Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 3, March : 2023

DATA MINING AND MACHINE LEARNING IN THE CONCEPT OF CYBER SECURITY

Abhimanyu Patra Ph.D Research Scholar, Department of Computer Science Engineering, Utkal University, Bhubaneswar- 751004, Odisha, India. Email: a_patra_2005@yahoo.com
Sarojananda Mishra Department of Computer Science Engineering & Application, Indira Gandhi Institute of Technology, Saranga-759146, Odisha, India. Email: sarose.mishra@gmail.com
Manas Ranjan Senapati Department of Information & Technology, Veer Surendra Sai University of Technology, Burla-768017, Sambalpur, Odisha, India. Email: manassena@gmail.com
Rajesh Kumar Behera Department of Mechanical Engineering, Krupajal Engineering College, Prasanti Vihar, Kausalya Ganga, Bhubaneswar-751002, Odisha, India.
Email: rajesh_k_behera@yahoo.co.in

Abstract
In this paper, a high potential routing protocol for cyber security that may be applied to Internet of Things applications based on WSNs that experience large traffic loads is examined. DM was the most efficient and advanced technology for identifying previously undiscovered patterns and trends that could be used to boost an organization's productivity. All businesses are growing more and more with the aid of data mining technologies. The finding of previously undiscovered and very profitable information in vast volumes of data is made possible through data mining. Finding new trends in massive sets of data was the primary objective of database knowledge exploration. It incorporates a number of fields, such as machine learning, artificial intelligence, and statics. Users can gain insight into the unprocessed data acquired from various Internet of things applications thanks to DM's ability to transform a big data collection into a logical framework and extract important information. The Internet of Things is a network of physical objects or items that have been fixed with network connections, electronics, and sensors to collect and distribute data. The agent at each CH then makes a request to its CMs to begin the local computation after the BS requests the Cluster heads to calculate tasks like help and trust. As a result, fewer and smaller messages are exchanged between Cluster heads and BS, from and to CMs, and between Cluster heads and cluster heads, which saves energy.

Keywords: Data Mining; WS; Cyber Security; Machine Learning; Massive Data; DM; CH; BS

1. Introduction

Data mining is the process of finding data "models". A derogatory word for trying to extract information from data that isn't sponsored by the data is "data dredging." [1] Nowadays, data mining is more akin to machine learning, and the majority of methods for uncovering odd events hidden within massive amounts of data rely on machine learning algorithms. As a result of recent advancements in communication technology, people and objects are becoming more interconnected. It is feasible to connect a number of devices that may communicate and exchange data due to the availability of the Internet. A contemporary concept called "data mining with the IoT" enables users to connect various sensors and smart devices to collect real-time data from the environment.

The Software Denied Networking, which was the target of a Distributed Denial of Service assault, has been the most severely affected by security risks. When the network was not properly safeguarded, distributed denial of service might overwhelm the overflow switch or administrator. Many documents exist that discuss how to defend Software Denied Networks from Distributed Denial of Service attacks. When a Distributed Denial of Service attack is discovered, IDSs are employed in the network to alert the controller and detect packets. Many academics were drawn to machine learning to identify distributed denial of service attacks. Hence, the research field was active in defending the software that prevented network issues. The goal of this study is to take into account the best machine learning technique to identify a distributed denial of service attack.

More common surveys on the topic can be found in [2], where specialists examined DM and machine learning approaches for analysing medical data. DM was a comprehensive process that could be used to all types of data. Although the ordering of DM approaches in this survey was dependent upon the categorization, mining pattern, and clustering, there are numerous surveys available on each of these techniques. For instance, a typical mining pattern over a stream of data was shown. [3] gives an overview of WSN clustering algorithms. These publications exclusively focus on network architecture and maintenance, not knowledge discovery, while examining clustering algorithms. In a review of classification techniques over data streams, the author looks at conventional classification methodologies.

DM has received a lot of attention in the data industry and society at large in recent decades due to the widespread accessibility of enormous amounts of information and the pressing need to convert the information into meaningful knowledge and information. The knowledge and experience collected can be used to improve customer retention, market research, fraud detection, scientific discovery, and production management [4]. In the era of the Internet of Things, when everything interacts and communicates with one another, a lot of data was produced. This data needs to be correctly mined and analyzed in order to enhance IoT functioning. The combination of data mining approaches with IoT will revolutionize the economies of all nations if we can successfully implement them. For organizational decision-making, data mining techniques should be combined with IoT. Asa result, the primary goal of this paper is to provide a comprehensive overview of a data mining system that has been examined for Internet of things applications. The structure outlined in this research can include a roadmap for investigators interested in using DM to solve IoT applications.

