A Deep Learning Approach for False Data Injection Attacks Detection in Smart Water Infrastructure

Davide Giannubilo[1,*], Tommaso Giorgeschi[1,*], Michele Carminati[1], Stefano Zanero[1] and Stefano Longari[1,**]

[1]*Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Via Ponzio 34/5, 20133, Milan, Italy*

Abstract

Cyber-Physical systems (CPS) represent a sophisticated integration of digital technologies with physical processes, particularly vital in critical environments such as smart water infrastructures, which require advanced monitoring and control systems to guarantee safe and resilient operations, especially in the context of attacks. This study introduces a novel unsupervised deep learning approach for detecting false data injection (FDI) attacks in smart water infrastructures. The method employs Long Short-Term Memory (LSTM) networks and Autoencoders to discern the legitimate behavior of time-series water level sensor data. We evaluate this approach using the Mincio River water system in Italy as a case study, employing publicly available data augmented with synthetic—yet realistic—random, replay, and advanced attack scenarios. The experimental results demonstrate the effectiveness of the proposed method in distinguishing anomalies from legitimate data, highlighting its potential for enhancing the security of smart water systems.

Keywords

Intrusion Detection Systems, Cyber Physical Systems, Critical Infrastructure Security

1. Introduction

Cyber-Physical systems (CPS) integrate the cyber domain, comprising networked components and servers, with the physical domain, consisting of sensors and actuators [1]. These systems monitor and control physical processes through feedback loops, where physical processes influence cyber operations [2]. Widely employed across critical infrastructure such as industrial control, automotive systems, smart grids, and water treatment [3], Cyber-Physical System (CPS)s form the backbone of modern society. Given that the disruption of the operations may severely affect society, the security of such systems is paramount. However, the integration of physical and cyber components opens such systems to cyber-physical attacks. These attacks manipulate control systems governing physical processes, such as power grids or transportation systems, by exploiting digital interfaces to cause tangible disruptions. For example, attackers could compromise smart water infrastructure to manipulate water flow or disrupt operations.

Water and river systems are critical applications of CPSs, particularly in managing resources and mitigating environmental risks. These systems are specifically at risk against False Data Injection (FDI) attacks, which would allow, for instance, an adversary to alter sensor readings at the critical station, leading to the premature release of excess water downstream. This could result in catastrophic flooding and significant infrastructure damage.

To protect such systems, robust anomaly detection methods are necessary. Early solutions relied on mathematical models, but recent advancements incorporate Machine Learning (ML) and Deep Learning (DL) techniques. These approaches, including Long Short-Term Memory (LSTM) networks and Autoencoder (AE), excel at capturing temporal and spatial correlations in sensor data. By learning

Joint National Conference on Cybersecurity (ITASEC & SERICS), February 03-8, 2025, Bologna, IT

** Corresponding author.

* The authors contributed equally to this work.

davide.giannubilo@mail.polimi.it (D. Giannubilo); tommaso.giorgeschi@mail.polimi.it (T. Giorgeschi); michele.carminati@polimi.it (M. Carminati); stefano.zanero@polimi.it (S. Zanero); stefano.longari@polimi.it (S. Longari)

0000-0002-7533-4510 (S. Longari)

legitimate sensor behavior from historical data, these methods can identify anomalies indicative of potential attacks.

Our work focuses on the design of an Intrusion Detection System (IDS) leveraging LSTM autoencoders to detect anomalies in smart water infrastructure systems. The system reconstructs expected sensor behavior and calculates anomaly scores to flag deviations. We validate our approach on a use case based on the Mincio River water infrastructure in Italy. This system monitors and controls water flow from the Garda Lake to the Po River using strategically located control points and dams. Sensors measure water height at 15-minute intervals, and a critical station regulates flow into an artificial canal based on real-time sensor data. We use publicly available data and simulated FDI attacks crafted for this purpose.

The contributions of this study are threefold:

- The introduction of a novel Intrusion Detection System (IDS) in smart water systems.
- The creation of a new dataset specifically for validating IDS approaches.
- The evaluation of the performances of the proposed algorithm on a real-world case study, the Mincio River infrastructure.

2. Primer on Smart Water Infrastructures

In recent years, research on water management systems has predominantly focused on traditional water infrastructures, which rely on fewer advanced technologies and smart devices. However, the advent of smart water infrastructures – a distinct category of cyber-physical systems – is transforming the manner in which water systems are managed. A smart water infrastructure is defined as a system that integrates a range of cutting-edge technologies, including sensors, real-time communication capabilities (via wireless networks, satellite communications, etc.), automated controls, and artificial intelligence. Such improvements facilitate more efficient monitoring and management of water resources, representing a significant upgrade over traditional systems.

The impetus behind the transition to smart water systems can be attributed to the pressing global challenges of rising water scarcity and the considerable financial burden associated with the provision of potable water to expanding populations. The deployment of smart water infrastructures offers a number of advantages over conventional systems. These include more accurate measurement of water consumption, improved water quality control, advanced flood monitoring, and effective prevention of water wastage. To illustrate, a typical smart water infrastructure may entail the collection of water from rivers or seas, followed by its transportation to a water treatment facility where it undergoes purification. Subsequently, the water is either stored or delivered to consumers via a water distribution system, with real-time monitoring and optimization facilitated by advanced technologies.

