Research Article

TU Delft OPEN Publishing

Safety and Security Dependencies for Gridshield: Qualitative and Quantitative Analysis

Reza Soltani [1,*], Baver Ozceylan [1], Milan Lopuhaä-Zwakenberg [1], Christina Kolb [2], Gerwin Hoogsteen [1]

[1]Department of Electrical Engineering, Mathematics and Computer Science, University of Twente, The Netherlands
[2]Department of Behavioural, Management and Social Sciences, University of Twente, The Netherlands

*Corresponding author: r.soltani@utwente.nl

Abstract: Present-day cyber-physical systems, such as the Smart Grid, lead to the integration of multiple sub-systems into one single intertwined system. Such systems are characterized by many inter-dependencies between these sub-systems. This makes it complex to correctly assess the impact of new defense mechanisms with respect to the safety and security of the system as a whole. Existing formalisms, such as fault and attack trees, cannot describe the full system complexity.

This paper presents a novel integrated model, namely the Attack-Fault-Defense Tree (AFDT), and tools to analyze such cyber-physical systems. The presented visual representation allows experts from various disciplines to discuss system dependencies together. In addition, we also present how minimum cut sets can be derived to formally quantify how the safety and security of the overall system is enhanced with the implementation of new defenses. We furthermore extend this to quantitative analysis by assigning safety and security metrics to these minimal cut sets. The presented AFDT is applied to the Gridshield concept, a novel defense mechanism to prevent grid overloading in power grids due to simultaneous charging of electric vehicles.

One sentence summary: We introduce Attack-Fault-Defense Trees (AFDTs) and apply qualitative and approximate quantitative analysis to the Gridshield smart-grid defense mechanism.

Keywords: Cyber-physical systems, Fault trees, Attack trees, Defenses, Risk mitigation, Protection schemes

Publishing history: Submitted: 5 February 2025; Revised: 17 April 2025, 28 October 2025; Accepted: 16 December 2025; Published: 16 June 2026

Cite as: Soltani, R., Ozceylan, B., Lopuhaä-Zwakenberg, M., Kolb, C., & Hoogsteen, G. (2026). Safety and Security Dependencies for Gridshield: Qualitative and Quantitative Analysis. *Journal of Progress in Safety and Security, 2*. https://doi.org/10.59490/pss.2.2026.8020

ISSN: 3050-4570
Vol. 2, 2026
DOI: 10.59490/pss.2.2026.8020

©2026 Authors. Published by TU Delft OPEN Publishing on behalf of the authors. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence.

1 Introduction

The ongoing green energy transition, which refers to the shift from traditional energy sources to renewable and environmentally friendly energy sources, integrates various sub-systems into a complex cyber-physical system. Electrification is the key to this green transition, which means that the reliance on the power grid infrastructure increases in different sectors. For example, the adoption of electric vehicles (EVs) in transportation and heat pumps in industry connects the supply chain to the power grid infrastructure. In addition to electrification, digitalization plays a crucial role in this green transition. A notable example is the rapid integration of cutting-edge information and communication technologies in power grids. Particularly, continuous monitoring and autonomous control are vital to balance demand and supply in power grids due to the uncontrollable nature of renewable energy sources.

Such intertwined systems greatly improve the efficiency and flexibility of energy use. However, the dependencies between these systems increase safety and security risks since any equipment failure or any cyber attack may propagate and cause catastrophic results due to the complex control interactions that depend on accurate data (Yadav et al., 2016). Hence, new concepts are being developed that ensure a safe and secure mass integration of novel technologies. Among these approaches are process-aware intrusion detection systems (Menzel et al., 2023) or novel independent last-resort defense mechanisms such as Gridshield (Tangerding et al., 2022), which must reliably avoid grid overloading under all circumstances.

Effectively assessing safety and security risks in a complex cyber-physical system requires a deep understanding of its sub-systems, and the interactions between them. This means that experts and engineers from different disciplines have to work together. The challenge here is to establish a common language and framework that enables seamless communication across disciplines. Graphical, tree-based formalisms, such as fault trees (FTs) for safety (Kabir, 2017) and attack trees (ATs) for security (Mauw & Oostdijk, 2006), are popular, as they are both intuitive and expressive (Nicoletti et al., 2023). FTs and ATs can be extended with system recovery (Budde et al., 2021) or defenses (Kordy et al., 2014) to address the system owner's risk mitigation strategies. However, currently, no formalism exists that considers both attacks, faults, and defenses (Budde et al., 2021). Furthermore, existing formalisms model dependencies in small and medium case studies. Large case studies on representing dependencies between safety and security and their visualizations are still missing (Nicoletti et al., 2023).

Table 1. Comparison of Contributions Across AFDT Research Papers

Paper	Soltani et al. (2024a)	Soltani et al. (2024b)	this extension
Focus	AFDT formalism	Gridshield case study	approximate quantitative analysis
Qualitative formalism	impact of defenses on MCS	-	-
Quantitative formalism	exact MCS metrics + defense impact	-	approximate MCS metrics + defense impact
Visualisation	MCS tables, effort/probability graphs	-	effort/probability histograms
Case study insights	-	qualitative study	quantitative study

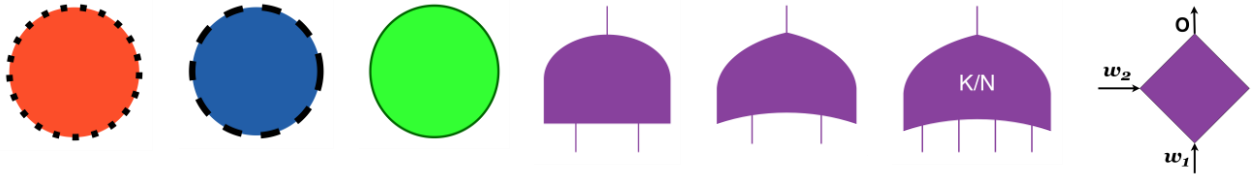
In our prior work (Soltani et al., 2024b), we proposed a model that integrates fault trees, attack trees, and defense mechanisms to form an Attack-Fault-Defense Tree (AFDT). Such a model is essential for engineers from different disciplines to identify, analyze, and mitigate safety and security risks in complex cyber-physical systems. The contributions of Soltani et al. (2024b) are summarized as follows:

- A systematic integration of attack-fault trees and attack-defense trees into a single formalism, the AFDT model, capturing safety, security, and defense mechanisms;
- An elegant visual representation that is understandable for experts with different backgrounds through the use of safety, security, and defenses;
- A qualitative analysis methodology based on minimal cut sets (MCSs), which assesses the influence of defense mechanisms on system vulnerability;
- Application of the AFDT framework to the Gridshield use-case, illustrating the model's applicability to a large, real-world cyber-physical system.

