Cybersecurity Education in Multi-Domain Operations: An Automotive Case Study for Defence Trustworthy AI*

Danilo Caivano[1,†], Christian Catalano[1,†], Gabriel Cellamare[1,†], Samuele del Vescovo[2,*,†] and Michele Scalera[1,†]

[1]Università degli studi di Bari Aldo Moro, Piazza Umberto I, 70121 Bari, Apulia, Italy
[2]Scuola IMT Alti Studi Lucca, Piazza S.Francesco, 19, 55100 Lucca, Italy

Abstract

The concept of Multi-Domain Operations (MDOs) has gained increasing recognition within both civilian and military strategic discourse, highlighting the necessity of integrating capabilities across multiple domains to achieve synergistic negative effects. So, the vehicles' attack surface is in expansion in unprecedented ways due to the integration of them in Smart City systems, exposing critical continental-scale networks that are essential for governmental and military operations. Vehicles can be perfect victims of attacks linked to future complex MDOs with the ultimate goal of affecting people-related effect dimension's. However, Multi-Surface Threats (MSTs) impacting several In-Vehicle systems simultaneously cannot be ruled out. One of these MDTs i.e. Adversarial Machine Learning (AML) can target In-Vehicle Machine Learning (ML) based security systems like Intrusion Detection Systems (IDSs) and networks for Traffic Sign Recognition Systems (TSRSs) simultaneously. So, the primary goal of this work is to investigate the potential relevance of specific hyperparameters associated with Decision Tree (DT)-based ensemble models on which the Supervised ML Intrusion Detection System (IDS) is made up for the CAN Bus Frame Detection task. These hyperparameters are evaluated in terms of their capacity to function as inherent defensive mechanisms (or deterrents) against a Black-Box AML attack i.e. the Zeroth Order Optimization (ZOO), particularly when conceptualized as the "Cyber" component within MDOs. The targeted IDS models represent Technology Transfer (TT) state-of-the-art approaches including Random Forests (RF) based on bagging trees, Gradient Boosting (GB) and Extreme Gradient Boosting (XGB). The number of bagging trees in RF and the number of boosting rounds in GB affect the time required to perform the attack. In contrast, the same parameters for the XGB does not exhibit the same influence. Thus, identifying optimal configurations for these parameters may serve as a concrete example of Trustworthy AI practice particularly useful to defeat Single Surface Threats (SSTs) as a part of MSTs in Multi-Domain (M-D) scenarios. The ultimate goal of this work is to contribute to proper education regarding the responsible development of AI/ML-based systems by industries and academies (civilian/military public/private) by evaluating the positive impact of the examinated values.

Keywords

Trustworthy AI, Automotive Cybersecurity, Multi-Domain Operations

1. Introduction

Multi-Domain Operations (MDOs) represent the contemporary paradigm in military strategy, aiming to generate synergistic effects through the integration of capabilities across multiple operational domains (i.e. Space, Air, Land, Sea and Cyber). This concept entails the execution of coordinated actions within diverse environments, assets, tools and methodologies, with the strategic objective of countering adversarial strengths and asymmetries [1, 2]. Such operations are structured to impose simultaneous operational and tactical challenges on opponents by leveraging a well-calibrated force posture and the coherent integration of resources across physical and informational environments. These actions are spatially and temporally distributed, while simultaneously creating complex multifaceted dilemmas for the adversary [3]. From the defence point-of-view, Multi-Domain Task Forces (MDTFs) are useful to

Joint Proceedings of IS-EUD 2025: 10th International Symposium on End-User Development, 16-18 June 2025, Munich, Germany.
*Corresponding author.
†These authors contributed equally.
danilo.caivano@uniba.it (D. Caivano); christian.catalano@uniba.it (C. Catalano); g.cellammare1@studenti.uniba.it (G. Cellamare); samuele.delvescovo@imtlucca.it (S. d. Vescovo); michele.scalera@uniba.it (M. Scalera)
0000-0001-5719-7447 (D. Caivano); 0000-0003-4038-2317 (C. Catalano); 0009-0001-8220-5135 (S. d. Vescovo); 0000-0002-2455-2032 (M. Scalera)

integrate defence resources across physical and informational environment.

