidence Ledger

Table 12 provides additional details for the extension studies summarized in Section 5.8. These studies exercise evidence surfaces distinct from the primary continuous-process evaluation, including source-semantic breadth, manufacturing/program-artifact grounding, controller-code enrichment, temporal guardrails, protocol evidence, and event/timing obligations. Accordingly, their results use study-specific denominators and are reported separately from the continuous-process grounding percentages and primary WT/WD validation-slice studies.

Table 12. Extension evidence ledger. Each row identifies the evaluated extension, its principal result, the XPHYSICS capability exercised, and the evidentiary scope of the result.

Extension	Key Result	What It Exercises	Why It Is Bounded	Paper Loc.
CISS observability	20 high-confidence threats; 13/20 full-WT grounded; avg score 2.45	Source semantics breadth; slice-minimality	Source semantics only, not perturbation replay	§5.1
TXT→S7 Fischertechnik	5/5 threats grounded; 102 tags; 26/28 components	Manufacturing/program-artifact grounding	Separate manufacturing denominator; not continuous-process matrix	§5.8, App. F
CrossPLC control study	MiniSWaT/SPHERE-WT; 6/6 role families; direct matrix delta 0	Controller-code enrichment; provenance	Optional enrichment; does not improve grounding percentages	§5.8, App. F
STL/RTAMT guardrail	WADI/WD: −63 accepted, −64 rejected; +73 accepted, +74 rejected	Formal slice-level temporal guardrail	One bounded WADI→WD branch; not full verification	§5.8, App. F
EPIC/Hydro scenarios	6 direct + 2 bridge scenarios; 0.8 semantic coverage; 3/3 strong support	Power/hydro domain breadth	Scenario support only; not EPIC attack-trace replay	§5.8, App. F
PowerDuck/ GOOSE	16 IPAL files; 1,133,589 packets; flooding/insertion/replay/suppression	Protocol-level evidence complement	Protocol complement to EPIC; not target replay	§5.8, App. F
VetPLC-style obligations	10 scenarios; 2 ST programs; 1 TON timer; 6 IF conditions; 12 variables	Event/timing obligation extraction	Obligation artifacts only; no VetPLC engine reproduction	§5.8, App. F

Figure 5 illustrates the TXT→S7 manufacturing/program-artifact grounding study. This study uses a manufacturing denominator distinct from the 78-threat continuous-process grounding corpus represented in Figure 4.

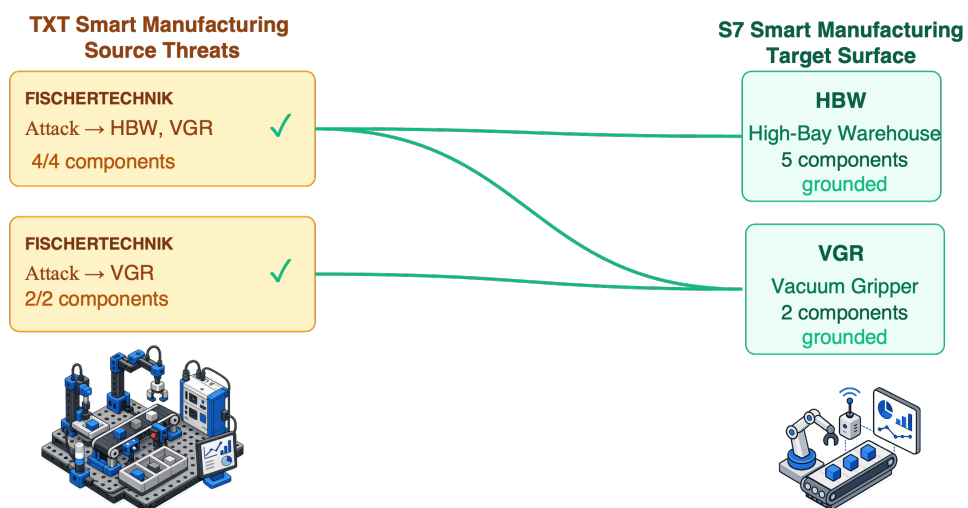


Fig. 5. TXT→S7 manufacturing/program-artifact grounding. Fischertechnik/TXT source abstractions are evaluated against a Siemens S7 target surface using the available program, tag, and state evidence. The study is reported under a manufacturing/program-artifact denominator separate from the continuous-process grounding matrix in Figure 4.

