Advances in Artificial Intelligence and Machine Learning; Research 4 (2) 2233-2243 Received 2-03-2024; Accepted 20-05-2024; Published 27-05-2024

Time-Series Analysis on AIDE IoT Attack Data Unraveling Trends and Patterns for Enhanced Security

Atdhe Buja ATDHE.BUJA@ACADEMYICT.NET
ICT Academy, Prishtina, Kosovo.

Melinda Pacolli PACOLLIMELINDA@GMAIL.COM
ECPD, Prishtina Kosovo.

Donika Bajrami DONIKA.BAJRAMI@ACADEMYICT.NET
ICT Academy, Prishtina, Kosovo.

Philip Polstra PPOLSTRA@COMMONWEALTHU.EDU
Bloomsburg University of Pennsylvania, PA, USA.

Akihiko Mutoh MUTOH@TSUKIJIHONGWANJI.JP
Tsukijihongwanji, Tokyo, Japan.

Corresponding Author: Atdhe Buja

Abstract

With the growing use of IoT devices, cyberattacks have increased against these devices, using various exploitable vulnerabilities. Cyberattacks are potentially ruinous events for business owners and the cost of a cyber-attack is not only financial, but companies also must spend time on recovering from the attacks. Through applied research we set out the latest findings and trends to provide new insights into the security of IoTs, focusing on the emerging nature of IoT attacks, frequency, and usual vulnerabilities, thus providing significance to cybersecurity and IoT defense, which will be beneficial to researchers, manufacturers, individuals, organizations, and governments. Although the literature on IoTs is quite rich to this day, however, there is currently no study that provides an in-depth analysis of patterns and frequencies of IoT attacks to discover insights into the most vulnerable times and systems. Compared to other related research on the security of IoTs, this applied research encompasses much more technical angles to the security of IoT. It begins with the comprehensive data acquisition from the Global Cyber Alliance's (GCA) Automated IoT Defense Ecosystem (AIDE) and preprocessing to ensure data integrity and relevance. Followed by a thorough feature selection and engineering process. Python scripts and libraries were used for efficient data analysis. Unraveling Trends and Patterns for Enhanced Security are sensitive areas that have remained untouched by previous research works. Time-Series Analysis on AIDE IoT Attack Data used in this study provided critical insights into attack patterns by fields such as timestamps, attack duration, and login credentials, implying a huge scientific and technological uncover of an intricate and evolving landscape of IoT security threats and attacks.

Citation: Atdhe Buja, et al. Time-Series Analysis on AIDE IoT Attack Data Unraveling Trends and Patterns for Enhanced Security. Advances in Artificial Intelligence and Machine Learning. 2024;4(2):128.

Keywords: IoT security, Predictive analysis, Attack trend analysis, Anomaly detection.

1. INTRODUCTION

Industrial Internet of Things (IIoT), Internet of Things (IoT) within the context of the Fourth Industrial Revolution, as we call Industry 4.0 can be used to make significant transformations of Cyber-Physical Systems (CPS) by using big data and analytics [1]. With the increased utilization of Internet of Things (IoT) devices across various sectors, there has been a similar increase in cyberattacks targeting these interconnected devices. Their exposure to the Internet prompts threats and attacks from hackers and Advanced Persistent Threats (APTs) [2]. The IoT is facing some recent attacks such as targeted ransomware [3]. Structured Query Language (SQL) Injection Attack (SQLIA) is still an intruder's exploit of choice to steal confidential data from [4]. Utilizing historical data and real-time information from IoT devices prevents unexpected failures and minimizes downtime [5]. Over 80% of organizations have applied the IoT, and nearly 20% of organizations have detected an IoT-based attack in the past three years [6]. It's highly challenging to manage IoT cybersecurity because of the converged solutions [7].

These attacks exploit various vulnerabilities essential in IoT systems, posing significant risks to businesses, individuals, and critical infrastructure. Understanding the complexities of IoT security threats is most important in today's digital landscape. Beyond financial losses, cyberattacks on IoT devices can have extensive consequences, disrupting operations, compromising sensitive data, and destroying user trust. Furthermore, the time and resources needed for recovery and mitigation further emphasize the seriousness of these threats. For this study, we had a chance to work with the Global Cyber Alliance (GCA) [8], a non-profit honeyfarm operator. In partnership, we analyze data from a large honey farm operator with 54,835,849 rows of data deployed worldwide.

