International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Task Scheduling with Dynamic Load Balancing for Cloud Computing

Abhishek Mandalik[1], Aditya Mane[2], Harshal Patil[3], Kaustubh Dalal[4], Prof. B.K. Bodkhe

[1] *Department of Computer Engineering, MES's Wadia College of Engineering, Pune*
[2] *Department of Computer Engineering, MES's Wadia College of Engineering, Pune*
[3] *Department of Computer Engineering, MES's Wadia College of Engineering, Pune*
[4] *Department of Computer Engineering, MES's Wadia College of Engineering, Pune*
[5] *Department of Computer Engineering, MES's Wadia College of Engineering, Pune*

---***---

Abstract - Using powerful cryptographic techniques and complex machine learning algorithms, this study explores an enhanced way for dynamic job scheduling, data recovery, and workflow management in cloud computing. To improve data recovery procedures and optimize job scheduling, we investigate the combination of the Random Forest, Q-learning from artificial neural networks (ANNs), and the J48 decision tree algorithms. To ensure data redundancy and integrity, we also use the Kochi matrix for effective data chunk generation and backup. The Advanced Encryption Standard (AES) is used to protect data. Our all-encompassing approach is to increase the effectiveness of resource allocation, hasten data recovery, and offer strong security in cloud environments. The efficiency of the suggested methodologies in improving overall system performance, security, and reliability is highlighted by the experimental findings, which show notable advancements over conventional approaches.

Key Words: Load Balancing, Machine Learning, ANN, Random Forest, AES, Q-learning

1.INTRODUCTION

The data management and processing landscape has been completely changed by cloud computing, which provides unmatched scalability, flexibility, and cost-effectiveness. This paradigm change lessens the need for large upfront expenditures in hardware infrastructure by enabling organizations to harness huge computational capabilities on-demand. Notwithstanding these benefits, cloud computing environments still have significant issues with data recovery, work scheduling, and security. To guarantee data integrity, optimize resource usage, and shield confidential information from unwanted access, these factors must be managed effectively.

A key function of cloud computing is job scheduling, which involves allocating different computational jobs to available resources in order to maximize performance metrics like throughput, execution time, and resource utilization. Because traditional scheduling algorithms cannot adjust to changing workloads and resource availability, they frequently perform poorly in dynamic and diverse cloud systems. To overcome these constraints, we apply the Random Forest algorithm, ANN Q-learning, and J48 decision tree algorithm in this study.

• J48 Decision Tree: This model for job scheduling is simple to understand and easy to use, which makes it useful for making decisions quickly.

• ANN Q-learning: Adaptively improves scheduling decisions based on input from the environment by incorporating the principles of reinforcement learning.

• Random Forest: Increases prediction accuracy and robustness against overfitting by combining several decision trees.

In Cloud Computing, Data recovery overcomes the drawbacks of conventional approaches, which can require high latency and resource consumption, by ensuring the timely and correct restoration of data following failures. We dynamically adapt data recovery algorithms based on past attempts to reduce downtime and improve reliability using ANN Q-learning. Machine learning approaches enhance process efficiency and are beneficial for workflow scheduling, which is crucial for coordinating interdependent operations in complicated systems. By creating more manageable, smaller data chunks for backup, enhancing redundancy, and facilitating effective recovery even in the event that some chunks are lost, the Kochi matrix improves data reliability. We incorporate the Advanced Encryption Standard (AES) to protect data from cyber threats, guaranteeing strong encryption both during storage and transmission.

2. LITERATURE SURVEY

1. This study examines a number of strategies and methods related to cloud computing workflow scheduling, data recovery, and dynamic job ordering. The most important contributions are emphasized, starting with load balancing techniques, which are essential for maximizing the use of cloud resources. Dynamic load balancing techniques for virtual machine instances were proposed by Ren et al. (2012) and Bhatia et al. (2012) with the goal of reducing makespan time and improving resource utilization. Heuristic and metaheuristic algorithms that are optimized for cloud characteristics like execution and response times are other noteworthy contributions. Scholars like Chakraborty et al. (2010, 2013) and Laber et al. (2014) have expanded on Murphy's seminal work on Q-learning (2003, 2005), tackling difficulties with statistical inference and non-linear models. ANN Q-learning has the potential to improve dependability and decrease downtime in data recovery. Workflow scheduling has been improved through machine learning integration, increasing task coordination efficiency. The Kochi matrix method significantly enhances data redundancy and recovery efficiency for data chunk formation and backup, while AES encryption remains critical for data security in cloud environments. Collectively, these contributions underscore advancements in cloud computing, focusing on performance, reliability, and security through innovative algorithms and methodologies.