A world where physical things are seamlessly integrated into the information network and can actively participate in business processes, according to S. Haller et al. [5]. Services are available to query the state of these "smart objects" and any information related to them through the Internet while taking security and privacy concerns into consideration.

A data estimating method called CARM, which refers to the current rule of association between the radars in the most recent window sliding, was studied by Jiang and Gruenwald [6]. The technique relied on CFI-stream, a common CARM for the data stream. It makes use of a memory data structure called the DIU to hold closed item-sets. The CARM's algorithm tests and window sliding updates the help for the CIS if a new transmission is received. If CRAM finds any missing values in the sensor readings, it produces the method that is closely related to the current sensor data rather than generating all conceivable association rules. Based on these guidelines and chosen closed item-sets that comprise item values, CRAM generates approximations of the values. It shows that in the DIU tree, there are currently 4 closed item sets: CBA, BA, DC and C with corresponding supports of 3, 3, 1, and 2 in the right upper corner. These frequently occurring item-sets produce a simple set of rules. This simple rule set can be used to infer all other rules.

In order to create an association rule using data from wireless sensor networks and a single scan database, Tanbeer et al. team[7] analysed a tree-based data structure called SP tree. An important idea was to design a PT based on canonical methodology, take into account the bandwidth of all monitoring radar data, and then rearrange the tree in high bandwidth order. In order to maintain the radar detection nodes at the top of the tree, the sensor pattern tree will be reorganised, giving the tree's structure a high degree of solidity. The FP-growth mining technique is used to identify the sensor sets that detect frequent events once the sensor pattern has been generated. To determine whether SP-tree outperforms PLT in terms of memory usage, experiments are conducted.

In order to locate the readings from the missing sensor, Halatchev and Gruenwald [8-11] investigated and developed a centralised strategy known as DSARM. It uses the rule of association, or algorithm for mining, to identify radars as related radars when they repeatedly record similar information in a window sliding, and then it measures information from a radar by data related radars. It was unable to directly apply a mining method like Apriori to sensor data due to the nature of radar data [12-15]. The DSARM system, which adapts the Apriori algorithm to the data stream acquired from sensor nodes, was created by the specialists as a result. The Department of Transportation in Austin, Texas, USA, gathered the source data for these simulation experiments using simulation software.

Umadevi and others [16] The basis for behavioural analytics, which tries to prevent harm, is data mining. Long-term advantages of machine learning include a probabilistic and prognostic approach. The detection of patterns, regularities, and abnormalities by machine learning and data mining techniques enables the avoidance of cybersecurity breaches. Mercy Beulah and others [17-21]. A variety of reasons, like the fact that it takes time to recognise compromising behaviour and that many users learn about hacking from a third party, define the significance of adopting machine learning. Automating the examination of security processes and real-time threat detection are crucial. These details increase the research's importance and main objective.

2. Methodology

With protocols that rely on cooperation, implicit trust has always been present. It involves DM networks and IoT routing procedures. IoT networks need to employ strong protection measures as they grow since they are more vulnerable to threats. Finding the right cryptography for wireless sensor networks is a big challenge because of the sensor nodes' constrained energy, processing, and storage resources. trustworthy minimum Trust in Deep Learning A brand-new energy-conscious routing method will be suggested for ad hoc networks called Secure Attacker Detection. The four key IoT requirements that DLTSAD targets are energy consumption, dependability, data aggregation, and attacker detection. DLTSAD is a potent routing method that improves hostile node detection and identifies paths that consume the least amount of total energy for E-E transversal packet.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 3, March : 2023

