

Design Considerations for Building Credible Security Testbeds: Perspectives from Industrial Control System Use Cases

Uchenna D. Ani¹, Jeremy M. D. Watson², Benjamin Green³, Barnaby Craggs⁴, and Jason R.C. Nurse⁵

ABSTRACT

This paper presents a mapping framework for design factors and an implementation process for building credible Industrial Control Systems (ICS) security testbeds. The security and resilience of ICSs has become a critical concern to operators and governments following widely publicised cyber security events. The inability to apply conventional Information Technology security practice to ICSs further compounds challenges in adequately securing critical systems. To overcome these challenges, and do so without impacting live environments, testbeds are widely used for the exploration, development, and evaluation of security controls. However, how a testbed is designed and its attributes, can directly impact not only its viability but also its credibility. Combining systematic and thematic analysis, and the mapping of identified ICS security testbed design attributes, we propose a novel relationship map of credibility-supporting design factors (and their associated attributes) and a process implementation flow structure for ICS security testbeds. The framework and implementation process highlight the significance of demonstrating some design factors such as user/experimenter expertise, clearly defined testbed design objectives, simulation implementation approach, covered architectural components, core structural and functional characteristics covered, and evaluations to enhance confidence, trustworthiness and acceptance of ICS security testbeds as credible. These can streamline testbed requirement definition, improve design consistency and quality while reducing implementation costs.

KEYWORDS

ICS Testbeds, Security Simulations, Security Modelling, Model Credibility, Cyber Security Simulations.

Introduction

Industrial Control Systems (ICSs) are essential components of critical national infrastructures (CNIs) that control societal services, e.g. power generation, water treatment, and transport infrastructure. Whilst advances in technology have improved ICS functionality through fundamental design, setup, and operational scope, the reality is much more challenging. ICSs can have real-world deployment life cycles measured in decades, leading to outdated and insecure legacy systems running alongside modern, more secure deployments. The security of ICSs has become a growing concern owing to observed challenges and real-world cyber-attacks [1], [2]. Where ICSs form a core component of CNI, they must provide high levels of system safety, availability, security, and operational resilience. These requirements can be tied to several factors, where the impact of failure from economic, environmental, human safety, and national security perspectives would be highly detrimental [3]. In addition, testing these requirements is often impracticable in real-life ICS operational environments because of the potential disruption to process functions that can occur [3], [4].

Background

Control system security research and development (R&D) has grown to the point where replicating ICS networks through modelling and simulation (M&S) is considered a viable alternative for exploring and addressing cybersecurity challenges [5]. This is, in part, due to the high cost of deploying and using real system hardware and software for testing purposes, and the obvious risks linked to conducting research-based tests upon live ICSs or Operational Technology (OT) [6]. In M&S, a model of an actual system or problem is used rather than directly working on the real (actual) physical system. This enables the creation of a replica environment where realistic operational and security scenarios can be formed, devices emulated, protocols

CONTACT - Uchenna D Ani, Email- u.ani@ucl.ac.uk, Affiliate - Department of Science Technology Engineering and Public Policy, University College London, Shropshire House (4th Floor) 11 - 20 Capper Street, London, WC1E 6JA.

¹<https://orcid.org/0000-0001-6064-480X>

tested, methods assessed, and data collected to inform insights, and drive pre-emptive and appropriate management/control decisions and actions [7].

Emulation of ICS can be approached in various ways providing exploratory platforms upon which experimentation and training can be performed safely, avoiding socio-economic impacts associated with performance degradation [8]. This process is more technically referred to as '*Simulation*', and is widely acknowledged to be effective in experimenting, studying, analysing, and developing ICS security solution best practices [9]. Whilst the term '*Simulation*' is commonplace, often such environments can also be referred to as '*Testbeds*'. This encompasses the range of setups that include hardware-in-the-loop (HIL) as a step between '*computer simulation*' and '*physical operational hardware simulation*'. These provide test platforms for executing activities and processes as if in a real-world environment. For clarity, and for the remainder of this paper, we will use testbeds as an all-inclusive term.

Literature show that a number of ICS-related testbeds have been developed and used by researchers to analyse or address security-related challenges [3], [10]–[13]. These contributions exemplify a range of approaches to the problem [9]. A key challenge is to model and simulate real control system conditions accurately and with enough detail to support confidence in the simulated system together with its attributes and processes, thus enhancing trustworthiness and acceptability. The confidence dimension of credibility depicts a state of certainty, either about the correctness of a hypothesis or prediction, or the effectiveness of a specified course of action [14]. The trustworthiness dimension describes the perceived goodness or unbiasedness of a testbed's source or in this case the designer(s). Thus, it relates closely to expertise – skills and know-how from which a security testbed comes into form. The acceptance property depicts the recognition of the correctness and reliability of a testbed [15].

In M&S, credibility is characterised by confidence, trustworthiness, and acceptance, and are often used interchangeably [16]–[18]. Relative to M&S testbeds, credibility refers to the attribute of a testbed information and development process involving the belief of the observer/user. Thus, credibility perception is intrinsically subjective though loosely tied to the information about a testbed's derivation process so that the reliability of the process only adds to credibility if the observer/user well-understands and appreciates it and its associated limitations [19]. Consequently, to trust the credibility of a security testbed and its development process, an observer or user must also trust that the developers have the appropriate competency to apply the process and do so correctly.

The testbed constructs in most publications seem focused on specific sectors or applications but lack detailed and sufficient views about the testbed use and results. In addition, the testbed constructs appear characterised by dissimilar approaches to developing and demonstrating security M&S research [20]. These introduce uncertainties and weak arguments for the reliability of each contribution, which is exacerbated by trade-offs between obtaining '*generality*' to a broader set of ICS applications or '*specificity*' to a specific pressing operational process or application problem at hand.

In this article, we draw from prior work [21] to interpret credibility as; *how well a security testbed system and scenario (test process, inputs and outputs) are able to reflect and advance confidence, trustworthiness and acceptability as a correct representation of a real system, and suitable to use for explorative security cases*. In other words, it can be referred to as the perceived quality of testbed structure and features that persuades believability. Thus, credibility relates to the non-inclusion or non-coverage of certain model design/simulation attributes that can add to a strong reflection of a real system. For ICS security testbeds, this is important as it can impact the accurate resolution of security and safety issues in real ICS. Simulation credibility can suggest how well a system, process, component, and/or outputs reliably re(presents) the actual system. One way of achieving this is through outlining the relevant design considerations or factors that support development and use, and which can support confidence and acceptability. We have not found any work that sufficiently addresses this. This presents another challenge when developing capabilities to support research objectives and when evaluating the quality of a testbed and related research. Thus, it is difficult to make statements or to demonstrate how previous work supports or improves confidence in actual ICS scenarios. It has been acknowledged that having and following a guiding structure is crucial in proving the relevance and significance of ICS testbeds and associated areas [22].

This paper aims to address the above challenge and need by combining a systematic review of existing ICS security testbed work and a focus group-based workshop involving participants interested in security M&S to identify relevant design factors that can provide guidance on security testbed development and use. We synthesise this work and propose a novel conceptual relationship map of credibility-supporting design factors

(and their associated attributes) and a process implementation flow structure, for ICS security testbeds. This is to help advance confidence, trust, and acceptance of ICS security testbed-related studies. Starting with an outline of research efforts related to reviews of ICS security testbeds and key characteristics of each work, the paper identifies some common design factors that need to be considered in developing an ICS security testbed and scenario. It describes the relevance of the factor and their attributes to the overall testbed security M&S scenario development, and how they contribute to credibility-building. A process implementation flow structure for ICS security testbeds is also developed to provide guidance on security testbed development and use.

The mapping structure and implementation process should not be construed as strictly sequential. The succession in the map using arrows is intended to show the relationships amongst design credibility factors and sub-attributes. The significant value of the relationship map is its ability to assist testbed developers and decision-makers in identifying suitable design factors and approaches peculiar to their requirements. The process implementation flow can assist with a step-by-step guide on how the attributes in the map can be adopted towards a credible ICS testbed implementation. Together, both structures can support establishing and(or) enhancing the credibility of security-related ICS testbed work. The concepts proposed can also be applied to other security testbed development areas such as wireless sensors and computer networks.

The remainder of this paper is structured into sections covering; an overview of related review works on ICS security testbeds, a description of the research methodology applied in this study, the results and analysis of both quantitative and qualitative research conducted, the mapping structure and implementation process proposed for demonstrating ICS security testbed credibility, and the conclusion and future work.

Related Work

In this section, we reviewed similar works that had considered the nature of security testbeds in relations to ICS or critical infrastructures. We explored the basic factors and attributes that underpinned each work's analysis in relations to design requirements, and associated link to credibility-building. The testbed publications reviewed included the works that are closely aligned to ICSs and/or cybersecurity including; the Internet of Things (IoT), cyber ranges, and Cyber Physical Production Systems (CPPS). These areas were selected because of their relevance to the research objective and provide a starting point towards identifying relevant design factors for building credibility in ICS security testbeds. Highlighting existing limitations also strengthen the motivation for our research.

Candell et al. [3] detail three design/case study ICS testbeds with security scenario demonstrations. For the scenarios, design attributes are not structurally defined from the outset. Instead, design details on process descriptions, components, architecture, protocols, modelling approach, and security scenarios appear unstructured across the paper. This makes it difficult to easily and clearly recognise attributes that might come across as potential design requirements. In addition, the work has a limitation in that it fails to consider credibility-supporting factors such as '*evaluation modes and outputs*' which can be useful in building credible ICS designs/testbeds [23], [24]. Not considering evaluation modes demonstrates a lack of corroboration by parties other than the researchers/authors on the quality and credibility of testbeds or related work. Although useful, the work appears to not sufficiently support any claimed credibility by the authors.

Gluhak et al. [25] performed a technology-based review of IoT experimental testbeds. They focused on Wireless Sensor Networks (WSNs), and examined the effort required when migrating from WSNs to a global networked infrastructure of IoT. They evaluated existing IoT testbeds based on design challenges and functional characteristics including scalability, heterogeneity, repeatability, federation, concurrency, and mobility. These reflect characteristics considered valuable in contributing to testbed design credibility.

Davis and Magrath [5] surveyed cyber ranges and computer network operations (CNO) testbeds from three broad classifications: Modelling and simulation, Ad-hoc or Overlay, and Emulations. One of the key conclusions is that simulation and emulation are the most common approaches for developing security testbeds due to a reduced cost of implementation, flexibility, scalability, and capacity for easy reconfiguration.

In Siaterlis and Genge [26], a comparative study of nine ICS testbeds is presented. An analysis was conducted using a coarse scale (1-3) to rate six key operational characteristics: Fidelity, Repeatability, Measurement accuracy, Safety, Cost effectiveness and Multiple critical infrastructures, and two sub-characteristics: Cyber and Physical. Besides failing to consider other crucial attributes like Scalability, Modularity, and Flexibility, the authors compared their work against others, but failed to provide any clear bench-marking requirements for the quality evaluations.

Holm et al. [11] surveyed 30 ICS testbeds proposed for scientific research. Most of the testbeds were designed for vulnerability analysis, test of defence mechanisms, and educational purposes. Pure simulation of ICS components appeared more common than virtualisation and hardware-based approaches. Testbed fidelity was heavily emphasised. However, factors like repeatability and safe execution were not well-addressed by the surveyed testbed articles.

In Salunkhe et al. [12], a conceptual design of CPPS testbeds is presented, based on a review of prior testbeds. These were analysed based on their application sectors, i.e., electrical grid, cybersecurity, network and communications, robotics and manufacturing, IoT, Web and Cloud computing, simulation-based, and others. Results showed cybersecurity to be the dominant area of interest.

Geng et al. [13] surveyed some common ICS security testbeds developed for research purposes based on models like the Purdue Enterprise Reference Architecture. The authors emphasised that when building a multi-level, high fidelity and strong interaction ICS testbed network for research purpose, it is necessary to simulate both information interactions of the network layer and business processes of the physical layer. Besides ensuring that testbeds are designed according to research requirements and realistic capabilities, relevant design properties highlighted include fidelity, repeatability, measurement accuracy, and safe execution. The author also highlighted that; a lack of hardware-software interaction, virtualisation being unsuited for closed source proprietary components, inadequate integrity of physical process, and insufficient diversity and heterogeneity of equipment form part of ICS testbed development challenges that need to be addressed [13].

Design considerations for security testbeds is clearly a topic of interest across relevant communities and stakeholders. However, using relevant requirements as a means to address design credibility, and how this may be achieved, is currently absent in community discussions and literature. While attributes that can pass as design considerations and development factors have been discussed directly or implicitly in several work, they are fragmented. This restricts the ability to identify a broader set of essential requirements or mappings to appropriate functionalities that can support the credibility of ICS security testbed designs and associated research activities. An outline of security testbed design essentials can help to streamline existing concepts and enable a pathway for suggesting a standardised evaluation or benchmarking approach for ICS security M&S testbeds.

Methodology

Recall that this study aims to identify the relevant design factors that can support the building of credible ICS security testbeds, and to understand how the factors might be considered in testbed development works. To achieve this aim, we adopted a multi-methods approach [27] where an initial literature survey method is used on an exploratory level for discovering and identifying the common factors in ICS security testbed works that can support credibility and acceptance. In addition, a second method is used to enable cross-validation and to generalise the findings from the survey in relations to the relevant factors and design conditions that can influence the credibility and acceptance of ICS security testbeds and associated research. Similar approach has been acknowledged as helpful in other areas; for exploring the risk assessment of knowledge exposure risk associated with 3D virtual reality environments [28], and for exploring system influences on patient safety in under-researched pre-hospital settings [29]. These have yielded valuable insights with strong confidence levels on finding.

We started by surveying and identifying relevant research contributions from related work, and then design factors from which attributes can be drawn. To provide a comprehensive view of the ICS security testbed space, we opted to conduct a systematic review [30]. This began with an unstructured survey involving online Google searches, applying related titles for a period covering 2008 to 2019. This was used to select relevant keywords. These were then applied to a structured search in the SCOPUS and Web of Science databases for relevant articles (focusing on finding related keywords within each article's title, keywords, or abstract). Both databases were chosen because together they enable access to more resources with strength of a wider coverage and resource concurrency [31]. The keyword search used was as follows: 'ICS Security Testbeds', 'ICS Testbeds', 'SCADA Security Testbeds', and 'SCADA Testbeds'. Finally, articles were selected if they contained text describing any variant of ICS security-related testbed design architectures, frameworks, or implementation, as research objectives or as a tool for validating practical security research scenarios.

Secondly, a qualitative study was also conducted involving a focus group workshop. The goal was to obtain a range of views and experiences from stakeholders on the significance of building credibility in ICS security testbeds, especially from a research and development perspective. This was broken down into two precise

questions: (i) *Is credibility an issue in the development, use, and utility of security testbed models for ICS? If Yes, why?* (ii) *Design considerations that can build/strengthen credibility of ICS security simulation testbeds and the research that uses them.* The focused group workshop lasted for three hours and comprised of 16 participants with ICS security modelling interests. These were drawn from academic, policymaking, and others (i.e., participants with multiple interests e.g., both in academia and policymaking). We termed those in the 'others' category as 'mixed-interest' group. Participants were asked to provide responses to specific questions. Inclusion criteria for participants was that they had experience or interests in designing, using, or regulating contexts related to ICSs M&Ss.

