IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304287 |

A Cyber Security Knowledge Graph for Advanced Persistent Threat Organization Attribution

Dr. M. Anantha Guptha[1], M. Surya Prakash[2], Orsu Shashikala[3], K. Sai Sumanth[4], Chapagalla Sivakumar[5], Yalamasetty Vineeth[6]

UG Student, Dept. of E.C.E., GATES Institute of Technology, Gooty, Anantapur (Dist.), Andhra Pradesh[2,3,4,5,6]

Associate Professor, Dept. of E.C.E., GATES Institute of Technology, Gooty, Anantapur (Dist.), Andhra Pradesh[1]

ABSTRACT: Open-source cyber threat intelligence (OSCTI) is becoming more influential in obtaining current network security information. Most studies on cyber threat intelligence (CTI) focus on automating the extraction of threat entities from public sources that describe attack events. The cybersecurity knowledge graph aims to change the expression of threat knowledge so that security researchers can accurately and efficiently obtain various types of threat information for preliminary intelligent decisions. The attribution technology can not only assist security analysts in detecting advanced persistent threats, but can also identify the same threat from different attack events.

Therefore, it is important to trace the attack threat actor. In this study, we used the knowledge graph technology, considered the latest research on cyber threat attack attribution, and thoroughly examined key related technologies and theories in the process of constructing and applying the advanced persistent threat (APT) knowledge graph from OSCTI. We designed a cybersecurity platform named CSKG4APT based on a knowledge graph. Inspired by the theory of ontology, we constructed CSKG4APT as an APT knowledge graph model based on real APT attack scenarios. We then designed an APT threat knowledge extraction algorithm for completing and updating the knowledge graph using deep learning and expert knowledge. Finally, we proposed a practical APT attack attribution method with attribution and countermeasures.

KEYWORDS: Python, Deep learning, Cyber security

I. INTRODUCTION

CSKG4APT (Cyber Security Knowledge Graph for Advanced Persistent Threat Organization Attribution) serves to provide an overview of the research focus, the significance of the problem addressed, and the objectives pursued in developing a comprehensive cyber security knowledge graph tailored for attributing Advanced Persistent Threats (APTs) to specific threat actors or organizations.

Networks always faces security issues with different types of attack in which some are permanent and some are non-permanent. APT (advance Persistent Attack) remain in network permanently. Existing algorithms on cyber threat intelligence (CTI) focus on automating the extraction of threat entities from public sources that describe attack events but this technique is not feasible so in propose paper author employing Knowledge Graph on APT attack dataset to discover APT attacks.

Building ontology-based knowledge graph from APT dataset to extract network features and then employing deep learning BI-LSTM with GRU layers algorithm to train a model on APT graph features and this model can be applied on any network test data to identify whether test data is normal or contains any APT attacks.

To implement this project author has used APT Text base network dataset and then apply BERT (bidirectional encoder representations from transformers) algorithm on text data to convert into numeric vector and this vector contains average frequency of each words from the dataset. This BERT vector will be input to BI-LSTM with GRU algorithm to train a model and this model will be applied on test data to calculate prediction accuracy, precision, recall and FSCORE

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304287 |

II. LITERATURE SURVEY

[1]Advanced Persistent Threat, 2020. [Online].
Available: https://en.wikipedia.org/wiki/Advanced_persistent_threatInformation and communication technology.

A **computer network** is a set of computers sharing resources located on or provided by network nodes. Computers use common communication protocols over digital interconnections to communicate with each other. These interconnections are made up of telecommunication network technologies based on physically wired, optical, and wireless radio-frequency methods that may be arranged in a variety of network topologies.
[2]T. Zhihong, "Detection and traceability of high covert unknown threats in cyberspace," *Inf. Commun. Technol.*, vol. 14, no. 06, pp. 4–7, 2020**.**

Open-source cyber threat intelligence (OSCTI) is becoming more influential in obtainingcurrent network security information. Most studies on cyber threat intelligence (CTI) focus on automating the extraction of threat entities from public sources that describe attack events.