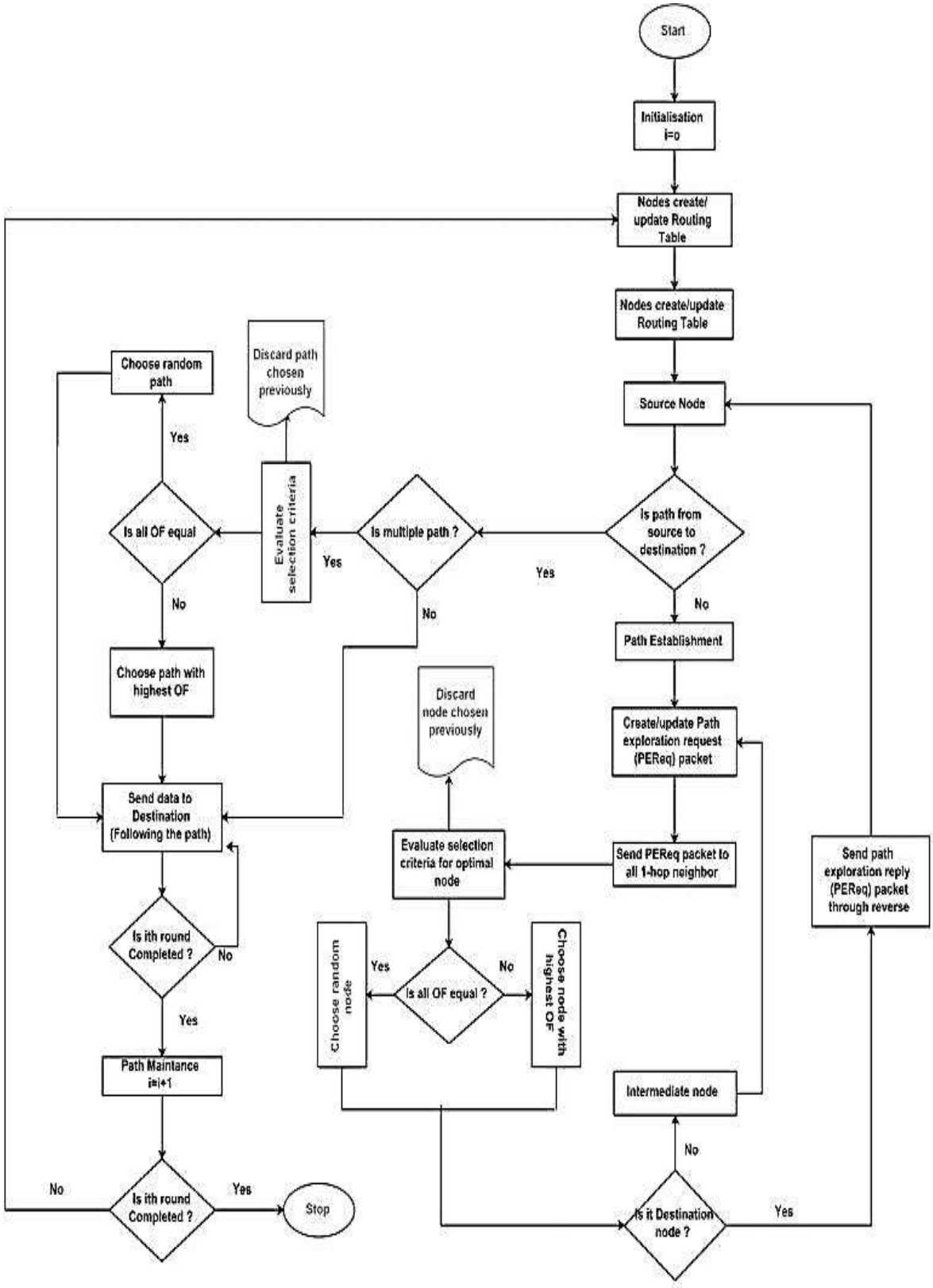

Figure 1: Classification of DM methods for sensor connection.

2.1 Proposed DLTSAD

To find more secure routes, algorithms consider the reliability of connections. Algorithms determine the mostpowerful routes. An Algorithm attempts to extend the lifetime of network by identifying paths that include nodes with high potential battery energy. Algorithm also determines with a greater

number of protection options.

2.1.1 Algorithm of DLTSAD

This module was created to propose DLTSAD [DATAGRAM LAYER TRANSPORT SECURITY ASSESSING DOG] focused on collaboration and routing, as well as attacker detection and prevention. In DLTSAD, we go over the strategies and contributions to trust-based protection. This shows how trust depend reasoning would permit each joint to measure the nodes actions and present a trust depend investigation of the DLTSAD protocol through particular trust language.