A fully realized smart water system is comprised of a multitude of components and devices, each of which plays a crucial role in ensuring the system's efficiency, reliability, and security. The process starts with a set of connected sensors, such as water pressure, flow, and height sensors, which allow for real-time assessments of the amount of water being provided downstream, alongside allowing the prediction of floods and droughts. The data retrieved from these components is then sent through SCADA or similar networks to command centers, where fully or partially automated control feedback loops are used to process the data to forecast demand, optimize water flow, and adjust operations. Such a decision-making process is then transformed into a set of instructions for the system's connected actuators, such as valves, dams, and gates, that - given the feedback provided by the sensors - allow automated regulation of the flow and pressure of water and its distribution in multiple emissaries.

3. Motivation and Threat Model

In **cyber-physical attacks**, the assets at risk encompass both the digital and physical elements of the system. To illustrate, manipulating water level sensor data through malware or network tampering may lead to inappropriate responses to water level changes. This may, in turn, cause flooding or structural

damage. Cyber-physical attacks require ad-hoc threat modeling and defensive measures to ensure their mitigation, especially due to many critical infrastructures comprising CPS. Fortunately, the predictable and structured nature of CPSs aids in developing certain security measures and mitigations, specifically in the field of attack and intrusion detection. In fact, these systems can leverage the regularity of communication patterns, sensor readings, and control signals inherent to CPS to detect anomalies that indicate intrusions or tampering. In particular, ML models can be trained on normal system behavior to identify deviations indicative of cyber intrusions or data manipulation, which might otherwise go unnoticed in conventional rule-based systems.

Threat Model. We consider an attacker with the capability to compromise the water level sensor data in a smart water infrastructure system, either through direct tampering with sensors, injecting false data into communication channels, or exploiting software vulnerabilities in the data acquisition system. Such an attacker could possess varying access levels, ranging from physical proximity to the sensor nodes for hardware-based tampering to remote access via network intrusion or exploitation of insecure communication protocols. The attacker's goal may include disrupting the system's functionality, such as causing overflows or shortages that can lead to physical damage, service outages, or safety hazards. Additionally, they might aim to manipulate reported metrics to mislead operators or decision-making algorithms, potentially inducing incorrect system responses or masking other malicious activities.

4. Related works

Earlier approaches to intrusion detection for CPSs have built mathematical models to detect anomalous behaviors. This kind of approach has been analyzed in several surveys, including Giraldo et al. [4] and Cardenas et al. [5]. Although mathematical approaches can achieve high detection accuracy, they often struggle when applied to complex cyber-physical systems [6]. Building accurate models for intricate physical processes can be challenging due to the need for substantial expertise and deep knowledge of system dynamics during the initial development phase [7, 8]. In recent years, several detection methods for CPSs have been developed using machine learning techniques that do not rely on expert knowledge or specific domain expertise, including methods for detecting misbehavior in individual sensors, where each sensor is analyzed by applying a separate instance of the model. An example of this approach is Process-Aware Stealthy Attack Detection (PASAD), which interprets time-series of sensor data through Singular Spectrum Analysis (SSA) in industrial control systems Aoudi et al. [7]. However, the original version of PASAD handles only univariate data, which limits its scalability to environments with multiple sensors [9]. As a result, researchers have devoted considerable effort to developing multi-sensor misbehavior detection methods. Zhang et al. [10] propose a multilayer data-driven cyber-attack detection system to improve the security of industrial control systems. Several of these methods rely on supervised machine learning, which is not applicable in the absence of labeled data [11]. In a real-world scenario, labeled datasets are not always available, so semi-supervised and unsupervised techniques are used to fill the gap [11, 12]. Recent unsupervised methods are based on ML techniques. Paffenroth et al. [13] introduce a methodology for detecting weak, distributed patterns in sensor networks through space-time signal processing. However, temporal information is crucial, as sensor observations are time-dependent, and historical data play a key role in reconstructing current states, helping to determine whether a system is operating normally or abnormally [14]. The latest applications of deep learning methods in multi-sensor domains have focused on capturing temporal dependencies within time-series data. The LSTM approach analyses time-series data to model temporal sequences and identifies long-term dependencies [15, 16, 17]. Zhu et al. [18] propose an approach for anomaly detection in complex time-series data. The method involves the use of an LSTM Encoder-Decoder architecture combined with adversarial training. Wei et al. [19] develop a deep learning approach for detecting anomalies in indoor air quality data by combining LSTM networks with an autoencoder architecture. Shrestha et al. [20] introduce a framework for detecting anomalies in smart electric grid systems. The authors present an anomaly detection system that employs LSTM and AEs to process sensor data from smart grids.

Detection in smart water infrastructures. Amin et al. [21, 22] propose a model for detecting anomalies in distributed control systems using a hydrodynamic model based on the Shallow Water Equations, they capture flow dynamics and account for propagation delays. Wei Gao et al. [23] develop an IDS for smart water utilities using a three-stage backpropagation neural network based on Modbus features. Their IDS monitors sensor and actuator data, focusing on water levels and valve settings. Raman et al. [24] introduce an anomaly detection method based on a Physics-based Neural Network (PbNN) approach combining deep CNNs with Industrial Control System (ICS) design knowledge, leveraging physical interactions to detect anomalies by comparing predicted and actual behavior in real time. Meleshko et al. [25] propose a hybrid anomaly detection method for wireless sensor networks in water management systems, employing classifiers like AdaBoost, Random Forest, and SVM to detect attacks on water level and flow sensors. Ramotsoela et al. [26] develop a behavioural intrusion detection system for water distribution systems using a voting-based ensemble of neural networks (ANN, RNN, LSTM, GRU, CNN). Nayak et al. [27] present a IDS for Smart Water Infrastructure (SWI) that combines fog computing with a fuzzy logic-based Intuitionistic System for feature selection, followed by a voting classifier using algorithms like Random Forest, SVM, and K-NN. Finally, Moazeni et al. [28] propose a deep learning approach for detecting FDI attacks in water distribution systems. They develop a supervised deep neural network optimized for identifying random FDIs targeting water level measurements.