Out of the 16 participants, 4 self-identified with policymaking, 6 with academia, and 6 with mixed interests. Answers were collected on boards using post-it notes for each interest group.

A thematic analysis method using Braun & Clarke's six-phase guide [32] was used to examine the data and to derive insights, as shown in Table 1. Data was collected from the sticky note responses of participant groups and analysed based on the research questions following a top-down (theoretical) theme approach. This involved combining semantic and latent-level evaluations to identify more specific patterns in group responses and exploring any underlying ideas and assumptions that may be associated with the themes.

Table 1: Thematic Analysis Process

Research Questions	Phases	Context	Description
<p>RQ1: <i>Is credibility an issue in the development, use, and utility of security testbed models for ICS? If Yes, why?</i></p> <p>RQ2: <i>Design considerations that can build/strengthen credibility of ICS security simulation testbeds and the research that use them.</i></p>	Phase 1	Familiarising with data	Made notes and jotted down early impressions on the value of credibility in ICS testbed M&Ss, and factors that can enhance credibility from participant post-it notes
	Phase 2	Generating initial codes	Coded data segments from written responses on post-it notes in order of relevance to RQs 1 and 2.
	Phase 3	Searching for themes	Examined codes and combined related codes into a single theme.
	Phase 4	Reviewing themes	Revised and grouped themes in terms of relevance to research questions
	Phase 5	Defining themes	Refined grouped themes and defined their essence and implication to study
	Phase 6	Write-up	Documented the results and interpretations based on research questions.

Results

From the systematic review, 112 articles were identified from the queried databases according to their match with applied search parameters and inclusion criteria. The relevance of each article was considered based on its title and abstract. Duplications were discarded. This left 57 articles found to contain substantial content on ICS security testbed use. These are presented Appendix A. Relevant design attributes that address ICS testbed design and security simulations were drawn from selected literature. The significance of identified attributes on credibility-building were also analysed comparatively with those obtained from thematic analysis of focus group responses.

Credibility-Supporting Design Factors

ICS security testbeds that can influence evidence-based decision-making on security policies and controls typically depend on the degree of conformity to real system that can be assured. This involves capturing some relevant and common elements and attributes that characterise a real ICS. These elements include user expertise, usage objectives, core characteristics, architectural component coverage, simulation approach, and evaluation approach as shown in Figure 1. Considering and including such factors can influence confidence in a testbed to reliably satisfy a specific intended purpose [24].

- **User/Experimenter Expertise**

From the systematic review, one of the factors that favour obtaining a representative ICS testbed system is domain knowledge [33], or rather – user expertise. As mentioned earlier, trustworthiness is a key credibility characteristic which also relates to expertise [15] or domain knowledge. In other words, to effectively evaluate and arrive at an overall assessment of testbed credibility, one does not only rely on an assessment of trustworthiness, but also of expertise of any agent involved [18]. Together, these characteristics provide superior information to persuade an understanding or perception around the state of credibility for any associated testbed. Relative to the study context, expertise or domain knowledge can refer to training or experiential competence in ICS operations (methods, constituents, and attributes) and Information/Cyber

security. A high domain knowledge can enable a more appropriate mental representations involving a deeper reflection of real concepts than low domain knowledge [34].

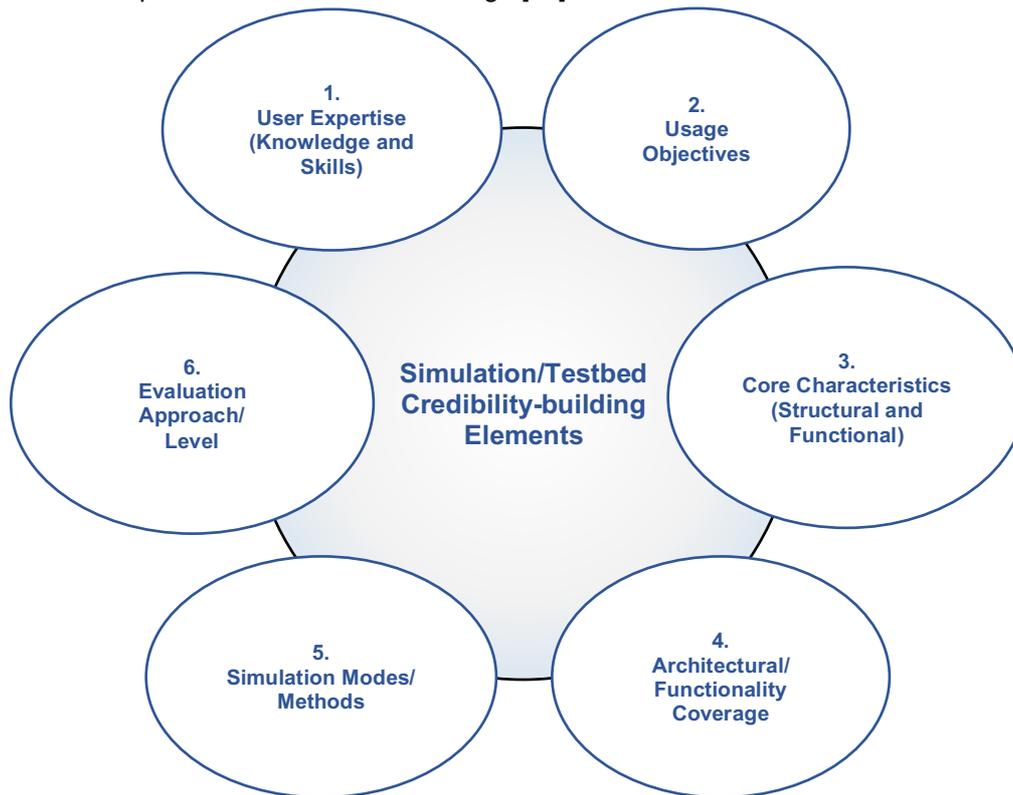


Figure 1: ICS Design Credibility-Building Themes

From the review results, ICS security testbed works that appear to capture either, a broad breadth of ICS functionality – (including all four functionality aspects), and large-scale and complex system application, with degree of physical/semi-physical M&S modes, appear to either come from academic research groups or national/international research laboratories or commissions [3], [26], [35], [36]. Often, these research formations involve a multi-disciplinary or multi-skilled team of researchers with expertise across the length of knowledge and skill specialties (e.g. ICS, IT/Cyber security, etc) relevant to achieve appropriate representations of project concepts and objectives. This characterises a concentration of domain knowledge in ICS and IT/Cyber security, which collectively enables obtaining simulation setups that mirror the nature of real ICS systems.

This also suggests that the composition of an ICS security testbed model may well depend on the degree of domain knowledge underpinning the testbed development, and the expertise of human experimenter to correctly define, design, integrate and configure testbed components, system and scenarios, documentation, evaluation and comprehension of the results of experimental security scenarios. When designing and building ICS testbeds, it is crucial and helpful from an output-reliability perspective to utilise specialists [1].

Thus, the quality of an ICS security testbed and the confidence it can instil can depend on the rigour of the operational theory underpinning the security testbed, and the expertise to properly apply the theories to logical and convincing outcomes [37]. However, the influence of expertise and knowledge in the credibility of security testbed M&S appears not to be well covered by existing research. This is despite the fact that experimenters not only choose components that can integrate well into a model of a planned system, but they also set-up and configure the components and process parameters, execute the processes, collect and analyse the results, and interpret outcomes.

People involved in ICS security testbed development have to make multiple decisions including; the security contexts and features that are important to be modelled, the appropriate M&S approach to use, the components and scale to adopt, the core characteristics relevant to the contexts adopted, and the level of evaluation necessary to validate the model. They must have, and employ, the requisite knowledge and skills to logically implement the outlined processes. The lack of enough expertise in any areas may result in errors in those aspects of the testbed development. This can cause overall quality degradation and cast doubts as to the credibility of the testbed.

- **Design/Usage Objectives**

Design considerations that are useful are driven by well-defined *usage objectives* [38]. Because of the trade-off between obtaining highly representative systems and their implementation costs, testbed design considerations and decisions on simulation need to be driven by an intended use [39]. Building an ICS testbed itself is typically not the final objective, rather, the testbed is a tool to explore and reach the greater objective, which in this context is related to security. Thus, it is crucial to be clearly aware and understand the objective(s) and associated constraints before engaging into practical design activities [40].

Design Objectives have been cited as a relevant and a key context in testbed preparation, which needs to also align with design configurations [11] as it contributes to credibility-building. Design objectives and configurations need to be articulated well ahead to provide direction and scope for the development process, as well as to support functional validity and credibility. For simulation testbeds, applicable objectives need to be well clarified, since a design setup can be valid for one objective but not for another [21].

- **Testbed Core Characteristics**

Security testbed reliability can also be supported by demonstrating certain *Core Operational Characteristics* [41], [42] that underpin the structure and operation of a testbed. The core characteristics can take structural and functional dimensions [43], [44]. These comprise of behavioural attributes that are expressed in testbed operations such as the ability to; reflect the real nature of a system (fidelity), add or remove components or test scenarios (modularity), and log status of test scenarios (monitoring and logging). These also cover attributes that refer to testbed performance indicators. These include the ability to; easily use the testbed (usability), adapt it to new applications or scenarios (adaptability), and be open to improvements and modifications (scalability). These features are normally off-shoots of functional features [45]. The relevance of these core operational characteristics in supporting testbed credibility has been acknowledged [11], [43], [44]. Thus, demonstrating these attributes within a simulation testbed design adds some assurances that can advance trustworthiness and acceptability of the testbed and associated research.

- **Testbed Architectural Component Coverage**

The credibility of ICS security testbeds can also be influenced by their *Architecture Components Coverage* [41], [42]. This refers to the common functionality coverage in an ICS setup comprising any combinations of the ICS functional areas; (i) Physical Process (PP), (ii) Field Devices (FD), (iii) Communications Gateway (CG), and (iv) Control Centre (CC) [3], [35].

PP functionality category is a part of the operational technology components that help to characterise physical process. This includes hardware machines, actuators, sensors and I/O ports connected to physical equipment to accomplish real input and output processes. The FD category of components refer to other operational technology infrastructures such as, remote terminal units (RTUs), programmable logic controllers (PLCs) and other control maintenance components used to automate control functions through the network. The CG category refer to local/wide area network components, routers, switches, satellite, etc, through which communication exchange is facilitated amongst networked components. Often, this is the bridge that allows OT infrastructures to communicate with IT components such as mail and file servers, messaging, and archival systems. The CC functionality category refer to industrial components that handle control initiations and monitoring such as Master terminal units (MTU), Human-machine-interface (HMI) device, engineering workstations, historians, including safety instrumented devices [46].

Architectural component coverage also covers aspects of communications protocols; consisting of either IP routable (e.g. Modbus TCP) and(or) IP non-routable (e.g., DeviceNet) protocols [3]. Incorporating more aspects of the basic architecture components and associated communication protocols helps to clarify security issues related to specific components and their implications across the entire ICS network. A broader coverage of components within a common architecture can enable the simulation of wider contexts and enable better realism of an ICS from architectural perspectives. These can also support attaining a more holistic expression of security tests, and insights into the entire system impacts. They lend credence to the resulting testbed and the research that uses it.

- **Testbed Simulation Approach**

The *Simulation Approach* adopted for a testbed also contributes to its perceived reliability [41], [42]. This refers to the structural and procedural formation of the components that constitute a simulation system testbed. Broadly, this can be classified into three: (i) Physical Simulation (PS) – involving purely real infrastructure

components, (ii) Semi-Physical Simulations (SPS), sometimes referred to as 'Hardware-in-the-Loop' (HIL) [47] – involving a combination of real, emulated and/or virtualised abstractions of ICS components (i.e., a mix of Emulation and implementation-based approaches), and (iii) Software-based Simulations (SBS) – involving the simulation of components on a single, purely software platform. Other terms for these categories include real system (hardware and software), computer emulations or virtualisation (including hardware-in-the-loop), pure software-based simulations [11] or live, virtual, and constructive (LVC) simulations [48], [49].

Real or live simulation involves actual/real-world control system components operating on/with actual/real-world ICS set-up and protocols. Despite using real components, this is considered a simulation because cyber-attack processes and scenarios are simulated, and not truly conducted against any live target adversary control system [49]. An example includes using actual operators, actual network devices, actual components, and actual non-emulated/simulated software. *Emulated or virtual* simulation involves actual ICS components interacting with limited or representative ICS system models and vice versa.

A *representative* simulation model is one that offers the operationally relevant partial or complete interactive interfaces, protocols, and features of the actual component or system. A simulation model is said to be *limited* if part of its components does not provide the relevant interactive interfaces, protocols, of the actual component. Examples include; having the emulators of components such as a PLC running on a virtual machine or replaying a logged real-life attack onto virtual or live systems. *Pure software-based or constructive* simulation approach involves the models of limited or representative components interacting with limited or representative system models. A typical example is simulating internet-scale traffic generation and background noise [49].

The choice of M&S approach can be influenced by factors including; the experience or expertise of humans involved [50], the desired degree of representation or capability of an actual system [51], the cost of development, and the budgeted development time [41]. In particular, the expertise of the human developer can affect how, and the degree of detail captured in a simulation testbed. Physical, real or live simulations typically enable the most representation of real system and data fidelity and is more likely to be credible than the other two approaches.

- **Testbed Evaluation Approach**

An ICS testbed's *Evaluation Process* can also influence design quality and credibility [23]. This refers to the procedures through which assessments are performed to determine how well a testbed's design or related outputs are correct, and(or) acceptable. The purpose of testbed evaluation is to demonstrate with appropriate evidence that a testbed set-up and its scenario results fit the use intended, and do not present any intolerable risks. Having such evidential information can support well-informed and confident decisions throughout a security M&S life cycle [19]. ICS testbed evaluation helps to clarify on the correctness of a testbed set-up, and its usefulness in addressing real-world industrial system needs.

We believe that to build or enhance credibility, simulation testbeds, scenarios, data, and results should rely on suitable evaluations. These should help demonstrate the fulfilment of relevant reliability factors including design objectives, structural, behavioural, and performance characteristics in line with intended use. Such evidence can exist in varied degrees, supporting a scale of credibility and acceptance. While evaluation proofs may be offered by testbed authors, they may be better accepted when coming from other sources, e.g., independent reviewers. However, the best evidences of credibility are likely to come from public institutions, standardisation or certification bodies such as National Institute of Standards and Technology (NIST), the UK's National Physical Laboratory (NPL) and The Institute of Engineering and Technology (IET).