Algorithm 1: Proposed algorithm

Input : A network with N nodes, E links, Source node (N_o), Destination node (N_d)
Output : Multiple optimal paths from source to destination
Parameters:
OF: Optimality factor
L_E: Estimated lifetime of node
R_c: Reliability of communication
T_I: Traffic intensity of node
i: Round of algorithm
Initialize :
$i \leftarrow 0$
$R_c \leftarrow 0$
$T_I \leftarrow 0$
$L_E \leftarrow$ Estimated lifetime of node

```
1  begin
2      while i≤100 && stopping criteria do
3          Calculate OF of path;
4          Create routing table of individual node;
5          Source node checks path in its routing table;
6          if path exists then
7              Call algorithm 2 for sending data;
8          else
9              Call algorithm 3 for path discovery and establishment;
10             Call algorithm 2 for sending data;
11             Call algorithm 4 for path maintenance;
```

Figure 2: Algorithms of DLSAD.

2.1.2 Framework of data mining for cyber security

DM was the most effective and evolving technologies for extracting previously unknown useful patterns andtrends in order to improve an organization's efficiency. All companies are increasingly expanding with the help of data mining capabilities. Data mining aids in the discovery of previously discovered and highly profitable information in large amounts of data. For eg., by identifying consumers' regular purchasing habits, a company canboost revenue by grouping products that are often purchased together, offering discounts on certain items, or eliminating duplicate items.

The main goal of database knowledge exploration was to find novel trends in large set of data. It combines a variety of domains, including statics, artificial intelligence and machine learning. DM converts a large data collection into a logical structure and extracts significant information, allowing users to gain insight into the raw data obtained from different Internet of things applications. As a result, the Internet of Things was a network of physical things or object that is fixed with network communication, electronics and sensors to capture and sharedata. The ideal link of DM and Internet of things yields a new cutting-edge technology that will help people from all walks of life. These

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 3, March : 2023

applications produce a massive amount of disparate data. Since data in Internet of things applications is continuously generated from various sources such as WSN, RFID, and so on.

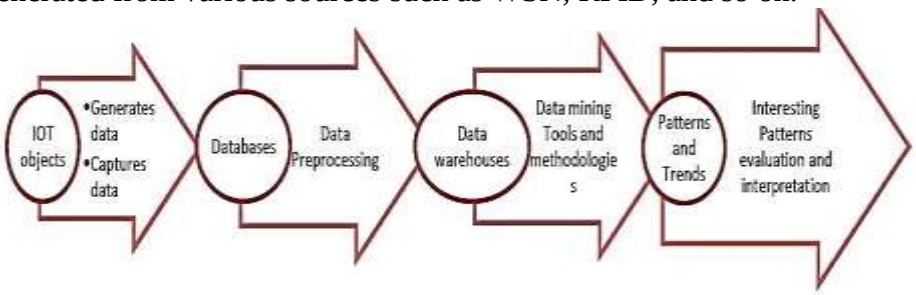

Figure 3: Data mining framework for IoT.

2.1.3 DMA Rules from Wireless Sensor Networks

A rule of associated could be referred as an allegation, such as A indicates B (A) B, where A and B denote theconsequent and antecedent item sets, respectively. That is, database record of transmission containing items fromitem set A should also contain things from item set B. An appropriate global computation in the implemented method was defining the rule of association in D. It was decomposed and distributed throughout the network, allowing calculation to be done locally and statistical summaries to be obtained and shared. To start the global computation, the BS asks the Cluster heads to compute tasks like help and trust, and then the agent at every CH gives a request to its CMs to start the local computation. This reduces the size and no. of message sent from and to CMs and Cluster heads, as well as between Cluster heads and BS, reducing the amount of energy consumed and extending the network's lifetime.

2.1.4 Advantages

1. Increasing the network's lifespan and achieving a substantial level of security.
2. It increases the performance of network and reduces the entire energy consumption. It also enhances thelifetime of network.
3. PDR and throughput ratio may be increased.
4. Decreased average E-E delay and overhead the routing messages.

3 Architecture / Input parameters

These procedures have 3 parameters to model the optimal factors for selection. They are reliability, lifetimenode and the probable traffic intensity.