2. The Advanced Encryption Standard (AES), which took the place of Triple DES in 2002, is described in the document. With a constant block size of 128 bits and key sizes of 128, 192, or 256 bits, AES encrypts data. There is a first round, multiple rounds of intermediate encryption, and a final round of encryption. The encryption key and the plaintext are XORed in the first round. Repeated based on key size, intermediate rounds consist of the following four steps: AddRoundKey (XORing with a round key), MixColumns (multiplying columns by a fixed polynomial in Rijndael's Galois field), SubBytes (using Rijndael's S-Box for confusion), and ShiftRows (shifting rows to the left). SubBytes, ShiftRows, and AddRoundKey are included in the final round; MixColumns are not included. The paper highlights that in order to protect against attacks particular to AES implementations, code must be thoroughly tested and supported.

3. The paper gives a thorough description of the WEKA data mining software, including information on its development, history, and main characteristics. In 1993, WEKA received funding from the New Zealand government with the goal of establishing a state-of-the-art machine learning center. From early C versions, it developed into a Java-based system with different user interfaces, preprocessing filters, and learning algorithms. Since WEKA 3.4, there have been significant changes that include support for PMML standards, preprocessing tools, and new algorithms. WEKA's capabilities have been extended through numerous initiatives, demonstrating its adaptability in domains such as distributed data mining, biology, and natural language processing. The paper outlines the contributions made by the WEKA community and its continued success as an adaptable, sustainable, open-source tool.

4. This paper addresses the challenging problem of assigning jobs to virtual machines (VMs) in order to decrease makespan. It does this by examining the use of Ant Colony Optimization (ACO) for cloud task scheduling. ACO is proposed as a scheduling strategy and contrasted with First-Come-First-Served (FCFS) and Round Robin (RR) algorithms. It is inspired by the pheromone communication and foraging behavior of ants. ACO performs better in simulations with CloudSim than in FCFS and RR. In the paper, cloud computing is defined as a distributed, parallel system in which businesses rent virtual machines (VMs). It also highlights the significance and difficulties associated with effective job scheduling. Details on ACO's components and operations are provided, along with a focus on the system's strengths in searching huge solution areas and adapting to changing situations.

5. The significance of energy prediction models in buildings for improving energy efficiency is covered in the document. The study discovered that although ANN performed marginally better than RF, RF has advantages when handling complicated data since it is simpler to tune and model using categorical variables. In addition to highlighting the importance of ensemble-based techniques like RF, the paper emphasizes the necessity of data-driven models in building energy applications. The study comes to the conclusion that ANN performed marginally better in this particular case study. Future studies should look into how temporal and spatial granularity affect prediction accuracy as well as how RF models could be used for near real-time HVAC management and optimization.

6. The development of Q-learning for finite-horizon sequential choice problems started with important research by Robins (1986-1993) and progressed with Murphy's work (2001, 2003), which combined the prospective outcomes framework with dynamic programming to enable recursive estimate of regret functions. In 2005, Murphy went on to expand this to batch Q-learning, preferring robustness and simplicity over Robins' intricate semiparametric estimators. Q-learning's usefulness has been increased by research into its statistical problems, especially with regard to the max operator in regression functions. This has led to improved techniques and nonlinear model extensions. The method's adaption for infinite-horizon problems, particularly in spatial-temporal decision-making and mHealth, emphasizes its robustness and expanding practical applications.

7. The article looks into several approaches and algorithms that are meant to improve efficiency and performance. Different types of load balancing algorithms—dynamic, static, or hybrid—are used for different tasks depending on the condition of the machine. Examples of these tasks include VM load balancing and Cloud Data Mining load balancing. Research focuses on enhancing fault sensitivity, throughput, implementation cost, and responsiveness, even in the face of enduring obstacles such scaling problems and expensive processing. The usefulness of metaheuristic techniques—like hybrid algorithms that use support vector machines—is emphasized. For reliable resource allocation, load balancing strategies should incorporate several Quality-of-Service measures.