5. Approach

In a multi-sensor cyber-physical system within a smart water infrastructure, the sensors are positioned at strategic points along rivers, reservoirs, or artificial canals to measure the height of the water, for instance. Usually, these points are sequential, and the temporal correlation between the measurements is essential for the system to function correctly. Our approach employs an LSTM-Autoencoder model to detect abnormal patterns within spatially and temporally correlated data, thus reinforcing the resilience of water distribution operations against both cyber and physical attacks. Figure 1 provides a comprehensive overview of the entire process, delineating the various phases involved. The initial stage is the preprocessing of the data, which involves the removal of any erroneous or anomalous measurements. Following the cleaning process, data are organized into fixed-time-length sequences, comprising all sensors simultaneously, to effectively capture both temporal and cross-sensor correlations. Once preprocessed and segmented into sequences, the time-series data from various sensors is passed through a stack of LSTM layers in the encoder phase to generate a latent representation. This representation is then decoded to attempt the reconstruction of the original input sequence. Once this process is complete, the reconstruction error, Mean Squared Error (MSE), for each sensor time series is calculated. Once the metrics have been calculated, these results are compared to a threshold. If the anomaly score exceeds the threshold, the data point is flagged as anomalous.

5.1. Preprocessing

During the data cleaning phase, any missing, invalid, or outlier entries are addressed to ensure data quality. When a missing value, an invalid reading, or an outlier occurs, the mean between the first preceding and first following valid measurements is calculated. This process allows the creation of a continuous and realistic data stream, accounting for outliers and inconsistencies without introducing anything that could affect the analysis. The resulting dataset, augmented with the synthetic attacks presented in the next section, provides a reliable foundation for the anomaly detection approach and can be found at [1].

Following the completion of the cleaning process, it is necessary to create time-series sequences for our DL architecture. These sequences are of a fixed length and are designed to capture the temporal and spatial correlations between sensors. Once these sequences are created, a 3D matrix (*n_rows* x *sequence_length* x *n_sensors*) is produced, whereby 'n_rows' represents the number of individual

[1] https://github.com/necst/ITASEC_SWI_dataset

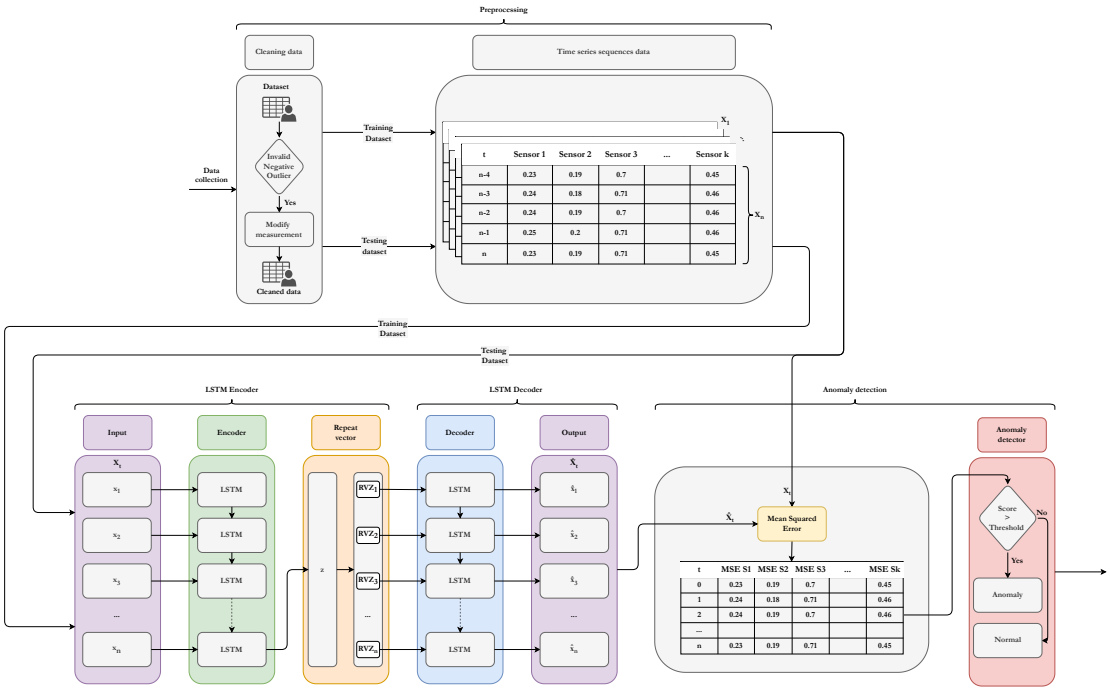


Figure 1: Overview of our approach.

measurements, 'sequence_length' is the size selected, and 'n_sensors' is the number of sensors available. In this manner, each row is comprised of a matrix (*sequence_length* x *n_sensors*), which correlates with a specific group of measurements.