A fundamental tenet of MDOs lies in the integration of synchronized kinetic and cyber capabilities across various operational domains, with the objective of imposing ”multiple dilemmas” on adversaries [4]. The main problems from the defensive point-of-view are inadequate synchronization mechanisms (across various domains). To overcome these limitations, the Cyber Social Security (CSS) framework has been proposed as a means of enabling the effective incorporation of cyber defence operations within Multi-Domain (MD) strategic architectures, giving rise to the CSS-MDO framework, designed for defense-oriented applications [5, 6]. The horizontal axis delineates the five active warfare domains, each capable of employing tailored tools, methodologies, and procedures that correspond to the Detection-Response-Prevention cycle (represented on the vertical axis). By defining these vertical operational tiers (in which the MDTFs can fall), the model provides a basis for managing cyber impacts in civilian contexts involving attacks aimed at the cognitive dimension of individuals [7].

A particular asset involving Cyber and Land domains in future MDOs is the Smart City. Specific vulnerable assets in the “dense network” of a Smart City’s asset are Connected and Autonomous Vehicles (CAVs) [8] originating the concept of the Internet of Vehicles (IoV). These are the core of the future evolution of shared mobility (as well as the evolution of electric mobility) useful to optimising any sustainable travel [9]. In this great climate of innovation, the attack surface exploitable by malicious actors will take forms difficult for any "cyber social" blue team to understand. For example, it would be very easy for any attacker (more or less skilled) to attack Controller Area Network (CAN) protocol-based In-Vehicles Networks (IVNs) [10] related to civilian-use vehicles by exploiting a set of activities (including military ones) conducted through different domains to perceive, understand, and orchestrate "dilemmas" [11]. Certainly, Artificial Intelligence (AI) and Machine Learning (ML) are powerful tools that could prevent these attacks ultimatly targetting the psychological and physical well-being of passengers [12]. Some of these approaches can be ML-based Intrusion Detection Systems (IDSs) [13] but Traffic Sign Recognition Systems (TSRS) too. Adversarial Machine Learning (AML) is considered one of the most AI/ML systems’ threats [14, 15], especially in Multi-Domain Threats (MDTs) scenario. In the context of evasion-based attacks, attackers can craft input data at testing (or deployment) time [16], modifying features such as pixels in an image or values in a CAN bus frame in an imperceptible way to humans, yet causes the targeted model to predict incorrect classifications [17, 18]. The Black-Box setting of these is regarded as both the most realistic and the most accessible from the attacker’s perspective, as it does not require any prior access to the internal details of the victim system [14, 19]. However, ispriating to the concept of MDTs (in which the attacker acts on multiple domain assets even at the same time), it is reasonable to imagine threats related to impacting attacks on different surfaces of the same asset during MDOs. This concept can be coined as Multi-Surface Treats (MSTs), recognized as (possible) parts of MDTs. Current literature on the application or conceptualization of Black-Box attacks within the CAN bus frame detection task remains limited and in an early stage of development (even in MSTs and MDTs scenarios).