8. The paper explores the J48 method, a C4.5 variation that is well-known for creating decision trees. They are highly regarded for their capacity to adapt to a wide range of input data and to withstand errors and missing values. Using J48 as an example, the univariate method divides data according to single attributes and gives priority to information acquisition, highlighting two important construction approaches. On the other hand, multivariate trees take into account several

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

characteristics at each split, effectively managing correlations, and being especially useful for large datasets. The paper goes into detail about how J48 is implemented using Weka software, including how to prepare datasets, set up algorithms, and analyse models. In conclusion, multivariate trees outperform univariate trees in terms of performance and interpretability by utilising attribute correlations and providing a thorough examination of the applications and relative benefits of decision tree approaches in data mining situations.

9. In this study, the Random Forest (RF) algorithm is applied to survival analysis, and its performance is compared with that of the traditional Cox model. The author highlights the limitations of the Cox model, including the assumption of proportionate risks and the need for explicit inclusion of non-linear components and interactions. On the other hand, RF is a non-parametric technique that uses predictor randomization and bootstrap aggregation to achieve higher predictive accuracy. The mechanics of RF are explained, along with how well it grows several decision trees with different predictor sets and aggregates the results. The study presents the application of RF to colon cancer data from the SEER database and shows that it performs similarly to the Cox model, albeit it is not as interpretable as linear models. The study concludes by highlighting the potential of RF in survival analysis and emphasizing its superiority over the Cox model in terms of flexibility and predictive power, while also recognizing the inherent trade-off between interpretability and complexity.

10. The paper examines a novel method of load balancing in communication networks with an emphasis on energy consumption issues. The research addresses issues created by the unpredictable nature of renewable energy sources by introducing the Energy Packet Network (EPN) paradigm. The study highlights the superiority of dynamic load balancing techniques, including receiver-initiated load balancing, over sender-initiated methods in circumstances with high traffic. A major contribution is the incorporation of energy limitations into load balancing. It studies layered network structures to get equitable load distribution. The efficacy of the suggested methodology is demonstrated by numerical studies, which also show performance gains and fair load balancing under different network scenarios.

3. PROPOSED SYSTEM AND ALGORITHM

Proposed System: The system aims to enhance classification process of Straggler nodes through different approaches like Random Forest, ANN, and J48. The detected nodes are then utilized as parameters for the different servers. On the basis of these parameters load is dynamically assigned to the nodes and time taken by each server is calculated.

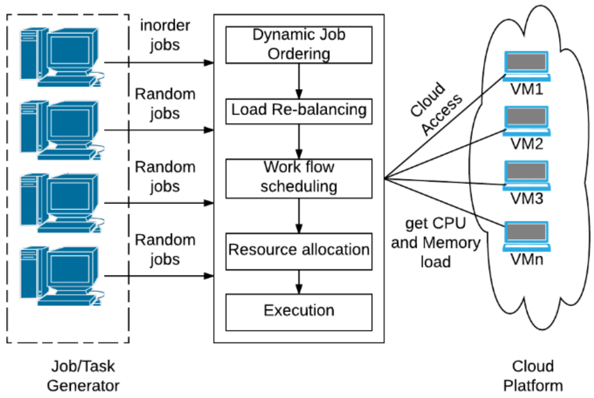

Fig -3.1: System Architecture

Algorithm Used: The system utilizes the RF (Random Forest), ANN (Artificial Neural Network), J48, Kochi Matrix, and AES (Advanced Encryption Standard) algorithms, these algorithms play a crucial part in the processes of dynamic job ordering and workflow scheduling, Data recovery, Chunk formation, and Data security.

A. Dynamic Job Ordering and Workflow Scheduling

J48 Algorithm: A decision tree-based classifier that is effective for job scheduling due to its simplicity and accuracy.

ANN Q-Learning: An adaptive learning algorithm that dynamically optimizes job scheduling based on past experiences.

Random Forest: An ensemble learning method that enhances prediction accuracy by aggregating multiple decision trees.

B. Data Recovery

ANN Q-Learning for Data Recovery: Utilizes the Q-learning reinforcement learning approach to optimize data recovery processes by learning from the cloud environment's dynamic nature.