References

- [1] Syed Ghazanfar Abbas, Muslum Ozgur Ozmen, Abdullellah Alsaheel, Arslan Khan, Z. Berkay Celik, and Dongyan Xu. 2024. SAIN: Improving ICS Attack Detection Sensitivity via State-Aware Invariants. In *33rd USENIX Security Symposium (USENIX Security 24)*. USENIX Association, Philadelphia, PA, 6597–6613. <https://www.usenix.org/conference/usenixsecurity24/presentation/abbas>
- [2] Sridhar Adepu, Ferdinand Brasser, Luis Garcia, Michael Rodler, Lucas Davi, Ahmad-Reza Sadeghi, and Saman Zonouz. 2020. Control behavior integrity for distributed cyber-physical systems. In *2020 ACM/IEEE 11th International Conference on Cyber-Physical Systems (ICCPs)*. IEEE, 30–40. <https://doi.org/10.1109/ICCPs48487.2020.00011>
- [3] Sridhar Adepu, Nandha Kumar Kandasamy, and Aditya Mathur. 2018. Epic: An electric power testbed for research and training in cyber physical systems security. In *International Workshop on Security and Privacy Requirements Engineering*. Springer, 37–52.
- [4] Chuadhry Mueeb Ahmed, Venkata Reddy Palleti, and Aditya P. Mathur. 2017. WADI: A Water Distribution Testbed for Research in the Design of Secure Cyber Physical Systems. In *Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks*. 25–28. <https://doi.org/10.1145/3055366.3055375>
- [5] Md Raihan Ahmed, Jainta Paul, Levi Taiji Li, Luis Garcia, and Mu Zhang. 2025. ICSTracker: Backtracking Intrusions in Modern Industrial Control Systems. In *2025 55th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. 193–207. <https://doi.org/10.1109/DSN64029.2025.00031>
- [6] Jarmo Alanen, Joonas Linnosmaa, Timo Malm, Nikolaos Papakonstantinou, Toni Ahonen, Eetu Heikkilä, and Risto Tiusanen. 2022. Hybrid ontology for safety, security, and dependability risk assessments and Security Threat Analysis (STA) method for industrial control systems. *Reliability Engineering & System Safety* 220 (2022), 108270.
- [7] Otis Alexander, Misha Belisle, and Jacob Steele. 2020. MITRE ATT&CK for industrial control systems: Design and philosophy. *The MITRE Corporation: Bedford, MA, USA* 29 (2020).
- [8] Georgios Bakirtzis, Bryan T. Carter, Carl R. Elks, and Cody H. Fleming. 2018. A Model-Based Approach to Security Analysis for Cyber-Physical Systems. In *2018 Annual IEEE International Systems Conference*. <https://doi.org/10.1109/SYSCON.2018.8369518>
- [9] Georgios Bakirtzis, Tim Sherburne, Stephen Adams, Barry M. Horowitz, Peter A. Beling, and Cody H. Fleming. 2022. An Ontological Metamodel for Cyber-Physical System Safety, Security, and Resilience Coengineering. *Software and Systems Modeling* 21 (2022), 113–137. <https://doi.org/10.1007/s10270-021-00892-z>
- [10] Georgios Bakirtzis, Brandon J. Simon, Aidan G. Collins, Cody H. Fleming, and Carl R. Elks. 2020. Data-Driven Vulnerability Exploration for Design Phase System Analysis. *IEEE Systems Journal* 14, 4 (2020), 4864–4873. <https://doi.org/10.1109/JSYST.2019.2940145>