This paper extends Soltani et al. (2024b) in several key aspects. First, we enhance the AFDT analysis by incorporating a structured quantitative framework that systematically evaluates the impact of defenses on safety and security metrics. While our prior work (Soltani et al., 2024a) also introduced quantitative metrics for AFDTs, it relied on exact values for failure probabilities and attacker efforts. In contrast, this work specifically addresses scenarios where such values are only approximately known, which is particularly relevant in the cybersecurity context, where precise probabilities and costs are difficult to estimate. We propose an approach that categorizes probabilities and efforts qualitatively (e.g., low, medium, high) and demonstrates how these approximate values can still yield meaningful quantitative insights. Additionally, we apply the resulting refined quantitative method to the Gridshield use case, providing a deeper risk assessment and illustrating the added value of combining qualitative and approximate quantitative analyses. This contribution improves the practical applicability of AFDTs, particularly for large-scale, real-world systems where obtaining exact numerical data is challenging. To represent the contributions of this work relative to our prior research, Table 1 presents a systematic comparison of the methodological advancements across the three papers.

The remainder of this paper is structured as follows. Section 2 presents the AFDT models and analysis methods. Section 3 applies the AFDT model to the Gridshield defense mechanism use-case and analyzes its effectiveness. In Section 4, we include qualitative analysis using safety and security metrics, and we apply it to the Gridshield case study. Furthermore, in the same section, this paper extends Soltani et al. (2024b) by applying quantitative safety and security metrics to the Gridshield case study, offering a more actionable risk assessment. Conclusions and future work are presented in Section 5.

2 Model and Method



(a) BAS (b) BCF (c) BDS (d) AND (e) OR (f) VOT(k/n) (g) INH

Figure 1. Attack-Fault-Defense tree elements

In this section, we introduce the AFDT concept and its corresponding notation. A toy example is presented for illustration, followed by a discussion on the qualitative analysis methodology. Figure 1 illustrates various elements of AFDTs, with Fig. 1a–1c showing the different types of leaves in the form of circular nodes. The interconnection among leaves happens using logical symbols called gates.

These gates represent discrete events and can have inputs from individual leaves or other gates. The gate types AND (Fig. 1d), OR (Fig. 1e), VOT(k/n) (voting, Fig. 1f), and INH (inhibition, Fig. 1g, see 2.2), are considered in the analysis. Gates activate when the conditions for their specific logic are met.

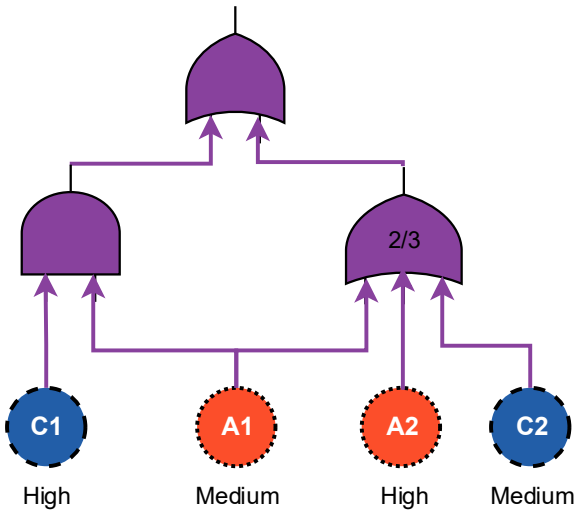


Figure 2. An AFT example. The system fails if either C_1 and A_1 both occur, or if at least 2 of A_1, A_2, C_2 occur.

2.1 Modeling safety and security

The intricate relationship between safety and security in the cyber-physical realm involves trade-offs (Soltani et al., 2021a), where measures enhancing safety may compromise security, and vice versa (Soltani et al., 2021b). Safety and security are commonly investigated. When it comes to safety, FT analysis is often used to figure out how things can go wrong, like component failures spreading throughout a system. On the security side, AT analysis helps understand how attackers might break things down into smaller steps and elements to achieve their goals. Mathematically, FTs and ATs are similar: both consist of a top-level event (TLE), representing the outcome we aim to prevent. The TLE forms the root of a hierarchical diagram, whose leaves (basic component failures (BCFs) for FTs and basic attack steps (BASs) for ATs) represent discrete, atomic events affecting the system's

integrity. Intermediate events are AND, OR, or VOT(k/n)-gates, that activate when resp. all, one, and k of its inputs are activated. The system's integrity is compromised when the TLE is activated.

When both BCFs and BASs are present, the model is an Attack-Fault Tree (AFT), which models safety, security, and their interaction simultaneously (Nicoletti et al., 2023).

Figure 2 illustrates an example of AFT with two BCFs and two BASs. The TLE is activated when both component failure C_1 and attack A_1 occur, or as a result of failure C_2 combined with attacks A_1 and A_2, with the condition that at least two out of these three events must transpire.

2.2 Attack-Fault-Defense Tree for comprehensive risk analysis

AFTs allow us to model the interaction between safety and security risks, but they do not incorporate risk mitigation strategies by the system owner. To this end, we introduced AFDT (Soltani et al., 2024a), which integrates fault trees, attack trees, and defense mechanisms into a unified model for safety and security analysis in cyber-physical systems. The paper defines the semantics and cut set metrics for AFDTs, focusing on qualitative analysis through MCSs. AFDT extends the AFT framework with defense actions that can stop risks from propagating. It also extends the concept of Attack-Defense Trees (ADTs) (Kordy et al., 2014) from security. We highlighted the need for a common modeling language that enables interdisciplinary collaboration for risk assessment in our prior work (Soltani et al., 2024a), and our primary emphasis was qualitative analysis of AFDT.

The relation between the different tree-based modeling formalisms is shown in Fig. 3.

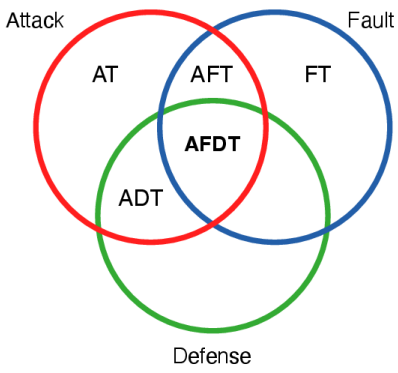


Figure 3. Venn diagram representing the tree-based formalisms discussed in this paper.

Compared to AFTs, AFDTs have two new elements. The basic defense step (BDS) is a new leaf that represents atomic actions by the defender to increase the system's resilience. Meanwhile the INH-gate serves as a countermeasure that, upon activation, thwarts the propagation of a safety/security event.

It is equipped with one input (arrow A in Fig. 1g) and a corresponding trigger (arrow B in Fig. 1g). With an active trigger, the INH gate is consistently deactivated. When the trigger is inactive, the defense gate's output mirrors the input. The mathematical formalization of the INH gate is expressed as $\neg B \wedge A$.

