Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 6 (2023)

Cybernetic Sentinel: Harnessing Generative AI for Centralized IoT Security Orchestration

S. Balachandar[1], Dr. R. Chinnaiyan[2]

[1]*Research Scholar, VTU-RC, Department of MCA, CMR Institute of Technology, Bengaluru - 560037*

[2]*Professor & Supervisor, VTU-RC, MCA, Department of MCA, CMR Institute of Technology, Bengaluru - 560037*

Abstract: Centralized IoT (Internet of Things) Security Analysis using a chatbot involves leveraging a chatbot to streamline and enhance the security analysis process for IoT devices. The logs collected from different IOT gateways, IOT Edge servers or direct devices are the key data sources for analysis. The chatbot (Virtual Assistant) is an application deployed in the cloud environment which constantly helps the device users, command center team or any regulator to query the device characteristics and its metadata and current data pattern. Some of the models deployed on top of this log data will also help the bot to answer whether any packet flooding or anomalous data pattern that we observe in a defined time window.

Keywords: *chatbot, GPT, TCP packet, IOT Edge, Cloud Platform, Distributed Database, IOT Gateway, anomaly detection*

1. Introduction

As per Forbes by the end of 2024, more than 207 billion IOT devices connected to network which brings more value to business and consumer IOT markets.[1]. It alarms that there many security threats, Vulnerabilities may arise in device memory, firmware, physical networks, web applications and network services. Its important to safeguard sensitive data and critical infrastructure used by the IOT network. The data emitted from different networks like Edge[2] or Cloud which needs a centralized storage to collect those logs and messages which not only helps to protect these devices from any unknown attacks[3] (e.g., Denial of services, Botnets, etc..) but also helps to understand the security risk from different device end points, network zones. Employing chatbot[4] will help us to answer the queries from these logs that we collected in the central storage to get the details what we look from the security perspective but also it helps to get any additional data that are required to predict any security attacks soon. We are going to analyze different models that we used to build chatbots. In section II we analyze the reviews from different authors and publishers mentioned about how chatbot models and IOT Security risks and models will play vital role for understand the patterns of attacks or to prevent any vulnerabilities. We will discuss the high level approach and components in Section III. We will elaborate different IOT security models are feasible with chatbot based approach in the Section IV. In section V we will mention about problem relevancy with high level solution mapping with known challenges. We will share the deployment considerations and issues in section VI. In section VII we will present our recommendations and conclusion.

Problem Statement: To collect and query different IOT device metadata details and understand the security related data points is always a challenge in a large scale centralized IOT environment (e.g., Smart City or any Industry who is using thousands of devices).

2. Literature Survey

Employing chatbots will help us to get many benefits from security aspect, however the deployment of functional or AI based chatbots in cloud environment must undergo lot of challenges. The chatbot scope needs to be narrowed down to analyze data that are required for chatbot's entities, context, intents, actions.

- **"ROHAN KAR" AND "RISHIN HALDAR"** researched on November 2016 ABOUT **"APPLYING CHATBOTS TO THE INTERNET OF THINGS: OPPORTUNITIES AND ARCHITECTURAL ELEMENTS[5]"** they articulated about the technology centric challenges for IoT, Integration of Chatbot in IOT cloud platform and they also mentioned potential areas like stronger AI based agents, Cyber physical IOT Systems, Wisdom of Things and Semantic Web. It is important to evaluate AI based agents that will overcome some of the challenges that they posted (Cognitive Burden, Support Challenges, Search and Discoverability)
- Prasannjeet Singh, Mehdi Samanazari, Francesco Vitale, Francesco Flammini, Nicola Mazzocca, Mauro Caporuscio and Johan Thornadtsson released their paper about "USING LOG ANALYTICS AND PROCESS MINING TO ENABLE SELF-HEALING IN THE INTERNET OF THINGS[6]" . It talks more out process mining which combines the log analytics with machine learning model. It supports early diagnosis, prognosis, and subsequent automated repair to improve the resilience of IoT devices within possibly complex cyber-physical systems. They leveraged Event Logs emitted from IOT Devices and their research is more toward effective Process Discovery and Conformance Checking to support Self-Healing in IoT. It corelates with my topic of How those logs what we collect can help on investigating the security events captured which not only safeguard the device but also ensure that it can be self-healed once we instruct a chatbot to either turn off the device or update the device properties in run-time to prevent any major attacks.
- Partha Pratim Ray explained about "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope[7]". His research explores the various ways ChatGPT has been used in different areas like scientific research, spanning from data processing and hypothesis generation to collaboration and public outreach. It also examines the potential challenges and ethical concerns surrounding the use of ChatGPT in research, while highlighting the importance of striking a balance between AI-assisted innovation and human expertise. This relates to my study on how conversational bot or virtual assistant needs to make a trade-off carefully for analyzing the IOT data where some of the IOT data captures privacy information which needs strong anonymization.
- Ayan Chatterjee, Bestoun S. Ahmed talked about "IoT anomaly detection methods and applications: A survey[8]". Their paper talks about how two different types of data (IOT Application and Network) used. Also it mentions about wavelet autoencoder anomaly detection technique. Also, high detection accuracy is obtained by combining IOT Edge and Cloud data. The HDoutliers algorithm is a powerful unsupervised method for discovering abnormalities in high-dimensional data based on a distributional model that allows outliers to be tagged with a probability.