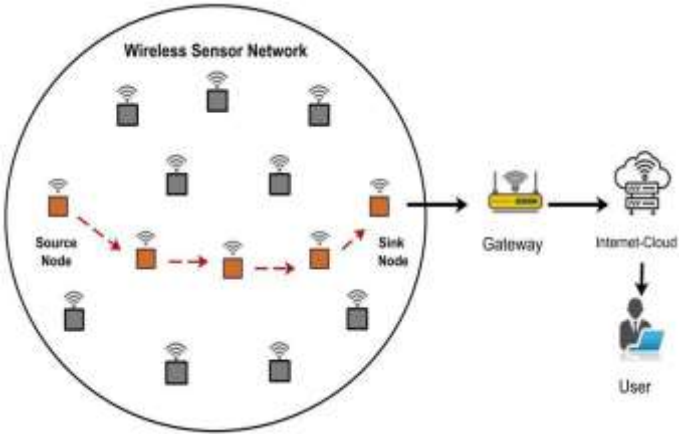

Figure 4: Architect of WSN.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 3, March : 2023

3.1 Procedure Cyber Attack Model

1. Key generation is a crucial step in which we must produce both a public and a private key.[22] Thecommunication should be decoded by the private key of receiver and encoded by the PK of sender.
2. The PK was produced by the following formula.Q = d * P
3. ‘d ’ was the arbitrary no. which was chosen inside the range of 1 to n-1. P was the curve point.
4. d was the private key and Q was the PK (public key).

4 Results and Discussion

4.1 Encryption / Decryption:

In encryption, the message must be represented on the curve. Encrypted document contains deep dataimplementation. Consider the point ‘M' on the curve ‘E' for ‘m.' Choose ‘k' at random from the list;
[1 – (n-1)]
Two cipher texts will be produced and it refers as C1 and C2.
C1 = k*P
C2 = M + k*Q
C1 and C2 will be sending.
Decryption means get back the message m which was given to customer.M = C2 – d * C1
M was the original message which was send to all.

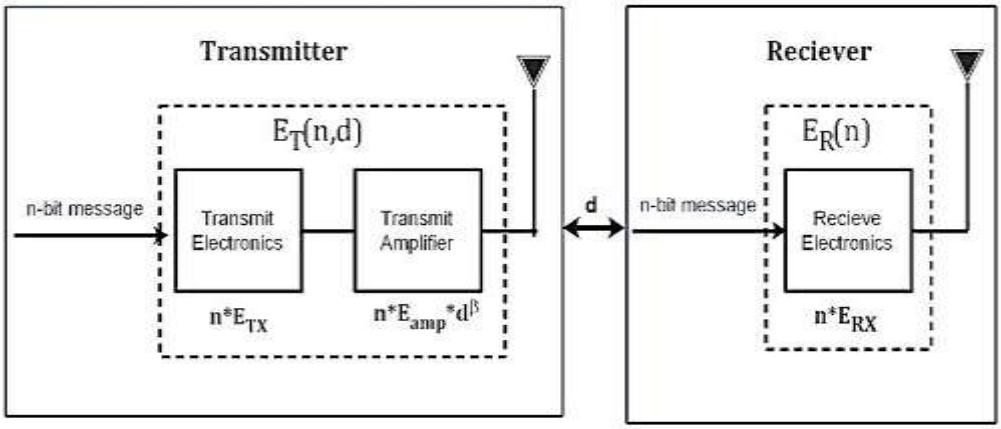

Figure 6: Proposed protocols.

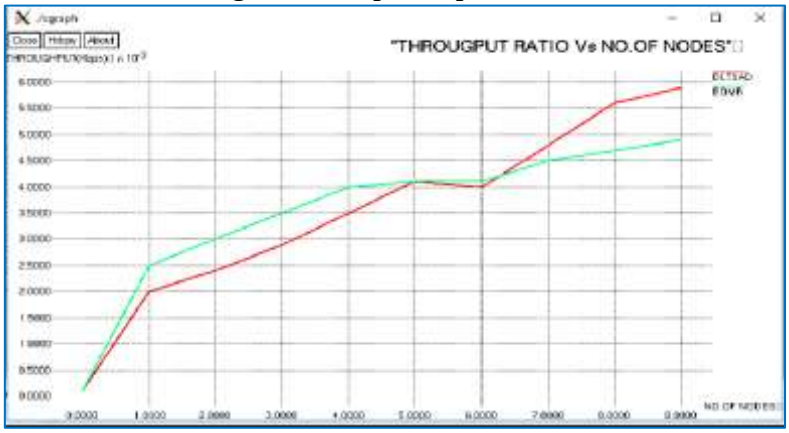

Figure 7: Throughput ratio.

4.2 PDR

PDR refers to the proportion of total packets given to total packets sent from a source node to a destination node in a network. The maximum number of data packets should be sent to the target. As the PDR value rises, thenetwork output rises with it. PDR is determined after a comparison without and with a black hole threat in the network. Packet delivery ratio was identified to be very low during the attack relative to the ratio before the attack,implying that fewer packets enter the sink node.