While existing studies provide a valuable understanding of various aspects of IoT security, they often lack a thorough examination of temporal trends and vulnerabilities. This identifies the need for applied research aimed at unraveling the emerging nature of IoT attacks and identifying strategies for advanced defense. In response to this gap, this paper sets out a unique view to address the following research objectives:

- To provide new insights into the emerging nature of IoT attacks, including trends, frequencies, and common vulnerabilities.
- To conduct a thorough analysis of attack patterns using time-series analysis techniques applied to data acquired from the GCA Automated IoT Defense Ecosystem (AIDE).
- To contribute to the enhancement of cybersecurity and IoT defense measures by uncovering critical insights into IoT attack trends and vulnerabilities.

By undertaking a tough investigation of IoT attack data and employing advanced analytical techniques, this research aims to illuminate previously unexplored areas of IoT security. Through our findings, we seek to empower researchers, manufacturers, individuals, and organizations with actionable intelligence to support their cybersecurity defenses against evolving threats.

2. METHODOLOGY

The methodology employed in this study adhered to tough scientific principles to ensure the validity and reliability of the research findings [9, 10]. Each step of the methodology was based on the data science lifecycle to address the research objectives effectively and to maximize the insights derived from the data. Further, the methodology is unfolded in this section on the steps commonly associated with data science and cybersecurity. This methodology employs systematic approaches to gather, preprocess, analyze, and interpret data from the GCA Automated IoT Defense Ecosystem (AIDE) [8], a large honey farm operator, aiming to uncover patterns and trends in IoT attack data. Below is a high-level overview of the steps carried out.

2.1 Data Acquisition

The first phase of the methodology involved the collection of data from an operational honey farm by found collaboration with GCA [8] serves as a centralized repository of IoT attack data, providing researchers with access to a distinct and extensive dataset. The selection of AIDE as the data source was based on its reputation for providing high-quality, real-world data on IoT security incidents, therefore ensuring the integrity and relevance of the analysis. The utilization of tools and software has supported obtaining data and structuring them in datasets CSV. Python appears as the primary programming language for implementing our methodology. Python was utilized to efficiently obtain data in chunks from the AIDE, which contained a significant dataset (54,835,849 records) for a certain period (1st May 2023 – 31st July 2023). The dataset used was acquired from the GCA AIDE, an integrated repository of IoT attack data. Every record in the dataset holds information on specific events related to IoT attacks, as well as timestamps, attack duration, login attempts, and credentials. Then, the dataset is powerful, as documented by the variety of timestamps and login combinations, giving valuable insights into the growing IoT attacks.

This approach includes breaking down the dataset into smaller chunks or batches, retrieving each chunk sequentially, and processing them iteratively. Python's versatility and scalability allowed the development of custom scripts to automate the retrieval process, ensuring optimal performance and resource utilization.

2.2 Preprocessing Techniques

Once the data is acquired, it goes through preprocessing to ensure its integrity, and relevance, and prepare it for analysis. This involved several steps, including data cleaning, normalization, and outlier detection. By utilizing these preprocessing techniques, we were able to mitigate potential biases and ensure the robustness of the following analysis.

2.3 Feature Selection and Engineering

A crucial aspect of the methodology involved the selection and engineering of relevant features for analysis. Given the intricacy of IoT attack data, it was required to identify relevant features

that could provide significant insights into attack patterns and trends. Through a blend of statistical analysis, and domain expertise, we identified a subset of features that were considered most instructive for achieving the research objectives. Various Python libraries were utilized to streamline the analysis process. Custom Python scripts were developed to automate repetitive tasks and refine the analysis pipeline. These scripts outlined various analytical procedures, allowing for efficient processing of large volumes of data. Then, visualization libraries such as Matplotlib and Seaborn were employed to generate insightful visualizations, supporting the interpretation and communication of analysis results. Overall, the methodology employed in this study encompassed robust data acquisition, thorough preprocessing, reflective feature selection and engineering, and efficient data analysis using Python scripts and libraries.