The cybersecurity knowledge graph aims to change the expression of threat knowledge so that security researchers can accurately and efficiently obtain various types of threat information for preliminary intelligent decisions.
[3]L. Yue , "Overview of network security threat intelligence sharing and exchange," *Comput. Res. Develop.*, vol. 57, no. 10, pp. 2052–2065, 2020.

The emerging threats in cyberspace are endangering the interests of individuals, organizations and governments with complex and changeable attack methods. When traditional network security defense methods are not strong enough, the threat intelligence sharing and exchange mechanism has brought hope to the protection of cyberspace security.

Cybersecurity threat intelligence is a collection of information that can cause potential harm and direct harm to organizations and institutions. This information can help organizations and institutions study and judge the cybersecurity threats they face, and make decisions and defenses accordingly.

[4]Z. Zhu and T. Dumitras, "ChainSmith: Automatically learning the semantics of malicious campaigns by mining threat intelligence reports," in *Proc. IEEE Eur. Symp. Secur. Privacy*, 2018, pp. 458–472. Modern cyber attacks consist of a series of steps and are generally part of larger campaigns. Large-scale field data provides a quantitative measurement of these campaigns. On the other hand, security practitioners extract and report qualitative campaign characteristics manually. Linking the two sources provides new insights about attacker strategies from measurements

III. EXISTING SYSTEM

In today's time the world has faced many health issues because of unavailability of natural food. Also, the artificial fertilizers and chemicals are used for the fast growth of plants due to which the soil loses its original nature and it takes years to regain that back. Farmers have no alternative than to grow plants artificially with the help of chemicals and fertilizers available.

IV. PROPOSED SYSTEM

Networks always face security issues with different types of attack in which some are permanent and some are non-permanent. APT (advance Persistent Attack) remain in network permanently. Existing algorithms on cyber threat intelligence (CTI) focus on automating the extraction of threat entities from public sources that describe attack events but this technique is not feasible so in propose paper author employing Knowledge Graph on APT attack dataset to discover APT attacks.Building ontology based knowledge graph from APT dataset to extract network features and then employing deep learning BI-LSTM with GRU layers algorithm to train a model on APT graph features and this model can be applied on any network test data to identify whether test data is normal or contains any APT attacks.To implement this project author has used APT Text base network dataset and then apply BERT (bidirectional encoder representations from transformers) algorithm on text data to convert into numeric vector and this vector contains average frequency of each words from the dataset.

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304287 |

This BERT vector will be input to BI-LSTM with GRU algorithm to train a model and this model will be applied on test data to calculate prediction accuracy, precision, recall and FSCORE.Open-source cyber threat intelligence (OSCTI) is becoming more influential in obtaining current network security information. Most studies on cyber threat intelligence (CTI) focus on automating the extraction of threat entities from public sources that describe attack events. The cyber security knowledge graph aims to change the expression of threat knowledge so that security researchers can accurately and efficiently obtain various types of threat information for preliminary intelligent decisions.

BLOCK DIAGRAM

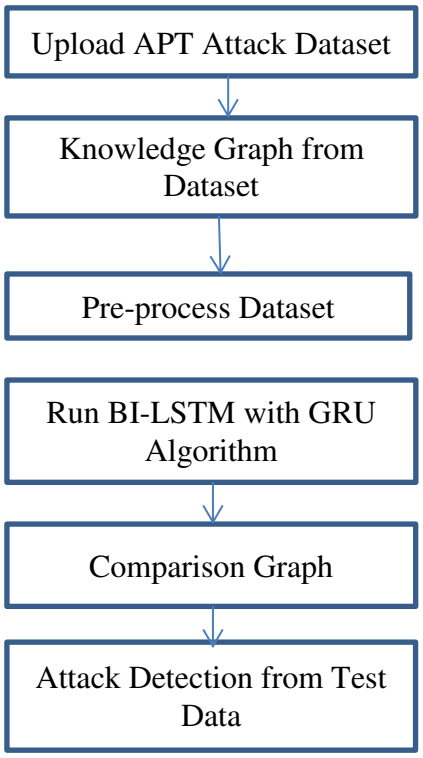

Figure 1: Flow chart of working

The attribution technology can not only assist security analysts in detecting advanced persistent threats, but can also identify the same threat from different attack events. Therefore, it is important to trace the attack threat actor.
Proposed paper apply knowledge graph technology, considered the latest research on cyber threat attack attribution, and thoroughly examined key related technologies and theories in the process of constructing and applying the advanced persistent threat (APT) knowledge graph from OSCTI.