The AFT of Fig. 2 is extended into an AFDT in Fig. 4 by incorporating defenses. The risk posed to the TLE consisting of C_1 and A_1 can now be prevented by defense D_1, while the joint propagation of A_1, A_2, and C_2 can be prevented by D_2. Meanwhile, D_2 can be disabled by the failure C_3. All disabling relations are depicted by INH-gates. Note that some inhibiting actions are beneficial for the attacker, while others are beneficial for the defender. We color gates and propagations purple when beneficial to the attacker and green when beneficial to the defender, so that the distinction between positive and negative effects on a system is made clear visually. In more complicated systems, the distinction between positive and negative consequences not be trivial to determine. For instance, locking the door to one's house at night decreases the probability of a successful break-in (increased security) but makes escaping the house in the case of a fire more difficult (decreased safety). In general, adding defenses can introduce new vulnerabilities. Thus, a consistent color coding is not possible in general, but we will make use of it when possible to facilitate understanding.

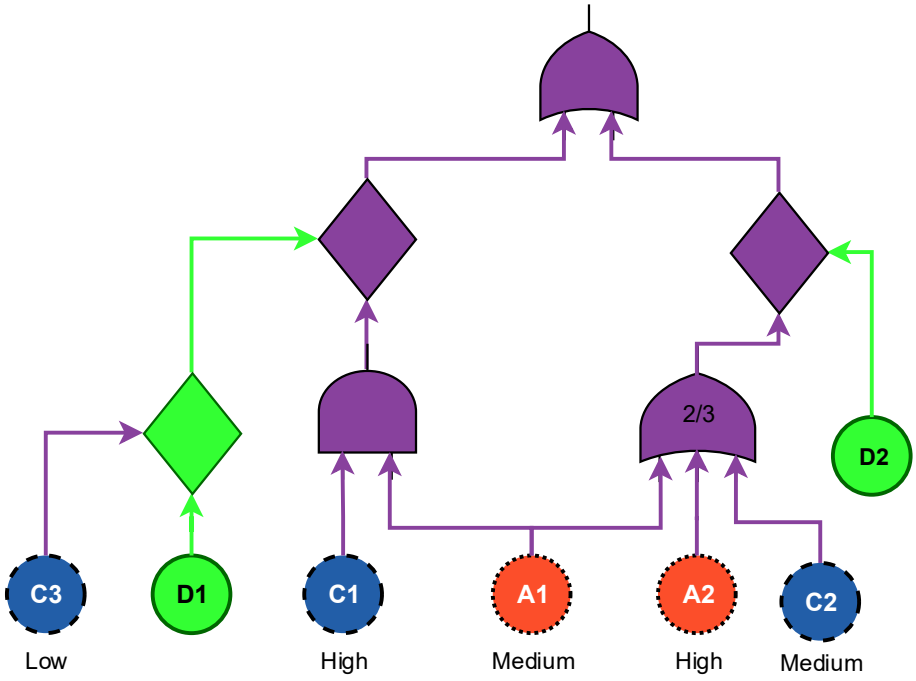


Figure 4. An exemplary depiction of AFDT. The probability of each BCF, and the required effort of each BAS, is depicted below it; see Section 4 for details.

2.3 Qualitative analysis of AFDT

AFDTs give a comprehensive, system-wide overview of how risks from safety and security can coalesce into system-level failures. It visualizes how attacks and failures propagate to a higher level and lead to the failure of the TLE, and how this propagation can be prevented by defenses. One of the most common ways of qualitative risk analysis on AFTs is finding its minimal cut sets. In an AFT, a cut set denotes any combination of components whose failure can trigger system failure, without necessarily being the most minimal combination. Conversely, an MCS is defined as a set of BCFs/BASs that, when combined, activate the TLE, and the removal of any element renders the TLE inactive. For instance, the AFT in Fig. 2 has four MCSs: {C_1, A_1}, {A_1, A_2}, {A_1, C_2}, and {A_2, C_2}. MCSs furnish information concerning the system's vulnerability and delineate the most concise path to the occurrence of the TLE. Each MCS contains a certain number of elements, and as this number

increases, propagating TLE becomes more challenging. Examining MCS gives valuable insights into the system's reliability, as the overall reliability of the system can be augmented by mitigating the attack or failure probability associated with these cut sets.

Identifying MCS in an AFDT differs from an AFT due to the presence of defenses that can prevent attack propagation. In an AFDT, the set of MCSs heavily relies on the activation of defenses: every subset of defenses that the defender chooses to activate results in a different set of MCSs. Thus, if there are n defenses, there are 2^n collections of MCSs. In theory, the collection can be an empty set ∅ as well, i.e., the activated defenses result in the system having no vulnerabilities. In Fig. 4, there are two BDSs; therefore, we have four collections of MCSs. Table 2 provides an overview of various MCSs corresponding to the activation status of each defense in Fig. 4. As indicated in Table 2, the activation of defense D_2 results in the elimination of three MCSs, indicating its effectiveness. Defense D_1 affects only one MCS, albeit increasing its element count. The simultaneous activation of both defenses results in a three-element MCS, enhancing system reliability. Note that both increasing the size of an MCS and removing it altogether increase a system's reliability.

MCS analysis in AFDT offers an overview of each defense's impact and how it relates to system reliability. When data is available, MCS analysis facilitates quantitative analysis, by considering the impact of each defense on MCS parameter, such as occurrence probability.

Table 2. MCS analysis of the toy example AFDT

No defense	$\{D_1\}$	$\{D_2\}$	$\{D_1, D_2\}$
$\{C_1, A_1\}$	$\{C_3, C_1, A_1\}$	$\{C_1, A_1\}$	$\{C_3, C_1, A_1\}$
$\{A_1, A_2\}$	$\{A_1, A_2\}$	X	X
$\{A_1, C_2\}$	$\{A_1, C_2\}$	X	X
$\{A_2, C_2\}$	$\{A_2, C_2\}$	X	X

3 Gridshield use-case

We demonstrate the usability of AFDTs on the Gridshield concept (Tangerding et al., 2022; van Sambeek et al., 2023) for smart charging of electric vehicles. Gridshield is a defense mechanism in the cyber-physical domain that overrides the maximum charging power of EV chargers. The general concept is as follows. A sender module continuously monitors the load on a critical asset and broadcasts a control message using a local LoRa communication network in case of overloading. Subsequently receiver modules, embedded within the chargers, adjust the charging power based on the received message. A change in the load on the asset is then observed at the monitored asset and subsequently following control messages can be dispatched by the sender. Moreover, Gridshield is proposed as a last-resort defense mechanism that must prevent power grid service disruptions. That is, it is an additional defense mechanism next to existing physical protection schemes (e.g., fuses) and existing digital technologies, such as Dynamic Load Management Systems (DLMS) that are often employed with charging stations.

