International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2025: 7.089

Forecasting Cyber Threat Trends: A Hybrid Statistical-Deep Learning Approach

Jayashree P. Darade[1], Dr. Anita Chaware[2]

P.G. Department of Computer Science, SNDT Women's University, Santacruz, India
Email: *jayashreedarade1[at]gmail.com*

P.G. Department of Computer Science, SNDT Women's University, Santacruz, India
Email: *anita.chaware[at]computersc.sndt.ac.in*

Abstract: ***The fast growth of technology has resulted in significant rise in cyber-attacks. It creates substantial challenges for organizations and governments worldwide. Traditional cybersecurity measures are not sufficient to address growing attack patterns. This study presents a forecasting framework using a Hybrid Statistical Model ARIMA and Deep Learning Model LSTM to predict year wise cyber-attack trends and number of affected users across different attack types. The linear component of attack pattern is captured by ARIMA and the nonlinear residual pattern are modeled using LSTM networks. This combination improves the prediction accuracy to 100%. Historical Time Series cyber-attack data from 2015-2024 is used for training and evaluation. The hybrid model gives higher performance with MSE, RMSE and MAPE validating its effectiveness. The framework provides forecasting of next five years for actionable insights for cybersecurity planning, resource allocation and risk mitigation.***

Keyword: Cybersecurity, ARIMA, LSTM, Time Series Prediction, Cyber-attacks

1. Introduction

A time series refers to the sequence of observations recorded in chronological order at consistent time intervals like hourly, daily, weekly, monthly or yearly. Time series forecasting relies on the idea that historical pattern and behaviors present in the data is used to predict upcoming values of the series [1]. Cybersecurity involves safeguarding information system and digital infrastructure and protecting data from evolving cyber threats through advanced technologies, strategic approaches and collaborative efforts, including AI and ML based threat detection and response mechanisms [2]. Cyber-attacks have become global challenge driven by digital acceleration, extensive acceptance of cloud computing and increasing interconnectedness of modern systems. As dependence of cyberspace continues to grow, organizations face a number of cyber threats including ransomware, malware, man-in-the-middle, phishing, DDoS, SQL Injection, etc. [3]. There is a need for robust and comprehensive strategies to reduce risks, safeguard sensitive information and maintain data integrity and system reliability [4]. Cyber security is a challenge in maritime sector also, so some studies addresses this research gap and surveys the methods, algorithms, tools and architectures used for cyber security monitoring in the maritime sector [5]. Existing research focus on the effectiveness of AI-driven approaches in enhancing the overall cyber security education lifecycle [6]. Time series forecasting models offer a systematic approach in analyzing historical cyber-attack data and predicting future trends. Statistical models such as the ARIMA are effective in addressing linear temporal dependencies and long term trends in time series data. ARIMA model face difficulties in representing complex linear patterns commonly observed in cybersecurity datasets. On the other hand, deep learning model, LSTM networks are well suited for modeling nonlinear relationships and temporal dependencies but when it directly applied to non-stationary data without explicitly accounting for linear components, their performance may be limited [7]. To overcome these limitations, this study presents a hybrid ARIMA-LSTM forecasting framework for year-wise prediction of cyber-attack trends based on number of affected users across multiple attack types. In the proposed approach ARIMA is first applied to model the linear and trend components of time series, while LSTM networks are trained on the residuals to capture the linear patterns. This hybrid approach increases the strength of statistical and deep learning methods which results in improved forecasting accuracy. The proposed model is evaluated using historical cyber-attack data of multiple years. The performance is accessed using standard error matrices like Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). The result demonstrate that hybrid approach effectively captures both linear and non-linear components of cyber-attack time series, making it a trustworthy tool for proactive cyber security planning and decision making.