C. Chunk Formation and Backup

Kochi Matrix: A mathematical structure used for efficient chunk formation and backup, ensuring data redundancy and reliability. Backup Strategy: Combining Kochi matrix with cloud storage strategies to create robust data backups.

The process is divided into two parts- Admin side and User side

Admin Side-

1. Input the VM (Virtual Machine) logs into the Q-learning algorithm to label the nodes.
2. Convert the input data based on parameters such as time taken and CPU load as labeled data containing the labels- Straggler and Non-Straggler
3. Classifier is started using different algorithms of J48, ANN and Random Forest and segregates the nodes as Straggler and Non-Straggler.
4. The output is stored into a text file that is stored in the required location.
5. Weka software is used to display the performance of different classifiers in bar graph form.

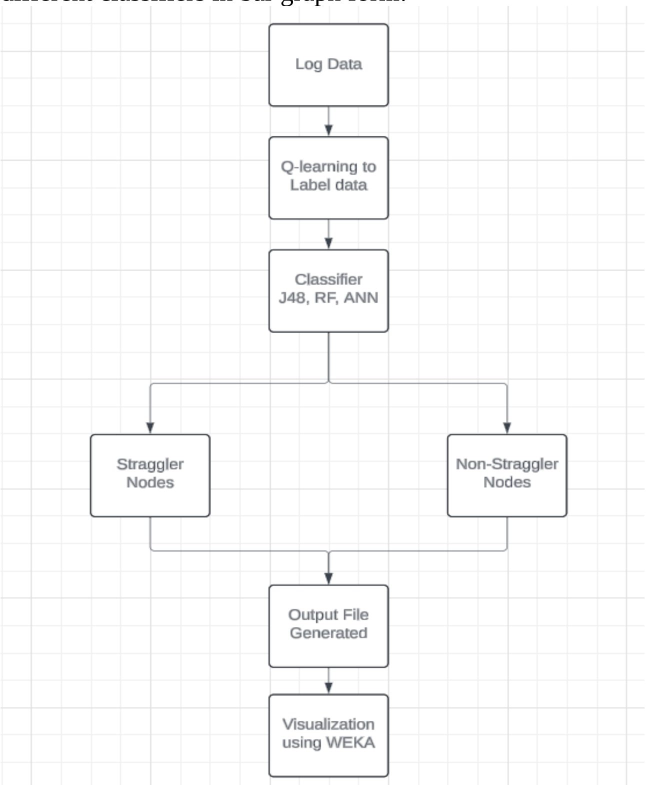

Fig -3.2: Admin Side

1. Upload the input files.

2. Assigning the files to servers based on the rewards assigned to them.

3. Files are divided into chunks with the help of Kochi Matrix Algorithm.

4. Display the time required for each Job.

5. The chunks are recombined to form the original files available for download.

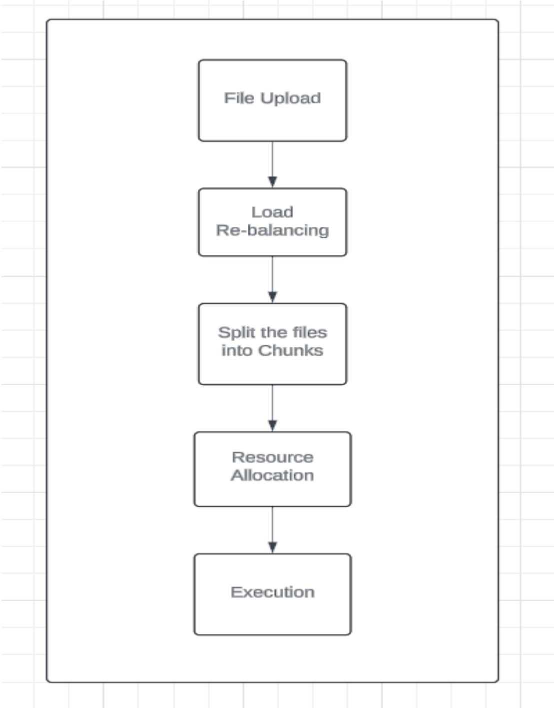

Fig -3.3: User Side

D. AES (Advanced Encryption Standard)

The AES algorithm is the symmetric algorithm that is secure enough to provide security for confidential data operating on plaintext of 128-bit with variable key length of 128, 192 and 256-bit [3]. The number of rounds performed is 10, 12 and 14 respectively. AES is one of the strongest algorithms until now and we can use only one key at the sender and receiver side, hence the privacy made by the key is secured.