The three contexts of testbed M&S evaluation described can be more technically termed as: *verification*, *validation*, and *accreditation* [52]. Verification describes the process of clarifying that an ICS testbed model implementation and its associated data correctly represent the developer's specifications. Validation defines the process of determining the degree to which an ICS testbed model and its associated data provide a correct representation of the real-world ICS system from the perspective of the intended uses of the testbed. Accreditation describes the official certification that a testbed simulation model or a federation of testbed and its associated data is acceptable for use for a specific purpose [53]. Each of the evaluation categories seek to answer a unique question that captures a specific testbed simulation idea. Verification answers; *Was the testbed built right?* Validation answers; *Was the right testbed built?* Accreditation answers; *Is the built testbed believable enough to be used?* That an evaluation process can transition from verification to validation and then to accreditation reflects an incremental appraisal process, whose results can provide stronger evidence(s)

and ground(s) to persuade confidence, belief, trustworthiness, and acceptability of the testbed simulation outputs.

Arguably, persuading credibility in ICS security testbeds and associated research can involve demonstrating that component setups, functional and application approaches, experimental processes and results are clear and sound. Typically, these include maintaining a set of associated documentation including records demonstrating that testbed systems conform to design goals, architecture components, functionality set-ups and applications, experimental scenarios and measurement outcomes, and evaluation procedures, as applicable in a real-world context. The broad set of documentation should cover context breakdowns of testbed process descriptions including process model schematics, protocols, logical architectures (zones and enclaves), physical architectures, control strategies and parameters, module and component descriptions, assembly details, measurement data collections, evaluation metrics (security and operational).

Relevant testbed details need to also include the security requirement descriptions following prescribed/guiding security standards, such as the standard series of ISA99, Industrial Automation and Control System Security [54], or similar contexts in NIST 800-82 [46], and NIST Advanced Manufacturing series 200-1 [55]. The intention is to help users/experimenters become familiar with the technologies, test and evaluation processes involved, and to serve as reference manual. Indicating the qualification and expertise of the human experimenter can support confidence and acceptance in the associated research and its outputs. This can also contribute to improved rigour in supporting decision makers' assessments.

Quantitative analysis of reviewed works on credibility-supporting design factors for ICS security testbeds

This section presents a quantitative analysis of existing literature that bothered ICS security testbed research and the recognition of relevant credibility-factors therein. In analysing the semantic description of security design objectives, eight broad themes have emerged. These include: Threat Analysis, Vulnerability Analysis, Attack Analysis, Impact Analysis, Defence Mechanism Test/Analysis, Education and Training, Creation of Policies and(or) Standards, and Performance/Quality of Service Analysis. Their occurrence across the reviewed work is summarised in Figure 2. 'Attack Analysis' (68.42%) and 'Defence Mechanism Tests/Analysis' (59.65%) are the two most common objectives for designing and using ICS security testbeds for research. These are closely followed by usage objectives related to the *analysis of attack impacts on the systems* with 43.86% of works. Other dominant design objectives for ICS security testbeds include: Vulnerability Analysis, and Education & Training.

For core structural/functional characteristics, a total of fourteen distinct key operational characteristics were found across the projects as shown in Figure 3. It showed that 71.93% of the reviewed works contained statements and descriptions that emphasise the relevance one or more of the fourteen core ICS testbed structural/functional characteristics. Researchers more commonly emphasised on *fidelity* (45.61%) than the other characteristics. Other highly emphasised characteristics desirable for ICS security testbeds included *flexibility/adaptability* with 26.32% proportion, *scalability/extensibility/Reusability* with 24.56% proportion of reference work, and *repeatability/reproducibility* characteristics with 21.05%, with the remaining characteristics having fewer reference results.

From Figure 4, analysis of the adoption of simulation approaches showed that 19.30% of studied work combined Software-based and Semi-physical (emulation, virtualisation, or HIL) to realise the desired ICS security testbeds systems, processes, and tests. A similar proportion (19.30%) also used purely physical approach – real hardware and software as applicable in actual ICS systems. 15.79% combined elements of all three methods: real hardware and software with virtual/emulated parts and software-defined components (SBS + SPS + PS) – combined. The same proportion (15.79%) also used solely software-based simulation, and another also combined Semi-physical (emulation, virtualisation, or HIL) with Physical (Hardware or Software) methods respectively. A smaller proportion (14.29%) of works used a form of Semi-physical method alone – involving either emulation, virtualisation, or HIL techniques. Often, the works in this category were found to only simulate just a part, not all of the functionality setup or components in an ICS architecture.

For architecture components coverage, results in Figure 5 show that more than half (52.63%) of the reviewed works defined and(or) adopted design component structures that covered all four (field devices/process (FD), physical process (PP), communications gateway components/processes (CG), and control centre components/processes (CC)) broad functional areas of ICS described earlier. Between 3 and 14 works covered a combination of three ICS function areas. On average, 3 works covered a combination of

two function areas. CGs appear to be the ICS function area most explored, with a coverage of 96.49%. This is followed by CC components (92.98%) and PP components (84.21%).

Concerning evaluation processes, results in Figure 6 show that up to half (50.88%) of the reviewed works lacked information relating to any form of evaluation to verify, validate or accredit their works. A third (33.33%) of works mentioned evaluation approaches that relate to design/scenario comparisons against either; common standards and reference model documentation, prior works done by same authors or others, or certain real ICS system setups. These may pass as a validation process that dwells or abstracts from ideas or concepts from independent parties. Examples of standards referenced include: NIST SP 800-82 R2 [46], PERA Reference Model [56], and IEC 60870-5-104 TCP/IP Communications [57]. A little as 10.52% of works used a verification approach – showing at completion that their works satisfied prescribed design objectives. No work demonstrated any form of accreditation, neither did any project demonstrate evaluation to a level suitable for accreditation.

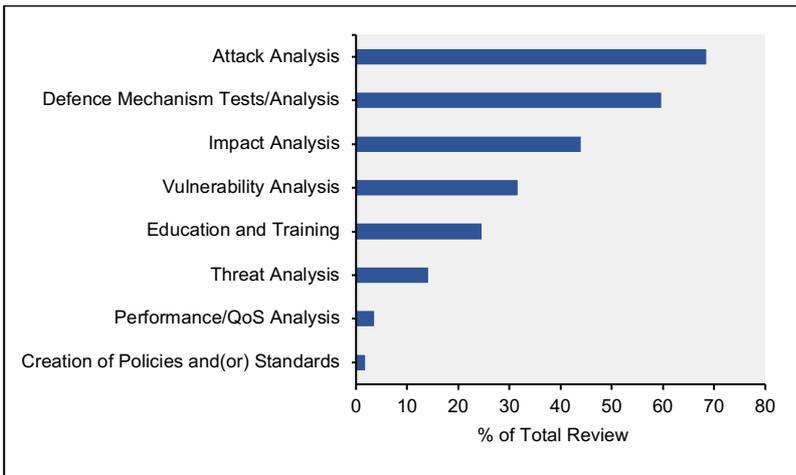


Figure 2: Analysis of ICS Security-Related Design Objectives

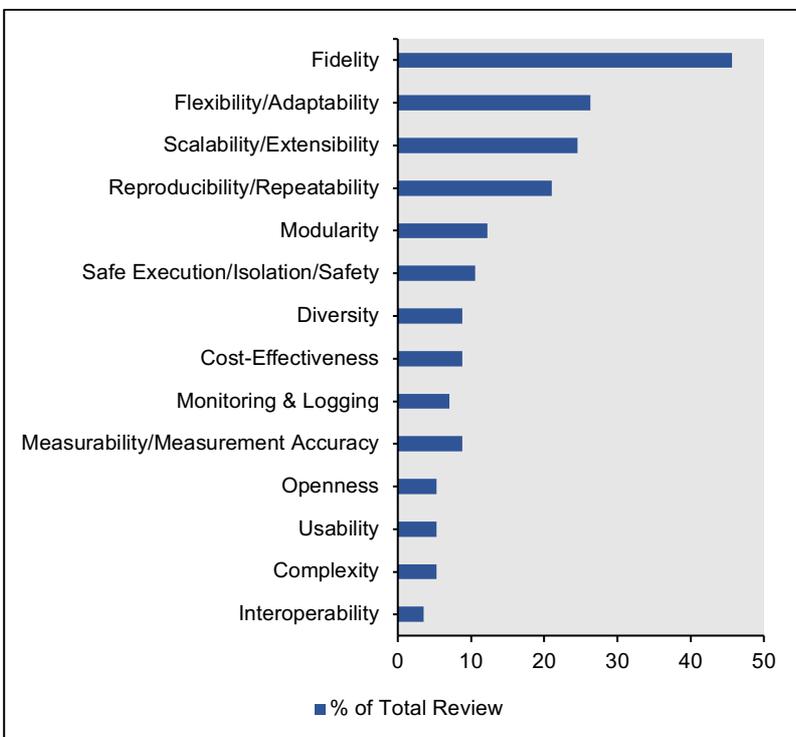


Figure 3: Analysis of ICS Security-Related Testbed Core Characteristics

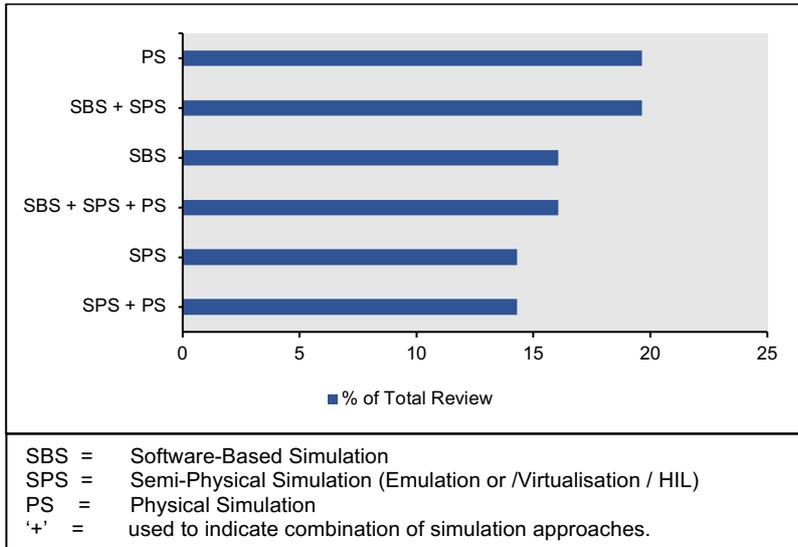


Figure 4: Analysis of ICS Security-Related Design/Simulation Approach

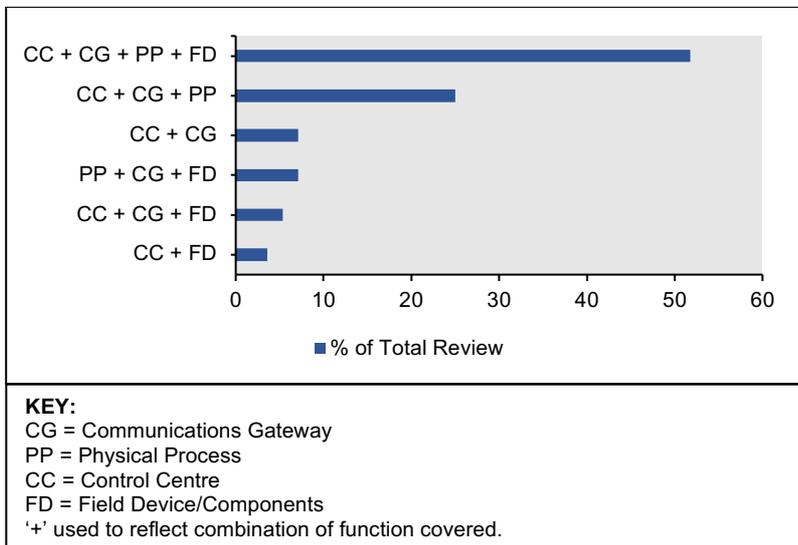


Figure 5: Analysis of ICS Security-Related Design/Simulation Coverage

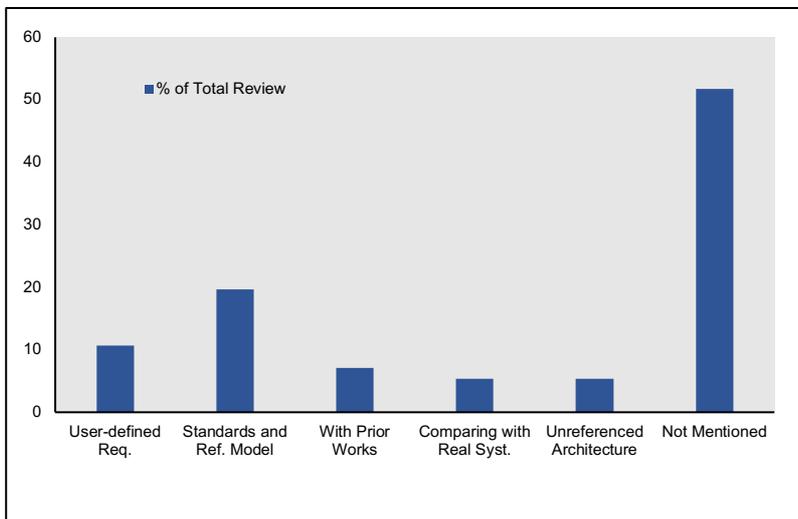


Figure 6: Analysis of ICS Security-Related Testbed Evaluation Approach

Qualitative Analysis of Focus Group on credibility-supporting design factors for ICS security testbed

This section presents the results of thematic analysis of the responses and feedback from the focus group workshop.

For the first questions (RQ1) on whether credibility is an issue in the development, use, and utility of security testbed models for ICS, we find the responses from participants to be unanimously affirmative. There is common agreement that demonstrating credibility is crucial in ICS security testbed-related work. It was common view that clear guidelines on how credibility may be built in ICS security M&S are currently lacking, and not being emphasised enough to command attention and response of testbed security experiment developers

To support the opinion for the importance of demonstrating credibility in the development, use, and utility of security testbed models for ICS, participants identified *“building or enhancing trust”*, and *“supporting real-world applications”* as key drivers for engagement. There is a need to trust and accept as reliable, the design, structure and process implementations of an ICS security testbed, the research that uses it, and any associated results. Being able to depend on the testbed to demonstrate the functionalities and processes expected in real system domains is also vital.

The capacity to *‘enable analysis’* also resonated as a common theme that was emphasised during focus group discussions. Analysis dimensions highlighted in this regard include: *“behavioural impact analysis”*, *“accident impact analysis”*, and *“modular-based analysis”*. These highlight capability benefits that can be gained from using ICS testbeds for security analysis. Thus, the criticality of ICSs to societal function supports the need to ensure a significant degree of certainty and accuracy about any analysis context engaged. The mentioned analysis dimensions can also pass as potential **design objectives** for a security testbed and were found useful to support the context being studied.

On the second question (RQ2) on the design considerations that can build/strengthen the credibility of ICS security simulation testbeds and the research that use them, disparate feedbacks were aggregated from the three stakeholder categories. Policymakers identified available institutional resource capability (cash and skills), demonstrating a shared/cascaded development responsibility, design interoperability and flexibility. Interoperability resonated in the points from the academic group alongside ‘demonstrating object-oriented scenario setup’, ‘capturing system layer-based simulation’, and ‘ability to simulate automated load, failure handling, and decision-making’. The Mixed Interest group also acknowledge the importance of experts (knowledge and skills) along with an ability to replicate real world scenarios.