2. Literature Review

Forecasting cyber threat trends is essential for cybersecurity risk mitigation. The changing and non-stationary nature of cyber-attack data increases the challenges for traditional forecasting methods. The hybrid approached which integrates different models offer improved capability to capture both linear patterns and complex temporal dependencies. Time series is a series of data collected over a period of time. For the study different domains literatures were reviewed and mentioned here. The data collected over time can be linear or non-linear. There are different models of forecasting available depending upon nature of the data [8]. Social Cybersecurity focuses on human-centric cyber threats within online network. It includes attack detection using machine learning and network analysis techniques, supported by publically available datasets and tools for addressing existing challenges and future research direction [9]. Anomaly detection in cyber physical energy systems analyzes sensor time-series data using predictive models and heuristic segmentation to identify the trend anomalies, improve reliability, security and

real time operational performance [10]. Threat intelligence data forms a temporal sequence reflecting evolving cyber risks. Statistical time series models analyze trends and volatility to support future threats and support timely data driven security decisions [11]. Phishing attack data increases over a period of time with noisy and complex patterns. Deep learning based time series forecasting models capture temporal dependencies to predict future phishing trends and enable cyber security planning [12]. Cyber warfare and cyber terrorism involve evolving digital threats [13]. Time series based intelligent detection system analyze network traffic patterns using machine learning algorithms to accurately identify DDoS attack achieving high detection accuracy [14]. LSTM is a recurrent neural network used for historical time series prediction. It incorporates multiple memory cells to capture complex features in time series data [15], [16]. LSTM is a robust and superior choice for modeling non-linear dynamic systems, effectively mitigating traditional RNN challenges [17]. Time series forecasting using ARIMA supports vegetable sales prediction, while linear programming optimizes pricing strategies and decision making in agricultural supply chain [18]. Hybrid ARIMA-LSTM model capture linear trends and non-linear volatility in stock market price time series achieving improved prediction accuracy across diverse sectors and varying economic conditions [19]. Short term load forecasting models combine ARIMA-LSTM to capture a linear trends and non-linear pattern in time series data, achieving improved prediction accuracy across diverse power consumption datasets [20]. As per the reviewed literature there is a need of accurate prediction of cybersecurity threats. ARIMA and LSTM models are performing very well in diverse domains and giving better performance. Combination of both the models are not used in the cyber security domain. So we have proposed a Hybrid ARIMA-LSTM Model for cybersecurity threats trends in this paper.

3. Methodology

A) Autoregressive Integrated Moving Average (ARIMA):
ARIMA is a statistical time-series forecasting model used to analyze historical data and predict future values. It is particularly effective for capturing linear trends and temporal dependencies in time-dependent data. It consists of three components.

1) Autoregression (AR):
Autoregression in ARIMA is given by parameter 'p'. It captures how current observations depends on their past values.

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t$$

Here, Y_t - current observation, 'c' - constant, ϕ_1, ϕ_2 are autoregressive parameters and ϵt - error term at time 't'

2) Differencing (I):
Differencing in ARIMA is represented by 'd'. It is use to make the series stationary.

$$Y'_t = Y_t - Y_{t-1}$$

Here, Y'_t - differenced series at time 't', Y_t - original series at time 't' and Y_{t-1} - the value of the series at the previous time stamp.

3) Moving Average(MA):
Moving Average component, denoted by 'q', represents the dependence of current observation on the past forecast errors.

$$Y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

Here, Y_t - the current observation, C - constant, ϵ_t - error term at time 't' and θ_1 to θ_q are moving average parameters.

B) Long Short Term Memory(LSTM):
LSTM networks is a type of Recurrent Neural Network used effectively for time series forecasting. The heart of LSTM Network is its cells. LSTM network includes three primary gates, forget gate, input gate and output gate. Gates are responsible to retain or discard the information to learn long term dependencies effectively in the data.

The equation for gates in LSTM are:

$$i_t = \sigma\,(w_i\,[h_{t-1}, x_t] + b_i)$$

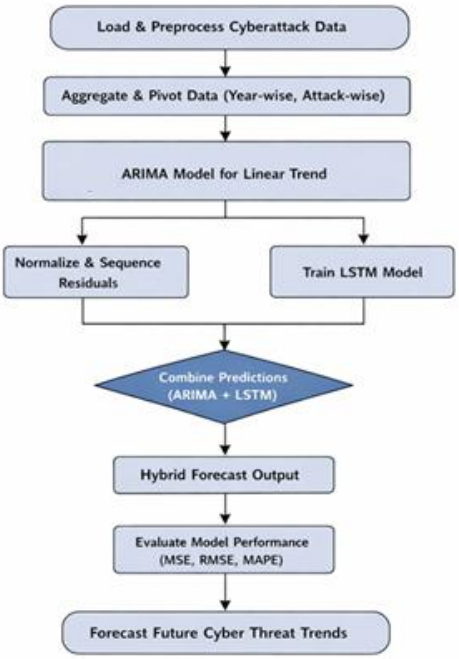
$$f_t = \sigma\,(w_f\,[h_{t-1}, x_t] + b_f)$$