PROPOSED DESIGN – CENTRALIZED IOT SECURITY ANALYSIS USING CHATBOT

Here we are collecting the event data from different IOT devices or IOT gateways through a streaming platform (e.g. Kafka[9]) and its continuously streams and store the data in to a Distributed SQL database "YugabyteDB[10]". YugabyteDB is helping to store all device metadata like device id, certificate details, manufacturer, firmware details, etc. The data will be transferred from YugabyteDB over to Elastic Search[11] Database continuously in real-time to help searching, analyzing and discover the data coming from different IOT gateways and IOT sensors. We also deploy a custom anomaly detection model to identify the data anomaly coming from different devices data. The chatbot application built it using GPT3 model which continuously learns the get the prompts based on the user's questionnaire about IOT Event Log data. The below diagram (Figure1) shows the end-to-end data flow from IOT Devices to Chatbot App.

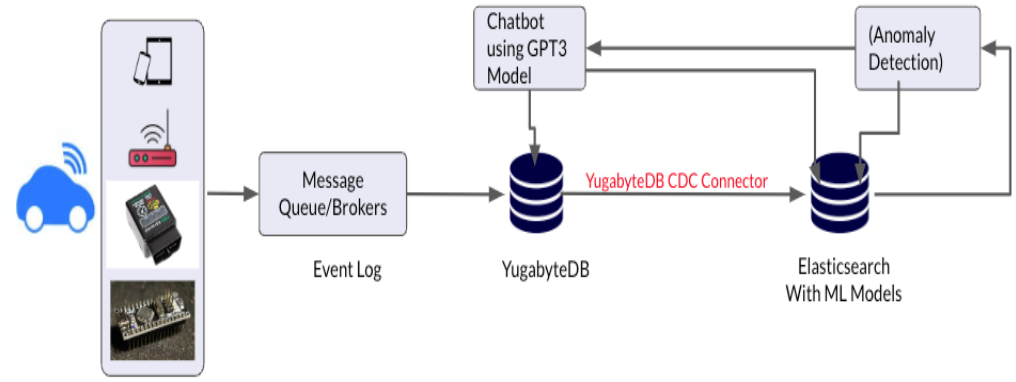

Figure-1 Proposed Architecture for IOT Security Analysis using Chatbot

System/Application	What it does	Remarks
IOT Devices/IOT Gateway/IOT Edge Servers	Stores/Streams the events from different devices, Gateways and edge server in the form of JSON/CSV data	Data Origination Point
Message Queue/Brokers	Streamed Data gets ingested through messaging platform	Data Collection Point
Distributed Database	Data coming from message brokers gets stored into Distributed database (YugabyteDB) which helps to powers global edge and streaming applications. It enables hybrid environments, eradicates data silos, and facilitates data collection from the edge and aggregation in the cloud	Data Store
Search Engine	Data Stored in database gets into Search Engine for data discovery, analysis and helping the chatbot or any virtual assistant to query the historical data and helps to build a Textual model to build Q&A for the IOT Security aspect. We can also build the Machine Learning Model using the historical data that we persist in Elastic Search Engine.	Historical Data Store
Chatbot	Helps to query the data from either YugabyteDB for transactional data or Elastic Search for historical and machine learning model output (e.g. anomaly behavior)	Virtual Assistant

Based on the above problem statement, its better to combine the strengths of both qualitative and quantitative approaches, mixed-methods research provides a more nuanced and robust understanding of research problems. Hence we decided to take mixed-methods research type.