Therefore, this paper presents an empirical investigation into the role of specific hyperparameters associated with Decision Tree (DT)-based ensemble models i.e. Random Forest (RF), Gradient Boosting (GB), and Extreme Gradient Boosting (XGB) used as the core of supervised ML-based IDS in the context of CAN Bus Frame Detection task. It is assumed that the IDS is installed onboard the vehicle and it is subjected to a Black-Box Adversarial Machine Learning (AML) attack i.e. the Zeroth Order Optimization (ZOO) (in a pure evasive Black-Box setting). This type of attack is conceptualized as a Single-Surface Threat (SST) seen as a part of a complex MST, falling into the Cyber component of a complex MDO. The core of the analysis lies in evaluating how variations in selected hyperparameters affect the time required to generate adversarial examples for each targeted ML model. Experimental findings indicate that the number of bagging trees in RF and the number of boosting rounds in GB models have a significant impact on the time needed for the attack. Conversely, the same does not hold for the boosting rounds in XGB. These hyperparameters, in the cases of RF and GB, can thus be interpreted as intrinsic defense or deterrence mechanisms against the ZOO attack. Appropriately tuning these hyperparameters may exemplify a trustworthy AI-by-design approach for In-Vehicle ML systems’ robustness [20, 16]. In particular, the work underscores the relevance of robustness [21] and security [22] properties in the

design and deployment of ML models within adversarial environments [23]. Moreover, the secondary goal of this work is to qualitatively identify the positive impact resulting from educating MDTFs about the best practices programming (i.e. appropriate values for the previously mentioned hyperparameters) related to ZOO countermeasures (in a pure Black-Box scenario) on the "Detection", "Response" and "Prevention" axes of the CSS-MDO framework useful to defeat MDTs and MSTs in future cyber social scenarios [5, 24]. The central aim of this work is to improve a multidimensional national deterrence strategy [7].

In summary, the research questions (RQs) are:

- RQ1: "Can the hyperparameters realted to the number of bagging trees in Random Forest (RF), the number of boosting rounds in Gradient Boosting (GB), and the number of boosting rounds in Extreme Gradient Boosting (XGB) affect the time needed to generate ZOO adversarial examples, when applied to supervised ML-based Intrusion Detection Systems (IDS) in the context of the CAN Bus Frame Detection Task (in a Black-Box attack scenario)?"
- RQ2: "Is it possible to qualitatively quantify the (positive) impact of these values on the "Detection", "Response" and "Prevention" axes of the CSS in MDOs framework (for educational purpouse)?"

2. Methodology

This phase addresses all the RQs. In order to answer to RQ1, considering the possibility of MSTs (i.e. impacting several In-Vehicle systems simultaneously), an example of that could be a simultaneous threat on the In-Vehicles networks protected by an IDS for CAN bus frame detection and on the network enabling the TSRS. Figure 1 describe the current attack scenario. Regarding the first SST i.e. the CAN based, it is reasonable to think that an ideal scenario for any Vehicle-SOC involves deploying a ML-based IDS that maximizes the time required by an adversary to generate adversarial examples. Based on this premise, certain hyperparameters of ensemble-based ML models, specifically those pertaining to RF, GB, and XGB may serve as intrinsic defense mechanisms by influencing the computational cost of adversarial example generation. This strategy aims to enhance organizational resilience by increasing the adversarial effort necessary to breach the system. The methodological approach adopted in this analysis entails, for each ML model (RF, GB, and XGB), the measurement of the time (in seconds) required to generate 92,270 adversarial examples for incrementally varied hyperparameter values. These include the number of bagging trees in RF, and the number of boosting rounds in GB and XGB, respectively. Each observation is captured following approximately five minutes of computation.

In order to answer to RQ2, considering the critical nature of the scenario that surrounds this research work, an impact (positive) analysis realted to the MDTF education (acting along the CSS-MDO framework vertical axes) about the best programming practices that improve the resilience of ML models to Black-Box attacks is an important milestone to underline the right importance of these. The actual qualitative analysis is based on the Land and Cyber domains. This analysis comes from an high-level qualitative risk assessment related to this SST. This second analysis is adopted considering multiple hypothetical negative consequences: the potential to incite a climate of terror through anomalous vehicle behavior and the cognitive disruption of civilian and military operators, the reputational demage to the national infrastructure and institutions [25], and the inherently risk-averse perspective guiding the human evaluator point-of-view. Accordingly, this RQ seeks to underscore this critical need for future research.