$$o_t = \sigma\,(w_o\,[h_{t-1}, x_t] + b_o)$$

where, i_t - input gate, f_t - forget gate, o_t - output gate, σ - sigmoid function, w_x - weight for respective gate (x), h_{t-1} - output of previous LSTM block at timestamp $t-1$, x_t - input at current time stamp and b_x represents biases for the respective gates.

The input gate controls how much of the new information should be added to the cell state. The forget gate determines which part of the previous cell state should be retained or discarded that allows model to keep only relevant information. The output gate regulates how much of the current cell state is passed on to the next layer or time stamp. Together these gates allow LSTM to maintain long term memory and learn dependencies over time in sequential data.

C) ARIMA-LSTM Model:



The proposed ARIMA-LSTM framework forecasts cyber trends by integrating statistical and deep learning techniques. Firstly, year-wise cyber- attack data is loaded and

preprocessed to remove inconsistencies. After that data is aggregated and pivoted to form multivariate time series based on attack types. ARIMA model is applied to each attack category like DDoS, Malware, Man in the Middle, Phishing, Ransomware, SQL injection to capture linear trends after logarithmic transformation. The residual obtained from ARIMA are normalized and structured into time dependent sequences. These residual sequences are used to train an LSTM network capable of learning complex temporal dependencies. The final forecast is generated by combining ARIMA predictions with LSTM predicted residuals by capturing both linear and non-linear dynamics. Model performance is evaluated using MSE, RMSE and MAPE. The trained hybrid model is then used to forecast future cyber threat trends for proactive cyber security planning. Following are the graphical representation of actual versus predicted attacks based on attack category.