3. TIME-SERIES ANALYSIS OF AIDE IOT ATTACK DATA

In this section, we engage time-series analysis methods to examine the rich dataset obtained from the GCA AIDE. The boosting occurrence of IoT devices in various sectors has accompanied to the increase in cyber-attacks targeting these devices. As the IoT environment continues to expand and progress, understanding the dynamics over time of IoT attacks has turn vital for effective cybersecurity defense strategies. Our research treats this acute issue by focusing on the time-series analysis of IoT attack data, providing precious insights into attack patterns, trends, and vulnerabilities. Our objective is to unravel patterns and trends in IoT attacks, using key features such as timestamps (@timestamp), attack duration (startTime, and endTime), and login credentials (loggedin, and credentials). Time-series analysis involves the inspection of data points collected and recorded at regular time intervals.

The analysis is led by the research question:

- What are the trends in attack occurrences over time?

In our study, we engaged several time-series analysis techniques to gain insights into the temporal dynamics of IoT attacks. We use a methodology based on time-series analysis techniques, including decomposition, autocorrelation analysis, and smoothing. These techniques allow for a thorough investigation of temporal relationships and patterns in the IoT attack data, letting us discover hidden insights and trends. We selected decomposition because it gives a systematic approach to analyzing the complex and complicated nature of IoT attack data, allowing us to identify recurring patterns and anomalies. We choose autocorrelation analysis because it allows us to notice recurrent patterns and identify potential predictive factors. The smoothing technique supports us reveal long-term trends and variations in IoT attacks, smoothing the identification of meaningful patterns and changes over time.

3.1 Time-Series Decomposition

We decompose the IoT attack data into its components, including trend, seasonality, and noise, using techniques like seasonal decomposition of time series (STL). The majority of methods focus on learning the temporal patterns of the signals to detect anomalies [11]. We inspected the first few

rows of the dataset which provides insights into the temporal variability of attack occurrences, which is essential for time-series decomposition analysis. Then, we applied the method (data.describe) to result in summary statistics for the numerical columns, offering insights into the dataset distribution and central tendencies shown in FIGURE 1. The dataset contains 54,835,849 records, each capturing specific events related to IoT attacks. Key fields ('@timestamp', 'startTime', 'endTime', 'loggedin', and 'credentials') provide vital insights into attack timestamps, duration, and login attempts. Particularly, the dataset's dynamic nature, as exhibited by the unique timestamp and login combinations, emphasizes the evolving landscape of IoT attacks. The summary statistics reveal interesting patterns in the data. Refer to, the timestamp (47,177,755) field shows a diverse range of unique values, suggesting temporal variability in attack occurrences. As well, the loggedin (3,596,305) field emphasizes common vulnerabilities exploited by attackers, with certain login credentials recurring frequently, certainly indicative of common exploit vectors. Also, the credentials (38,136,609) field underlines the frequency of attempted logins without successful authentication, emphasizing the continuity and sophistication of IoT attack vectors. While the analysis provides valuable insights into the dataset structure, FIGURE 2 time-series plot showing the daily attack counts in attack frequency over time. However, FIGURE 3 demonstrates how the frequency of attacks has changed over time.

	@timestamp	startTime	endTime	loggedin	credentials
count	54835849	54835849	54835849	23439682	54835849
unique	47177755	54835602	54835502	90776	197184
top	2023-05-04T13:21:23.744Z	2023-06-10T03:46:13.929604Z	2023-05-20T10:45:54.840120Z	['root', '3245gs5662d34']	[]
freq	79	2	2	3596305	38136609

Figure 1: Summary statistics of the dataset distribution.