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304287 |

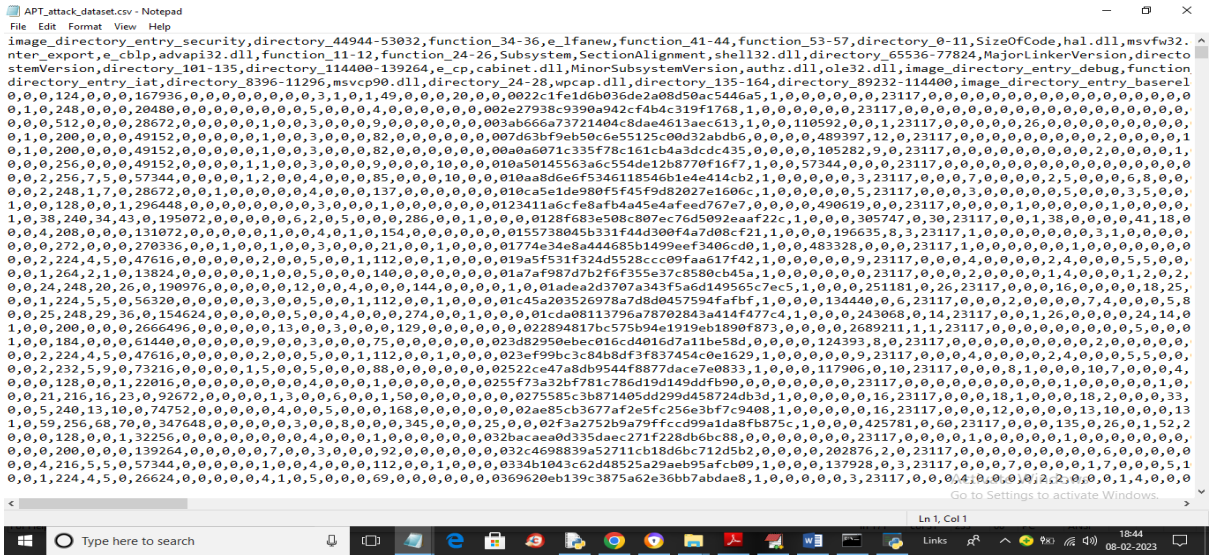

Figure 2: Dataset Frequency

Above dataset screen first 4 rows contains words from the dataset and remaining words contains average frequency of each words under that column word name. This dataset will be input to deep learning algorithm to train a APT attack detection model.