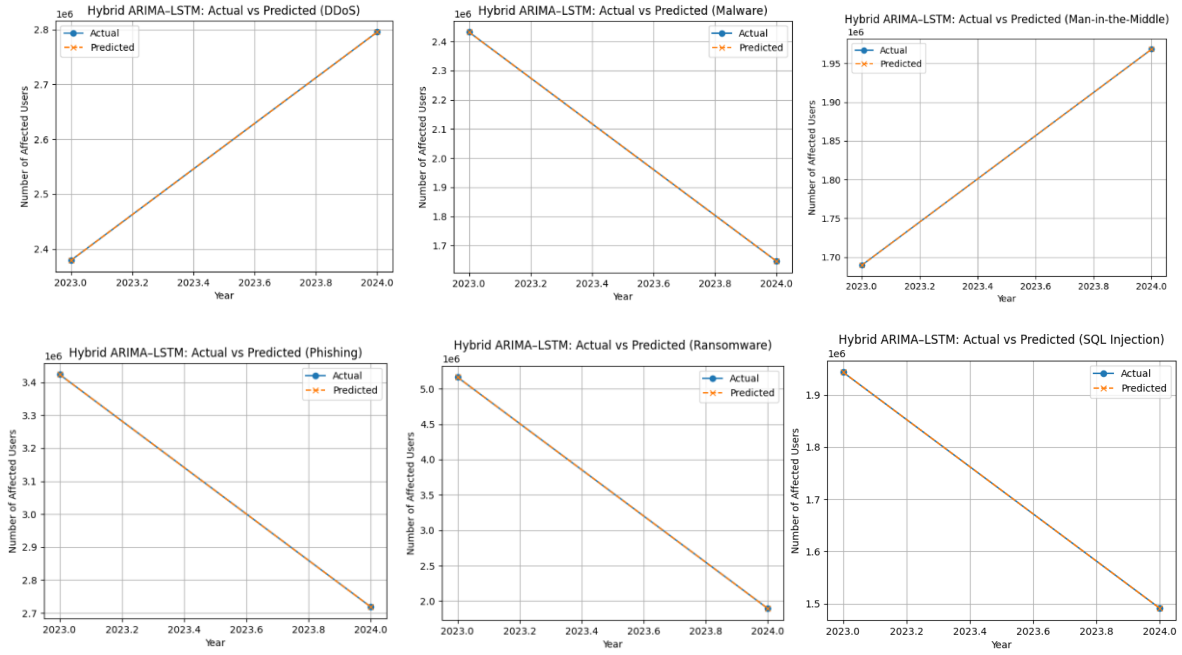


Figure 1: Actual versus predicted attacks based on attack category

4. Results & Discussion

The hybrid ARIMA-LSTM model was evaluated using year-wise cyberattack data across multiple attack categories. Model performance is evaluated on the basis of Mean Squared Error(MSE), Root Mean Squared Error(RMSE) and Mean Absolute Percentage Error(MAPE). Based on MAPE accuracy is calculated. As per the experiment the hybrid model achieved MSE is 0.17, RMSE is 0.41 and MAPE is 0.00%, which gives forecasting accuracy of 100%. The result shows the high precision and robustness of proposed framework in modeling cyber threat trends.

5. Conclusion

In this study we included the idea of combining statistical and deep learning model to get better forecasting results. We developed a hybrid ARIMA-LSTM framework for forecasting cyber threat trends using year-wise, attack-wise affected user data. By combing the strength of ARIMA in modeling linear temporal patterns with the capability of LSTM to capture nonlinear dependencies, the proposed approach effectively addressed the non-stationary nature of cyber- attack data. Experimental results demonstrated high predictive accuracy with minimum forecasting error. The framework provides forecasting of next five years for actionable insights for cybersecurity planning, resource allocation and risk mitigation.

6. Future Scope

Future work may focus on incorporating real-time data streams, additional threat indicators and advanced deep learning architecture to further enhance forecasting performance and scalability.

References

[1] Vijay Kotu, Bala Deshpande, Chapter 12 Time Series Forecasting,Editor(s): Vijay Kotu, Bala Deshpande, Data Science (Second Edition), Morgan Kaufmann,2019, Pages 395-445, ISBN 9780128147610, https://doi.org/10.1016/B978-0-12-814761-0.00012-5.

[2] Wasyihun Sema Admass, Yirga Yayeh Munaye, Abebe Abeshu Diro,Cyber security: State of the art, challenges and future directions,Cyber Security and Applications,Volume 2,2024,100031, ISSN 2772-9184, https://doi.org/10.1016/j.csa.2023.100031.

[3] *Cyber Security Challenges*. (2025, August 20). SentinelOne. Retrieved December 17, 2025,from https://www.sentinelone.com/cybersecurity-101/cybersecurity/cyber-security-challenges/

[4] W. A. Busari and A. A. Bello, "Security, Trust, and Privacy in Cyber-physical Systems (CPS)," 2024 2nd International Conference on Cyber Physical Systems, Power Electronics and Electric Vehicles (ICPEEV),

Hyderabad, India, 2024, pp. 1-6, doi: 10.1109/ICPEEV63032.2024.10932087.

[5] R. Vaarandi, L. Tsiopoulos, G. Visky, M. Ur Rehman and H. Bahşi, "A Systematic Literature Review of Cyber Security Monitoring in Maritime," in IEEE Access, vol. 13, pp. 85307-85329, 2025, doi: 10.1109/ACCESS.2025.3567385.

[6] S. Jawhar, J. Miller and Z. Bitar, "AI-Driven Customized Cyber Security Training and Awareness," 2024 IEEE 3rd International Conference on AI in Cybersecurity (ICAIC), Houston, TX, USA, 2024, pp. 1-5, doi: 10.1109/ICAIC60265.2024.10433829.