The responses from participants were aggregated and harmonised into similar themes following the analysis steps outlined in Table 1. The resulting common themes expressed attributes related to testbed M&S *design, process, structure, organisation, application, and capability*. For example, responses such as, *“capturing system layer-based simulation”*, *“including computational infrastructure”*, and *“including object-oriented scenario setup”* related to design factors. Concerning capability factors, one response read *“expert opinion is important”*. Responses related to structural factors include: design *“flexibility”*, *“design interoperability”*, and *“design fidelity”*.

Discussion on Enabling Credibility Factors in ICS Security Testbeds

It is found from the combined review of existing work that the following factors contribute to the trustworthiness of ICS security testbeds and(or) associated research: (i) clearly defined security-related design objectives and security scenarios; (ii) the type(s) of simulation approach(es) and the degree of abstractions involved; (iii) the scope of architecture components covered; (iv) the reflection of core characteristics; and (v) the testbed evaluation methods. The recurrence of these attributes in literature gives an idea of their relevance too.

For most of the ICS testbed work studied, not involving evaluation to a level that can support accreditation supports the argument that insufficient emphasis has been given to the significance of building credibility. The lack of a form of evaluation characterises works in this area, which is probably influenced by the significant lack of emphasis in existing standards and best practice guidelines. It seems that researchers and experts do not perceive the need to address such contexts and attributes in their work as necessities to demonstrating quality and stimulating acceptance.

This leaves experimenters short in addressing open questions on how to prove that associated testbeds produce claimed objective parameters and/or valid results, without showing how the verification and/or validation was achieved. Often, since these concerns do not appear to be raised publicly or by industry bodies, they are often swept aside or ignored. Although testbeds tend to have more documentation that serves as a

reference guide enabling users to familiarise themselves with the technologies, understanding key contexts that support credibility is crucial to guide researchers and experimenters, since typically user documentation is not of a standard fit for publication. This is perhaps a limitation of existing work, which exposes researchers to the risks of missing valuable information that could underpin credible designs and experiments, particularly those that contribute to repeatability and measurability. A possible solution lies in identifying the factors that are most important to specific sectors and/or applications. This should be explored while considering the trade-off between specificity and generality of a testbed's purpose and following a structured approach in selecting and implementing attributes that can inform the credibility of the setup and/or associated research.

Thematic analysis suggests a strong emphasis on *impact*, evidenced by authors' concerns and focus on demonstrating and learning from negative impacts before they happen, and the quest to achieve resilience. Losses that arise from compromising impacts and the need to reduce or completely avoid system consequences is an associated reason for emphasis. Although significant, these contexts represent just one out of the range of potential design objectives or benefits of the testbed security modelling approach.

There are overlapping views between the themes and codes in thematic analysis and the factors identified from systematic study. Although the terms used to describe contexts appear to vary in both views, the semantics point in a similar direction. Results from the time-constrained focus group are not as detailed and encompassing as those from the systematic study. However, the data available still demonstrate common ideas. For example, response codes under *capability factors* can be linked to *human expertise* in knowledge and skills. Indeed, the opinion of experts depends on their knowledge and skills in the context considered. This can in turn affect the potential credibility level. Responses under *design factors* can be related to both *testbed design objectives*, while combining the responses under *design*, *process*, and *organisational factors* point to *architectural design attributes*. Responses under *structural factors* can link to *core operational characteristics*. Thus, there is a degree of coherence between the two perspectives concerning the perceived factors that contribute to building or enhancing credibility in ICS testbed security modelling.

Mapping Credibility Characteristics to Security Testbed Configurations

We have developed a novel mapping structure that outlines the relationships between testbed design factors and demonstrate how the identified design factors co-relate to support credibility. As shown in Figure 7, considering these elements and features together can greatly support building a compelling argument about ICS security testbed design, setup, and its use and utility for security analysis. Such a narrative can provide a wider understanding of an ICS testbed's composition, functionalities, abstraction, simplifications, assumptions (where available) and test/experimental results, underlining the need for a reliable representation of the real system being modelled or analysed.

- **User/Experimenter Competence (Knowledge and Expertise)**

Clearly domain knowledge is valuable in building credibility of simulation setups [33]. The degree of knowledge and expertise available to the experimenters influences the quality of decisions made concerning the realisation of a security testbed, spanning defining design objectives through to evaluating their implementation[34]. From the design stage of ICS testbeds, it is crucial to ensure that selected and purchased components matches defined requirements, and able to support desired functional/operational objectives with high degree of realism. Appropriate specialist skills (e.g., in engineering and control) beyond computer science remits is required to correctly connect ICS-OT devices [1].

Resource requirements for security testbed M&S, e.g., experimentation time, budget, and available technology infrastructure, can also influence the parameters of choice, and the level of fidelity achievable. These requirements make it difficult to achieve a generic testbed setup for a span of skills categories, especially for low-skilled users. Documentation that clarifies the context and appropriate level of user expertise can inform confidence in, and the reliability of inferences drawn from experiments. The likelihood is that testbed research from more experienced researchers will potentially provide more depth of analysis and give great confidence in reliability [37].

The knowledge and experience of testbed developers/users can help to identify core characteristics that need to be captured in specific testbed modelling contexts and scenarios. Once identified, these core characteristics can help guide the characterisation of design objectives and define a range of scenarios to explore. Consequently, relevant metrics and measurement approaches can also be determined. Expertise also informs the appropriate design architecture, components, and simulation methods (including associated

sub-attributes). The knowledge and experience of experimenters also contributes to the level of evaluation that can be undertaken.

As a minimum, modern ICS Security testbed experimenters require expertise in both ICS and IT systems development, together with an ability to: adopt appropriate modelling approach, techniques and tools, configure test applications, execute test scenarios, collect and interpret results [1], [58]. These steps require experience as they are susceptible to errors arising from insufficient knowledge and skills. One way is to engage external professional expertise in areas where researchers have limited experience and aptitude, by employing experts in ICS technology development to handle ICS operations design and implementation, while security researchers focus on the security-related experimentation. Demonstrating the involvement of a multi-disciplinary team of experts in ICS and IT security demonstrates better chances of minimising errors or limitations in the accuracy towards appropriate and credible security testbed design models and test scenarios. These can support task handling based on specialisation, supporting depth and quality of task development which can persuade confidence in the results obtained.

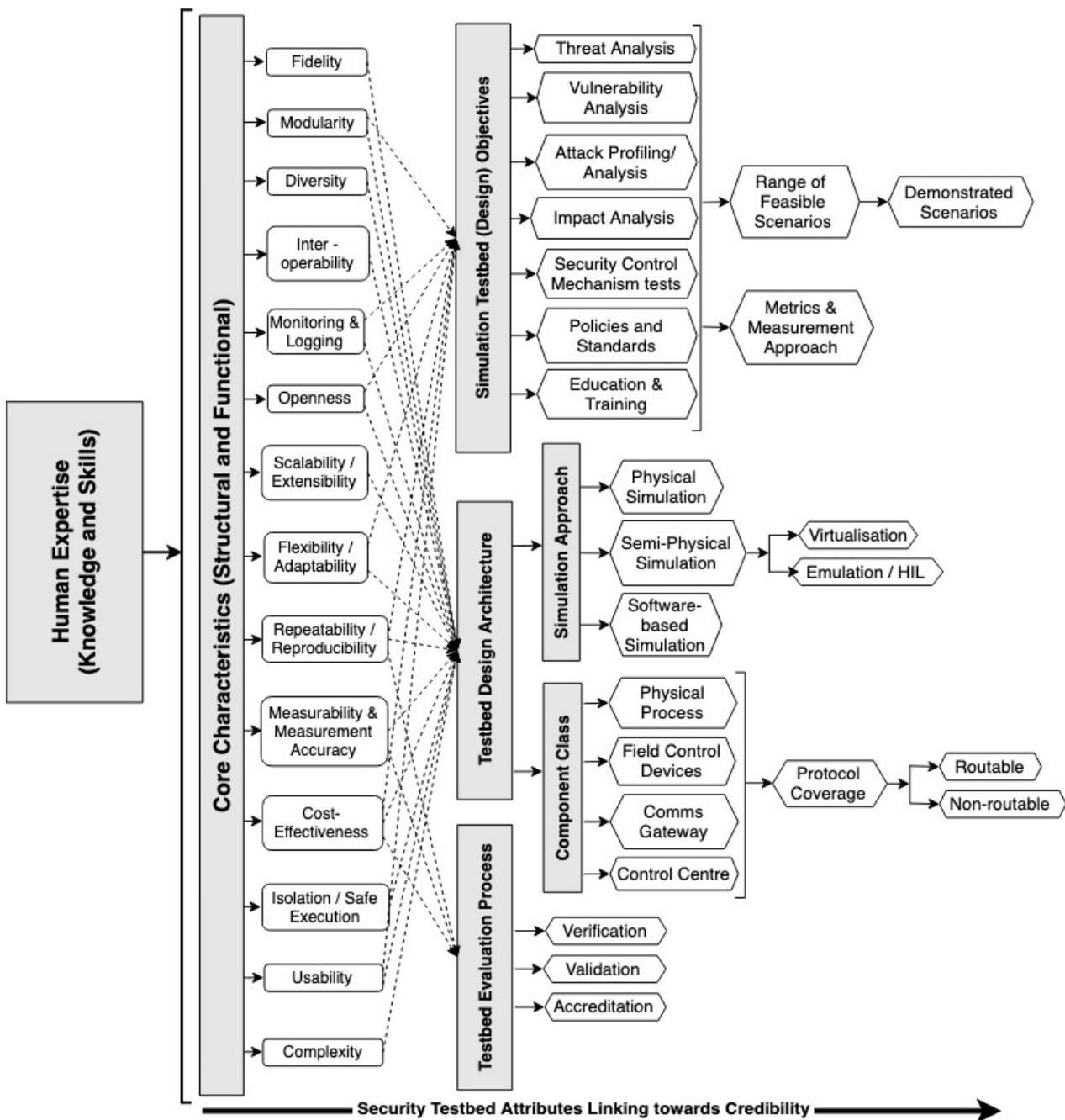


Figure 7: Mapping Structure for ICS Testbed Credibility Demonstration

- **Core characteristics (Structural and Functional)**

We believe that the core characteristics outlined are important as they individually contribute to measures that help establish or advance credibility. It seems that there are characteristics that contribute more than others to overall credibility. The ranking of importance can vary across functionalities and application domains, and often based on targeted usage objectives. However, results from analysis and occurrence frequencies of characteristics can provide suggestions on how the relevance of these characteristics is viewed by the security simulation design community. The number also provides a way of potentially ranking characteristics. For example, the requirement for demonstrating the *fidelity* of a simulation testbed and/or its use seem of greatest significance. This is apparently followed (in order) by *scalability (extensibility)*, *flexibility (adaptability or controllability)*, *repeatability (reproducibility)*, *modularity*, *measurability/measurement accuracy*, *cost-effectiveness*, *safe execution/isolation*, *diversity*, and *usability*. To build credibility, it is important for ICS testbed system and associated research to consider and demonstrate these characteristics. Evidencing as many as possible of these characteristics improves the confidence of decision-makers and other stakeholders to consider and accept the results of testbed designs, thereby improving their value.

Fidelity refers to the degree of correlation between security simulation or test predictions and real world observations [1], [11], [22], [59]. It quantifies the degree of representativeness between a testbed setup and an actual system, in terms of tools (hardware and software technologies), functionalities and tasks. The degree of fidelity can typically be determined by the simulation approach adopted – either software-based simulation (SBS), semi-physical simulation (SPS), physical simulation (PS), or combinations of these. Physical simulation is typically considered to have the highest fidelity while purely software-based simulations, the least [48].

Scalability or Extensibility refers to the characteristic to grow the size of a testbed setup (network) and functionality [22], [48], [60]. This can be demonstrated by the ability to add or migrate components (e.g., sensors & actuators) to existing operational testbed subsystems, thereby increasing capabilities or functionality without significant re-organisation or re-design. Examples of how this may be achieved are demonstrated thus; for software-based simulation approaches using *SciLab simulators* to add Field Devices; for semi-physical simulation techniques using *Virtual Machines* to emulate Control Centre components; and for physical simulations – using *real subsystems such PLCs* as Field Devices [1], [22].

Flexibility or Adaptability describes the ability to easily and swiftly re-define and repurpose a simulation system and setup for alternative use cases [35], [48], [61]. It can also be viewed as ‘*controllability*’ – emphasising the ability and extent to enable the control of environment variables. And of course, to be controllable, a systems needs to be observable [6]. This can be theoretically expressed in design/simulation objectives and practically demonstrated in design/simulation architecture. For example, an ability to show that a simulation system initially purposed for security vulnerability analysis can be easily re-structured to perform security impact analysis. This attribute can promote continuous innovation with integrated knowledge and with accumulated testbed design experiences [62].

Repeatability or Reproducibility refers to the characteristic whereby similar outputs/outcomes are obtained from identically replicated designs/testbed setups. Exact copies of designs and testbed setups or security test scenarios should produce identical or statistically consistent results [11], [47]. One way this characteristic can be obtained is through full documentation of design and process configurations, as well as security scenarios [1], [22], [47]. Other researchers can thereby obtain consistent results by applying the same configurations to directly recreate and test scenarios.

Modularity describes a design capability that allows easy adaption to changing requirements, including complexities and flexibility in industrial operations [60], [63], [64]. It involves developing ICS testbed structures that can accommodate continuous improvements. It is typified by a design that can accommodate real components, emulated nodes, and network simulators (data traffic) such as the OPNET modeller, which can enable a typical system-in-the-loop (SITL) capability [65]–[67]. Such design concepts and provisions can improve system understanding, reduce complexity, increase flexibility, and facilitate the reuse of components [68]. Implementing modularity can provide a structured approach and an action path that realises, through validated module re-use, incremental credibility of a security simulation testbed with respect to environment, data, and results.

Cost-effectiveness is a property that relates to achieving testbed design objectives and scenarios within financial budgets that are affordable for research purposes [11], [69]. The emphasis is on using smaller budgets/costs (in comparison to actual system costs) to achieve the same design objectives and scenarios (including architectural setups and configurations) as the real. This can be achieved through setups that

simulate numerous components and services consolidated into a single portable testbed system [67]. For example, using virtual machines and other virtual infrastructures to emulate control workstations, servers and other ICS components [67], [70], which can result in a cost-efficient alternative to using real and proprietary hardware workstation and server systems. This is typically subjective and varies across projects, depending on budgetary availabilities and research/test requirements. Often, a trade-off and balance is required between the cost of constructing testbeds and the fidelity of the system, and the decision that needs to be informed [69].

Measurability and Measurement Accuracy describes the ability to ensure that the process of testing or replicating cyber security scenarios via testbeds can be quantified, and that such measurements do not interfere with corresponding outputs [11], [59]. This can be demonstrated by including tools (e.g. sensors) or features for verifying attributes like traffic flows and response times amongst components. The capability to show and document data and values associated with these features also needs to be demonstrated [26], [69] typically at the documentation stages of an evaluation procedure (e.g. verification).