V. EXPERIMENTAL RESULT

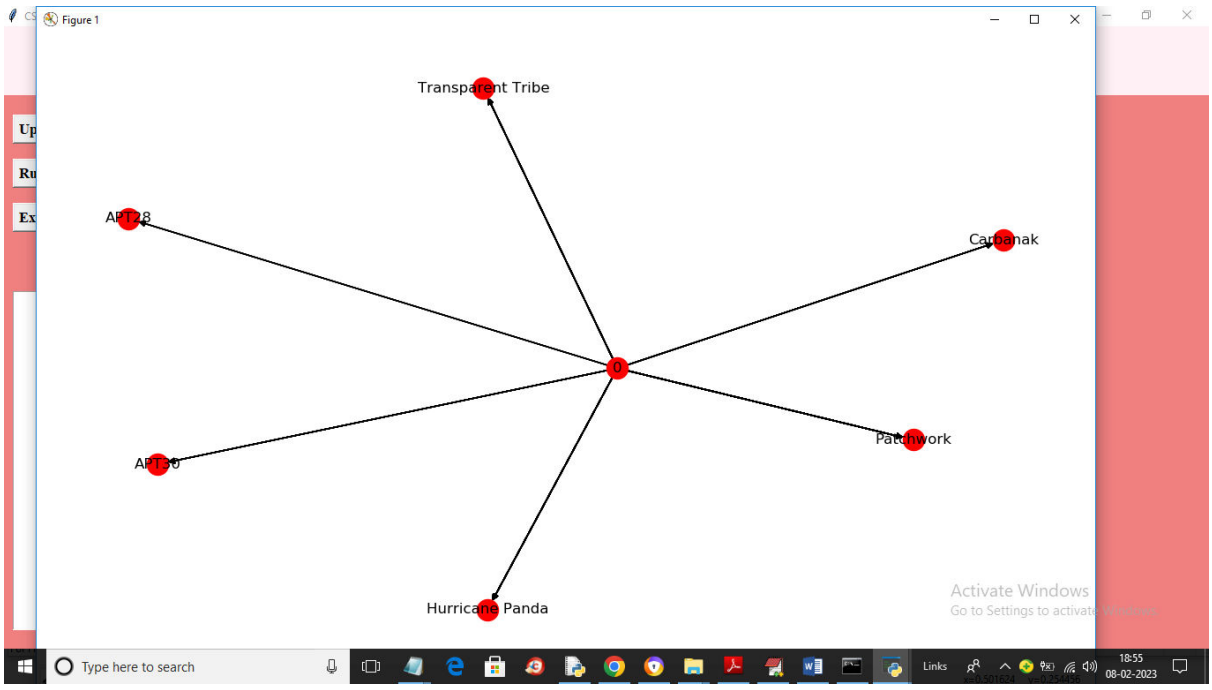

Figure 3: Dataset we got knowledge graph with various attacks

In above screen dataset processing completed and we can see dataset contains 1415 records and then application using 80% (1132 records) dataset for training and 283 (20% records) dataset values for testing.ADVANTAGES AND APPLICATIONS.

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304287 |

To implement this project author has used APT Text base network dataset and then apply BERT (bidirectional encoder representations from transformers) algorithm on text data to convert into numeric vector and this vector contains average frequency of each words from the dataset. This BERT vector will be input to BI-LSTM with GRU algorithm to train a model and this model will be applied on test data to calculate prediction accuracy, precision, recall and FSCORE

1. It can be implemented in gardens or nurseries even in balconies with minimum cost and resources. Also helps in proper utilization of the available resources and helps in avoiding wastage of electricity and water.

2. Can be easily configured and scaled up to work on larger fields.

3. Provides a user-friendly interface hence will have a greater acceptance by the technologically unskilled workers. The system is more compact compared to the existing ones, hence is easily portable and low cost. Advancements in sensor technology and control systems allow for optimal use of resources. Our aim is to design and develop newer techniques that will allow garden automation to flourish and deliver to its full potential. Thus it can be concluded that the proposed project will lessen labour, conserve water, increase crop yield, provide maximum automation and benefit the society by adopting the fast-growing IoT (Internet of Things) to implement newer and sustainable ways of farming.

VI. CONCLUSION AND FUTURE SCOPE

The development and implementation of CSKG4APT (Cyber Security Knowledge Graph for Advanced Persistent Threat Organization Attribution) represent a significant stride forward in the realm of cybersecurity attribution. This groundbreaking framework, designed to address the complexities of attributing Advanced Persistent Threats (APTs) to specific threat actors or organizations, holds substantial promise in fortifying cyber defense strategies and mitigating sophisticated cyber threats. In essence, CSKG4APT stands as a pioneering framework that significantly contributes to the advancement of cyber threat attribution capabilities. Its ability to connect disparate pieces of cyber threat intelligence, extract actionable insights, and attribute APT attacks to specific threat actors or organizations marks a crucial step forward in bolstering cybersecurity defenses against evolving and sophisticated cyber threats. The future scope of a study on CSKG4APT (Cyber Security Knowledge Graph for Advanced Persistent Threat Organization Attribution) encompasses various aspects related to building, implementing, and utilizing a knowledge graph for enhancing cyber threat attribution specifically concerning Advanced Persistent Threats (APTs). The scope of the study involves a comprehensive exploration of building a cyber security knowledge graph specifically tailored for attributing APTs to threat actors or organizations. It encompasses both theoretical research and practical implementation to advance the field of cyber threat attribution.