3. Conclusion & Future Work

MDOs integrate capabilities across all the active warfare domains to attack victims through synchronized cross-domain impacts. Smart Cities and especially CAVs represent critical assets in this scenario, vulnerable to cyberattacks based on Black-Box AML. The target of these can be the ML-based IDS for CAN-based IVNs security. These attacks pose significant risks within MDTs and especially MSTs scenarios (considering other In-Vehicle user-assistance systems). Despite growing concerns, the literature

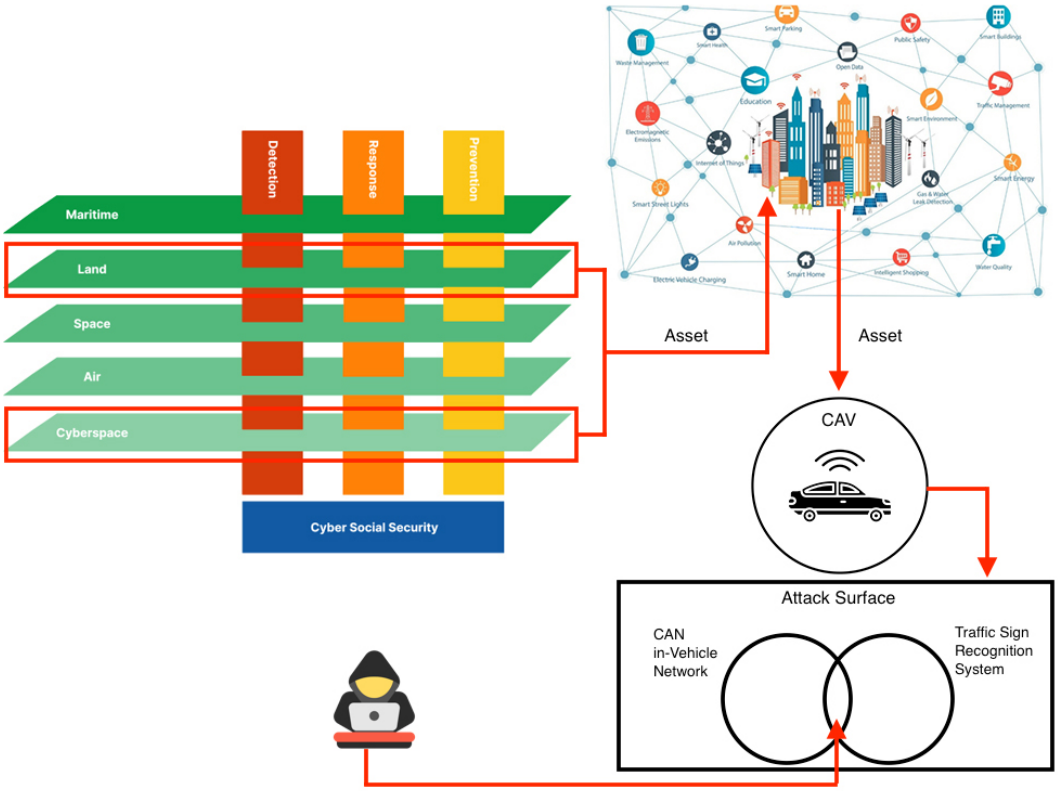

Figure 1: MST under exam

on Black-Box AML targeting CAN frame detection remains limited and underdeveloped (even in a M-D scenario). Therefore, this paper has the goal of explore the possible influence of some hyperparameters related to DT-based state-of-the-art Ensemble models (i.e. RF, GB, XGB) underlying an IDS, victim of the ZOO attack (in a purely Black-Box scenario) seen as a MST, on the time needed to the generation of adversarial examples is evaluated (RQ1) in the MDTs scenario. In addition, the impact of such on the CSS framework for MDOs is qualitatively evaluated (RQ2). Some future directions are: assess the impact of this attack also on a TSRS in a MDTs componing future MDOs.