[7] Landauer, M., Skopik, F., Stojanović, B. et al. A review of time-series analysis for cyber security analytics: from intrusion detection to attack prediction. Int. J. Inf. Secur. 24, 3 (2025). https://doi.org/10.1007/s10207-024-00921-0

[8] Fisher, A., Hodgdon, T., & Lewis, M. (2024). Time-series forecasting methods : a review. https://doi.org/10.21079/11681/49450

[9] Aos Mulahuwaish, Basheer Qolomany, Kevin Gyorick, Jacques Bou Abdo, Mohammed Aledhari, Junaid Qadir, Kathleen Carley, Ala Al-Fuqaha, A survey of social cybersecurity: Techniques for attack detection, evaluations, challenges, and future prospects,Computers in Human Behavior Reports, Volume 18, 2025, 100668, ISSN 2451-9588

[10] Y. Bao, Z. Wei, Z. Wang, Y. Zhao, Y. Li and Y. Liu, "Anomaly Detection for Sensor Data in Cyber-Physical Energy Systems Based on Time-Series Prediction and Heuristic Segmentation," 2025 40th Youth Academic Annual Conference of Chinese Association of Automation (YAC), Zhengzhou, China, 2025, pp. 3008-3013, doi: 10.1109/YAC66630.2025.11150096.

[11] D. Li, Z. Ran, Y. Yao and X. Zhou, "Threat Intelligence Disclosure Trend Analysis Model Based on Time Series," 2021 2nd International Conference on Electronics, Communications and Information Technology (CECIT), Sanya, China, 2021, pp. 1245-1250, doi: 10.1109/CECIT53797.2021.00221.

[12] M. S. I. Alsumaidaie , K. M. A. Alheeti and A. K. Al-Aloosy, "Intelligent Detection System for a Distributed Denial-of - Service (DDoS) Attack Based on Time Series," 2023 15th International Conference on Developments in eSystems Engineering (DeSE), Baghdad & Anbar, Iraq, 2023, pp. 445-450, doi: 10.1109/DeSE58274.2023.10100180

[13] S. Kumar and A. K. Keshri, "Cyber Security against Cyber Warfare and Cyber Terrorism," 2025 International Conference on Innovations in Intelligent Systems: Advancements in Computing, Communication, and Cybersecurity (ISAC3), Bhubaneswar, India, 2025, pp. 1-6, doi: 10.1109/ISAC364032.2025.11156392.

[14] M. S. I. Alsumaidaie, K. M. A. Alheeti and A. K. Al-Aloosy, "Intelligent Detection System for a Distributed Denial-of - Service (DDoS) Attack Based on Time Series," 2023 15th International Conference on Developments in eSystems Engineering (DeSE), Baghdad & Anbar, Iraq, 2023, pp. 445-450, doi: 10.1109/DeSE58274.2023.10100180.

[15] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[16] Cell-expanded Long Short-term Memory. (2022). 2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposiumc on Advanced Intelligent Systems (SCIS&ISIS). https://doi.org/10.1109/scisisis55246.2022.10001924

[17] Sahu, R., Srivastava, S., & Kumar, R. (2023). Modelling of a Non-Linear Dynamic System Using Long Short-Term Memory. 1016–1021. https://doi.org/10.1109/icccis60361.2023.10425027

[18] Z. Dong and H. Hao, "Vegetable Sales Forecasting and Pricing Strategy Planning Based on ARIMA Algorithm and Linear Programming Model," 2024 4th International Symposium on Computer Technology and Information Science (ISCTIS), Xi’an, China, 2024, pp. 376-381, doi: 10.1109/ISCTIS63324.2024.10698868.

[19] H. Margaretha, P. M. Tauhalomoan, P. Widjaja and F. V. Ferdinand, "Predictive Analysis of Stock Prices Using ARIMA and Hybrid ARIMA-LSTM Models in the Indonesian Market," 2025 22nd International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Bangkok, Thailand, 2025, pp. 1-6, doi: 10.1109/ECTI-CON64996.2025.11101223.

[20] S. Chen, R. Lin and W. Zeng, "Short-Term Load Forecasting Method Based on ARIMA and LSTM," 2022 IEEE 22nd International Conference on Communication Technology (ICCT), Nanjing, China, 2022, pp. 1913-1917, doi: 10.1109/ICCT56141.2022.10073051.