5.2. Attacks Generation

Based on similar approaches in the literature, we design three synthetic attacks: random, replay, and gradual decrement attacks. Attack intervals are chosen to ensure each starting point supports the full duration of the attack sequence. For each of these attacks, we devise a scenario where an attacker selects a sensor before a dam and moments when there is a high water level. Once chosen the target, he or she injects lower-than-actual values into one or more sensors, causing the control system to interpret the water level as low and, consequently, to open the dam wider than necessary.

Random attacks aim at generating random values for a fixed-length sequence. Within each chosen interval, the water height values are deliberately altered. Specifically, the attack simulates intentional deviations by replacing the sensor readings with new, randomly generated values that fall within a predefined range, as shown in Eq. 1.

$$\text{Random Range} = min\_value + \frac{min\_value + max\_value}{3} \tag{1}$$

These new values are specifically chosen within this range to ensure low water height readings while still remaining within the acceptable range for each sensor.

Replay attacks simulate a scenario in which historical values of sensor data are reused with the intention of misleading the control system. Within each selected interval, sensor values are replaced with prior valid readings from within a defined range, effectively replaying earlier water-level data. Initially, we tested a range defined as the one for the random attack (Eq. 2). However, we found that no historical values fell within this narrower range. As a result, we adjusted it as shown in Eq. 2.

$$\text{Replay Range} = min\_value + \frac{min\_value + max\_value}{2.9} \tag{2}$$

This modification provides realistic but deceptive data that subtly mislead the system. This approach maintains realistic fluctuations within the targeted range, subtly introducing misleading data into the system.

Gradual decrement attacks simulate a slow and progressive reduction in sensor values as an adversarial attempt to bypass the intrusion detection process. For each identified interval, we decrease each subsequent reading by a small, predefined amount until reaching a target threshold. This target is set just above the minimum measurable water level, ensuring the data remains plausible. Once the target threshold is reached in every targeted sensor, the sensor value remains constant at this target level for the remaining duration of the attack. This gradual decrease avoids causing abrupt changes, which may increase the likelihood of bypassing simple detection mechanisms and obscuring the true water level trends over time.

5.3. Training and parameter tuning

Our methodology entails comprehensive training and parameter tuning to ascertain the optimal configuration for our LSTM-based model in anomaly detection within smart water infrastructures. To achieve this, a range of hyper-parameters is experimented with, including the number of units in each LSTM layer, regularisation parameters, dropout rates, sequence length, and batch size. This tuning process is crucial for achieving an appropriate balance between the model's complexity and its capacity to generalize effectively across different operational states.

In particular, the effectiveness of models with varying architectures, including 2, 4, and 6 LSTM layers, is evaluated in order to assess their capacity to capture the temporal dependencies inherent in multi-sensor time-series data. Each model variant is subjected to hyper-parameter tuning, whereby parameters such as the number of LSTM units per layer and regularisation penalties are iteratively adjusted with the objective of minimizing validation loss. These evaluations are presented in Section 7.

As a loss function, we empirically found that the MSE was the most effective one, which measures the average squared difference between the model's predictions and the actual values. As an evaluation metric, we use Mean Absolute Error (MAE), which calculates the mean of the absolute differences between predicted and actual values. MAE ($MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$) (where n is the total number of data points, y_i is the actual value of one data point, and $\hat{y}_i|$ is the predicted value for the same data point) provides an indication of the model's average deviation without excessively penalizing larger errors, making it a useful metric for understanding how much the model's predictions deviate from the actual values in general. Finally, an early stopping technique is employed to prevent overfitting, thereby ensuring that the model retains the optimal weights once the validation performance has stabilized.

6. Use Case: the Mincio River Water Infrastructure

Our use case, developed under the SARIL European project [29], takes into consideration the water infrastructure that traverses the Mincio River, which flows through the city of Mantua. It functions as a cyber-physical system with the objective of monitoring and managing the river's flow. As illustrated in Figure 2, the river originates from Garda Lake and traverses Mantua. The system comprises a network of strategically located control points, each of which is equipped with a small dam and a collection of sensors that continuously monitor water levels. The sensors are responsible for measuring the height of the water, thereby providing the system with real-time data that is essential for the effective management of the water flow, the maintenance of optimal levels, and the assurance of the system's responsiveness to environmental changes. This configuration enables precise control and timely adjustments across the infrastructure, thereby augmenting its ability to prevent overflow and manage resources efficiently.

The "Pozzolo" station represents the most critical juncture, where the river's course divides into two branches. This bifurcation occurs in front of a dam that regulates the flow of water in downstream areas. On the opposite side of the bifurcation, there is an artificial channel designed as a bypass to prevent flooding and facilitate the transfer of water to other areas. The functionality of the dam is contingent upon the real-time data collected by the water level sensors, which are indispensable for the precise