Safe execution or Isolation of tests describes a characteristic that ensures cyber security scenarios and activities are performed in a secure and isolated approach and environment, such that they do not increase risk or impact safety in the real environment [71]. This can easily be demonstrated using network segmentation approaches [36] to separate plant networks from enterprise networks and processes. The use of access control policies at various network layers is another approach typically implemented at communication gateway components such as access point devices [71].

Usability refers to the ability for a testbed to be readily employed by reasonably skilled operators, with little likelihood of simulation misuse [11], [48]. This is essential to cope with different skill sets of potential users. Usability can be demonstrated through adopting design and developing structures using components that enable human-centred user interfaces [72]. For example, using virtualisation and VLANs to enable the easy integration of testbed components in the CG section of ICS architectures [1], [22]. Usability consideration is crucial in ICS security testbed-related work as the absence of it can lead to user/experimenter frustrations and reduced productivity [73]. In the context of security, poor or the lack of usability can cause confusion, frustration, and an inaccurate or inadequate configuration of security scenarios, tools, and functionality. This can lead to users/experimenters undermining correct security attributes, and eventually limiting system/model effectiveness [74]. One way of ensuring usability is by following common usability guidelines such as International Standards Organisation published guide for creating usable user interfaces [75] or adopting appropriate and effective design patterns such as Tidwell's effective interaction design patterns [76], or the more contextual usable cybersecurity guidelines by Nurse et al [77].

Diversity refers to the ability of an ICS testbed to incorporate a varied range of components without undermining the capacity for scalability as discussed earlier. An effective testbed needs to be able to mirror a variety of ICS setups [1], [22], [78]. This includes demonstrating where feasible and necessary; market-driven heterogeneity in components (vendor products e.g., Siemens, Allen Bradley, Schneider), protocols (e.g., TCP, UDP, OPC, Modbus/TCP, DNP3, EthernetIP) and processes (e.g., manufacturing, assembly, traffic control, water treatment) employed in a testbed setup. This can provide valuable ICS security insights from legacy, contemporary, and future outlooks, and deployments that reflect industrial practices, enabling a variety of experimental setups and scenarios. These can help advance system credibility for practical applications [35].

Interoperability refers to the ability of combinations of SBS, SPS, and PS testbed simulation components to interface, communicate, exchange and use information to achieve desired objectives. This can typically be demonstrated in the development of hybrid ICS testbeds and security experiments involving PS components such as Control workstations that connect to IED interfaced with SPS techniques such as virtual machine servers and virtual communication components [67].

Monitoring & Logging describes the ability to observe and record process execution and to optimise event logging for security purposes [1], [22]. One way of achieving this is through implementing a measurement enclave with Syslog tools and traffic monitoring systems to keep track of operational activities [35]. This can be better achieved through automated granular data flows – understanding data sources and pathways to help resolve undesirable impacts on process functionality [79].

Complexity and Openness describe two related attributes identified as valuable in modern ICS testbed designs. *Openness* defines the capability of a testbed simulation setup to support remote access or data openness [1], [22]. While *complexity* ensures that architectures are represented in a transparent manner such that a single point of data access or extraction can be enabled from different network zones or segments of the ICS [1], [22], [78]. Complexity may be achieved by demonstrating the coverage of a wider scope of

operations to include depth of details, for example, demonstrating the M&S of interdependency relationships amongst components/functionality groups in the design architecture, as well as multiple impacts contexts including, functional, economic, physical impacts. Openness can be demonstrated by the use of open and generic file formats or open/common technology standards and testbed parameter settings [7] in both design architecture and simulation approaches.

As shown in Figure 7, the mapping can be quite complex with a range of one-to-many connections that may not be clear at first glance. Contextual description may be needed to clarify the specific attributes involved as presented in section 5.1. Most (14 arrows) of the characteristics appear to map to the '*simulation design architecture*' factor and associated attributes, compared with '*design objectives*' and '*testbed evaluation process*' which had fewer (7 and 3 arrows connections respectively) mappings. This suggests that the most significant task and proportion of effort for establishing credibility in ICS security testbeds and associated research lies in simulation design architecture. This is exemplified by the simulation approach adopted and the design components and functionalities (hardware and software) employed. The design architecture needs to be carefully considered to ensure capture of the necessary credibility-supporting characteristics and backed with sufficient evaluation processes to maximise system trustworthiness as being representative of the real world.

- **Design Objectives**

Design objectives are a vital consideration as pointed out in Section 4.1, since ICS security testbed design architecture, attributes and decisions must be driven by usage intentions. A majority of ICS security-related research activities seem to focus on investigating cyber-attack feasibilities, the capability of security controls and defence mechanisms, and the analysis of attack impacts, instanced by successful attacks or failed security controls.

However, it is also important to understand the existence and nature of vulnerabilities in industrial control systems and components. Often this builds on the assumption that vulnerabilities nearly always exist in ICSs, since they are by default presumed to lack security. This means that experimenters typically focus on understanding random attack modes, often through penetration testing, and on ascertaining robustness against specific attacks given certain security measures applied. Along these ideas, there is a risk of losing sight of the susceptibilities that may emerge due to system complexity, interdependencies, and cascading impacts. Yet these are the types of insights that are needed to support better security decision-making, and that should be considered in emerging and future testbed security analysis work. Formulating system-level and organisational security policies and standards is another security testbed design priority that requires more attention in order to re-focus the technical community – from attributing greater relevance to tasks related to establishing security than those of security governance and standardisation.

In addition, clearly defining testbed design objectives serves to resolve the typical tension between the high cost of deploying security testbed components, and the degree of similitude to the real system. For example, a testbed to determine vulnerabilities in ICS sub-systems like PLCs may not require the significant implementation and representation effort of an entire industrial architecture set-up, which can be expensive and unnecessary. Model approaches that involve combining the target component/module with virtual/software-based components can also be considered. Both measures can significantly reduce the cost of testbed development. Articulating specific high-level security-related design objective(s) well ahead of implementation can help with the decisions related to the control choice(s) to be made. For example, Fovino et al [36] described the analysis of cyber-attacks and impacts as objectives for their testbed-related security study. They further narrowed these objectives to encompass "SCADA system phishing with DNS poisoning, DoS Worm, and Modbus/DNP3 protocol worm". Similarly, Bergman et al [71] described their objective to aim at supporting the analysis of security control and impact. They clarified this further by indicating that in the context of the objective mentioned, their work explored "*testing the impacts of network segmentation and SSL encryption using OpenVPN*". Providing this level of specificity proved necessary to support a clearer understanding of what the work entailed in scope and design.

- **Simulation Approach**

For the simulation approach, all three schemas – real/live, emulated/virtual, and software-based/constructive, appear to have significant support in the user community. However, combining multiple simulation approaches seems more popular than using each alone. Besides aiming for a greater similitude to real system, another motivation involves exploring combinations that enable one approach to cover for the limitations of another.

Often, the choice of method(s) to combine is influenced by the degree of fidelity desired, the cost and affordability of development involved, and the time available for testbed development. The requirement for researcher/user expertise is a cross-cutting theme. Achieving high realism involves using infrastructures that work in the real environment; these are often expensive. Also, longer periods may be required to complete set-ups and configure the system and test processes. Often, decisions need to be weighted by trade-offs between attributes based on the defined objectives.

- **Architecture Components**

From results, the coverage of communication gateway (CG) components seems to dominate other architecture functional groups. This may be due to the mature nature of research and development in digital system/network communications gateway infrastructure, which increases its popularity over other functionality groups. The coverage frequency of component classes may also be driven by the degree of class criticality in testbed considerations.

Often, a broader coverage of functional areas, component classes and applicable routing protocols in a testbed architecture can depend on the objective(s) and scope of desired test scenario(s). Recapping the earlier scenario of assessing the vulnerabilities of a single FD device, this is unlikely to require building an entire SCADA system network but could involve a more direct approach of executing security audits on the desired device without necessarily embedding it in an operational ICS network. The hardware-in-the-loop (HIL) approach is a good technique to use. Another example involves structures that include CG components such as routers and switches where routable protocols are required. However, when tests do not involve the flow of data over router-based sub-systems and connections, then non-routable networks and protocols are appropriate. It is safer to consider routable protocols (e.g., Modbus TCP and DNP3) in ICS security testbed designs since they are by design better adapted to secure configuration across network paths than the non-routable protocols (e.g., DeviceNet), which are better placed to handle perimeter-based security. Notwithstanding the choice, incorporating all four functionality classes into an ICS security testbed design can allow for wider contexts to be mirrored, which typically provides a better representation of an ICS from both architectural and operational perspectives.

- **Design Evaluation Process**

The existence and rigour of evaluation process affects the credibility of ICS security simulation testbeds. The three levels of evaluation – *verification*, *validation*, and *accreditation (VV&A)* – indicate the possibility of a scale of credibility.

Verification enables project experimenters to double-check that adopted/defined security design/simulation requirement(s), system functions and processes, and any associated data, correctly represent the experimenter's conceptual description and specifications. It spans a range of contexts such as: (i) *verifying testbed structure and defined objectives* – justifying that testbed attributes are identified and usage objectives are clearly defined with sufficient accuracy; (ii) *verifying design problems* – proving that the problem adopted contains the actual environment problem, and is sufficiently well-formed to allow sufficiently credible solutions to be obtained; (iii) *verifying functions* – activities that demonstrate that testbed system functions and operations accurately mirror known real system behaviours relative to defined objectives; (iv) *verifying solutions* – activities demonstrating that the outputs and results reflect the known outcomes in real systems subject to the same parameters and operational conditions [37]. *Verification* is performed by the security testbed experimenters – a way of self-corroboration – to support credibility by demonstrating that predefined requirements in design, functionality, and outputs are well-satisfied. Techniques that can be applied for this can include: Desk checking, model review, result analysis, instrumentation-based testing, functionality testing, and sensitivity analysis [80].

Validation often comes from parties non-affiliated to the context being validated and seeks to establish the extent to which an ICS security testbed and associated data, mirror the real-world intended use scenario. Typically, it may involve the work of third parties in repeating processes and methodologies of verification, to ensure agreement between the observed or known behaviour of real system components and processes with that of the testbed simulation system. It also includes ascertaining whether any difference(s) between the two are acceptable given the testbed's intended use. This crucially helps to timely identify and rectify issues and avoid unnecessary misuse of evaluation time and other resources by users. *Validation* can be done using different methods, including based on historical data, through comparison with other testbed simulators, from expert judgements, parameter sensitivity analysis, or predictions [81].

Accreditation aims to certify that a testbed set-up and/or associated data are acceptable for a defined purpose. An accreditation evaluation leads to a recommendation that a certifying official or body authorises that a specific testbed M&S set-up or tool can be used for the purpose it is designed [82]. Thus, an accreditation often relies on the evidence(s) from verification and validation, and from additional in-depth and multi-level evaluations to establish more concrete attestation by an official certification body (typically government-based) following an independent assessment. In testbed model accreditation, the acceptability criteria are identified well ahead, and then the evidential knowledge from verification and validation processes is applied to ascertain how the intended use of the testbed model is impacted [83]. This way, accreditation not only focuses on the intended use, but also on the requirements for adopting testbed models.

Not all testbed M&S structures require an evaluation to accreditation level, so this should be pursued only if necessary. We believe that the level of evaluation necessary to support credibility and adoption can depend on factors such as project costs, available time and resources. In a resource-constrained setting, the cost of an evaluation process that involves VV&A activities on ICS security testbeds M&S can be prohibitively high compared to what is available or affordable. The unavailability of appropriate reference data, information from development products, documentation of past evaluations, can increase evaluation costs. Thus, it is advisable that the extent to which evaluation investments are made be weighed against potential risks of reaching a weak conviction, a bad decision, and huge scale of adverse impact. Such risks may stem from unreliable testbed M&S results or uncertainty in the evaluation process and life cycle. For example, although accreditation can afford greater credibility, it can take longer, and extend project schedules to achieve the in-depth documentation and assessments that may be required by certification bodies. This can be costly to achieve and may be less of an issue in the case of validations or verification – with the possible consequence of impacting perceived credibility of relevant testbeds or related work.

Unarguably, it is beneficial to pursue transitioning from verification to validation, and finally accreditation, as this provides a stronger evidence base and grounding for credibility. However, testbed developers and decision-makers responsible for evaluation investments need to consider risk possibilities related to defects in testbed set-ups, simulation scenarios and processes, hardware components, software elements, data, or even misjudged testbed capabilities by users. The level of tolerance considered acceptable should be part of the determining factors for the level of evaluation to reach.

Credibility-building process

Besides understanding the crucial design factors that can help to improve credibility of testbed simulation systems and their relationship as shown in Figure 7, it is crucial to adopt a structured approach in applying the factors into design implementation processes. The process can guide ICS security testbed experimenters on the steps to follow in their M&S process to persuade confidence and acceptability. This is demonstrated in Figure 8.

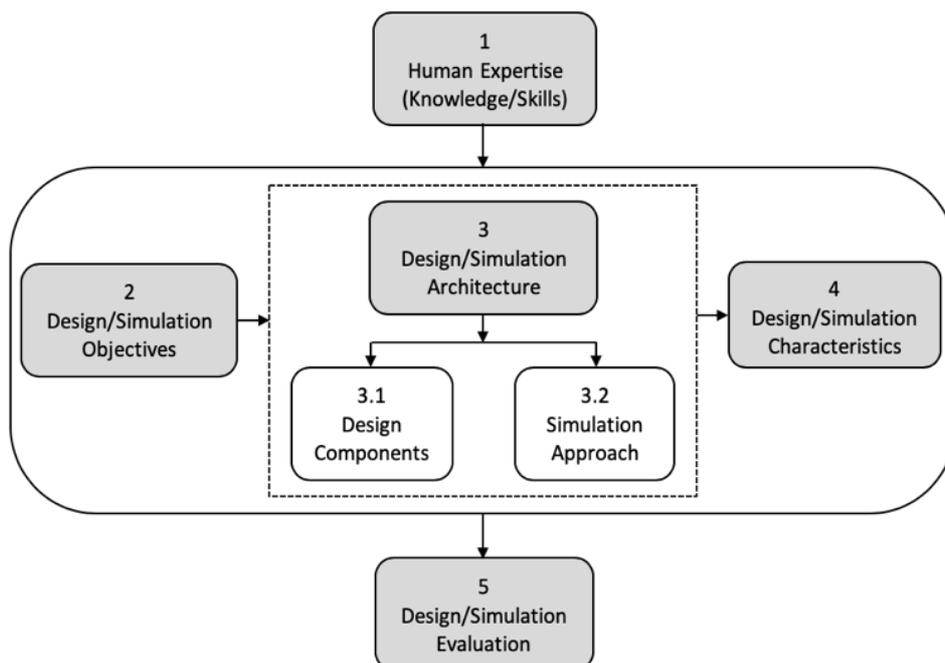


Figure 8: ICS Testbed Credibility-building Process

The first step in the credibility-building process should start with human expertise. Typically, based on the expertise of the developers involved, it is essential to have a testbed development plan – covering aspects of ICS and the IT and how they are integrated. Expertise in these areas could be demonstrated and inferred from the quality of context, design setup, and descriptions provided, which should reflect the real system. Testbed security design and simulation objectives need to be clearly defined and described in context to set the target scope.