- [11] Bryan Carter, Stephen Adams, Georgios Bakirtzis, Tim Sherburne, Peter Beling, Barry Horowitz, and Cody Fleming. 2019. A Preliminary Design-Phase Security Methodology for Cyber-Physical Systems. *Systems* 7, 2 (2019), 21. <https://doi.org/10.3390/systems7020021>
- [12] Bryan Carter, Georgios Bakirtzis, Carl R. Elks, and Cody H. Fleming. 2018. A Systems Approach for Eliciting Mission-Centric Security Requirements. In *2018 Annual IEEE International Systems Conference*. <https://doi.org/10.1109/SYSCON.2018.8369539>
- [13] Yutong Cheng, Osama Bajaber, Saimon Amanuel Tsegai, Dawn Song, and Peng Gao. 2025. CTINexus: Automatic Cyber Threat Intelligence Knowledge Graph Construction Using Large Language Models. <https://doi.org/10.48550/arXiv.2410.21060> arXiv:2410.21060 [cs].
- [14] Alessandro Di Pinto, Younes Dragoni, and Andrea Carcano. 2018. TRITON: The first ICS cyber attack on safety instrument systems. *Proc. Black Hat USA* 2018 (2018), 1–26.
- [15] Nicolas Falliere, Liam O Murchu, and Eric Chien. 2011. W32.Stuxnet Dossier. https://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/w32_stuxnet_dossier.pdf.
- [16] Cody H. Fleming, Carl Elks, Georgios Bakirtzis, Stephen C. Adams, Bryan Carter, Peter A. Beling, and Barry Horowitz. 2021. Cyberphysical Security Through Resiliency: A Systems-Centric Approach. *Computer* 54, 6 (2021), 36–45. <https://doi.org/10.1109/MC.2020.3039491>
- [17] Luis Garcia, Jelena Mirkovic, David Balenson, Erik Kline, Yuri Pradkin, David Choffnes, Daniel Dubois, Terry Benzel, Srivatsan Ravi, Joseph Barnes, et al. 2025. SPHERE CPS Enclave: A Reconfigurable Testbed for Industrial Control System Security Experimentation. In *Proceedings of the ACM/IEEE 16th International Conference on Cyber-Physical Systems (with CPS-IoT Week 2025)*. 1–2.
- [18] Milapji Singh Gill, Tom Jeleniewski, Felix Gehlhoff, and Alexander Fay. 2025. Representing Time-Continuous Behavior of Cyber-Physical Systems in Knowledge Graphs. arXiv:2506.13773 [cs.AI] <https://arxiv.org/abs/2506.13773>
- [19] Moses Ike, Kandy Phan, Keaton Sadoski, Romuald Valme, and Wenke Lee. 2022. SCAPHY: Detecting Modern ICS Attacks by Correlating Behaviors in SCADA and PHYsical. *arXiv preprint arXiv:2211.14642* (2022). <https://arxiv.org/abs/2211.14642>
- [20] iTrust, Centre for Research in Cyber Security. 2019. CISS 2019 Dataset. <https://itrust.sutd.edu.sg/>. Critical Infrastructure Security Showdown (CISS), Singapore University of Technology and Design.
- [21] Rick A. Jones and Barry M. Horowitz. 2011. System-Aware Cyber Security. In *Proceedings of the 2011 Eighth International Conference on Information Technology: New Generations*. <https://doi.org/10.1109/ITNG.2011.158>
- [22] Rick A. Jones and Barry M. Horowitz. 2012. A System-Aware Cyber Security Architecture. *Systems Engineering* 15, 2 (2012), 225–240. <https://doi.org/10.1002/sys.21206>
- [23] Pavel Kozak, Ivo Klaban, and Tomáš Šlajs. 2023. Industroyer cyber-attacks on Ukraine’s critical infrastructure. In *2023 International Conference on Military Technologies (ICMT)*. IEEE, 1–6.