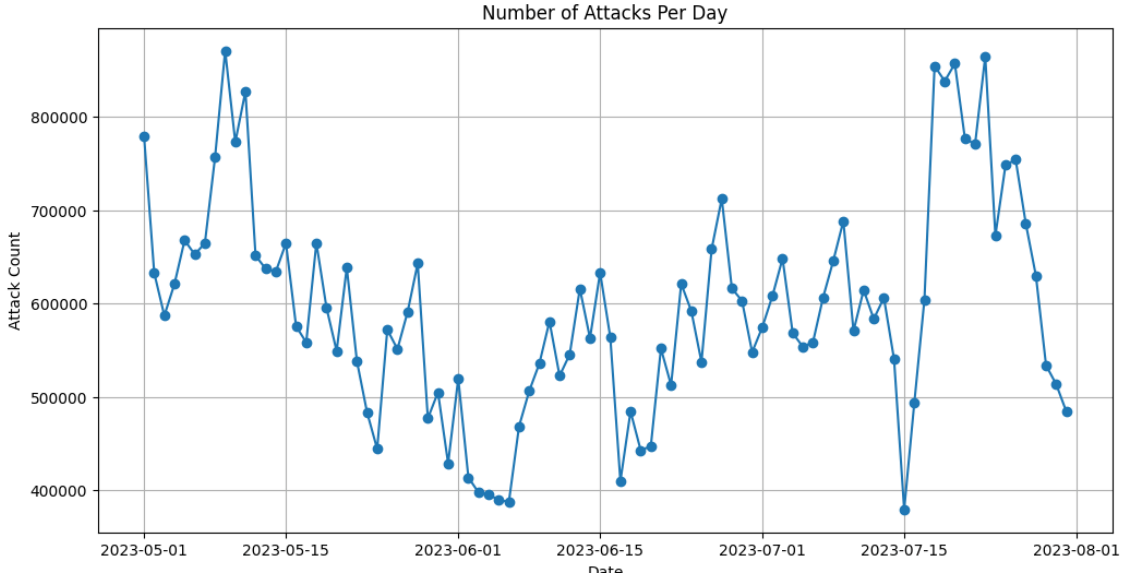

Figure 2: Time-series daily attack counts the fluctuation in attack frequency over time.

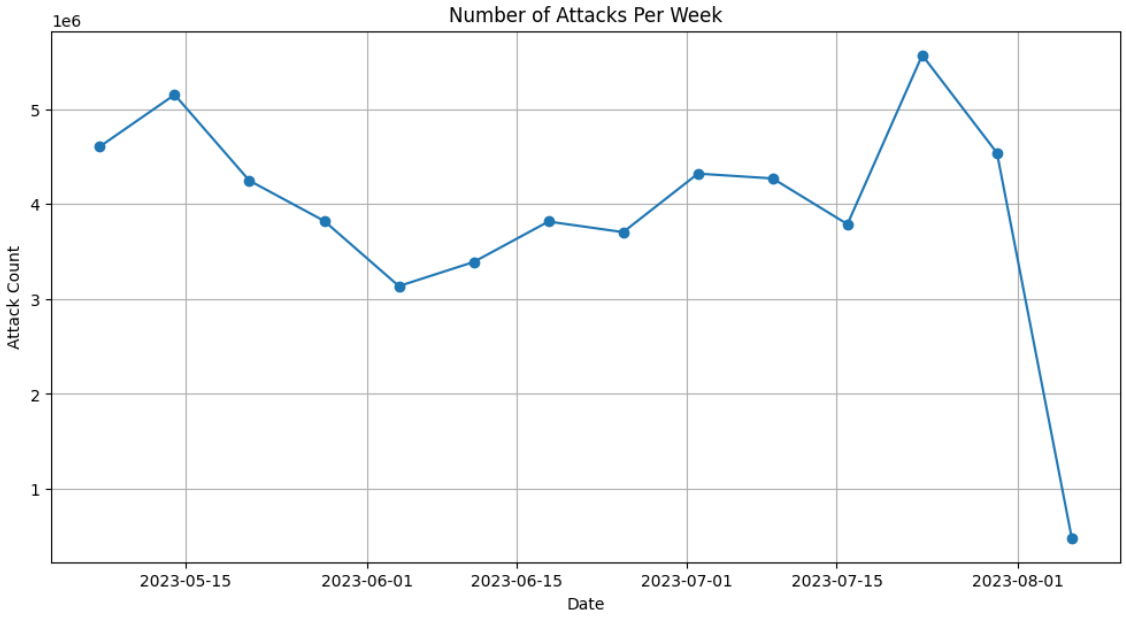

Figure 3: Time-series weekly attack counts the fluctuation in attack frequency over time.