The second step in the process involves defining the security design objective of interest. This should describe in clear context the security purposes for which the testbed design and scenarios are being implemented. The objectives can be singular, or a range of target security capacities or intents that could be achieved with the design. Expert knowledge and skills, coupled with clearly defined design/simulation objectives, contribute to defining an appropriate architecture.

The third step involves defining the design/simulation architecture. The architecture will encompass hardware, software, and protocol components and sub-systems. It also includes the simulation approach adopted, or any combinations thereof which are relevant for the intended test strategy and cope. These are aimed at persuading a high degree of confidence, trustworthiness acceptance of outcomes – credibility, as shown in Figure 7.

The fourth steps involve identifying and demonstrating core design/simulation characteristics that have been deemed relevant and considering any trade-offs. This gives better idea on the context and scope of security modelling and simulation adopted.

Once initial design and security test(s) have been completed, it is crucial to engage a final step of evaluating security scenario outcomes against initial characteristics intended in the plan. Evaluation must also include checking that the setup/simulation outcomes satisfy target objective(s), design architectures, and core functional characteristics. Evaluation can be done by various agents; from the experiment developers to external individuals or standards organisations, in order to corroborate initial claims on the reliability of the security testbed and/or results. We believe that increased expression of agreement with security testbed research by experts normally not actively involved in the initial testbed development process can, in turn, facilitate trustworthiness and credibility, and provide a platform for promoting widespread acceptance. The underlying ideas is to obtain a shared concurrence in the reliability of security testbed design and research concepts, attributes, and outputs.

Summary

The drivers that underpin design considerations in existing research indicate clear tensions between the '*generality to suit a broad set of ICS domains and/or applications*' and '*specificity to solve impending simulation challenges peculiar to domain or application*'. Common interests tend to focus on the latter (i.e., specificity).

The *generality* attribute can enable a capacity to mirror contexts and applications pertaining to multiple ICS domains without the need for significant re-configuration. Often this leads to a downside resulting from a lack of depth in system and process replication and analysis. Only high-level views of the system are captured, making up for 'breadth', but lacking in 'depth' of context coverage. *Specificity* to particular security modelling simulation problems enables the adoption of design attributes which favour a more focused coverage and in-depth analysis down to a detailed level possible. This enables a more tailored and a better understanding of simulation systems behaviour and performance. The appears more common in the community of designers and users perhaps because of the less demanding requirements related to; engaging a more narrowed area and view, and a lower level of expertise and specialisation.

Despite the underlying use of similar hardware and software infrastructure in various ICS domains, design architectures, protocols, functionalities, operations often vary amongst sectors. Each sector application usually involves complex component and process interactions that require specialised knowledge and skills to implement. It is rarely feasible to find expertise in depth that spans multiple infrastructures and that would enable a broader ICS testbed implementation that is fit for multiple purposes. This would enable in-depth modelling and simulation of multi-modal sector applications, architectures, components, protocols, processes, interactions and complexities. Specialisation seem to occur because developers and users engaged in ICS security testbed design and simulation typically have expertise (deep knowledge and skills) in a specific ICS domain or a narrow area of application.

Conclusion and Future Work

Several factors need to be considered when evaluating the reliability of simulation systems, the research that use them and their outcomes to guide the perception of credibility. Our research has explored literature and interacted with stakeholders to identify relevant factors that can provide guidance on ICS security testbed development and use, and which can support the decision on testbed credibility. Our work contributes novelty by developing mapping framework outlining relevant ICS security testbed design factors and associated attributes, and how they co-relate to support building credibility. Also, this is used by following a testbed credibility-building process (see Figure 2) presented, which provides a structured approach to consider/apply the credibility-building factors into design implementation.

Demonstrating credibility in ICS security simulation testbeds is an issue of concern, and the requirements to support this need to be streamlined. Building or enhancing credibility typically arises mainly from architectural coverage, characterised by the adopted implementation approach, selected components, and the demonstration of a reasonable degree of evaluation. These need to be engaged through a structured process, from defining security testbed design/simulation objectives to evaluating the work using the most feasible/available approaches. ICS security researchers and developers must strive to achieve fundamental architectures that are representative of real-world systems and can allow appropriate, yet realistic testing.

The expertise (domain knowledge and experience) of researchers and developers is crucial relative to achieving defined objectives and scenarios. Clear security-related design objectives defined from the outset can help drive the testbed development process, maintain a focused direction, and contribute reliability to the outcome. Clarifying the testbed simulation approach provides a path to understanding the tools and techniques adopted and their simulation capabilities. It also provides the information needed to reproduce and validate simulation testbed designs/systems and associated research. A clear outline of the architectural composition, and the adopted testbed simulation approach, increases the potential for demonstrating scientific rigour and repeatability, adding credibility to claims of quality and fidelity. Demonstrating evaluation procedures across verification, validation, and(or) accreditation can help attest to the satisfaction of quality, value, and acknowledgement in communities beyond the immediate designers, developers, and researchers. Including evaluation details can help resolve queries related to if and how a security testbed was validated and persuade a wider acceptance of a claimed credibility state. Having simulation systems and testbeds subjected to this type of multi-level evaluation process against available credibility criteria, can evidence quality and trustworthiness for critical decision-making.

It is beneficial to capture the core characteristics within a testbed setup. However, choosing the most important compliance characteristics within a specific project will depend on the project's core objectives and scope. Trade-offs may be needed, and considering the available resources/capabilities, certain characteristics may be incorporated or maximised at the expense of others. New attributes can also be considered based on emerging interests and evolving dynamics in the system or context of application. The novel relationship mapping approach can promote effective and well-organized procurement of systems and sub-system components guided by clearly defined design requirements; responding to system and functional dynamics, and the endorsement of the relevant community of stakeholders. It can thereby streamline the task of setting requirements and reduce the costs of both infrastructure development and sub-system integration. It can lead to greater consistency and efficiency in developing research related to ICS security testbeds, building on what already exists. Most conveniently, by combining this with the growing trend and capability for federating ICS security testbeds, as has been keenly advocated and explored in recent publications, the potential is increased for testbed availability and interoperability. Furthermore, a federation architecture/system can minimise the diversity in design structures between different and physically dispersed testbed infrastructures. For future work, we will explore how ICS testbeds are evaluated in practice, and how credibility may be tested.

Acknowledgments

The Research leading to the results presented in this paper is from the Analytical Lenses for Internet of Things Threats (ALloTT) research theme under the PETRAS Cyber Security of the Internet of Things project Hub. The project was supported by the UK Engineering and Physical Sciences Research Council (EPSRC), grant number EP/N02334X/1.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Dr Uchenna D. Ani is a Research Fellow with the PETRAS National Centre of Excellence for IoT Systems Cybersecurity, Department of Science, Technology, Engineering and Public Policy (STeAPP), University College London (UCL). He holds a PhD in the Cyber Security for Industrial Control System Environments. His research interests include; ICS and IoT technical and socio-technical security modelling, simulations, and risk analysis, digital forensics, and data-driven cyber security with applications to manufacturing, transportation, and energy systems critical infrastructures.

Professor Jeremy McKendrick D. Watson is a Professor of Engineering Systems and Director/Principal Investigator, PETRAS National Centre of Excellence for IoT Systems Cybersecurity, Department of Science, Technology, Engineering and Public Policy (STeAPP), University College London (UCL). He holds a PhD in Synthesized Speech Aids for the Vocally Handicapped. His research interests include: interactions in, and the design of, socio-technical critical infrastructure systems (including Internet of Things), cybersecurity (including cyber–physical) emerging technology identification, development and deployment, and strategic innovation processes for security.

Dr Benjamin Green is an academic fellow within the department of Computing and Communications, and a member of the Security Lancaster Research Institute, at Lancaster University. His research interests span a range of themes, all of which are aligned to cyber security challenges found within critical infrastructure, with a focus on industrial control systems.

Dr Barnaby Craggs is a Lecturer in the Department of Computer Science, the Bristol Cyber Security Group, University of Bristol, United Kingdom. His research interests are in Security Ergonomics and Cyber Physical Systems Security. His focus lies in the potential risks to both human safety and systems availability that arise from information-based decision processes. Dr Craggs is the lead researcher for the Bristol Cyber Security Group's portfolio of cyber physical projects.

Dr Jason R. C. Nurse is an Associate Professor in Cyber Security at the School of Computing at the University of Kent, and a Visiting Academic at the University of Oxford. His research focuses on the interaction between users and aspects of cyber security, privacy, and trust across the broader spectrum of modern technologies in use today. His interests encompass topics such as security and privacy in the IoT, usable security and security awareness programs in organizations and for the public, technical and psychological aspects of cybercrime, identity security, and privacy risks in cyberspace. He holds a PhD in Internet Security from the University of Warwick.

ORCID

Uchenna D Ani  <https://orcid.org/0000-0001-6064-480X>

References

- [1] J. Gardiner, B. Craggs, B. Green, and A. Rashid, "Oops I Did it Again : Further Adventures in the Land of ICS Security Testbeds," in *Proceedings of the 2019 Workshop on Cyber-Physical Systems Security and Privacy*, 2019.
- [2] Q. S. Qassim, N. Jamil, M. Daud, A. Patel, and N. Ja'afar, "A review of security assessment methodologies in industrial control systems," *Inf. Comput. Secur.*, pp. 1–15, 2019.
- [3] R. Candell, T. Zimmerman, and K. Stouffer, "An industrial control system cybersecurity performance testbed (NISTIR 8089)." National Institute of Standards and Technology, 2015.
- [4] A. Rashid, J. Gardiner, B. Green, and B. Craggs, "Everything is awesome! or is it? Cyber security risks in critical infrastructure," in *Critical Information Infrastructures Security - 14th International Conference, CRITIS 2019*, 2019, pp. 3–17.
- [5] J. Davis and S. Magrath, "A Survey of Cyber Ranges and Testbeds," Edinburgh South Australia, 2013.
- [6] M. Krotofil, J. Larsen, A. Isakov, A. Winnicki, D. Gollmann, and P. Gurikov, "Rocking the Pocket Book: Hacking Chemical Plants for Competition and Extortion," in *DefCon Conference (DefCon 23)*, 2015, pp. 1–52.
- [7] M. Almgren *et al.*, "RICS-el: Building a national testbed for research and training on SCADA security (short paper)," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11260 LNCS, pp. 219–225, 2019.
- [8] H. Christiansson and E. Luijff, "Creating a European SCADA Security Testbed," in *Critical Infrastructure Protection. ICCIP 2007. IFIP International Federation for Information Processing*, vol. 253, E. Goetz and S. Sheno, Eds. Springer, Boston, MA, 2008, pp. 237–247.
- [9] M. Frank, M. Leitner, and T. Pahi, "Design Considerations for Cyber Security Testbeds: A Case Study on a Cyber Security Testbed for Education," in *15th International Conference on Dependable, Autonomic and Secure Computing, IEEE 15th International Conference on Pervasive Intelligence and Computing, IEEE 3rd International Conference on Big Data Intelligence and Compu*, 2018, vol. 2018, pp. 38–46.
- [10] H. Holm, M. Karresand, A. Vidstrom, and E. Westring, "A Survey of Industrial Control System Testbeds," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, NordSec 20., vol. 9417, S. Buchegger and M. Dam, Eds. Springer International Publishing Switzerland, 2015, pp. 11–26.
- [11] H. Holm, M. Karresand, A. Vidström, and E. Westring, "A Survey of Industrial Control System Testbeds," in *In: Buchegger S., Dam M. (eds) Secure IT Systems. Lecture Notes in Computer Science*, vol. 9417, Switzerland: Springer, Cham, 2015, pp. 11–26.
- [12] O. Salunkhe, M. Gopalakrishnan, A. Skoogh, and Å. Fasth-Berglund, "Cyber-Physical Production Testbed: Literature Review and Concept Development," *Procedia Manuf.*, vol. 25, pp. 2–9, 2018.
- [13] Y. Geng, Y. Wang, W. Liu, Q. Wei, K. Liu, and H. Wu, "A survey of industrial control system testbeds," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 569, no. 4, pp. 1–10, 2019.
- [14] D. R. Danielson and S. Y. Rieh, "Credibility: A Multidisciplinary Framework," in *Annual Review of Information Science and Technology*, vol. 41, B. Cronin and D. Shaw, Eds. American Society for Information Science and Technology in Wiley Online Library, 2008, pp. 307–364.
- [15] S. Tseng and B. Fogg, "Credibility and Computing Technology," *Commun. Acm*, vol. 42, no. 5, pp. 39–44, 1999.
- [16] S. Y. Harmon and S. Youngblood, "Evolving the validation process maturity model (VPMM)," in *Proceedings of the 2008 Summer Computer Simulation Conference (SCSC '08)*, 2008, vol. 97457, no. 541, pp. 327–332.
- [17] W. L. Oberkampf, M. Pilch, and T. G. Trucano, "Predictive Capability Maturity Model for Computational Modeling and Simulation," California, 2007.
- [18] B. J. Fogg and H. Tseng, "The elements of computer credibility," in *CHI 99 Human Factors in Computing Systems Conference*, 1999, no. May, pp. 80–87.
- [19] SISO, "Reference for Generic Methodology for Verification and Validation (GM-VV) to Support Acceptance of Models , Simulations and Data," vol. 3, no. October. Simulation Interoperability Standards Organization (SISO), Inc, Florida, USA, pp. 1–48, 2013.
- [20] B. Craggs, A. Rashid, C. Hankin, R. . Antrobus, O. Şerban, and N. Thapen, "A Reference Architecture

for IIoT and Industrial Control Systems Testbeds.," in *2nd Conference on Living in the Internet of Things*, 2018.