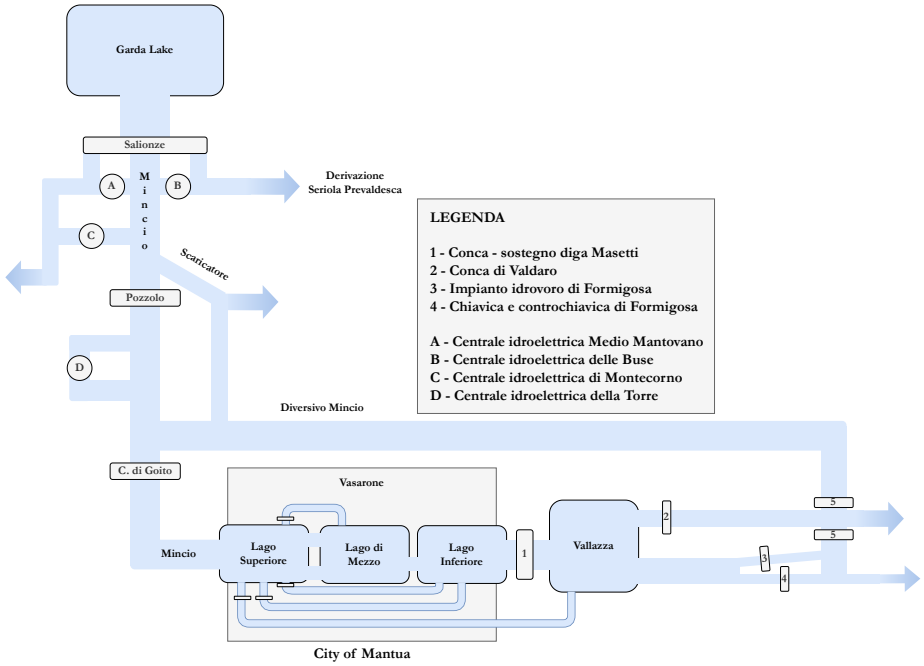


Figure 2: Diagram representing the Mincio Infrastructure.

and responsive regulation of the water flow. The proportion of water to be directed into the canal and the amount to be allowed to flow through the main river downstream of Pozzolo is determined by the control system based on the water levels detected by the sensors. The city of Mantua is crossed by three basins called respectively *Lago Superiore*, *Lago di Mezzo*, and *Lago Inferiore*.

Within the stations, there are sensors that measure water levels every 15 minutes. These measurements are stored in a public repository accessible via the Agenzia Interregionale per il fiume PO (AIPO) website [30]. This accessibility provides open access to both real-time and historical information.

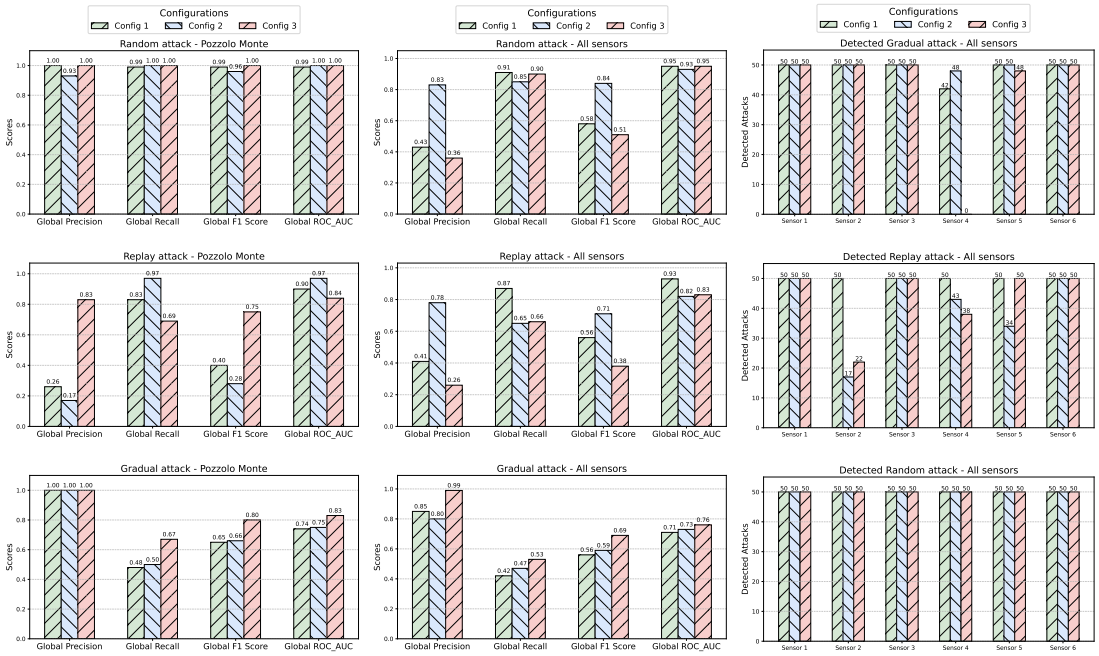
While other cybersecurity measures are in place, the absence of an IDS renders the infrastructure incapable of automatically discerning anomalous behavior and identifying potential attacks. It is, therefore, potentially possible for an adversary that obtains control of the sensor data to exploit these conditions, potentially implementing an FDI attack.

7. Experimental Evaluation

The assessment of our algorithm focuses on demonstrating the efficacy of LSTM models in anomaly detection within a multi-sensor system and on presenting the reasoning behind the design choices and parameter tuning presented in Section 5. For our evaluation, we used precision, recall, F1-Score, ROC-AUC, and the average detection time of an attack.

We collect data from a public repository [30], comprising water level measurements recorded every 15 minutes across various sensors. We divide this dataset into training and testing sets. Since our approach is unsupervised, the testing dataset consists of both real measurements and manipulated data generated through the attacks we have implemented. For each attack type, we generate 50 attack instances on different sensors to thoroughly assess detection capabilities. Our dataset can be accessed at [1].