3.2 Autocorrelation Analysis

Autocorrelation analysis is pivotal in discerning temporal patterns and dependencies within the AIDE IoT attack dataset. Using the detailed statistical insights, we delve into the core fields of the dataset (@timestamp, startTime, endTime, loggedin, and credentials) to unravel essential temporal relationships and cyclic trends. The dataset, containing significant (54,835,849) records, shows a diverse temporal landscape, as evidenced by the count of unique timestamps, start times, and end times. Particularly, each session or attack instance typically displays a unique start (54,835,602) and end time (54,835,502), emphasizing the dynamic nature of the attack environment. Regardless of the dataset's extensive temporal coverage, the frequency of the most common timestamp (79), start time (2), and end time (2) remains relatively low, suggesting a lack of recurring patterns based on these temporal indicators. Also, the statistical analysis of successful login attempts (loggedin) and credential usage (credentials) illuminate prevalent attack methodologies and vulnerabilities. The dataset outlines a diverse of successful login combinations, with the most frequent combination ('root', '3245gs5662d34') characteristic of a commonly exploited vulnerability. On the contrary, the credentials field mostly contains empty lists ([]), referring to either a lack of attempted credential usage or a failure to capture credential data. FIGURE 4 presents a thorough distribution of records, unique values, and frequency of the most frequent value for each field within the AIDE IoT attack dataset. The bars show the total count of records (@timestamp, startTime, endTime, loggedin, and credentials), emphasizing their comparable volume and underscoring the dataset's scale.

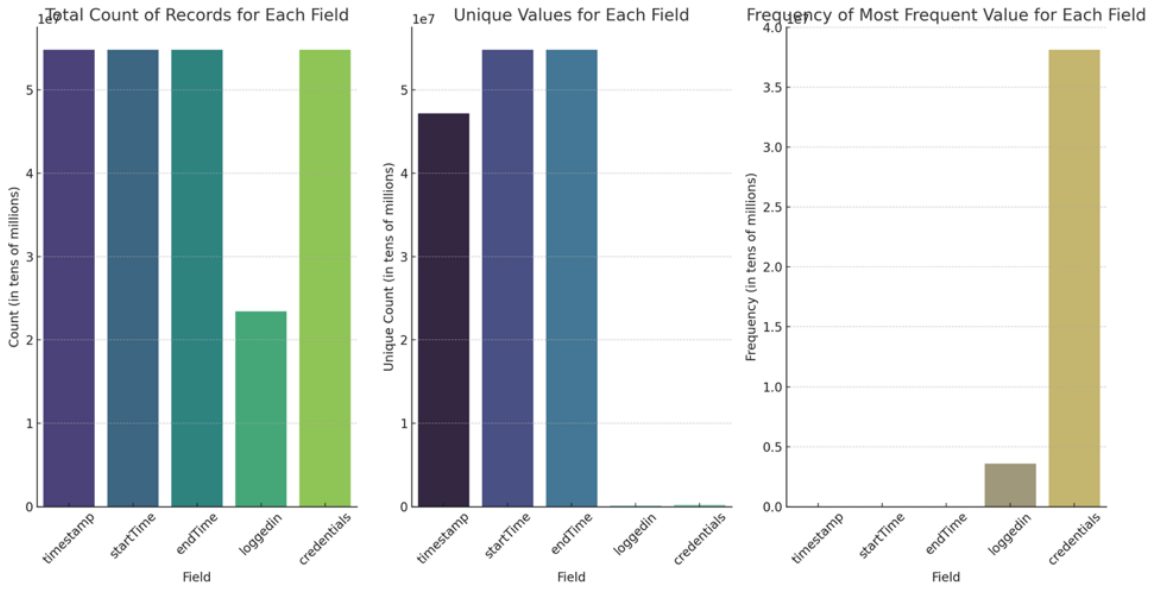

Figure 4: Distribution of records, unique values, and frequency of the most frequent value for each field.