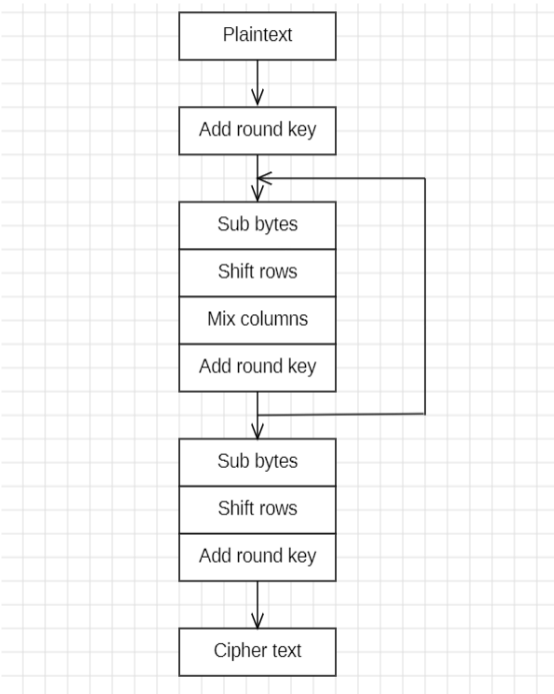

Fig -3.4: AES

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

4. SYSTEM FLOW DIAGRAM

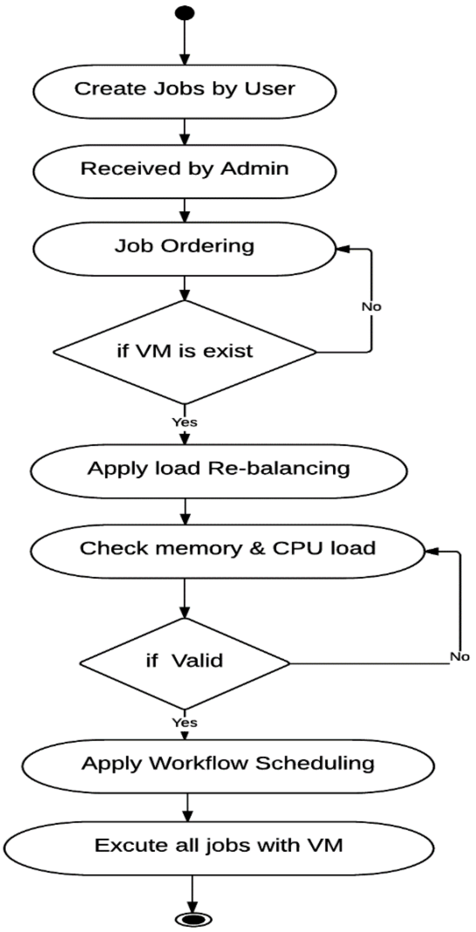

Fig -4.1: System Flow Diagram

4. IMPLEMENTATION

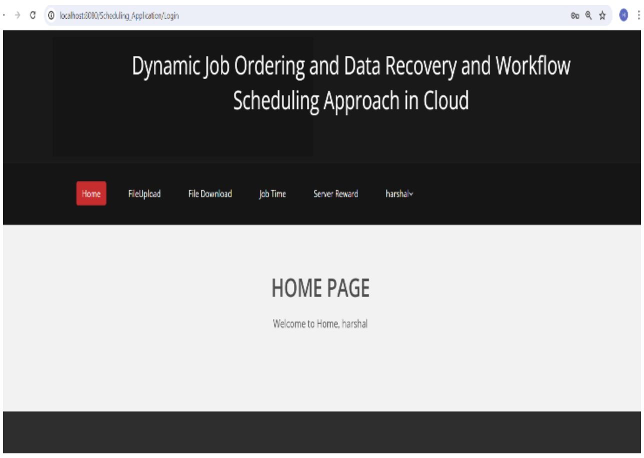

Fig -5.1: User Interface

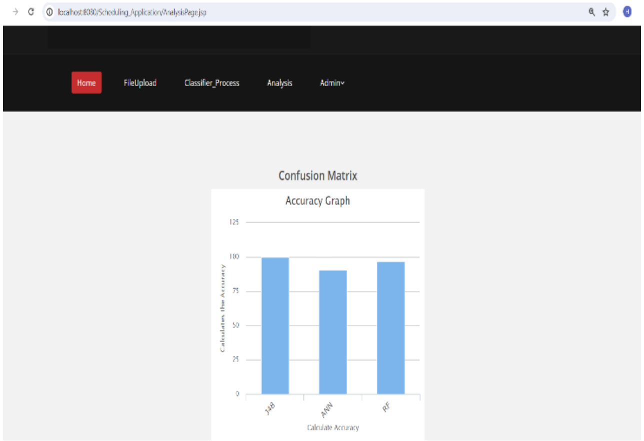

Fig -5.2: Admin Side

Fig -5.3: User Side

3. CONCLUSION

In conclusion, this research provides a comprehensive and efficient methodology for addressing dynamic job ordering, data recovery, workflow scheduling, chunk formation, and data security in cloud computing environments. By integrating advanced machine learning algorithms such as J48, ANN Q-learning, and Random Forest with robust cryptographic techniques like the Kochi matrix method for data chunk formation and AES encryption for data security, we effectively tackle several critical challenges inherent to cloud computing. This integrated approach ensures that resources are allocated efficiently, data management processes are optimized, and security measures are robust, which collectively enhance the overall performance of cloud systems.

Our experimental results substantiate the effectiveness of the proposed methodology, demonstrating significant improvements over traditional approaches in terms of system performance, reliability, and security. The use of machine learning algorithms has proven to be particularly beneficial in enhancing workflow scheduling and data recovery processes, while the implementation of cryptographic techniques has strengthened data security and integrity. These advancements ensure that cloud computing environments can handle large-scale data and complex workflows with greater efficiency and reduced downtime, addressing both operational and security concerns.