To select the optimal architecture for our LSTM-AE model, we conduct initial tests exploring various configurations. First, we employ *KerasTuner* to determine the best settings for LSTM units per layer, dropout rate, and L2 regularisation, evaluating these configurations based on validation loss. Notably, this tuning process is conducted separately for models with 2, 4, and 6 layers, allowing us to identify



(a) Evaluation results in Pozzolo Monte sensor.

(b) Evaluation results in multisensors context.

(c) Number of detected attacks in multisensors context.

Figure 3: Detection performances depending on the design configuration.

the top-performing configuration for each model depth. From these three optimal configurations (one for each layer setup), we test different validation split values, finding that a validation split of 0.2 yields the best results, meaning 20% of the training dataset is allocated for validation while 80% is used for the actual training.

After identifying the top three models, we conduct extensive tests to evaluate the performance of each architecture in terms of sequence length, batch size, and anomaly detection parameters, specifically percentile, Z-Score threshold, and lower and upper quantiles across all types of attacks. We implemented (random, replay, and gradual decremental). Additionally, for each type of attack, we test both attacks targeting all sensors and attacks focusing solely on the Pozzolo Monte sensor, which is the most critical one. To evaluate the best setup regarding sensor reading sequence length and batch size, we perform over a thousand runs for each of the three best-performing models. Clearly, it is not feasible to present all these results, and considering the limitations of space and readability, we just focus on the three best configurations to highlight the key differences.

To achieve a balanced detection across all three types of attacks, we focus on identifying configurations that yield effective results for all of them. We evaluate the configurations on recall, precision, F1-Score, and whether the complete attack was detected at least once. From this analysis, we identify three optimal configurations: **Configuration 1:** 2-layers, batch size of 32, sequence length of 6, using Mahalanobis Distance with a percentile threshold of 98.5. **Configuration 2:** 4-layers, batch size of 32, sequence length of 6, using Z-Score with a threshold of 2. **Configuration 3:** 6-layers, batch size of 64, sequence length of 4, using Euclidean Distance with a percentile threshold of 98.5.

7.1. Results

Figures 3a and 3b show the variations in precision, recall, F1-Score, and ROC-AUC for each configuration during attacks on single or all sensors. Figure 3c displays the performance of the configurations in detecting 50 attacks of each type to all the sensors. An attack is considered detected if at least one of the manipulated measurements is classified as anomalous.

All three configurations perform well in detecting single-sensor random attacks, but while all three

maintain a high recall in multi-sensor attacks, configuration 2 achieves good precision, resulting in the highest F1 score. Note that, nonetheless, as shown in the graphs, all three configurations successfully detect all the random attacks.

Regarding replay attacks, the performances drop significantly. Configuration 3 appears to be effective against single-sensor attacks, but its F1 score drops significantly when evaluating multi-sensor ones. Configuration 1 appears to be detecting all attacks, but this is primarily due to its high recall combined with very low precision, which results in the system effectively triggering alerts for the 50 real attacks while also producing a number of false positives. Configuration 2, on the other hand, misses a few attacks, although only on three specific sensors.

Finally, all three configurations achieve good precision but a relatively low recall, consequently affecting the overall F1-Score, leading to moderate values. The ROC-AUC stabilizes around 0.7 across all configurations, indicating similar performance. In detecting the 50 gradual decrement attacks, configuration 2 proves to be the most effective, despite missing only a few attacks, specifically on sensor 4. In contrast, configuration 3 fails to detect all attacks on sensor 4 and misses some on sensor 5, while configuration 1 misses a few attacks on sensor 4.

Discussion. Based on the experimental results, **configuration 2** appears to be the most effective in terms of overall performance. Specifically, it demonstrates superior performance in detecting both random and replay attacks. While its performance in identifying gradual decrement attacks is slightly lower than that of configuration 3, it still manages to detect almost all attacks, unlike configuration 3. As observed in the experiments, increasing the complexity of the attack type leads to a reduction in performance, particularly in terms of recall. Nonetheless, configuration 2, when faced with complex attacks like replay or gradual decrement, still manages to identify the vast majority of the attacks implemented.

8. Conclusion

This research introduced an innovative approach for detecting anomalies in cyber-physical systems, focusing on a smart water infrastructure use case along the Mincio River. By utilizing advanced deep learning techniques, specifically Long Short-Term Memory (LSTM) networks and Autoencoder models, the study demonstrated the feasibility of accurately capturing temporal dependencies within time-series data, enabling the effective detection of False Data Injection (FDI) attacks. Extensive evaluations of multiple model configurations were conducted to identify the most effective design, showcasing the method's ability to detect a variety of attack types. The deployment of this model within the Mincio River infrastructure illustrated its potential for real-world application while offering valuable insights into the broader domain of anomaly-based intrusion detection for critical infrastructure systems. Future work will explore the potential integration of distributed ledger technologies, inspired by Maffiola et al. [31], to enhance security and transparency in cyber-physical domains. Additionally, federated learning will be investigated to facilitate distributed and secure training within CPS environments, leveraging collaborative frameworks to strengthen anomaly detection capabilities [32].