3.3 Moving Average and Exponential Smoothing

These techniques help us unravel shifts in the data and highlight fundamental trends or patterns. FIGURE 5 describes the Moving Average and Exponential Smoothing of IoT attack data over time. To preprocess the data, we observe the fields of the dataset (Timestamp, LoggedIn, and Credentials). Binning or Aggregation was engaged, grouping the data into time intervals and counting occurrences of login attempts within each bin. Addressing potential limitations such as data gaps and ensuring a continuous representation of attack counts. The blue line represents the Original Counts, showing the raw data distribution over the entire period, signifying variations in the attack frequency. The red line means the Moving Average, computed to smooth out short-term fluctuations and highlight longer-term trends in attack activity. The red line (moving average) is 0 representing missing data (NaN values) which is a limitation. This choice prioritizes clarity by indicating periods with no recorded data points. It makes simple visualization by placing the moving average at zero when data is unavailable. Also, the green line represents Exponential Smoothing, providing a weighted average of attack counts to highlight recent trends while attenuating noise. The exponential smoothing line (400,000 counts axis) suggests a possible average level of attack counts around that value. FIGURE 5 provides insights into the temporal patterns of IoT attacks. The Original Counts uncover fluctuations in attack activity, with peaks and curves correlating with periods of heightened and reduced attack volumes, respectively. The Moving Average smooths out short-term variations, uncovering fundamental trends in attack activity. The placement of the red line at 0 emphasizes periods with no recorded attacks, signifying potential gaps in data collection or periods of decreased attack activity. Exponential Smoothing gives further insights by highlighting recent trends while lowering the effects of noise. The relatively stable distribution of the green line indicates a consistent

level of attack activity over time, with minor fluctuations around the smoothed trend. Overall, FIGURE 5 exhibits the application of moving average and exponential smoothing for smoothing time series data mean IoT attack counts.

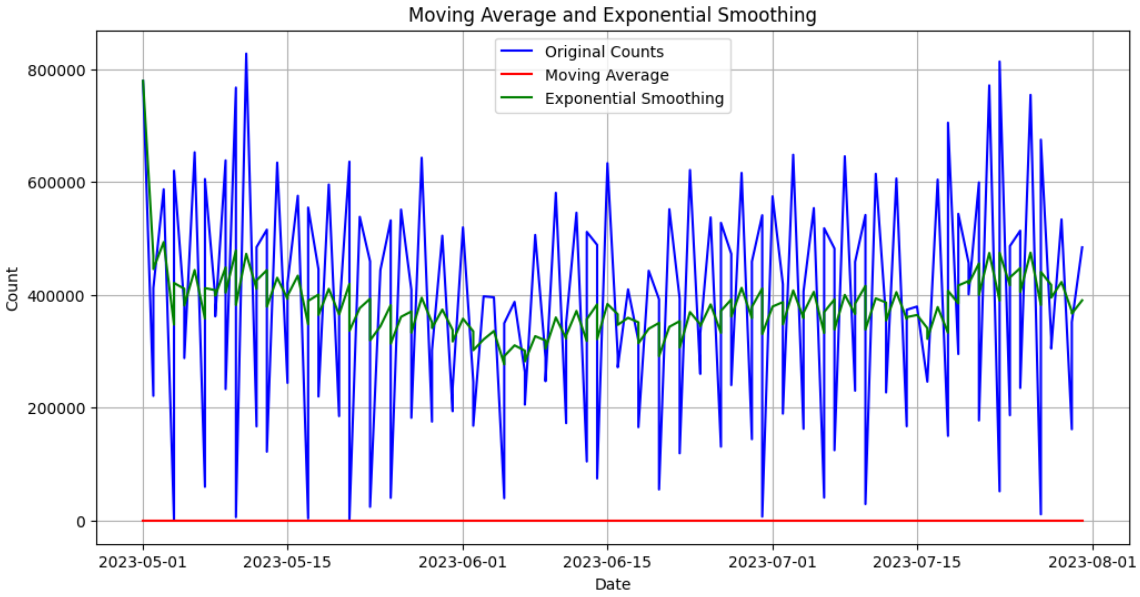

Figure 5: Time-series weekly attack counts the fluctuation in attack frequency over time.

4. DISCUSSION

The study presented provides valuable insights into the emerging trends and vulnerabilities in IoT security threats through a comprehensive analysis of attack data obtained from the GCA AIDE. Utilizing systematic preprocessing techniques, including data cleaning, normalization, and outlier detection, we prepared the data for analysis, mitigating potential biases and ensuring the robustness of the later findings. Time-series decomposition techniques, such as seasonal decomposition of time series (STL), give insights into the fundamental trends, seasonality, and noise present in the data. Autocorrelation analysis explained temporal relationships and cyclic trends within the dataset, emphasizing the dynamic nature of IoT attack occurrences over time.