Acknowledgments

This work was partially supported by the following projects: SERICS - ”Security and Rights In the CyberSpace - SERICS” (PE00000014) under the MUR National Recovery and Resilience Plan funded by the European Union - NextGenerationEU; Casa delle Tecnologie Emergenti del Comune di Bari – “Bari Open Innovation Hub” – CUP J99J19000300003.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] F. T. and, Nato’s approach to multi-domain operations: From the perspective of the economics of alliances, Defence and Peace Economics 35 (2024) 281–294.

URL: https://doi.org/10.1080/10242694.2023.2235502. doi:10.1080/10242694.2023.2235502. arXiv:https://doi.org/10.1080/10242694.2023.2235502.

[2] C. Pardo, F. J. Pino, F. García, M. Piattini, M. T. Baldassarre, S. Lemus, Homogenization, comparison and integration: A harmonizing strategy for the unification of multi-models in the banking sector, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6759 LNCS (2011) 59 – 72. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-79960278328&doi=10.1007%2f978-3-642-21843-9_7&partnerID=40&md5=b84e6a16df8aa4edd29f46db95f3cc02. doi:10.1007/978-3-642-21843-9_7.

[3] A. Gilli, M. Gilli, G. G. and, Nato, multi-domain operations and the future of the atlantic alliance, Comparative Strategy 44 (2025) 73–91. doi:10.1080/01495933.2024.2445491. arXiv:https://doi.org/10.1080/01495933.2024.2445491.

[4] F.-S. Gady, A. Stronell, Cyber capabilities and multi-domain operations in future high-intensity warfare in 2030, Cyber Threats and NATO 2030: Horizon Scanning and Analysis (2020) 151.

[5] V. S. Barletta, M. Calvano, A. Sciacovelli, Cyber social security in multi-domain operations, in: 2024 IEEE International Workshop on Technologies for Defense and Security (TechDefense), 2024, pp. 41–46. doi:10.1109/TechDefense63521.2024.10863352.

[6] M. T. Baldassarre, M. Piattini, F. J. Pino, G. Visaggio, Comparing iso/iec 12207 and cmmi-dev: Towards a mapping of iso/iec 15504-7, 2009, p. 59 – 64. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-78049481915&doi=10.1109%2fWOSQ.2009.5071558&partnerID=40&md5=b3721f95d723e13988b78319b6df10a2. doi:10.1109/WOSQ.2009.5071558.

[7] S. G. della Difesa Italiana, The italian defence approach to multi-domain operations (approccio della difesa alle operazioni multidominio), https://www.difesa.it/assets/allegati/31787/2.1defence_approach_to_mdos.pdf, 2022.

[8] M. A. Richter, M. Hagenmaier, O. Bandte, V. Parida, J. Wincent, Smart cities, urban mobility and autonomous vehicles: How different cities needs different sustainable investment strategies, Technological Forecasting and Social Change 184 (2022) 121857. URL: https://www.sciencedirect.com/science/article/pii/S004016252200381X. doi:https://doi.org/10.1016/j.techfore.2022.121857.

[9] H. Olufowobi, G. Bloom, Chapter 16 - connected cars: Automotive cybersecurity and privacy for smart cities, in: D. B. Rawat, K. Z. Ghafoor (Eds.), Smart Cities Cybersecurity and Privacy, Elsevier, 2019, pp. 227–240. doi:https://doi.org/10.1016/B978-0-12-815032-0.00016-0.

[10] K. N, V. Ravi, V. Sowmya, Unsupervised intrusion detection system for in-vehicle communication networks, Journal of Safety Science and Resilience 5 (2024) 119–129. URL: https://www.sciencedirect.com/science/article/pii/S2666449624000070. doi:https://doi.org/10.1016/j.jnlssr.2023.12.004.

[11] F. Tommasi, C. Catalano, M. Fornaro, I. Taurino, Mobile session fixation attack in micropayment systems, IEEE Access 7 (2019) 41576–41583. doi:10.1109/ACCESS.2019.2905219.