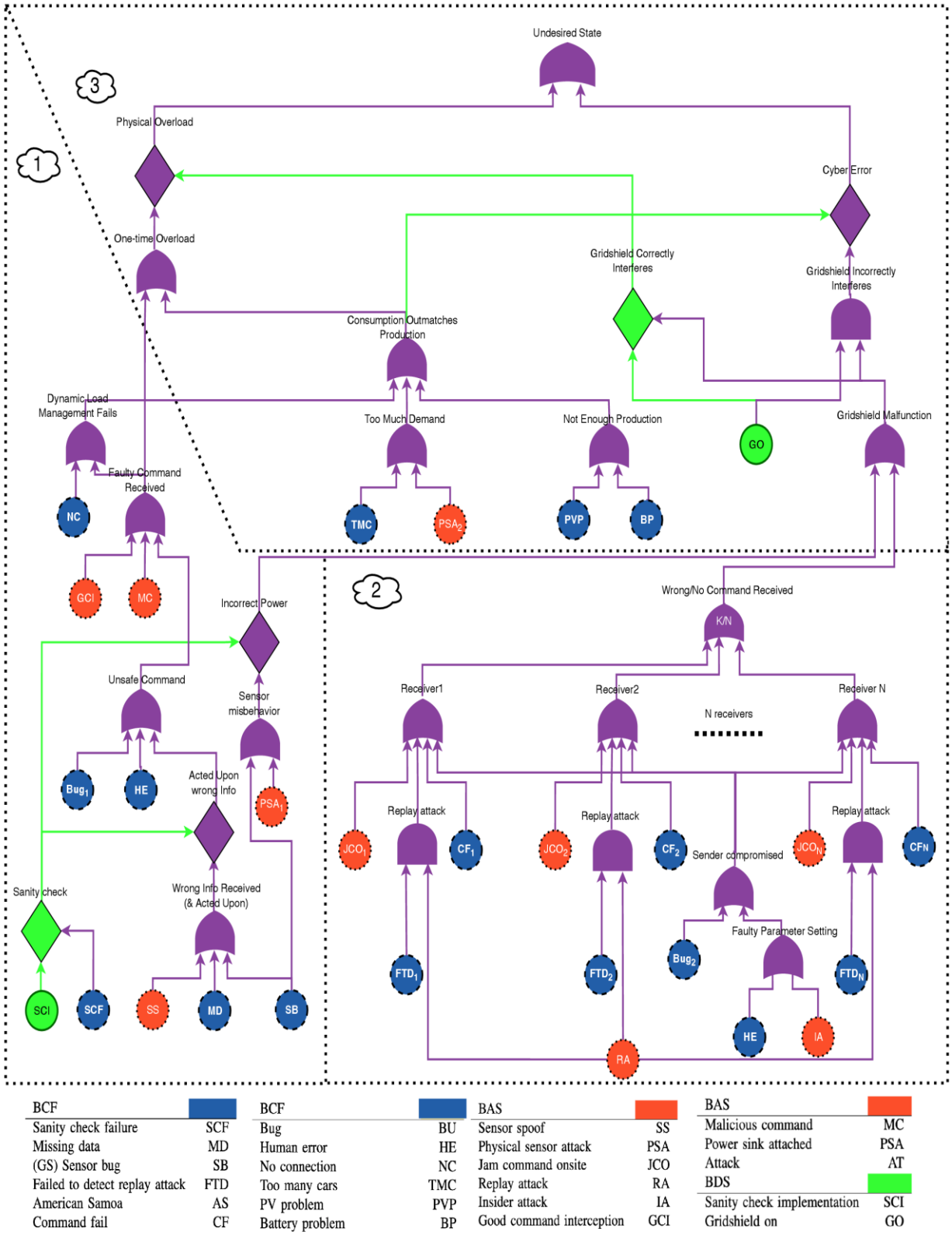


Figure 5. The Gridshield AFDT

3.1 Model

An AFDT was created based on the Gridshield concept in consultation with experts from different fields (e.g., software engineering, electrical engineering, etc.) through various iterations and discussions. This subsection presents and explains various aspects of the resulting Gridshield AFDT model, which is presented in Fig. 5.

In Fig. 5 three parts are depicted that we first briefly introduce from a system perspective before zooming into detail. Fig. 5 part 1 (left bottom) models the DLMS. This DLMS, often supplied by charging station manufacturers, performs load balancing to ensure that the total power consumption of EVs does not exceed grid limitations. The purple lines that leave box 1 indicate propagation of failures in this DLMS operation to the other parts of the system. Fig. 5 part 2 (bottom right) models the Gridshield system with the different charging stations. The effects of all individual Gridshield charging stations is aggregated, on the basis of which system malfunction may propagate to the top. The top part (Fig. 5 part 3) takes the input of the first two parts and based on that evaluates whether the combined effect, based on the (mal)functioning of the DLMS and Gridshield, and the actual load in the system, leads to a malfunction of the physical system through overloading. Note that it is possible that, despite the malfunctioning of the digital systems, still no overload occurs due to a limited physical power load on the system.

3.1.1 Dynamic Load Management

The left-hand side of the AFDT (Fig. 5 part 1) models the risks attached to the existing DLMS. These systems can prevent connection overloading during times of simultaneous charging and/or react to external energy market incentives for economic benefit. However, failure of DLMS is possible, such as entities that wish to manipulate the electricity markets or destabilize the grid balance (MC). Another risk is that DLMS may respond to wrong sensor readings or incomplete data (MD). The latter may be because of the limited local scope of many DLMS (i.e., only the local chargers) and therefore not being aware of global problems in the grid. Furthermore, many DLMS use cloud services that may malfunction (e.g., network disconnection), which results in lack of control. It is expected that basic sanity checks w.r.t. sensor readings are implemented in DLMS (SCI BDS).

3.1.2 Gridshield sender and receivers

The right-bottom side of the AFDT (Fig. 5 part 2) models the implementation of Gridshield. Herein a replication of the receiver modules can be observed at the bottom, leading to the Wrong/No Command Received VOT-gate. The receiver modules can fail in their task when they receive no commands (JCO) or malicious commands (RA), which may only affect a number of modules. A potential system failure is only propagated when k out of n receiver modules fail and form a critical mass. This k/n ratio depends on the grid capacity and power drawn by chargers. The sender module can also be compromised, e.g., by IA or HE_2. Such faults automatically propagate to all receiver modules. Another common failure is when the Gridshield modules are compromised through remote access (discussed further in Subsection 3.2.2).

Furthermore, the sender model is also implicitly modelled as the Gridshield Correctly Interferes INH-gate higher up in the tree. This is a result of the fact that correct operation of the whole Gridshield solution is normally assumed through correctly operating receiver modules. The Gridshield sender

can malfunction when it receives wrong sensor readings of the load on the asset under consideration. If the sender is unable to communicate commands, then this manifests as a JCO BAS at all receiver modules.

3.1.3 Top Level Event

The TLE to prevent is asset overloading, which is shown in the top of the model (Fig. 5 part 3). The left-hand side indicates when a potential overloading occurs when sub-systems fail, but Gridshield interferes correctly. Two things are important to observe here: 1) Gridshield can replace the defenses offered by DLMS, and 2) it has a higher level view, allowing it to also counteract malfunction of other assets such as energy production (e.g., from photovoltaics) and storage (e.g., a battery). The right-hand side provides a path for failure if overloading occurs and Gridshield does not intervene correctly, but this also requires the DLMS to fail simultaneously. Note that other events can also lead to a failure, such as a power cable malfunction due to digging. These are left out as they are not within the scope of a cyber-physical system and the risks for which Gridshield is developed.