Research Type: Mixed-Methods Research [14]

Complexity of the System: The described research involves a multi-faceted system that includes data collection from IoT devices, streaming and storage in YugabyteDB, real-time transfer to Elastic Search, deployment of an anomaly detection model, and interaction with a GPT-3 based chatbot. A mixed-methods approach allows for a comprehensive understanding of this complex system.

Quantitative Aspect - Database Performance: The quantitative component can focus on assessing the performance metrics of YugabyteDB and Elastic Search. This includes metrics like data transfer rates, storage efficiency, and query response times. Quantitative data helps in objectively measuring the efficiency and effectiveness of the chosen technologies.

Qualitative Aspect - User Interaction and Anomaly Detection: The qualitative component can explore user experiences with the chatbot (e.g. GPT3 Model), understanding how well it serves their needs and the effectiveness of the anomaly detection model. This involves gathering insights through user interviews, feedback, and observations. Qualitative data provides a deeper understanding of user perceptions and system usability.

Mixed-Methods - Integration and System Dynamics: Combining qualitative and quantitative data allows for a holistic analysis of the integration between components and the overall system dynamics. This approach is particularly useful when dealing with emerging technologies like IoT, where understanding user experiences and system performance is equally important.

Continuous Learning and Adaptation: The mixed-methods approach aligns well with the continuous learning aspect mentioned in the description. It allows researchers to adapt the research design based on emerging insights, ensuring that the study remains responsive to the evolving nature of the IoT data, anomaly patterns, and user interactions.

Comprehensive Insights: By combining quantitative metrics on database performance with qualitative insights into user interactions, the research can provide comprehensive insights into the strengths, weaknesses, and potential improvements of the entire system. This aligns with the complexity of the research objectives.

3. Summary

In summary, a mixed-methods approach is chosen to leverage the strengths of both qualitative and quantitative research in addressing the multi-dimensional aspects of the research problem, providing a well-rounded understanding of the entire IoT data processing and interaction system.

4. References

[1] Applying Chatbots to the Internet of Things: Opportunities and Architectural Elements, DOI:10.14569/IJACSA.2016.071119, November 2016International Journal of Advanced Computer Science and Applications 7(11)

[2] G Sabarmathi, R Chinnaiyan (2019), Envisagation and Analysis of Mosquito Borne Fevers: A Health Monitoring System by Envisagative Computing Using Big Data Analytics, Lecture Notes on Data Engineering and Communications Technologies book series (LNDECT, volume 31), 630-636. Springer, Cham

[3] G. Sabarmathi and R. Chinnaiyan, "Investigations on big data features research challenges and applications," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, 2017, pp. 782-786.

[4] G. Sabarmathi and R. Chinnaiyan, "Reliable Machine Learning Approach to Predict Patient Satisfaction for Optimal Decision Making and Quality Health Care," 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2019, pp. 1489-1493

[5] Hari Pranav A;M. Senthilmurugan;Pradyumna Rahul K;R. Chinnaiyan , "iot and Machine Learning based Peer to Peer Platform for Crop Growth and Disease Monitoring System using Blockchain," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-5, doi:

[6] https://enterprisetalk.com/featured/top-five-edge-computing-trends-to-watch-out-for-in-2024/#:~:text=As%20per%20a%20recent%20report,pivotal%20in%20the%20coming%20years.