Anomaly detection is the identification of patterns, where there are elements that make challenging the interpretation [12]. Leveraging statistical learning methods aid uncover flag deviations as anomalies [13, 14]. IoT devices and systems have limited processing resources, resulting in limitations, particularly for anomaly detection [15]. One of the foremost findings of our analysis was the identification of distinct patterns in attack frequency over time. Visualizations (FIGURE 2, and FIGURE 3) of time-series data displayed fluctuations in attack activity, periodic spikes in login attempts, and potential anomalies indicative of coordinated attack campaigns. Moving average and exponential smoothing techniques were engaged to smooth out short-term fluctuations and underline longer-term trends in attack activity. Although these techniques proved effective in unveiling trends

in attack behavior, they also introduced potential limitations and biases. The approach to fill NaN values in the moving average with 0 was made to sort clarity and simplicity. Regardless of limitations, our analysis underlines moving averages and exponential smoothing as tools for detecting temporal patterns in IoT attack data. By leveraging these techniques, cybersecurity practitioners can gain key insights into attack trends, identify potential threats, and advance their proactive defense strategies.

Our analysis using time-series techniques unveiled key insights into the temporal dynamics of IoT attacks. Here's a sort of the key findings:

- Temporal Variability: The summary statistics (FIGURE 1) exposed a high number of unique timestamps, login attempts (loggedin), and attempted credentials (credentials), emphasizing the highly dynamic nature of IoT attacks.
- Attack Frequency: The time series (FIGURE 2 and FIGURE reffig3) exhibited fluctuations in attack frequency over time, suggesting no consistent seasonal patterns.
- Autocorrelation Analysis: Since the dataset demonstrated a vast temporal scope, the analysis of timestamps, start times, and end times (FIGURE 4) did not uncover recurring temporal patterns. Though, the analysis of successful login attempts (loggedin) highlighted commonly exploited vulnerabilities (root, and 3245gs5662d34).

These findings convey significant weight in the realm of IoT security. The dynamic nature of attack occurrences emphasizes the constant evolution of attack vectors and vulnerabilities. The lack of clear seasonal patterns indicates attackers may be opportunistic, exploiting newly discovered vulnerabilities or targeting specific campaigns. The identification of commonly exploited login credentials emphasizes the need for robust password management practices and ongoing vulnerability assessments for IoT devices.

Moreover, the effectiveness of smoothing techniques in disclosing underlying trends underlines the importance of data preprocessing and feature engineering for time-series analysis of IoT attack data. By smoothing out short-term fluctuations, these techniques allow the identification of long-term trends and potential baselines for attack. Furthermore, the analysis focused on a specific dataset from a single honey farm. Examining data from diverse sources and locations could give a more thorough understanding of global IoT attack trends. Future research could go into specific attack vectors and vulnerabilities identified through login attempt analysis.

5. CONCLUSION

In this study, we looked into the complex landscape of IoT security threats, focusing on the emerging trends, frequencies, and vulnerabilities related to attacks on IoT devices. Through extensive data analysis and the application of advanced time-series techniques, we unravel key insights into the temporal dynamics of IoT attacks, providing valuable intelligence for enhancing cybersecurity defenses. The findings of our analysis emphasize the dynamic nature of IoT attack occurrences, with fluctuations in attack frequency over time indicating the evolving tactics and strategies engaged

by attackers. Utilizing techniques such as time-series decomposition, autocorrelation analysis, and moving averages, we discovered temporal patterns and anomalies in attack data and illuminated critical vulnerabilities and attack vectors.