Acknowledgments

The presented work was performed in the context of the Horizon Europe project SARIL [29], which is funded by the European Union under grant agreement ID 101103978. Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency. Neither the European Union nor the granting authority can be held responsible for them."

Declaration on Generative AI

Generative AI tools such as Grammarly and ChatGPT 4o were utilized solely for proofreading and grammar refinement in the preparation of this manuscript. The authors retain full responsibility for the content presented in the final version.

References

[1] R. Baheti, H. Gill, Cyber-physical systems, The impact of control technology 12 (2011) 161–166.

[2] S. Han, M. Xie, H.-H. Chen, Y. Ling, Intrusion detection in cyber-physical systems: Techniques and challenges, IEEE Systems Journal 8 (2014) 1052 – 1062. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84913586160&doi=10.1109%2fJSYST.2013.2257594&partnerID=40&md5=b266efa26c010ca73a63315023a5a230. doi:10.1109/JSYST.2013.2257594.

[3] S. Z. Yong, M. Q. Foo, E. Frazzoli, Robust and resilient estimation for cyber-physical systems under adversarial attacks, in: 2016 American Control Conference (ACC), 2016, pp. 308–315. doi:10.1109/ACC.2016.7524933.

[4] J. Giraldo, D. Urbina, A. Cardenas, J. Valente, M. Faisal, J. Ruths, N. O. Tippenhauer, H. Sandberg, R. Candell, A survey of physics-based attack detection in cyber-physical systems, ACM Comput. Surv. 51 (2018). URL: https://doi.org/10.1145/3203245. doi:10.1145/3203245.

[5] A. A. Cárdenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang, S. Sastry, Attacks against process control systems: risk assessment, detection, and response, in: Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security, ASIACCS '11, Association for Computing Machinery, New York, NY, USA, 2011, p. 355–366. URL: https://doi.org/10.1145/1966913.1966959. doi:10.1145/1966913.1966959.

[6] C. Feng, V. R. Palleti, A. Mathur, D. Chana, A systematic framework to generate invariants for anomaly detection in industrial control systems., in: NDSS, 2019, pp. 1–15.

[7] W. Aoudi, M. Iturbe, M. Almgren, Truth will out: Departure-based process-level detection of stealthy attacks on control systems, in: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, CCS '18, Association for Computing Machinery, New York, NY, USA, 2018, p. 817–831. URL: https://doi.org/10.1145/3243734.3243781. doi:10.1145/3243734.3243781.

[8] Y. Chen, C. M. Poskitt, J. Sun, Learning from mutants: Using code mutation to learn and monitor invariants of a cyber-physical system, in: 2018 IEEE Symposium on Security and Privacy (SP), 2018, pp. 648–660. doi:10.1109/SP.2018.00016.

[9] W. Aoudi, M. Almgren, A scalable specification-agnostic multi-sensor anomaly detection system for iiot environments, International Journal of Critical Infrastructure Protection 30 (2020) 100377. URL: https://www.sciencedirect.com/science/article/pii/S187454822030041X. doi:https://doi.org/10.1016/j.ijcip.2020.100377.

[10] F. Zhang, H. A. D. E. Kodituwakku, J. W. Hines, J. Coble, Multilayer data-driven cyber-attack detection system for industrial control systems based on network, system, and process data, IEEE Transactions on Industrial Informatics 15 (2019) 4362–4369. doi:10.1109/TII.2019.2891261.

[11] J. Suaboot, A. Fahad, Z. Tari, J. Grundy, A. N. Mahmood, A. Almalawi, A. Y. Zomaya, K. Drira, A taxonomy of supervised learning for idss in scada environments, ACM Comput. Surv. 53 (2020). URL: https://doi.org/10.1145/3379499. doi:10.1145/3379499.

[12] L. Erhan, M. Ndubuaku, M. Di Mauro, W. Song, M. Chen, G. Fortino, O. Bagdasar, A. Liotta, Smart anomaly detection in sensor systems: A multi-perspective review, Information Fusion 67 (2021) 64–79. URL: https://www.sciencedirect.com/science/article/pii/S1566253520303717. doi:https://doi.org/10.1016/j.inffus.2020.10.001.

[13] R. Paffenroth, P. du Toit, R. Nong, L. Scharf, A. P. Jayasumana, V. Bandara, Space-time signal

processing for distributed pattern detection in sensor networks, IEEE Journal of Selected Topics in Signal Processing 7 (2013) 38–49. doi:10.1109/JSTSP.2012.2237381.

[14] Y. Hao, H. Cao, A. Mueen, S. Brahma, Identify significant phenomenon-specific variables for multivariate time series, IEEE Transactions on Knowledge and Data Engineering 33 (2021) 1019–1031. doi:10.1109/TKDE.2019.2934464.

[15] S. Tariq, S. Lee, Y. Shin, M. S. Lee, O. Jung, D. Chung, S. S. Woo, Detecting anomalies in space using multivariate convolutional lstm with mixtures of probabilistic pca, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 2123–2133. URL: https://doi.org/10.1145/3292500.3330776. doi:10.1145/3292500.3330776.

[16] S. Longari, D. H. N. Valcarcel, M. Zago, M. Carminati, S. Zanero, Cannolo: An anomaly detection system based on LSTM autoencoders for controller area network, IEEE Trans. Netw. Serv. Manag. 18 (2021) 1913–1924. URL: https://doi.org/10.1109/TNSM.2020.3038991. doi:10.1109/TNSM.2020.3038991.