3.2 Analysis

In this section, we perform a theoretical analysis of the Gridshield AFDT using MCSs and highlight insights that the AFDT gave in discussions regarding a practical design choice.

3.2.1 MCS analysis

As mentioned in Section 2.3, MCSs furnish information concerning the system's vulnerability and delineate the most concise path to TLE activation. Examining MCS gives valuable insights into the system's reliability, as the system's overall reliability can be augmented by mitigating the attack or failure probability associated with these MCSs. In addition, the size of the MCS is crucial, as MCSs with one element represent a single point of failure. Furthermore, we can comprehend the impact of the defenses through how they impact MCSs.

Table 3 indicates the MCSs related to the Gridshield AFDT of Fig. 5. For convenience, we consider the case $n = 4$, $k = 3$ in that Table. As it illustrates, when no defense is active there are twelve MCSs with only one element, indicating twelve single points of failure. This decreases to nine single points of failure by activating the defense SCI, as three of the one-element cut sets receive an additional element.

By activating GO the number of single points of failure decreases to five, which shows the positive effect of this defense. However, GO activation also generates new MCSs: out of five MCS of size one, four of them are generated by GO activation; in addition to $2^k\binom{n}{k} = 32$ MCSs of size $k = 3$, and $(3^k - 2^k)\binom{n}{k} = 76$ MCSs of size $k + 1 = 4$. Here, k refers to the number of failed receivers in the VOT(k/n)-gate. These new minimal cut sets arise from the fact that Gridshield may also interfere incorrectly, which requires faulty behavior from k of its n receiver modules: in each of these k receiver modules i either JCO_i or CF_i occurs ($\binom{n}{k}2^k$ possibilities); or RA occurs, an in at least one i FTD_i occurs ($\binom{n}{k}(3^k - 2^k)$ possibilities). Nonetheless, the scale of these MCSs is such that the likelihood of failure is minimal. Please note that upon the activation of the GO, many size-one MCSs are deleted because they cease to be MCSs. For instance, NC can lead to TLE failure only if SB fails or a PSA

attack occurs (that deactivates GO). Consequently, NC is no longer considered an MCS since SB and PSA become the new MCSs.

We have the fewest size one MCS (only three) when both defenses are engaged. There are two size two MCSs, $2^k\binom{n}{k} = 32$ MCSs of size $k = 3$, and $(3^k - 2^k)\binom{n}{k} = 76$ MCSs of size $k + 1 = 4$. In fact, the efficiency of the defenses is demonstrated by the fact that when both defenses are activated, we have the least single point of failure.

From the MCS analysis for the Gridshield AFDT of Fig. 5, it can be concluded that Bug_2, HE_2, and IA are the most critical elements since they are the only single points of the failure. As such, they are the most important elements that should be paid attention to, in order to reduce the probability of the TLE's failure.

Table 3. MCS analysis of the Gridshield AFDT. We consider $n = 4$ receivers, $k = 3$ of which need to fail to cause Wrong/no command received.

With no defense	While SCI is activated	While GO is activated	While both defenses are activated
{NC}	{NC}	X	X
{GCI}	{GCI}	X	X
{MC}	{MC}	X	X
{Bug_1}	{bug_1}	X	X
{HE_1}	{HE_1}	X	X
{SS}	{SCF, SS}	X	X
{MD}	{SCF, MD}	X	X
{SB}	{SCF, SB}	{SB}	{SCF, SB}
{TMC}	{TMC}	X	X
{APS}	{APS}	X	X
{PVP}	{PVP}	X	X
{BP}	{BP}	X	X
X	X	{PSA}	{SCF, PSA}
X	X	{Bug_2}	{Bug_2}
X	X	{HE_2}	{HE_2}
X	X	{IA}	{IA}
X	X	{JCO_1, ..., JCO_k}, {CF_1, ..., CF_3}, ...	{JCO_1, ..., JCO_3}, {CF_1, ..., CF_3}, ...
X	X	{RA, FTD_1, ..., FTD_3}, ...	{RA, FTD_1, ..., FTD_3}, ...
size 1 (12 ×)	size 1 (9 ×), size 2 (3 ×)	size 1 (5 ×), size 3 (32 ×), size 4 (76 ×)	size 1 (3 ×), size 2 (2 ×), size 3 (32 ×), size 4 (76 ×)

3.2.2 Practical process

Different stakeholders have had different views on how to implement the Gridshield defense system. One design decision is whether to include internet connectivity or not. The original Gridshield concept with only local LoRa connectivity has as an advantage that it cannot be compromised remotely. The disadvantage is that firmware updates or parameter change cannot be easily executed either. Incorporating such remote administrative access resulted in the introduction of the IA nodes in the scheme. While proper security and access control remains necessary, it cannot be guaranteed that no vulnerability exists that can be exploited. Moreover, the existence of such a vulnerability in a centralized control system for all Gridshield modules in the field might even lead to a massive attack

on all deployed Gridshield modules. Hence, remote access does enhance maintainability, but adds a significant risk. However, the practical effect of an IA may be limited as it still requires simultaneous malfunctioning of the DLMS and overproduction to trigger the TLE. Alternatively, chargers may be shut off, which will lead to inconvenience (i.e., EVs not being charged), but no harm to grid assets or system stability.

3.2.3 Remote access architecture decision

The AFDT analysis directly influenced a critical architectural decision regarding remote administrative access. Initial stakeholder discussions revealed two competing designs, local-only vs. remote-enabled. In a local-only design, Gridshield modules operate autonomously with firmware updates requiring physical access. This eliminates the IA (Insider Attack) node entirely from the AFDT, removing it as a single point of failure. On the other hand, a remote-enabled design supports remote administration for efficient updates and parameter tuning, introducing the IA vulnerability but improving maintainability. Based on this analysis, the remote-enabled design with additional compensating controls and monitoring was selected. This decision exemplifies how AFDT analysis transforms abstract security concerns into quantifiable engineering trade-offs based informed multi-stakeholder views. Furthermore, the identification of Bug_2 and HE_2 as persistent single points of failure (present in both designs) led to prioritizing e.g., code reviews for the sender component, as these vulnerabilities cannot be mitigated through architectural choices alone.

3.3 Discussion

The Gridshield AFDT has helped in the discussions on design choices of the Gridshield implementation with various stakeholders from different disciplines. The analysis do indicate that Gridshield implements all essential functionality that DLMS already implements and could in theory replace the DLMS functionality. However, MCS analysis also indicates that there is a huge benefit in combining both systems to further enhance the security. The resulting AFDT can also serve as a basis to design a test protocol to validate the complete system by utilizing the tree structure, in which independent triggers and responses of defense mechanisms may be observed and therefore allow for separate validation.