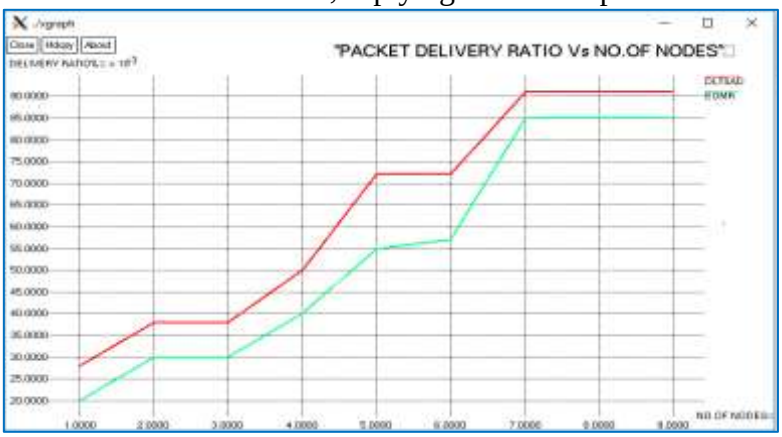

Figure 8: PDR.

4.3 Energy Consumption

Energy characterization is important for assessing the needs of an intensive data process which work effectivelyon mobile gadgets. An experimental analysis of the energy consumption of DM algorithms operating on mobile gadgets is presented in this paper.

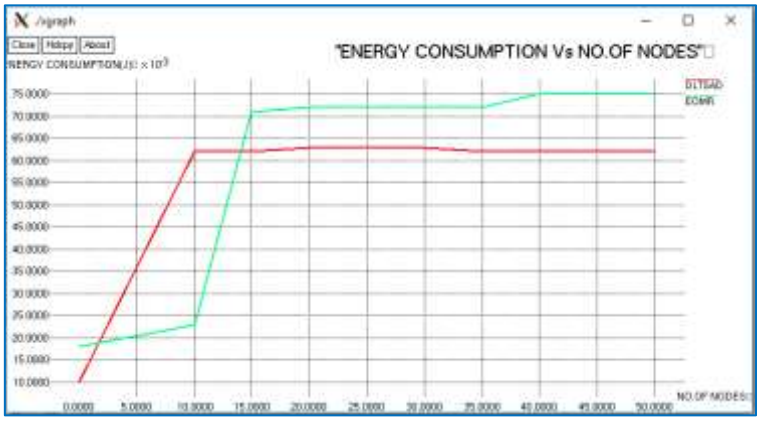

Figure 9: Energy Consumption.

5 Conclusion

The growing demand for DM techniques in the field of wireless sensor networks prompted the growth of aslew of algorithm. These algorithms addresses specific problems associated to the form and implementation of WSNs. This paper examines a high potential routing protocol for cyber security that can be used in WSN based Internet of things applications with heavy traffic loads. DM was the most effective and evolving technologies for extracting previously unknown useful patterns and trends in order to improve an organization's efficiency. All companies are increasingly expanding with the help of data mining capabilities. Data mining aids in the discoveryof previously discovered and highly profitable information in large amounts of data. The main goal of database knowledge exploration was to find novel trends in large set of data. It combines a variety of domains, including statics, artificial intelligence and machine learning.

Reference

[1] A. Boukerche and S. Samarah, “An efficient data extraction mechanism for mining association rules from wireless sensor networks,” in Proceedings of the IEEE International Conference on Communications (ICC ’07), pp. 3936–3941, June 2007.

[2] A. Rozyyev, H. Hasbullah, and F. Subhan, "Indoor child tracking in wireless sensor network using fuzzy logic technique," Research Journal of Information Technology, vol. 3, no. 2, pp. 81– 92, 2011.

[3] Halatchev, B.Mukherjee, and D. Ghosal, "Wireless sensor network survey," ComputerNetworks, vol. 52,no. 12,pp. 2292–2330, 2008.

[4] J. Gama, P. P. Rodrigues, and L. Lopes, "Clustering distributed sensor data streams using local processing and reduced communication," Intelligent Data Analysis, vol. 15, no. 1, pp. 3–28, 2011.

