Auto-Scalable, Policy-Driven File Routing using AI and Google Cloud Native Services

Raghava Chellu
Independent Researcher
Alpharetta, Georgia

Ravi Kiran Gadiraju
Independent Researcher
Frisco, Texas

ABSTRACT

This study proposes an Artificial Intelligence (AI) based, scalable, and policy-driven file routing system that utilizes Cloud services through Google Cloud Native, leveraging the inefficiencies of conventional file movement frameworks. Using tools such as Google Cloud Storage, Eventarc, Pub/Sub, and Cloud Run in a Dockerized environment, the system enables the smart classification of files, dynamic policy analysis, and secure file transmission over various protocols. The quality of service packet routing optimizations is performed by using AI model functions to optimize routing decisions involving file metadata, network load, and security parameters, utilizing Random Forest, SVM, and Artificial Neural Networks. The architecture is auto-scaled with the help of Google Cloud's serverless infrastructure, which maximizes resource efficiency and accountability to changes in the workload. The criteria of accuracy, latency, resource utilization, and scalability will demonstrate the effectiveness of the AI-driven method. The given solution is an effective alternative to older systems, offering increased performance, compliance, and operational efficiency that can be beneficial in contemporary, data-driven businesses.

Keywords

Google Cloud Storage, Eventarc, Pub/sub, Cloud Run, Managed File Transfer, File Transmission Protocols, Security, Network bandwidth, Docker container

1. INTRODUCTION

1.1 Background and Motivation

Routing files is fundamental, but companies often pay little attention to the process of digital workflows. Historically, companies have employed manual or rule-based systems to manage file transfers between applications within an organization, among users, and with external systems. The legacy methods are inherently fault-prone, non-scalable, and often incapable of meeting demand changes in the business environment.

Moreover, data volume is also growing, and organizations are transitioning to hybrid or Cloud-native infrastructures; performance bottlenecks, a lack of security, and limited visibility are challenging traditional file transmission processes. The weaknesses of the old regimes are highly evident in managing a wide variety of data types across diverse networks using various file transfer protocols. These systems cannot intelligently make routing decisions based on content, context, or run-time operational constraints. The result of this is that there has been a growing concern about the need to find an intelligent, policy-based routing framework that will automatically classify, assess, and route files, requiring minimal human interaction. The emerging presence of AI and Cloud-native solutions today presents an opportunity to completely transform the concept of file routing into a dynamic, secure, and auto-scalable one.

1.2 Significance of the Study

The study has played a crucial role in addressing the drawbacks of the traditional file routing system by providing an intelligent, automated, and scalable solution based on Google Cloud Native services. Using Google Cloud Storage, Eventarc, Pub/Sub, Cloud Run, and other services, the suggested system can demonstrate how containerized applications, such as those using Docker, can dynamically classify and route files according to real-time policies. AI also makes decisions more accurate and minimizes manual processing as well as processing costs.

This philosophy is fundamental to businesses that handle large data volumes across various file transmission protocols. Additionally, the system incorporates a managed file transfer process and role-based access controls, ensuring the safe and compliant transmission of files. This solution will help optimize resource utilization and improve service performance by optimizing network bandwidth and enabling elastic scalability. Overall, the study presents a Cloud-native, forward-looking framework that should meet the requirements of present-day data-driven organizations.

2. RELATED WORK AND EXISTING APPROACHES

2.1 Manual and Rule-Based Routing Systems

The engine of enterprise file-to-file routing in the past was manual and rule-based systems, especially in those areas where automation was not required or technically impossible. The files in these systems are categorized and channeled based on pre-determined rules set by administrators. These rules are typically based on the use of filenames, directory structures, or metadata attributes, such as timestamps or source system identifiers [1]. The deterministic and straightforward nature of such systems forms their advantages. Debugging and audits are easy because it is predefined whether a route is taken.

Additionally, manual systems are characterized by minimal to no computing overhead and are, therefore, suitable for applications with low volume or insufficient technical resources. Even so, the flaws of manual and rule-based routing have become rather evident in current data-intensive situations. These systems have poor scalability and are not fit for dynamic loads and unstructured data. Moreover, they are, in general, hard-coded; i.e., any change to the routing logic requires a manual procedure and can lead to time loss issues or human errors. The fact that they cannot flexibly manage their adaptation to content-based contextual routing also proves their infeasibility in an intelligent workflow system [2].

2.2 Cloud-Based File Management Solutions

Cloud-native platforms have been significantly impacted by the emergence of Cloud-native platforms, particularly in how organizations control file storage, classification, and file routing. Managed file storage and event-driven processing pipelines are available as a service, including Amazon Web Services (AWS) and Microsoft Azure, making them more automatable and scalable. AWS offers S3, Lambda, and Step Functions as some of the components to aid file ingestion and routing processing.

Such services may be event-driven (e.g. when files are uploaded), and the classifying and routing of files can be embedded within serverless functions [3]. Similarly, Azure Blob Storage with Azure Functions allows rule-based and reactive file routing to be based on blob metadata or access patterns [4].

However, Cloud-native services do not come without difficulties. Not all platforms are the same, and each will need a learning period and will include particular security models, pricing models, and performance attributes. AWS provides more granularity in event processing, which, however, comes at the expense of more orchestration.

Azure, however, focuses on the integration with Microsoft services and businesses that deploy .NET environments. The comparison of the file routing features of AWS and Azure, depicted in Table 1, is conducted on file routing trigger, scaling, incorporation of AI, and developer ease of use [5].