The current model does not incorporate weights on the different risks and their impact. This is a deliberate choice since these values are often also hard to obtain. However, qualitatively looking at the resulting AFDT, one thing is clear: attacks on the Gridshield setup require the attacker to be physically on location, except for the insider attack. Therefore, a widespread attack, leading to grid destabilization, is very unlikely. Furthermore, the existence of various DLMS vendors limits the risk that both Gridshield and DLMS malfunction simultaneously.

4 Towards a quantitative analysis

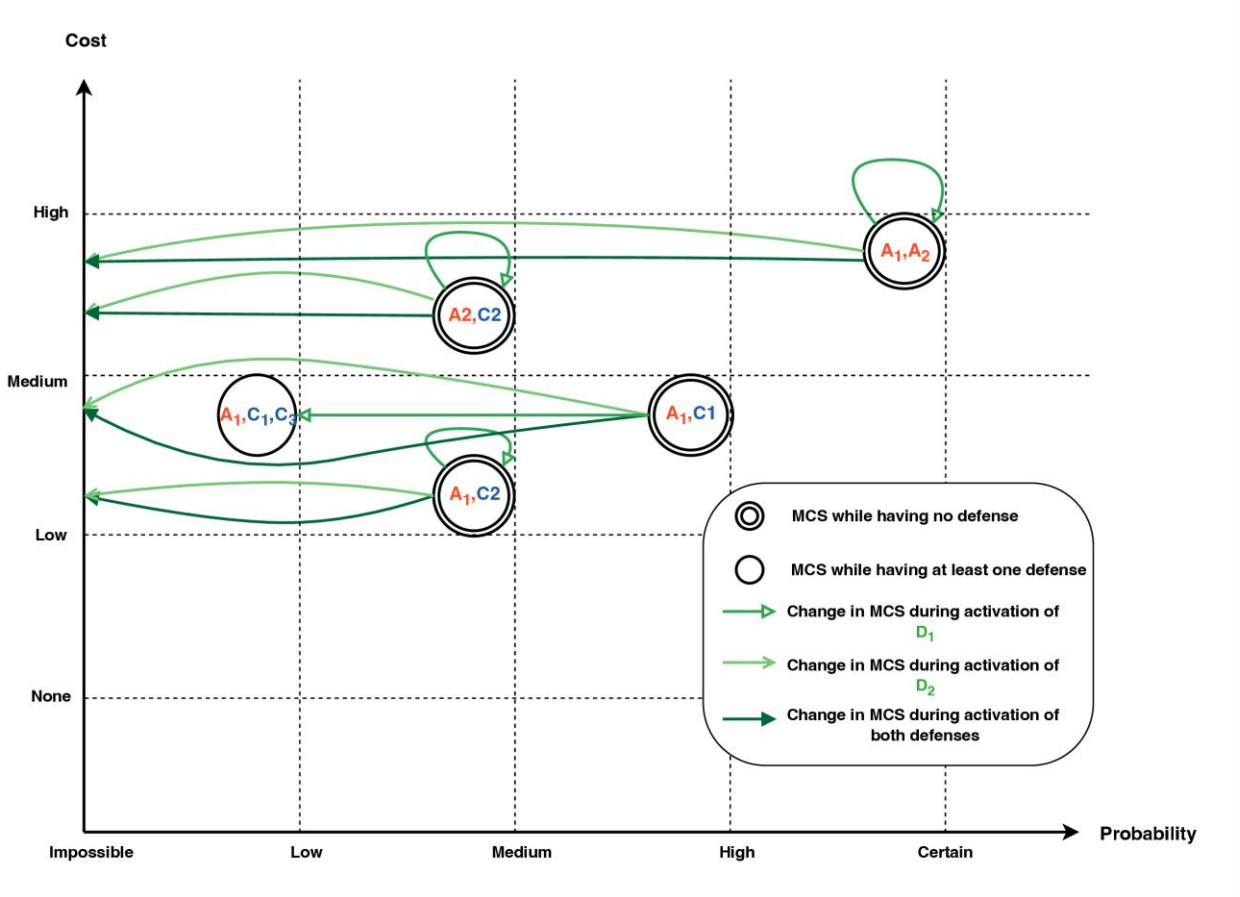


Figure 6. The impact of defenses on safety and security metrics in the AFDT of Fig. 4.

The methodology of Section 2 allows us to analyze complex systems such as Gridshield qualitatively. Unlike Section 3, which focuses solely on qualitative MCS analysis, this section introduces a systematic classification of MCSs with respect to the metrics *probability* and *attacker effort*. This new categorization helps in identifying high-risk failure scenarios and highlights the effectiveness of different defensive strategies. In contrast to our prior work in Soltani et al. (2024a), where exact numerical values for probabilities and efforts were assumed, here we specifically consider cases where these values are only approximately known (e.g., categorized as low, medium, high). This extension is especially relevant for cybersecurity contexts, where precise data is often unavailable, yet meaningful quantitative assessments are still necessary. Additionally, the impact of activating specific defenses is now assessed more quantitatively, providing a more precise understanding of their contribution to system resilience.

To each BCF C we assign an occurrence probability $p(C)$, and to each BAS A we assign the attacker effort $e(A)$ it takes to complete it. Ideally, these metrics are expressed as real numbers; however, the data required to estimate such parameters precisely is often lacking, especially in cybersecurity (Tsakalidis et al., 2018). Instead each $p(C)$ and $e(A)$ takes values `low/medium/high` in our approach, which we assign based on expert assessments. Given these values, we assign probability and effort values to each MCS as follows: For an MCS S, let S_{BCF} be its set of basic component failures, and let S_{BCF} be its set of basic attack steps (so $S = S_{\text{BCF}} \cup S_{\text{BAS}}$). We then define

$$p(S) = \min_{C \in S_{BCF}} p(C),$$

$$e(S) = \max_{A \in S_{BAS}} e(A).$$

The intuition is that an MCS S requiring multiple component failures is at least as rare as the rarest failure in S, and an MCS requiring multiple attack steps takes as least as much effort as the most 'expensive' step in S.

We set $e(S) = \text{none}$ if S does not contain any BASs, as then S's occurrence is purely probabilistic and independent of attacker effort. Likewise, we set $p(S) = \text{certain}$ if S does not contain any BCFs, as S' occurrence then only depends on attacker actions and not on probabilistic events. Note that certain here does not mean that S will always occur, but rather that it will occur any time an attacker successfully performs its constituent BASs. This aligns with the standard deterministic approach in cost-based attack tree models (Kordy et al., 2011; Lopuhaä-Zwakenberg & Stoelinga, 2023; Mauw & Oostdijk, 2006).