- [21] A. M. Law, "How to build valid and credible simulation models," in *2009 Winter Simulation Conference (WSC)*, 2009, pp. 24–33.
- [22] B. Green, A. Le, R. Anrobus, U. Roedig, D. Hutchison, and A. Rashid, "Pains, Gains and PLCs: Ten Lessons from Building an Industrial Control Systems Testbed for Security Research," in *10th USENIX Workshop on Cyber Security Experimentation and Test (CSET '17)*, 2017, pp. 1–8.
- [23] C. McLean, Y. T. Lee, S. Jain, and C. Hurchings, "Modeling and Simulation of Critical Infrastructure Systems for Homeland Security Applications," *NIST Special Publications*. National Institute of Standards and Technology, p. 86, 2011.
- [24] Government Office for Science, "Computational Modelling: Technical Futures," London, 2018.
- [25] A. Gluhak, S. Krco, M. Nati, D. Pfisterer, N. Mitton, and T. Razafindralambo, "A Survey on Facilities for Experimental Internet of Things Research," *Commun. Mag. IEEE*, vol. 49, no. 11, pp. 40–48, 2011.
- [26] C. Siaterlis and B. Genge, "Cyber-Physical Testbeds," *Commun. ACM*, vol. 57, no. 6, pp. 64–73, 2014.
- [27] S. C. Petter and M. J. Gallivan, "Toward a framework for classifying and guiding mixed method research in information systems," in *Proceedings of the Hawaii International Conference on System Sciences*, 2004, vol. 37, no. C, pp. 4061–4070.
- [28] G. C. Peng, M. B. Nunes, and F. Annansingh, "Investigating information systems with mixed-methods research," in *Proceedings of the IADIS International Workshop on Information Systems Research Trends*, 2011, pp. 1–20.
- [29] M. Johnson *et al.*, "Multiple triangulation and collaborative research using qualitative methods to explore decision making in pre-hospital emergency care," *BMC Med. Res. Methodol.*, vol. 17, no. 1, pp. 1–11, 2017.
- [30] M. J. Grant and A. Booth, "A typology of reviews: an analysis of 14 review types and associated methodologies," *Health Info. Libr. J.*, vol. 26, pp. 91–108, 2009.
- [31] L. Salisbury, "Web of Science and Scopus: A comparative review of content and searching capabilities," *Advis. Rev.*, no. July, pp. 5–19, 2009.
- [32] V. Braun and V. Clarke, "Using thematic analysis in psychology Using thematic analysis in psychology," *Qual. Res. Psychol. ISSN*, vol. 3, no. 2, pp. 77–101, 2006.
- [33] R. V Rogers, "What makes a modeling and simulation professional?: The consensus view from one workshop," in *Proceedings of the 1997 Winter Simulation Conference*, 1997, pp. 1375–1382.
- [34] S. Karanam, G. Jorge-Botana, R. Olmos, and H. van Oostendorp, "The role of domain knowledge in cognitive modeling of information search," *Inf. Retr. J.*, vol. 20, no. 5, pp. 456–479, 2017.
- [35] R. Candell Jr., T. A. Zimmerman, and K. A. Stouffer, "An Industrial Control System Cybersecurity Performance Testbed." National Institute of Standards and Technology (NIST), pp. 1–47, 2015.
- [36] I. N. Fovino, M. Masera, L. Guidi, and G. Carpi, "An experimental platform for assessing SCADA vulnerabilities and countermeasures in power plants," in *3rd International Conference on Human System Interaction, HSI'2010 - Conference Proceedings*, 2010, pp. 679–686.
- [37] S. R. Blattnig, L. L. Green, J. M. Luckring, J. H. Morrison, and R. K. Tripathi, "Towards a Credibility Assessment of Models and Simulations," in *49th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, 2008, pp. 1–16.
- [38] M. J. McDonald and B. T. Richardson, "Position Paper : Modeling and Simulation for Process Control System Cyber Security Research , Development and Applications," 2009.
- [39] M. Govindarasu and C. Liu, "Cyber Physical Security Testbed for the Smart Grid: Fidelity, Scalability, Remote Access, and Federation," in *National CPS Energy Workshop*, 2013.
- [40] F. Sauer, M. Niedermaier, S. Kießling, and D. Merli, "LICSTER – A Low-cost ICS Security Testbed for Education and Research," in *Workshops in Computing series at 6th International Symposium for ICS & SCADA Cyber Security Research 2019 (ICS-CSR 2019)*, 2019, pp. 1–12.
- [41] W. Zhao, Y. Peng, and F. Xie, "Testbed Techniques of Industrial Control System," in *3rd International Conference on Computer Science and Network Technology, ICCSNT 2013*, 2013, pp. 61–65.
- [42] R. B. Vaughn and T. Morris, "Addressing Critical Industrial Control System Cyber Security Concerns via High Fidelity Simulation," in *11th Annual Cyber and Information Security Research Conference on - CISRC '16*, 2016, pp. 1–4.
- [43] W. Knowles, D. Prince, D. Hutchison, J. F. P. Disso, and K. Jones, "A survey of cyber security management in industrial control systems," *Int. J. Crit. Infrastruct. Prot.*, vol. 9, pp. 52–80, 2015.

- [44] S. McLaughlin *et al.*, "The Cybersecurity Landscape in Industrial Control Systems," *Proc. IEEE*, vol. 104, no. 5, pp. 1039–1057, 2016.
- [45] S. Siboni, V. Sachidananda, A. Shabtai, and Y. Elovici, "Security Testbed for the Internet of Things," 2016.
- [46] K. Stouffer, V. Pillitteri, S. Lightman, M. Abrams, and A. Hahn, "Guide to Industrial Control Systems (ICS) Security - NIST.SP.800-82r2." NIST, US Department of Commerce, Gaithersburg, Maryland, pp. 1–247, 2015.
- [47] G. Koutsandria, R. Gentz, M. Jamei, A. Scaglione, S. Peisert, and C. McParland, "A Real-Time Testbed Environment for Cyber-Physical Security on the Power Grid," in *First ACM Workshop on Cyber-Physical Systems-Security and/or PrivaCy - CPS-SPC '15*, 2015, no. June 2017, pp. 67–78.
- [48] H. Kavak, J. J. Padilla, D. Vernon-Bido, R. J. Gore, and S. Y. Diallo, "A Characterization of Cybersecurity Simulation Scenarios," in *19th Communications and Networking Simulation Symposium (CNS'16)*, 2016, pp. 1–8.
- [49] J. M. Couretas, "Cyber Modeling & Simulation (M&S) for Test and Evaluation (T&E)," in *An Introduction to Cyber Modeling and Simulation*, J. G. Behr and R. Diaz, Eds. New Jersey, USA: John Wiley & Sons, Inc., 2019, pp. 125–136.
- [50] GSE Systems, "Fidelity Matters: What 'High-fidelity' Really Means," *Simulation & Training Blog*, 2017. [Online]. Available: <https://www.gses.com/blog/simulation-training/high-fidelity/>. [Accessed: 14-Jun-2018].
- [51] J. M. Couretas, "Cyber Model-Based Evaluation Background," in *An Introduction to Cyber Modeling and Simulation*, J. G. Behr and R. Diaz, Eds. New Jersey, USA: John Wiley & Sons, Inc., 2019, pp. 89–100.
- [52] DoD, "Department of Defense Modeling and Simulation Best Practices Guide," no. October. US Department of Defense, 2010.
- [53] DoD, "M&S VV&A RPG Core Document: Introduction." US Department of Defence, pp. 1–34, 2011.
- [54] ISA99, "ISA-62443 [multiple parts], Security for Industrial Automation and Control Systems," *ISA99 Standards Portal*, 2015. [Online]. Available: http://isa99.isa.org/ISA99 Wiki/WP_List.aspx. [Accessed: 21-Jun-2018].
- [55] T. Hedberg and M. Helu, "Design and Configuration of the Smart Manufacturing Systems Test Bed," *NIST Advanced Manufacturing Series 200-1*. National Institute of Standards and Technology (NIST), pp. 1–37, 2017.
- [56] T. J. Williams, "The Purdue enterprise reference architecture," *Comput. Ind.*, vol. 24, no. 2–3, pp. 141–158, 1994.
- [57] P. Matoušek, "Description and analysis of IEC 104 Protocol," Brno, Czech Republic, 2017.
- [58] Y. Tao, W. Xu, H. Li, and S. Ji, "Experience and lessons in building an ics security testbed," in *1st International Conference on Industrial Artificial Intelligence, IAI 2019*, 2019, pp. 1–6.
- [59] T. Alves, R. Das, and T. Morris, "Virtualization of Industrial Control System Testbeds for Cybersecurity," in *2nd Annual Industrial Control System Security Workshop on - ICSS '16*, 2016, pp. 10–14.
- [60] A. Hahn, A. Ashok, S. Sridhar, and M. Govindarasu, "Cyber-physical security testbeds: Architecture, Application, and Evaluation for Smart Grid," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 847–855, 2013.
- [61] R. Neg, A. Kumar, S. K. Shukla, and A. Dayal, "A SCADA test bed For Cyber Security Education & Research," Cyber Security Centre, Indian Institute of Technology, Kanpur, Kanpur, India, 2016.
- [62] J. Kim, K. Kim, and M. Jang, "Cyber-Physical Battlefield Platform for Large-Scale Cybersecurity Exercises," in *International Conference on Cyber Conflict, CYCON*, 2019, vol. May, pp. 1–19.
- [63] I. Ahmed, V. Roussev, W. Johnson, S. Senthivel, and S. Sudhakaran, "A SCADA System Testbed for Cybersecurity and Forensic Research and Pedagogy," in *2nd Annual Industrial Control System Security Workshop on - ICSS '16*, 2016, pp. 1–9.
- [64] A. Gilchrist, "Introducing Industry 4.0," in *Industry 4.0: The Industrial Internet of Things*, First., Bangken, Thailand: Apress, 2016, pp. 195–215.
- [65] B. Van Leeuwen, V. Urias, J. Eldridge, C. Villamarin, and R. Olsberg, "Cyber Security Analysis Testbed: Combining Real, Emulation, and Simulation," in *44th Annual 2010 IEEE International Carnahan Conference on Security Technology*, 2010, pp. 1–6.
- [66] B. Van Leeuwen, V. Urias, J. Eldridge, C. Villamarin, and R. Olsberg, "Performing cyber security analysis using a live, virtual, and constructive (LVC) testbed," in *Proceedings - IEEE Military Communications Conference MILCOM*, 2010, pp. 1806–1811.

- [67] V. Urias, B. Van Leeuwen, and B. Richardson, "Supervisory Command and Data Acquisition (SCADA) system cyber security analysis using a live, virtual, and constructive (LVC) testbed," in *EEE Military Communications Conference MILCOM*, 2012, no. Lvc, pp. 1–8.
- [68] H. El Haouzi, A. Thomas, and J. F. Pétin, "Contribution to reusability and modularity of manufacturing systems simulation models: Application to distributed control simulation within DFT context," *Int. J. Prod. Econ.*, vol. 112, no. 1, pp. 48–61, 2008.
- [69] H. Gao, Y. Peng, Z. Dai, T. Wang, X. Han, and H. Li, "An Industrial Control System Testbed Based on Emulation, Physical Devices and Simulation," in *IFIP Advances in Information and Communication Technology*, Critical I., J. Butts and S. Sheno, Eds. Arlington, Virginia, USA: Springer Berlin Heidelberg, 2014, pp. 79–91.
- [70] J. Stites, A. Siraj, and E. L. Brown, "Smart Grid Security Educational Training with ThunderCloud: A Virtual Security Test Bed," in *13 Information Security Curriculum Development Conference - InfoSecCD '13*, 2013, pp. 105–110.
- [71] D. C. Bergman, D. Jin, D. M. Nicol, and T. Yardley, "The virtual power system testbed and inter-testbed integration," in *2nd conference on Cyber Security Experimentation and Test*, 2009, no. August, p. 5.
- [72] A. Almalawi, Z. Tari, I. Khalil, and A. Fahad, "SCADA-VT-A framework for SCADA security testbed based on virtualization technology," in *Conference on Local Computer Networks, LCN*, 2013, pp. 639–646.
- [73] M. Reul, "Bringing usability to industrial control systems," *Conf. Hum. Factors Comput. Syst. - Proc.*, pp. 3335–3340, 2009.
- [74] A. L. Stephano and D. P. Groth, "USEable security: Interface design strategies for improving security," in *Proceedings of the 3rd International Workshop on Visualization for Computer Security, VizSEC'06. Co-located with the 13th ACM Conference on Computer and Communications Security, CCS'06*, 2006, pp. 109–116.
- [75] ISO, "ISO 9241: Ergonomic requirements for office work with visual display terminals." ISO/IEC, 1998.
- [76] J. Tidwell, *Designing Interfaces*, 2nd ed. O'Reilly Media, Inc., 2010.
- [77] J. R. C. Nurse, S. Creese, M. Goldsmith, and K. Lamberts, "Guidelines for usable cybersecurity: Past and present," in *Proceedings - 2011 3rd International Workshop on Cyberspace Safety and Security, CSS 2011*, 2011, pp. 21–26.
- [78] C. Hankin *et al.*, "Open Testbeds for CNI." RITICS Project Portal, London, UK, p. 16, 2018.
- [79] B. Green, M. Krotofil, and D. Hutchison, "Achieving ICS Resilience and Security through Granular Data Flow Management," in *2nd ACM Workshop on Cyber-Physical Systems Security and Privacy - CPS-SPC '16*, 2016, pp. 93–101.
- [80] O. Balci, "How to assess the acceptability and Design of Experiments," in *21st conference on Winter simulation*, 1989, pp. 62–71.
- [81] MITRE, "Verification and Validation of Simulation Models," *SE Life-Cycle Building Blocks Articles*, 2019. [Online]. Available: <https://www.mitre.org/publications/systems-engineering-guide/se-lifecycle-building-blocks/other-se-lifecycle-building-blocks-articles/verification-and-validation-of-simulation-models>. [Accessed: 04-Oct-2019].
- [82] D. A. Cook and J. M. Skinner, "How to perform credible verification, validation, and accreditation for modeling and simulation," *CrossTalk J. Def. Softw. Eng.*, no. 5, pp. 20–24, 2005.
- [83] S. Youngblood, M. Stutzman, D. K. Pace, and P. P. Pandolfini, "Risk Based Methodology for Verification, Validation, and Accreditation (VV & A) M&S Use Risk Methodology (MURM)," no. NSAD-R-2011-011. The Johns Hopkins University Applied Physics Laboratory, pp. 1–131, 2011.

Appendix A.