[17] S. Longari, C. A. Pozzoli, A. Nichelini, M. Carminati, S. Zanero, Candito: Improving payload-based detection of attacks on controller area networks, in: S. Dolev, E. Gudes, P. Paillier (Eds.), Cyber Security, Cryptology, and Machine Learning - 7th International Symposium, CSCML 2023, Be'er Sheva, Israel, June 29-30, 2023, Proceedings, volume 13914 of *Lecture Notes in Computer Science*, Springer, 2023, pp. 135–150. URL: https://doi.org/10.1007/978-3-031-34671-2_10. doi:10.1007/978-3-031-34671-2_10.

[18] H. Zhu, S. Liu, F. Jiang, Adversarial training of lstm-ed based anomaly detection for complex time-series in cyber-physical-social systems, Pattern Recognition Letters 164 (2022) 132–139. URL: https://www.sciencedirect.com/science/article/pii/S0167865522003129. doi:https://doi.org/10.1016/j.patrec.2022.10.017.

[19] Y. Wei, J. Jang-Jaccard, W. Xu, F. Sabrina, S. Camtepe, M. Boulic, Lstm-autoencoder-based anomaly detection for indoor air quality time-series data, IEEE Sensors Journal 23 (2023) 3787–3800. doi:10.1109/JSEN.2022.3230361.

[20] R. Shrestha, M. Mohammadi, S. Sinaei, A. Salcines, D. Pampliega, R. Clemente, A. L. Sanz, E. Nowroozi, A. Lindgren, Anomaly detection based on lstm and autoencoders using federated learning in smart electric grid, Journal of Parallel and Distributed Computing 193 (2024) 104951. URL: https://www.sciencedirect.com/science/article/pii/S0743731524001151. doi:https://doi.org/10.1016/j.jpdc.2024.104951.

[21] S. Amin, X. Litrico, S. Sastry, A. M. Bayen, Cyber security of water scada systems—part i: Analysis and experimentation of stealthy deception attacks, IEEE Transactions on Control Systems Technology 21 (2013) 1963–1970. doi:10.1109/TCST.2012.2211873.

[22] S. Amin, X. Litrico, S. S. Sastry, A. M. Bayen, Cyber security of water scada systems—part ii: Attack detection using enhanced hydrodynamic models, IEEE Transactions on Control Systems Technology 21 (2013) 1679–1693. doi:10.1109/TCST.2012.2211874.

[23] Wei Gao, T. Morris, B. Reaves, D. Richey, On SCADA control system command and response injection and intrusion detection, in: 2010 eCrime Researchers Summit, IEEE, 2010, pp. 1–9. URL: http://ieeexplore.ieee.org/document/5706699/. doi:10.1109/ecrime.2010.5706699.

[24] M. R. G. Raman, A. P. Mathur, A hybrid physics-based data-driven framework for anomaly detection in industrial control systems, IEEE Transactions on Systems, Man, and Cybernetics: Systems 52 (2022) 6003–6014. doi:10.1109/TSMC.2021.3131662.

[25] A. Meleshko, V. Desnitsky, I. Kotenko, E. Novikova, A. Shulepov, Combined approach to anomaly detection in wireless sensor networks on example of water management system, in: 2021 10th Mediterranean Conference on Embedded Computing (MECO), 2021, pp. 1–4. doi:10.1109/MECO52532.2021.9460237.

[26] T. D. Ramotsoela, G. P. Hancke, A. M. Abu-Mahfouz, Behavioural intrusion detection in water distribution systems using neural networks, IEEE Access 8 (2020) 190403–190416. doi:10.1109/ACCESS.2020.3032251.

[27] O. Nayak, J. Lachure, R. Doriya, Fog enabled cyber-physical attack detection using ensemble

machine learning, in: 2022 1st International Conference on Sustainable Technology for Power and Energy Systems (STPES), 2022, pp. 1–6. doi:10.1109/STPES54845.2022.10006594.

[28] F. Moazeni, J. Khazaei, Detection of random false data injection cyberattacks in smart water systems using optimized deep neural networks, Energies 15 (2022). URL: https://www.mdpi.com/1996-1073/15/13/4832. doi:10.3390/en15134832.

[29] European Climate, Infrastructure and Environment Executive Agency, SARIL: Sustainability And Resilience for Infrastructure and Logistics networks, 2024. URL: https://saril-project.eu/.

[30] AIPO - Agenzia Interregionale per il fiume Po, AIPO Servizio di Piena, 2024. URL: https://idrometri.agenziapo.it/map/map2d.

[31] D. Maffiola, S. Longari, M. Carminati, M. Tanelli, S. Zanero, GOLIATH: A decentralized framework for data collection in intelligent transportation systems, IEEE Trans. Intell. Transp. Syst. 23 (2022) 13372–13385. URL: https://doi.org/10.1109/TITS.2021.3123824. doi:10.1109/TITS.2021.3123824.

[32] G. Digregorio, E. Cainazzo, S. Longari, M. Carminati, S. Zanero, Evaluating the impact of privacy-preserving federated learning on CAN intrusion detection, in: 99th IEEE Vehicular Technology Conference, VTC Spring 2024, Singapore, June 24-27, 2024, IEEE, 2024, pp. 1–7. URL: https://doi.org/10.1109/VTC2024-Spring62846.2024.10683636. doi:10.1109/VTC2024-SPRING62846.2024.10683636.