When more detailed quantitative data is available, one can estimate each $p(C)$ and $e(A)$ as real numbers, and compute

$$p(S) = \prod_{C \in S_{BCF}} p(C), \qquad e(S) = \sum_{A \in S_{BAS}} e(A),$$

as is done in Soltani et al. (2024a). In principle, one could also use numeric estimates for the probability/effort values of low/medium/high, and use these formulas to obtain probability/effort values for MCS. However, the downside of this approach is that without grounding in quantitative data, the numeric values of $p(C)$ and $e(A)$ may be pure guesswork, and the resulting MCS values can give a misleading sense of certainty about the system's resilience. This is especially relevant for the security part, as accurate numeric values to attacker effort are hard to impossible to obtain. We therefore do not translate the values low/medium/high to real numbers. We also do not use summation and multiplication of these non-numerical values, as that would require us to introduce rules like "low effort + low effort = medium effort", which are arbitrary without proper quantitative grounding. Instead, the operators' min/max are unambiguous and defendable when only using expert-informed assessments of the BASs/BCFs.

An additional advantage of using min/max is that it gives a more conservative risk assessment. For probability, $p(S) = \min_{C \in S_{BCF}} p(C)$ reflects that the MCS occurrence is bottlenecked by its rarest component failure which is a standard assumption in reliability engineering where component failures are treated as independent events (Ruijters & Stoelinga, 2015). For effort, $e(S) = \max_{A \in S_{BAS}} e(A)$ captures that rational attackers must invest resources for the most expensive step, with other steps potentially executable in parallel or as byproducts. In other words, we may overestimate $p(S)$, and we may underestimate $e(S)$, and both are conservative in the sense that this analysis can make our system look less resilient than it is in reality.

We can also express the impact of defenses in terms of these metrics: defenses may add BASs or BCFs to a MCS, either increasing required effort or lowering probability. This is depicted in Fig. 6

for the AFDT of Fig. 4 (see Table 2). A defense eliminating a MCS S is depicted by setting $p(S)$ to impossible or alternatively by setting $e(S)$ to ∞. This graphical depiction shows the impact of a defense at a glance, by showing to what extent it moves MCS away from the bottom right corner.

Table 4. Metric values for basic component failures and basic attack steps.

BCF	Probability	BCF	Probability	BAS	Effort
SCF	L	TMC	H	SS	M
MD	H	PVP	L	PSA	L
SB	L	BP	H	JCO	M
FTD	M			RA	H
CF	M			IA	L
Bug	L			GCI	H
HE	M			MC	H
NC	L			APS	L

Table 5. Histogram for the quantitative analysis of Gridshield. For each set of defenses, we list the number of of MCS with each combination of *effort/probability* values, for $n = 4, k = 3$. Darker coloured cells denote higher risks; bluer cells have a larger security component, while redder cells have a larger safety component.

Effort	Probability							
	L	M	H	C	L	M	H	C
	no defense				SCI			
H	-	-	-	2	-	-	-	2
M	-	-	-	1	1	-	-	-
L	-	-	-	1	-	-	-	1
N	4	1	3	-	5	1	2	-
	GO				both			
H	-	76	-	-	-	76	-	-
M	-	24	-	4	-	24	-	4
L	-	-	-	2	1	-	-	1
N	2	5	-	-	2	5	-	-

4.1 Quantitative analysis of Gridshield

We apply this approach to quantitative analysis to the Gridshield AFDT. From consulting with experts we obtained probability values for BCFs and effort values for BASs; these are given in Table 4. Combining this information with Table 3 allows us to perform quantitative analysis.

Since this AFDT has many MCS, plotting them, and the effect of the defenses, in a graph will not give us a comprehensible overview. Instead, we create tables where we list the number of MCSs for each pair of *effort/probability* values; these are given in Table 5. Again, a defense's effect can be measured by to what extent it moves values away from the bottom right corner. For SCI, we see that it is effective, as it shifts two MCSs to low probability (from high and certain probability). The effect of GO is a bit more mixed: it removes many MCSs, including 3 no effort, high probability ones, but

it also introduces a low effort, certain probability MCS. Note that the new size k MCSs end up in different places in the table: for $k = 2$ for example,

- {JCO_1,JCO_2} has medium effort, certain probability;
- {CF_1,CF_2} has no effort, medium probability;
- {JCO_1,CF_2} has medium effort, medium probability.

This is not true for the MCS of size $k + 1$: {RA,FTD_1,FTD_2}, {RA,FTD_1,JCO_2}, {RA,FTD_1,CF_2} all have medium effort, low probability. If we apply both defenses, then the low effort MCS introduced by GO is moved to low probability. Thus, these tables give a clear systematic overview of the beneficial effects of both defenses, further strengthening our conclusions of Section 3.3.

These tables also show the limitations of this approach, as it cannot distinguish two MCSs where the BCFs have the same probability, but one has more BCFs than the other. In particular, in the GO section of Table 5, the values in the high-medium effort/high probability are $(3^k - 2^k)\binom{n}{k}$ and $(2^k - 2)\binom{n}{k}$ for general n and k. This means that for high n, increasing k would appear to make the system less reliable, as this increases the number of MCS. However, this analysis does not capture that these MCS will be less likely to occur as they require more BCFs/BASs, making the system more reliable. Therefore, this type of quantitative analysis should always be combined with qualitative analysis. Furthermore, with precise data, the mathematical treatment of risk assessment could move beyond min-max operations, incorporating summation $(+)$ and multiplication $(\times)$ to model cumulative and conditional risks more accurately. Specifically, if the exact attacker effort (e.g., cost) values are known, they can be combined additively across the steps in a cut set. Similarly, if the exact failure probabilities are available, they can be multiplied to compute the cumulative probability of system failure. This approach aligns with the methodology introduced in our prior work (Soltani et al., 2024b), where exact numerical values for costs and probabilities were used to enable comprehensive quantitative evaluation. In contrast, in this paper, we demonstrate how meaningful analysis can still be performed when only approximate or categorical values (e.g., low, medium, high) are available, which is especially relevant for cybersecurity contexts where precise data is often lacking. This would enable a more granular evaluation of attack and failure likelihoods, facilitating the identification of subtle dependencies and compounding effects within the system. As a result, integrating exact quantitative data would enhance predictive accuracy and decision-making, offering a more reliable framework for mitigating safety and security risks in cyber-physical systems.

5 Conclusion and Future Work

An integral model such as AFDT is required to assess the full complexity that present-day cyber-physical systems bring to technology that we all depend upon. This work extended previous work on AFDTs (Soltani et al., 2024a) by incorporating a structured quantitative framework, enabling a deeper evaluation of risk mitigation strategies. The extensions are tested on the same Gridshield model as previously considered by Soltani et al. (2024b). The presented model aids experts from different fields in uncovering complex dependencies with respect to safety and security and analyzing how innovations may affect the security and safety of the totally intertwined system using MCS analysis.

We have furthermore augmented this analysis quantitatively by introducing a quantitative dimension via integrating probabilistic and effort-based metrics, offering a more refined analysis of defense effectiveness in real-world smart grid scenarios, thereby enhancing the accuracy of safety and security assessments. We have shown so using the Gridshield use-case and quantified how its implementation strengthens overall safety and security.