REFERENCES

1. Advanced Persistent Threat, 2020. [Online].
2. Available: https://en.wikipedia.org/wiki/Advanced_persistent_threatInformation and communication technology.
3. 2.APT Annual Review, 2021. [Online]. Available: https://securelist.com/apt-annual-review-2021/105127
4. 3.T. Zhihong, "Detection and traceability of high covert unknown threats in cyberspace," *Inf. Commun. Technol.*, vol. 14, no. 06, pp. 4–7, 2020.
5. 4.L. Yue , "Overview of network security threat intelligence sharing and exchange," *Comput. Res. Develop.*, vol. 57, no. 10, pp. 2052–2065, 2020.
6. 5.Z. Zhu and T. Dumitras, "ChainSmith: Automatically learning the semantics of malicious campaigns by mining threat intelligence reports," in *Proc. IEEE Eur. Symp. Secur. Privacy*, 2018, pp. 458–472.
7. Y. Ghazi, Z. Anwar, R. Mumtaz, S. Saleem, and A. Tahir, "A supervised machine learning based approach for automatically extracting high-level threat intelligence from unstructured sources," in *Proc. Int. Conf. Front. Inf. Technol.*, 2018, pp. 129–134.
8. Y. Zhao, B. Lang, and M. Liu, "'Ontology-based unified model for heterogeneous threat intelligence integration and sharing," in *Proc. 11th IEEE Int. Conf. Anti-Counterfeiting Secur. Identification*, 2017, pp. 11–15.
9. Y. Guo , "CyberRel: Joint entity and relation extraction for cybersecurity concepts," in *Proc. Int. Conf. Inf. Commun. Secur.*, 2021, pp. 447–463.
10. G. Husari, E. Al-Shaer, M. Ahmed, B. Chu, and X. Niu, "TTPDrill: Automatic and accurate extraction of threat actions from unstructured text ofCTI Sources," in *Proc. 33rd Annu. Comput. Secur. Appl. Conf.*, 2017, pp. 103–115.

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304287 |

11. Z. Li, J. Zeng, Y. Chen, and Z. Liang, "AttacKG: Constructing technique knowledge graph from cyber threat intelligence reports," 2021, arXiv: 2111.07093.
12. 11.A. Singhal, "Introducing the knowledge graph: Things, not strings," Official Blog (of Google), 2012. Online]. Available: http://googleblog.blogspot.co.uk/2012/05/introducing-knowledge-graph-things-not.html
13. 12.W. Haofen, Q. Guilin, and C. Huajun, *Knowledge Graph: Method, Practice and Application*, Beijing, China: Publishing House Electron. Ind., 2019.
14. 13.STIX, 2022. [Online]. Available: https://oasis-open.github.io/cti-documentation/stix/intro
15. 14.CAPEC, 2019. [Online]. Available: http://capec.mitre.org/about/index.html
16. 15.C. N. Li and S. A. Thompson, "Mandarin chinese: A functional reference grammar," *J. Asian Stud.*, vol. 42, no. 3, pp. 10–12,1989.
17. 16.A. Alsaheel , "ATLAS: A sequence-based learning approach for attack investigation," in *Proc. 30th USENIX Secur. Symp.*, 2021, pp. 3005–3022.
18. 17.CYBOX, 2020. [Online]. Available: http://cyboxproject.github.io/sample
19. 18.S. Caltagirone, A. Pendergast, and C. Betz, "The diamond model of intrusion analysis," Center for Cyber Intelligence Analysis and Threat Research Hanover Md: Ft. Meade, MD, USA, 2013.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details