Overall, this work contributes valuable insights and practical solutions to the field of cloud computing, pushing forward the development of more resilient and efficient systems. By leveraging the strengths of both advanced machine learning and cryptographic techniques, we provide a holistic approach that meets the evolving demands of cloud computing. This research not only offers theoretical advancements but also practical applications that can be adopted by both academic researchers and industry professionals to enhance the reliability, performance, and security of cloud-based systems.

.

REFERENCES

1. Mohit Kumar , S.C. Sharma. Dynamic load balancing algorithm for balancing the workload among virtual machine in cloud computing https://doi.org/10.1016/j.procs.2017.09.141
2. Simon Heron, Advanced Encryption Standard (AES), https://doi.org/10.1016/S1353-4858(10)70006-4.
3. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA data mining software: an update. SIGKDD Explor. Newsl. 11, 1 (June 2009), 10–18. https://doi.org/10.1145/1656274.1656278
4. Muhammad Waseem Ahmad', Monjur Mourshed, Yacine Rezgui. Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption http://dx.doi.org/10.1016/j.enbuild.2017.04.038
5. Jesse Clifton and Eric Laber . Q-Learning: Theory and Applications https://doi.org/10.1146/annurev-statistics-031219-041220 .
6. Muhammad Junaid1 , Adnan Sohail 1 , Rao Naveed Bin Rais 2 , Adeel Ahmed 3 , Osman khalid4 , Imran Ali khan 4 , Syed Sajid Hussain 4 , and Naveed Ejaz1. Modelling an Optimized Approach for Load Balancing in Cloud https://doi.org/10.1109/ACCESS.2020.3024113
7. Abdullah, Ako. Advanced Encryption Standard (AES) Algorithm to Encrypt and Decrypt Data
8. Bhargava, Neeraj, Girja Sharma, Ritu Bhargava and Manish Mathuria. "Decision Tree Analysis on J48 Algorithm for Data Mining." (2013).
9. Steven J. Rigatti, MD, DBIM, DABFM. Random Forest https://doi.org/10.17849/insm-47-01-31-39.1
10. Eko A. Bušić , J. Doncel , J.M. Fourneau. Dynamic load balancing in energy packet networks https://doi.org/10.1016/j.peva.2024.102414