Table A1. ICS Security-related Testbed Works

	Authors	Paper Title	Institution	Country	Objectives	Approach	Landscape/ Coverage	Credibility Requirements	Evaluation/Validation
1	Giani et al 2008	A Testbed for Secure and Robust SCADA Systems	UC Berkeley	USA	VA, DMT	SBS, EM, PS	PP, FD, CG, CC	Not Mentioned	Not Mentioned
2	Hieb et al 2008	Security Enhancements for Distributed Control Systems	University of Louisville	USA	DMT	SBS	CC/PP	Not Mentioned	Not Mentioned
3	Queiroz et al, 2009	Building a SCADA Security Testbed	RMIT University	Australia	AA	SBS, EM	PP, CG, CC	Modularity, Fidelity Isolation, Reproducibility, Scalability, Flexibility, Fidelity	Base on Prior works
4	Bergman et al 2009	The Virtual Power System Testbed and Inter-Testbed Integration	University of Illinois at Urbana-Champaign	USA	IA, DMT	SBS, EMU	FD, CG, CC	Flexibility, Extensibility (Scalability) Adaptability (Flexibility)	Not Mentioned
5	Kush et al 2010	Smart Grid Test Bed Design and Implementation.	Queensland University of Technology	Australia	VA, TA, IA	Virtualisation	PP, FD, CG, CC	Repeatability, Safe Execution	Comparison with User-defined functional requirements
6	Chunlei et al 2010	A Simulation Environment for SCADA Security Analysis and Assessment	Tsinghua University of Beijing	China	AA	SBS, EM, PS	PP, FD, CG, CC	Unreferenced SCADA Reference Architecture	Unreferenced SCADA Reference Architecture
7	Fovino et al 2010	An Experimental Platform for Assessing SCADA Vulnerabilities and Countermeasures in Power Plants	European Commission Joint Research Centre	Italy	AA, IA	PS	PP, FD, CG, CC	Fidelity	Not Mentioned
8	Hahn et al 2010	Development of the PowerCyber SCADA Security Testbed	Iowa State University	USA	ET, AA,	EM	PP, CG, CC	Not Mentioned	Based on NERC & NIST Requirements
9	Stefanov and Liu, 2011	Cyber-Power System Security in a Smart Grid Environment	University College Dublin	Ireland	AA, TA, VA, IA	SBS	PP, CG, CC	Not Mentioned	Not Mentioned
10	Dondossola and Garrone, 2011	Cyber Risk Assessment of Power Control Systems – A Metrics weighed by Attack Experiments	Ricerca sul Sistema Energetico	Italy	AA, IA, DMT	SBS	PP, FD, CG, CC	Not Mentioned	Compliant with design standards (IEC 60870-5-104 TCP/IP Communications)
11	Morris et al 2011	A control system testbed to validate critical infrastructure protection concepts	Mississippi State University	USA	ET, VA, DMT	PS	PP, FD, CG, CC	Fidelity	Not Mentioned
12	Mallouhi et al 2011	A Testbed for Analysing Security of SCADA Control Systems (TASSCS)	University of Arizona	USA	DMT	SBS	FD, CG, CC	Not Mentioned	Not Mentioned
13	Jin et al 2011	An Event Buffer Flooding Attack in DNP3 Controlled Scada Systems	University of Illinois at Urbana-Champaign	USA	AA, VA	PS	CG, CC	Flexibility, Extensibility (Scalability) Usability, Scalability, Fidelity, Modularity	Based on a Real Design Testbed
14	Almalawi et al 2013	SCADA-VT–A Framework for SCADA Security Testbed Based on Virtualization Technology	RMIT University	Australia	AA, IA,	Virtualisation	PP, FD, CG, CC	Not Mentioned	Not Mentioned
15	Sayegh et al 2013	Internal Security Attacks on SCADA Systems	American University of Beirut	Lebanon	AA, VA	PS	PP, CG, CC	Not Mentioned	User-defined requirements

	Authors	Paper Title	Institution	Country	Objectives	Approach	Landscape/ Coverage	Credibility Requirements	Evaluation/Validation
16	Shahzad et al 2013	Secure Cryptography Testbed Implementation for SCADA Protocols Security	University Kuala Lumpur	Malaysia	DMT	SBS	Not Mentioned	Not Mentioned	Not Mentioned
17	Urias et al 2013	Supervisory Command and Data Acquisition (SCADA) system Cyber Security Analysis using a Live, Virtual, and Constructive (LVC) Testbed	Sandia National Laboratories	USA	VA, AA	SBS, EM/V, PS	PP, CC, CG	Modularity, Interoperability, Scalability, Cost-Effectiveness, Fidelity	User-defined requirements
18	Stites et al 2013	Smart Grid Security Educational Training with Thunder Cloud: A Virtual Security Test Bed	Tennessee Technological University	USA	ET, VA, AA	EM/V	PP, CC, CG	Cost-Effective	user-defined requirements
19	Hahn et al 2013	Cyber-Physical Security Testbeds: Architecture, Application, and Evaluation for Smart Grid	Iowa State University	USA	VA, IA, AA	SBS, EM, PS	PP, FD, CG, CC	Scalability, Modularity, Extensibility, Fidelity (Accuracy) Fidelity, Modularity, Repeatability, Measurability, Cost-Effective	Not Mentioned
20	Gao et al 2014	An Industrial Control System Testbed Based On Emulation, Physical Devices And Simulation	Technical Assessment Research Lab	China	VA, DMT,	SBS, EM, PS	PP, FD, CG, CC	Modularity, Repeatability, Measurability, Cost-Effective	compliant with ANSI/ISA-99 standard
21	McLaughlin et al 2014	Multi-attribute SCADA-Specific Intrusion Detection System for Power Networks	Queen's University Belfast	Ireland	AA, DMT	SBS	PP, CG, CC	Fidelity	User-defined requirements
22	Genge and Siaterlis, 2014	Cyber-Physical Testbeds - EPIC	European Commission Joint Research Centre	Italy	AA, IA, DMT, Network QoS Effects on cyber attacks	SBS, EM	PP, FD, CG, CC	Fidelity, Measurement Accuracy, Repeatability, Scalability, Safe Execution (Safety)	Compliant with design standards (IEEE 9, 30, 39 and 118)
23	Haney and Papa 2014	A framework for the design and deployment of a SCADA honeynet	The University of Tulsa	USA	DMT,	SBS, EM/V, PS		Not Mentioned	Not Mentioned
24	Candell et al 2015	An Industrial Control System Cybersecurity Performance Testbed	National Institute of Standards and Technology (NIST)	USA	DMT, IA	SBS, EMU/HIL, PS	PP, FD, CG, CC	Diversity, Flexibility, Scalability, Fidelity, Security Analysis, Extensibility	Compliant with NIST SP 800-82 Security guidelines
25	Singh et al 2015	A Testbed for SCADA Cyber Security and Intrusion Detection	Centre for Development of Advanced Computing Sapienza University of Rome, Arizona State University and Lawrence Berkeley National Laboratory	India	DMT, AA,	SBS, EM	PP, FD, CG	Not Mentioned	Not Mentioned
26	Koutsandria et al 2015	A Real-Time Testbed Environment for Cyber-Physical Security on the Power Grid	Centre for Development of Advanced Computing Sapienza University of Rome, Arizona State University and Lawrence Berkeley National Laboratory	Italy and USA	AA	SBS, EM, PS	PP, FD, CG, CC	Fidelity, Repeatability	Not Mentioned
27	Farooqui et al 2015	Cyber Security Backdrop: A SCADA Testbed	National University of Sciences and Technology	Pakistan	AA, IA	SBS	PP, CG, CC	Flexibility, Usability	Not Mentioned

	Authors	Paper Title	Institution	Country	Objectives	Approach	Landscape/ Coverage	Credibility Requirements	Evaluation/Validation
28	Jarmakiewicz et al 2015	Development of Cyber Security Testbed for Critical Infrastructure	Military University of Technology	Poland	DMT,	EM, PS	PP, CG, CC	Fidelity	Compliant with Standard
29	Ghassempour et al 2015	A Hardware-in-the-Loop SCADA Testbed	University of South Florida	USA	AA, DMT	SBS, EM/HIL,	CC, CG	Not Mentioned	Complaint with IEEE-C37.118 and Modbus protocol design
30	Krotofil et al 2015	Rocking the pocketbook: Hacking chemical plants for competition and extortion (Damn Vulnerable Chemical Process)	Hamburg University of Technology	Germany	AA, DMT	EMU/HIL, PS	PP, FD, CG, CC	Fidelity, Repeatability, Complexity, Measurability	Not Mentioned
31	Ghaleb et al 2016	SCADA-SST: A SCADA Security Testbed	King Fahd University of Petroleum & Minerals	Saudi Arabia	AA, DMT, IA	SBS/EM	PP, FD, CG, CC	Modularity, Extensibility, Reproducibility	Prior Works/Models
32	Hink and Goseva-Popstojanova, 2016	Characterization of Cyberattacks aimed at Integrated Industrial Control and Enterprise Systems: A case study	West Virginia University	USA	ET, VA, DMT, TA, IA,	PS	PP, CC, CG	Believed To Be Representative Of The Real Systems	Validation
33	Cruz et al 2016	A Cybersecurity Detection Framework for Supervisory Control and Data Acquisition Systems	University of Coimbra	Portugal	DMT, TA,	SBS, EM	PP, FD, CG	Fidelity, Repeatability, Data Accuracy	Validated by comparing normal and attack scenarios results
34	Mathur and Tippenhauer, 2016	SWaT: A Water Treatment Testbed for Research and Training on ICS Security	Singapore University of Technology and Design	Singapore	ET, IM, DMT,	PS	PP, FD, CG, IoT, HHC	Third Party Developers	Not Mentioned
35	Korkmaz et al 2016	ICS Security Testbed with Delay Attack Case Study	Binghamton University	USA	IA, DMT, ET	PS	PP, FD, CG	Fully Consistent With Industry Instrumentation Standard	Not Mentioned
36	Ahmed et al 2016	A SCADA System Testbed for Cybersecurity and Forensic Research and Pedagogy	University of New Orleans	USA	AA, ET, DMT	PS	PP, FD, CG, CC	Fidelity, Modularity	Not Mentioned
37	Neg et al 2016	A SCADA testbed for Cyber Security Education & Research	Indian Institute of Technology	India	ET, AA, VA, DMT, IA	EM, PS	PP, FD, CG, CC	Flexibility	Not Mentioned
38	Alves et al 2016	Virtualization of Industrial Control System Testbeds for Cybersecurity	University of Alabama in Huntsville	USA	AA, IA	EM/V, PS	PP, FD, CG, CC	Fidelity, Measurement Accuracy,	Not Mentioned
39	Sou pionis et al 2016	Cyber Security Impact on Power Grid Including Nuclear Plant	European Commission, Joint Research Centre (JRC)	Italy	AA, VA	SBS, EM/HIL,	PP, FD, CG, CC	Not Mentioned	Not Mentioned
40	Green et al 2017	Pains, Gains and PLCs: Ten Lessons from Building an Industrial Control Systems Testbed for Security Research	Lancaster University	England	ET, VA, AA, DMT, IA	EM, PS	PP, FD, CG, CC	Scalability, Diversity, Flexibility, Fidelity, Monitoring, Logging, Openness, Usability, Complexity	PERA Reference Model & Other prior testbed infrastructures
41	Koganti et al, 2017	A Virtual Testbed for Security Management of Industrial Control Systems	University of Idaho	USA	VA, AA, IA,	SBS, EM	PP, FD, CG	Not Mentioned	Not Mentioned
42	Rubio-Hernan et al 2017	Security of Cyber-Physical Systems from Theory to Testbeds and Validation	Universit´e Paris-Saclay	France	AA, ET	PS, EM	PP, CG, CC	Repeatability, Cost-Effective	Not Mentioned

	Authors	Paper Title	Institution	Country	Objectives	Approach	Landscape/ Coverage	Credibility Requirements	Evaluation/Validation
43	Teixeira et al 2018	SCADA System Testbed for Cybersecurity Research Using Machine Learning Approach	Federal Institute of Education, Science, and Technology, Sao Paulo	Brasil	AA, IA	PS	PP, FD, CG, CC	Not Mentioned	Comparison with real online system deployment
47	Almgren et al 2018	RICS-el: Building a National Testbed for Research and Training on SCADA Security	Chalmers University & Swedish Defence Research Agency	Sweden	AA, ET	EM/Virtualisation	PP, FD, CG, CC	Fidelity, Openness, Adaptability, Repeatability	Not Mentioned
49	Barrere et al 2018	CPS-MT: A Real-Time Cyber-Physical System Monitoring Tool for Security Research	Imperial College London & University College London	England	ET, DMT	SBS	CC, CG	Flexibility	Not Mentioned
50	Xie et al 2018	A Virtual Industrial Control System Testbed for Cyber Security Research	State Key Laboratory of Mathematical Engineering and Advanced Computing	China	AA	EM/Virtualisation	PP, CC, CG	Fidelity	Testbed architecture is based on prior work: Tennessee-Eastman Process
52	Kaouk et al 2018	A IoT-based control systems testbed for cybersecurity assessment of industrial and IoT-based control systems	University of Grenoble	France	AA, DMT, IA	PS, Virtualisation/HIL	PP, FD, CG,CC	Flexibility, Safety, Fidelity	Not Mentioned
53	Szanto et al 2018	A Testbed for Performing Security Experiments with Software-Defined Industrial Control Systems	University of Transylvania & University of Tîrgu Mureş	Romania	AA, DMT	SBS	CC, CG	Not Mentioned	Not Mentioned
44	Adepu et al 2018	EPIC: An Electric Power Testbed for Research and Training in Cyber Physical Systems Security	Singapore University of Technology and Design	Singapore	AA, VA, IA, DMT	PS	PP, FD, CC, CG	Not Mentioned	Not Mentioned
45	Kim et al 2019	Cyber-Physical Battlefield Platform for Larhe-Scale Cybersecurity Exercises	National Security Research Institute	South Korea	ET, AA, DMT	EM/V	PP, FD, CG, CC	Scalability/Extensibility, Reality(Fidelity), Flexibility/Reusability),	Real system design based on observed system and expert contributions.
46	Hui et al 2019	ICS Interaction Testbed: A Platform for Cyber-Physical Security Research	Queens University Belfast	Northern Ireland	IA, DMT	PS	PP,FD, CG,CC	diversity, logging	Industrial Best-practice standard
48	Gardiner et al 2019	Oops I Did it Again: Further Adventures in the Land of ICS Security Testbeds	University of Bristol & University of Lancaster	England	ET, VA, AA, DMT, IA	EM, PS	PP, FD, CG, CC	Scalability, diversity, flexibility, fidelity, Monitoring, Logging, Openness, usability, complexity, safety	Self-validation of past work by authors
51	Ashok et al 2019	A Multi-level Fidelity Microgrid Testbed Model for Cybersecurity Experimentation	Pacific Northwest National Laboratory	USA	AA, DMT	SBS, HIL/Virtualisation	PP,FD, CG,CC	Fidelity, flexibility	Based on Standard IEEE 37 node distribution feeder model
54	Sauer et al 2019	LICSTER: A Low-cost ICS Security Testbed for Education and Research	Hochschule Augsburg	Germany	AA, IA,	PS, EM/V, SBS	PP,FD, CG,CC	cost-effective, repeatability,	Unreferenced: Verification of process implementation, cybersecurity scenario feasibility, open research questions and existing threat model (STRIDE) comparative analysis.
55	Zhang et al 2019	Multilayer Data-Driven Cyber-Attack Detection System for Industrial Control	University of Tennessee, Knoxville	USA	AA, IA, DMT	PS, EM/V	PP, CC, CG	Not Mentioned	Not Mentioned

Authors	Paper Title	Institution	Country	Objectives	Approach	Landscape/ Coverage	Credibility Requirements	Evaluation/Validation
	Systems Based on Network, System, and Process Data							
56	Tao et al 2019 Experience and Lessons in Building an ICS Security Testbed	National Joint Engineering Lab for ICS Security & Shenyang Institute of Computing Technology, Chinese Academy of Sciences	China	AA, IA, DMT	PS	PP, FD, CC, CG	Diversity, scalability, fidelity, reproducibility, flexibility, logging	Reference to Purdue Enterprise Reference Architecture
57	Blazek et al 2019 Development of Cyber Cyber-Physical Security Testbed Based on IEC 61850 Architecture	Brno University of Technology	Czech Republic	P/QoS, VA	SBS, Virtualisation	PP, CC, CG	Isolation/Safety, fidelity, interoperability,	Designed based on IEC 61850 and NIST SP 800-161

AA = Attack Analysis

DMT/A = Defence Mechanism Tests/Analysis

IA = Impact Analysis

VA = Vulnerability Analysis

A&T = Education and Training

TA = Threat Analysis

P/QoS = Performance/QoS Analysis

CP&S = Creation of Policies and(or) Standards

SBS = Software-Based Simulation

SPS = Semi-Physical Simulation (Emulation or /Virtualisation / HIL)

PS = Physical Simulation

CG = Communications Gateway

PP = Physical Process

CC = Control Centre

FD = Field Device/Components