- [24] Robert Lee, Michael Assante, and Tim Conway. 2016. Analysis of the Cyber Attack on the Ukrainian Power Grid. https://www.nerc.com/pa/CI/ESISAC/Documents/E-ISAC_SANS_Ukraine_DUC_18Mar2016.pdf.
- [25] Sean Lyngaas. 2023. Federal investigators confirm multiple US water utilities hit by hackers. <https://www.cnn.com/2023/12/01/politics/us-water-utilities-hack/index.html>.
- [26] Aditya P. Mathur and Nils Ole Tippenhauer. 2016. SWaT: A Water Treatment Testbed for Research and Training on ICS Security. In *2016 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater)*. IEEE, 31–36. <https://doi.org/10.1109/CySWater.2016.7469060>
- [27] Bruno A Mozzaquatro, Raquel Melo, Carlos Agostinho, and Ricardo Jardim-Goncalves. 2016. An ontology-based security framework for decision-making in industrial systems. In *2016 4th International Conference on Model-Driven Engineering and Software Development (MODELSWARD)*. IEEE, 779–788.
- [28] Holger Neuhaus and Michael Compton. 2009. The semantic sensor network ontology. In *AGILE workshop on challenges in geospatial data harmonisation, Hannover, Germany*. 1–33.
- [29] Dejan Ničković and Tomoya Yamaguchi. 2020. RTAMT: Online robustness monitors from STL. In *International Symposium on Automated Technology for Verification and Analysis*. Springer, 564–571.
- [30] Jainta Paul, Lawrence Ponce, Mu Zhang, and Luis Garcia. 2024. Towards cross-physical-domain threat inference for industrial control system defense adaptation. In *Proceedings of the 2024 Workshop on Re-Design Industrial Control Systems with Security*. Association for Computing Machinery, 57–64. <https://doi.org/10.1145/3689930.3695210>
- [31] Frances Robles and Nicole Perlroth. 2021. ‘Dangerous Stuff’: Hackers Tried to Poison Water Supply of Florida Town. <https://www.nytimes.com/2021/02/08/us/oldsmar-florida-water-supply-hack.html>.
- [32] Burak Sahin, David Oygenblik, Mingxuan Yao, Yizhi Huang, Brendan Saltaformaggio, and Saman Zonouz. 2026. Fuzzing the Physical Space: Physics-Aware Testing of Black-Box Industrial Control Systems. In *2026 IEEE Symposium on Security and Privacy (SP)*. 1–18. <https://doi.org/10.1109/SP63933.2026.00178>
- [33] Guowei Shen, Wanling Wang, Qilin Mu, Yanhong Pu, Ya Qin, and Miao Yu. 2020. Data-Driven Cybersecurity Knowledge Graph Construction for Industrial Control System Security. *Wireless Communications and Mobile Computing* 2020, 1 (2020), 8883696.
- [34] Joe Slowik. 2018. Anatomy of an attack: Detecting and defeating crashoverride. *VB2018, October* (2018).
- [35] Muhammad Azmi Umer, Chuadhry Mujeeb Ahmed, Aditya Mathur, and Muhammad Taha Jilani. 2025. Attack Pattern Mining to Discover Hidden Threats to Industrial Control Systems. <https://doi.org/10.48550/arXiv.2508.04561> arXiv:2508.04561 [cs].
- [36] Corban Villa, Constantine Doumanidis, Hithem Lamri, Prashant Hari Narayan Rajput, and Michail Maniatakos. 2025. ICS-QUARTZ: Scan Cycle-Aware and Vendor-Agnostic Fuzzing for Industrial Control Systems. In *Proceedings 2025 Network and Distributed System Security Symposium*. Internet Society, San Diego, CA, USA. <https://doi.org/10.14722/ndss.2025.240795>
- [37] Konrad Wolsing, Eric Wagner, Luisa Lux, Klaus Wehrle, and Martin Henze. 2025. GeCos Replacing Experts: Generalizable and Comprehensible Industrial Intrusion Detection. In *34th USENIX Security Symposium (USENIX Security 25)*. USENIX Association. <https://www.usenix.org/conference/usenixsecurity25/presentation/wolsing>
- [38] Sven Zemanek, Immanuel Hacker, Konrad Wolsing, Eric Wagner, Martin Henze, and Martin Serror. 2022. PowerDuck: a GOOSE data set of cyberattacks in substations. In *Proceedings of the 15th Workshop on Cyber Security Experimentation and Test*. 49–53.
- [39] Mu Zhang, Chien-Ying Chen, Bin-Chou Kao, Yassine Qamsane, Yuru Shao, Yikai Lin, Elaine Shi, Sibin Mohan, Kira Barton, James Moyne, et al. 2019. Towards automated safety vetting of PLC code in real-world plants. In *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 522–538.