[7] https://kafka.apache.org/

[8] https://www.ameyo.com/blog/why-chatbots-can-be-used-as-internet-of-things-iot-interface/

[9] https://www.eccouncil.org/cybersecurity-exchange/network-security/guide-to-iot-security-protecting-critical-networks/

[10] https://www.elastic.co/

[11] https://www.sciencedirect.com/science/article/pii/S2352864817300214#s0310

[12] https://www.yugabyte.com/

[13] M Swarnamugi, R Chinnaiyan (2019), IoT Hybrid Computing Model for Intelligent Transportation System (ITS), Proceedings of the Second International Conference on Computing Methodologies and Communication (ICCMC 2018), 802-806.

[14] M. Caroline Viola Stella Mary, G. Prince Devaraj, et al., "Intelligent Energy Efficient Street Light Controlling System based on IoT for Smart City," IEEE Xplore

[15] M. Swarnamugi ; R. Chinnaiyan, "IoT Hybrid Computing Model for Intelligent Transportation System (ITS)", IEEE Second International Conference on Computing Methodologies and Communication (ICCMC), 15-16 Feb. 2018.

[16] M. Swarnamugi; R. Chinnaiyan, "Cloud and Fog Computing Models for Internet of Things", International Journal for Research in Applied Science & Engineering Technology, December 2017.

[17] Partha Pratim Ray, ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope".

[18] Prasannjeet Singh, Mehdi Saman Azari1, Francesco Vitale, Francesco Flammini1, Nicola Mazzocca, Mauro Caporuscio, Johan Thornadtsson, Using log analytics and process mining to enable self-healing in the Internet of Things, DOI:10.1007/s10669-022-09859-x

[19] S. Balachandar, R. Chinnaiyan (2019), Internet of Things Based Reliable Real-Time Disease Monitoring of Poultry Farming Imagery Analytics, Lecture Notes on Data Engineering and Communications Technologies book series (LNDECT, volume 31), 615- 620. Springer, Cham

[20] S.Balachandar , R.Chinnaiyan (2018), A Reliable Troubleshooting Model for IoT Devices with Sensors and Voice Based Chatbot Application, International Journal for Research in Applied Science & Engineering Technology,Vol.6,Iss.2, 1406-1409.

[21] S.Balachandar , R.Chinnaiyan (2018), Centralized Reliability and Security Management of Data in Internet of Things (IoT) with Rule Builder, Lecture Notes on Data Engineering and Communications Technologies 15, 193-201.

[22] S.Balachandar , R.Chinnaiyan (2018), Reliable Digital Twin for Connected Footballer, Lecture Notes on Data Engineering and Communications Technologies 15, 185-191.

[23] Chinnaiyan, R., Prasad, G., Sabarmathi, G., Swarnamugi, Balachandar, S., Divya, R. (2023). Deep Learning-Based Optimised CNN Model for Early Detection and Classification of Potato Leaf Disease. In: Bhateja, V., Yang, XS., Ferreira, M.C., Sengar, S.S., Travieso-Gonzalez, C.M. (eds) Evolution in Computational Intelligence. FICTA 2023. Smart Innovation, Systems and Technologies, vol 370. Springer, Singapore. https://doi.org/10.1007/978-981-99-6702-5_47

[24] Das, S. *et al.* (2023). Crowd Monitoring System Using Facial Recognition. In: Bhateja, V., Carroll, F., Tavares, J.M.R.S., Sengar, S.S., Peer, P. (eds) Intelligent Data Engineering and Analytics. FICTA 2023. Smart Innovation, Systems and Technologies, vol 371. Springer, Singapore. https://doi.org/10.1007/978-981-99-6706-3_50

[25] Chinnaiyan, R., Kondaveeti Sai, and P. Bharath. "Deep Learning based CNN Model for Classification and Detection of Individuals Wearing Face Mask." *arXiv preprint arXiv:2311.10408* (2023

[26] Sungheetha, Akey. "Optimized Deep Learning Models for AUV Seabed Image Analysis." *arXiv preprint arXiv:2311.10399* (2023).

[27] Sungheetha, Akey. "Emotion Based Prediction in the Context of Optimized Trajectory Planning for Immersive Learning." *arXiv preprint arXiv:2312.11576* (2023).

[28] Forbes - https://www.forbes.com/sites/bernardmarr/2023/10/19/2024-iot-and-smart-device-trends-what-you-need-to-know-for-the-future/?sh=2aab520f7f34