Our findings have not just enlightened the temporal dynamics of IoT attacks but have also contributed to the enhancement of IoT security measures. Utilizing the thorough analysis of IoT attack data using time-series techniques, we have discovered previously unrecognized patterns and trends that serve as valuable insights for cybersecurity practitioners and stakeholders. Our findings uncover critical insights, including the identification of repeated attack vectors such as credential and distributed denial-of-service (DDoS) attacks, the detection of anomalous login patterns reflective of unauthorized access attempts, and the detection of temporal changes in attack frequency correlated with global events and cybersecurity trends. By using those insights, industry organizations can proactively mitigate potential threats, sustaining the resilience of IoT environments against cyberattacks. Additionally, our study has accelerated the improvement and optimization of actual security protocols, empowering for more effective detection and response mechanisms. Also, our research has brought the development of novel defense mechanisms, such as anomaly detection algorithms and predictive analytics models. While our study provides a significant step forward in understanding IoT security threats, there are several directions for future research and exploration. Notably, future work could focus on feature engineering and the development of Machine Learning (ML) and Artificial Intelligence (AI) models for predictive analysis of IoT attacks within GCA AIDE. By benefiting from advanced analytical techniques and predictive modeling, researchers and cybersecurity practitioners can predict and mitigate emerging threats more effectively, thus advancing the resilience of IoT ecosystems against cyberattacks. In addition to enhancing predictive abilities, future research could go deeper into the investigation of specific attack vectors and vulnerabilities identified through login attempt analysis. By overseeing more granular analyses and clarifying the basic mechanisms thriving these attacks, we can strengthen IoT defenses and communicate the designing of targeted mitigation strategies. In summary, our study provides the basis for a more thorough understanding of IoT security threats and emphasizes the importance of continued research and innovation in this domain.

6. ACKNOWLEDGMENT

We would like to thank Global Cyber Alliance (GCA) for sharing the data with us. The ICT Academy supported the work under the Research & Innovation Department in partnership with GCA.

References

[1] Buja A, Apostolova M, Luma A, Januzaj Y. Cyber Security Standards for the Industrial Internet of Things (IToT)– A Systematic Review International Congress on Human-Computer Interaction, Optimization and Robotic Applications(HORA). 2022.

[2] Buja A, Apostolova M, Luma A. Enhancing Cyber Security in Industrial Internet of Things Systems: An Experimental Assessment 12th Mediterranean Conference on Embedded Computing (MECO). 2023.

[3] Al-Hawawreh M, den Hartog F, Sitnikova E. Targeted Ransomware: A New Cyber Threat to Edge System of Brownfield Industrial Internet of Things. IEEE Internet Things J. Aug. 2019;6:7137-7151.

[4] Uwagbole SO, Buchanan WJ, Fan L. Applied Machine Learning Predictive Analytics to SQL Injection Attack Detection and Prevention. INTERNATIONAL FEDERATION FOR INFORMATION PROCESSING. 2017.

[5] Mohan Raparthy E. Predictive Maintenance in IoT Devices using Time Series Analysis and Deep Learning. Dandao Xuebao J Ballistics. Dec. 2023;35:1-10.

[6] https://www.gartner.com/en/doc/iot-security-primer-challenges-and-emerging-practices.

[7] https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/cybersecurity-for-the-iot-how-trust-can-unlock-value.

[8] https://www.globalcyberalliance.org/

[9] https://research-methodology.net/research-methodology/research-types/applied-research/

[10] Nielsen C, Lund M, Montemari M, Paolone F, Massaro M, et al. Applied Research Methodology. Bus Models. 2018:16-21.

[11] Qin S, Chen L, Luo Y, Tao G. Multiview Graph Contrastive Learning for Multivariate Time-Series Anomaly Detection in IoT. IEEE Internet Things J. Dec. 2023;10:22401-22414.

[12] Cook AA, Misirli G, Fan Z. Anomaly Detection for IoT Time-Series Data: A Survey. IEEE Internet Things J. Jul. 2020;7:6481-6494.

[13] Li F, Shinde A, Shi Y, Ye J, Li XY, et. al. System Statistics Learning-Based IoT Security: Feasibility and Suitability. IEEE Internet Things J. 2019;6:6396-6403.

[14] Mohanta BK, Jena D, Satapathy U, Patnaik S. Survey on IoT Security: Challenges and Solution Using Machine Learning, Artificial Intelligence and Blockchain Technology. Internet Things. Sep. 2020;11:100227.

[15] Huc A, Trcek D. Anomaly Detection in IoT Networks: From Architectures to Machine Learning Transparency. IEEE Access. 2021;9:60607-60616.