For future work, the inclusion of conventional risk and impact analysis weighting should be included for further decision making. Furthermore, the AFDT can potentially also serve as a basis for the design of test vectors and verification methods of the modelled system.

CRediT authorship contribution statement

Reza Soltani: methodology, validation, formal analysis, investigation, writing. **Baver Ozceylan:** methodology, validation, investigation, writing. **Milan Lopuhaä-Zwakenberg:** conceptualization, methodology, formal analysis, writing, supervision. **Christina Kolb:** conceptualization, methodology, formal analysis, writing. **Gerwin Hoogsteen:** conceptualization, methodology, validation, writing, supervision.

Funding or Grant

This research is conducted within the SmoothEMS met GridShield project (MOOI32005) subsidised by the Dutch ministries EZK and BZK, and by the ERC Consolidator Grant CAESAR (864075), RUBICON (101187945), Marie Skłodowska-Curie grant MISSION (101008233), and ZORRO (KICH1. ST02. 21. 003).

Use of AI

During the preparation of this work, the author(s) occasionally used ChatGPT by OpenAI solely for improving sentence structure, grammar, language clarity, and readability of the manuscript. The AI tool was not used for generating scientific ideas, research contributions, experimental design, data collection, data analysis, interpretation of results, methodological decisions, or conclusions. All concepts, technical content, analyses, results, discussions, and intellectual contributions presented in this work were developed entirely by the author(s). After using the tool, the author(s) carefully reviewed, edited, validated, and approved all content and take full responsibility for the accuracy and integrity of the publication.

Declaration of competing interests

There are no conflicts of interest.

References

Budde, C. E., Kolb, C., & Stoelinga, M. (2021). Attack trees vs. fault trees: Two sides of the same coin from different currencies. In A. Abate & A. Marin (Eds.), *Quantitative evaluation of systems: 18th International Conference, QEST 2021, proceedings* (Vol. 12846, pp. 457–467). Springer. https://doi.org/10.1007/978-3-030-85172-9_24

Kabir, S. (2017). An overview of fault tree analysis and its application in model based dependability analysis. *Expert Systems with Applications, 77*, 114–135. https://doi.org/10.1016/j.eswa.2017.01.058

Kordy, B., Mauw, S., Radomirović, S., & Schweitzer, P. (2011). Foundations of attack-defense trees. In P. Degano, S. Etalle, & J. Guttman (Eds.), *Formal aspects in security and trust: 7th International Workshop, FAST 2010, revised selected papers* (Vol. 6561, pp. 80–95). Springer. https://doi.org/10.1007/978-3-642-19751-2_6

Kordy, B., Mauw, S., Radomirović, S., & Schweitzer, P. (2014). Attack-defense trees. *Journal of Logic and Computation, 24*(1), 55–87. https://doi.org/10.1093/logcom/exs029

Lopuhaä-Zwakenberg, M., & Stoelinga, M. (2023). Cost-damage analysis of attack trees. In *2023 53rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)* (pp. 545–558). IEEE. https://doi.org/10.1109/DSN58367.2023.00057

Mauw, S., & Oostdijk, M. (2006). Foundations of attack trees. In D. Won & S. Kim (Eds.), *Information security and cryptology – ICISC 2005: 8th International Conference, Seoul, Korea, December 1–2, 2005, revised selected papers* (Vol. 3935, pp. 186–198). Springer. https://doi.org/10.1007/11734727_17

Menzel, V., Arias, N. B., Hurink, J. L., & Remke, A. (2023). Securing smart grids locally using a power flow-based intrusion detection system. In *2023 IEEE Belgrade PowerTech* (pp. 1–9). IEEE. https://doi.org/10.1109/PowerTech55446.2023.10202762

Nicoletti, S. M., Peppelman, M., Kolb, C., & Stoelinga, M. (2023). Model-based joint analysis of safety and security: Survey and identification of gaps. *Computer Science Review, 50*, Article 100597. https://doi.org/10.1016/j.cosrev.2023.100597

Ruijters, E., & Stoelinga, M. (2015). Fault tree analysis: A survey of the state-of-the-art in modeling, analysis and tools. *Computer Science Review, 15–16*, 29–62. https://doi.org/10.1016/j.cosrev.2015.03.001

Soltani, R., Kang, E.-Y., & Heredia Mena, J. E. (2021a). Towards energy-aware cyber-physical systems verification and optimization. In M. Ganzha, L. Maciaszek, M. Paprzycki, & D. Ślęzak (Eds.), *Position and communication papers of the 16th Conference on Computer Science and Intelligence Systems* (Vol. 26, pp. 205–210). Polish Information Processing Society. https://doi.org/10.15439/2021F125

Soltani, R., Kang, E.-Y., & Heredia Mena, J. E. (2021b). Verification and optimization of cyber-physical systems: Preprint for FedCSIS [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2109.01574

Soltani, R., Lopuhaä-Zwakenberg, M., & Stoelinga, M. (2024a). Safety-security analysis via attack-fault-defense trees: Semantics and cut set metrics. In A. Ceccarelli, M. Trapp, A. Bondavalli, & F. Bitsch (Eds.), *Computer safety, reliability, and security: 43rd International Conference, SAFECOMP 2024, proceedings* (Vol. 14988, pp. 218–232). Springer. https://doi.org/10.1007/978-3-031-68606-1_14

Soltani, R., Özceylan, B., Lopuhaä-Zwakenberg, M., Kolb, C., & Hoogsteen, G. (2024b). Safety and security dependencies for Gridshield. In*2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)* (pp. 1–6). IEEE. https://doi.org/10.1109/ISGTEUROPE62998.2024.10863084

Tangerding, F., Varenhorst, I. A. M., Hoogsteen, G., Gerards, M. E. T., & Hurink, J. L. (2022). GridShield: A robust fall-back control mechanism for congestion management in distribution grids. In *2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)* (pp. 1–5). IEEE. https://doi.org/10.1109/ISGT-Europe54678.2022.9960301

Tsakalidis, G., Vergidis, K., Madas, M., & Vlachopoulou, M. (2018). Cybersecurity threats: A proposed system for assessing threat severity. In F. Dargam, P. Delias, I. Linden, & B. Mareschal (Eds.), *Decision support systems VIII: Sustainable data-driven and evidence-based decision support* (pp. 284–291). Springer.

van Sambeek, H. L., Zweistra, M., Hoogsteen, G., Varenhorst, I. A. M., & Janssen, S. (2023). GridShield—Optimizing the use of grid capacity during increased EV adoption. *World Electric Vehicle Journal, 14*(3), Article 68. https://doi.org/10.3390/wevj14030068

Yadav, S. A., Kumar, S. R., Sharma, S., & Singh, A. (2016). A review of possibilities and solutions of cyber attacks in smart grids. In *2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH)* (pp. 60–63). IEEE. https://doi.org/10.1109/ICICCS.2016.7542359