[5] Jiang and Gruenwald, S. A. Aljunid, B. Ahmad, A. Yahya, R. Kamaruddin, andM. S. Salim, "Wireless sensor actor network based on fuzzy inference system for greenhouse climate control," Journal of Applied Sciences, vol. 11, no. 17, pp. 3104–3116, 2011.

[6] Kiruthika , C & Nirmala Sugirtha Rajini,S(2014), "An Ill-identified Classification to Predict Cardiac Disease Using Data Clustering", International Journal of Data Mining Techniques and Applications, vol. 03,pp. 321-324, ISSN:2278-2419.

[7] L. T. Lee and C. W. Chen, "Synchronizing sensor networks with pulse coupled and cluster based approaches," Information Technology Journal, vol. 7, no. 5, pp. 737–745, 2008.

[8] M.Anita Priscilla Mary , M.S.Josephine , V.Jeyabalaraja & S.Nirmala Sugirtha Rajini(2020), "Identification and Performance valuation for Effective Utilization of Electrical Energy Resource using K Means Clustering Algorithm", International Journal of Advanced Science and Technology, Vol. 29, No. 9s, (2020), pp.55-62.

[9] S. Uma Devi & S. Nirmala Sugirtha Rajini (2019), " Detection of Traffic Violation Crime Using Data Mining Algorithms", Jour of Adv Research in Dynamical & Control Systems, Vol. 11, No.9, pp. 982-987.

[10] Mercy Beulah, E,Nirmala Sugirtha Rajini,S & Rajkumar, N (2016), "Application Of Data Mining In Healthcare: A Survey", Asian Journal of Microbiology, Biotechnology & Environmental Sciences, vol.18, no. 4, pp. 999-1001, ISSN-0972-3005.

[11] O. Horovitz, S. Krishnaswamy, and M. M. Gaber, "A fuzzy approach for interpretation of ubiquitous data streamclustering and its application in road safety," Intelligent Data Analysis, vol. 11, no. 1, pp. 89–108, 2007.

[12] P. K. Biswas and S. Phoha, "Self-organizing sensor networks forintegrated target surveillance," IEEE Transactions on Computers, vol. 55, no. 8, pp. 1033–1047, 2006.

[13] R. Szewczyk, E. Osterweil, J. Polastre, M. Hamilton, A. Mainwaring, and D. Estrin, "Habitat monitoring with sensor networks," Communications of the ACM, vol. 47, no. 6, pp. 34–40, 2004.

[14] S. H. Chauhdary, A. K. Bashir, S. C. Shah, and M. S. Park, "EOATR: energy efficient object tracking by auto adjusting transmission range in wireless sensor network," Journal of Applied Sciences, vol. 9, no. 24, pp. 4247–4252, 2009.

[15] S.Umadevi, S. Nirmala Sugirtha Rajini, A. Punitha & Viji Vinod (2020), "Performance Evaluation Of Machine Learning Algorithms In Dimensionality Reduction", International Journal of Advanced Science and Technology, Vol. 29, No. 9s, pp. 3845-3853

[16] S.Umadevi, S. NirmalaSugirtha Rajini,A. Punitha & Viji Vinod, (2020), " Dimensionality Reduction in Machine Learning Technique using Principal Component Analysis", Test Engineering and Management, Jauary - February 2020 ISSN: 0193 - 4120 Page No. 14546 - 14552 .

[17] T. Arampatzis, J. Lygeros, and S. Manesis, "A survey of applications of wireless sensors and wireless sensor networks," in Proceedings of the 20th IEEE International Symposium on Intelligent Control (ISIC '05), pp. 719–724, June 2005.

[18] T. Yairi, Y. Kato, and K. Hori, "Fault detection by mining association rules fromhouse-keeping data," in Proceedings of the 6th International Symposium on Artificial Intelligence, Robotics and Automation in Space, pp. 18–21, 2001.

[19] Tanbeer, Gruenwald "Monitoring forest cover changes using remote sensing and GIS: a global prospective," Research Journal of Environmental Sciences, vol. 5, pp. 105–123, 2011.

[20] Y.-C. Tseng, M.-S. Pan, and Y.-Y. Tsai, "Wireless sensor networks for emergency navigation," Computer, vol. 39, no. 7, pp. 55–62, 2006.

[21] Z. A. Aghbari, I. Kamel, and T. Awad, "On clustering large number of data streams,"

[22] Intelligent Data Analysis, vol. 16, no. 1, pp. 69–91, 2012.