[12] E. U. A. for Cybersecurity, G. Dede, R. Naydenov, A. Malatras, C. E. C. C. de Investigación, R. Hamon, H. Junklewitz, I. Sanchez, E. C. J. R. Centre, Cybersecurity Challenges in the Uptake of Artificial Intelligence in Autonomous Driving, EUR (Luxembourg. Online), Publications Office of the European Union, 2021. URL: https://books.google.it/books?id=9oZbzgEACAAJ.

[13] A. Alfardus, D. B. Rawat, Intrusion detection system for can bus in-vehicle network based on machine learning algorithms, in: 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2021, pp. 0944–0949. doi:10.1109/UEMCON53757.2021.9666745.

[14] B. Wu, Z. Zhu, L. Liu, Q. Liu, Z. He, S. Lyu, Attacks in adversarial machine learning: A systematic survey from the life-cycle perspective, 2024. arXiv:2302.09457.

[15] C. Xie, Z. Cao, Y. Long, D. Yang, D. Zhao, B. Li, Privacy of autonomous vehicles: Risks, protection methods, and future directions, 2022. arXiv:2209.04022.

[16] and European Union Agency for Cybersecurity, A. Malatras, I. Agrafiotis, M. Adamczyk, Securing machine learning algorithms, 2021. URL: https://op.europa.eu/publication-detail/-/publication/

c7c844fd-7f1e-11ec-8c40-01aa75ed71a1. doi:doi/10.2824/874249.

[17] F. Aloraini, A. Javed, O. Rana, Adversarial attacks on intrusion detection systems in in-vehicle networks of connected and autonomous vehicles, Sensors 24 (2024). URL: https://www.mdpi.com/1424-8220/24/12/3848. doi:10.3390/s24123848.

[18] S. Longari, F. Noseda, M. Carminati, S. Zanero, Evaluating the robustness of automotive intrusion detection systems against evasion attacks, in: Cyber Security, Cryptology, and Machine Learning: 7th International Symposium, CSCML 2023, Be'er Sheva, Israel, June 29–30, 2023, Proceedings, Springer-Verlag, 2023, p. 337–352. URL: https://doi.org/10.1007/978-3-031-34671-2_24. doi:10.1007/978-3-031-34671-2_24.

[19] S. Kotyan, A reading survey on adversarial machine learning: Adversarial attacks and their understanding, 2023. arXiv:2308.03363.

[20] E. U. A. for Cybersecurity (ENISA), Artificial intelligence and cybersecurity research, 2023. URL: https://www.enisa.europa.eu/publications/artificial-intelligence-and-cybersecurity-research. doi:10.2824/808362.

[21] H.-L. E. G. on AI European Commission, Ethics guidelines for trustworthy ai, 2024. URL: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

[22] N. I. of Standards, Technolgy, Ai fundamental research - security, 2023. URL: https://www.nist.gov/artificial-intelligence/ai-fundamental-research-security.

[23] S. Goellner, M. Tropmann-Frick, B. Brumen, Responsible artificial intelligence: A structured literature review, 2024. URL: https://arxiv.org/abs/2403.06910. arXiv:2403.06910.

[24] M. T. Baldassarre, V. S. Barletta, D. Caivano, D. Raguseo, M. Scalera, Teaching cyber security: The hack-space integrated model, in: Italian Conference on Cybersecurity, volume 2315, 2019. URL: https://ceur-ws.org/Vol-2315/paper06.pdf.

[25] V. S. Barletta, D. Caivano, C. Catalano, S. D. Vescovo, Black-box adversarial ml attacks on ids and multi-domain impact analysis for threat intelligence in automotive scenarios, in: 2024 IEEE International Workshop on Technologies for Defense and Security (TechDefense), 2024, pp. 132–137. doi:10.1109/TechDefense63521.2024.10863442.