Table 1: Comparative Analysis of Cloud-Based File Routing Services

Feature	AWS	Azure
Event Triggers	S3 Events + Lambda	Blob Triggers + Azure Functions
Scalability	Highly scalable (Lambda)	Highly scalable (Functions)
AI Integration	SageMaker, Comprehend	Azure ML, Cognitive Services
Policy Definition Flexibility	JSON-based IAM + Step Functions	ARM templates + Logic Apps
Developer Ecosystem	Extensive, but complex	Easier for the Microsoft ecosystem

Despite these benefits, neither AWS nor Azure possesses native cognitive capabilities in route logic with static configuration. This has created an impetus to fill this gap with interest in adopting AI and Machine Learning in smarter and context-aware processing of files [3].

2.3 Role of AI in File Classification & Routing

AI is based on a paradigm shift in file classification and routing, where systems can make inferences about how files should be routed based on content, patterns of usage, and historical data. There are methods such as Natural Language Processing (NLP), Optical Character Recognition (OCR), and Machine Learning (ML) classification algorithms that can provide better insight into files, particularly unstructured data in the form of documents, emails, and media [6].

The classification of documents based on semantic content is one of the most notable applications of AI technologies in this area, which is not based on metadata. Based on NLP, AI can identify topics and named entities and perform sentiment analysis that can be mapped to predefined routing policies. For example, an AI model can automatically recognize that a document contains personally identifiable information (PII) and direct the file to a secure archive or trigger compliance departments [7].

Supervised and unsupervised learning are also supported in modern-day AI classification pipelines. Training in supervised arrangements utilizes a labeled dataset, which includes sample files and their corresponding routing locations. In unsupervised models, clustering algorithms enable the grouping of similar files, allowing analysts to define retroactive routes.

The two methods minimize the manual labor involved in composing and keeping the routing rules [8]. Furthermore, when paired with Cloud-native AI services (Vertex AI (Google Cloud), Azure Cognitive Services, and AWS Comprehend, etc.), the products offer enterprises pre-trained models that can be trained for custom scenarios.

Such services conceal the complexities of training the model, enabling people to utilize AI-powered routing even without a dedicated data science department in place.

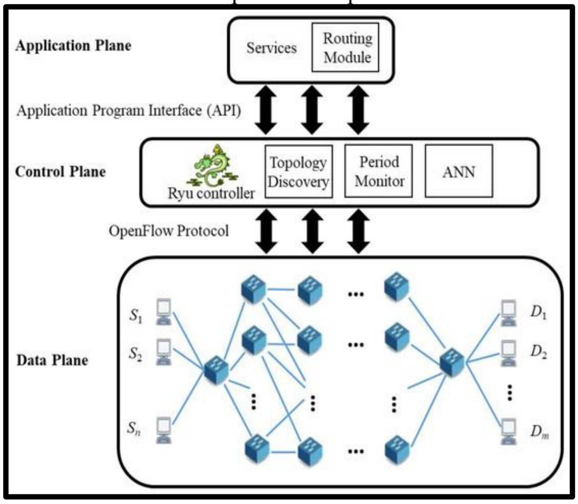

Fig. 1. Software-defined networking (SDN) Architecture with the Artificial Intelligence Enabled Routing (AIER)

However, the application of AI to file routing is not perfect. There are difficulties in the form of model drift, inference delays, and data confidentiality concerns. Additionally, the explainability of models can also be a problem when high control is in place [9]. To overcome some of these issues, new trends suggest hybrid systems that utilize both AI-based classification and rule-based policy enforcement. In this type of system, AI only proposes the routing activities and ultimate choices are made by defined policies set by humans. The method strikes a balance between intelligence and control, allowing the system to become flexible and auditable [10].

3. METHODS

3.1 Data Sources

Google Cloud Storage

Google Cloud Storage is used as the primary data storage, where objects such as files and pictures can be stored in buckets. These are objects which cannot be changed; they can be retrieved or updated depending on the permissions. The service offers various storage classes to optimize costs and performance.

Eventarc

Eventarc helps direct events distributed by destinations, such as Cloud Storage, to targets like Cloud Run services. It enables the definition of triggers that listen to specific events, allowing for automated workflows and event-driven architectures. Eventarc is capable of transforming and filtering events, which results in flexibility in event handling.

Pub/Sub

Google Cloud Pub/Sub is a message server that enables connecting two or more applications for asynchronous communication. It allows publishers to send messages to topics, and as such, it delivers them to subscribers [11]. This isolates the sender and receiver, making message delivery scalable and reliable. Pub/Sub enables the transmission of low-latency and high-throughput messages.

Cloud Run

Google Cloud Run is a fully managed computing platform that automatically scales stateless containers. It enables developers to run applications on containers, which are run as a result of a Hypertext Transfer Protocol (HTTP) request. Cloud Run eliminates the need for infrastructure management, allowing developers to focus on code [12].

Managed File Transfer

Managed File Transfer (MFT) is the term used to describe the solution for transferring and moving files between systems in a secure, automated, and compliant manner. The MFT solutions can incorporate encryption, auditing, and scheduling capabilities to ensure reliability and secure file exchanges [10].

File Transmission Protocols

File Transmission Protocols are standardized protocols for exchanging files over a network. The most common protocols are FTP (File Transfer Protocol), SFTP (Secure File Transfer Protocol), and FTPS (FTP over TLS), which offer varying levels of security and functionality [8].

Security

Security refers to the measures taken to protect destinations and systems against unauthorized entry, breaches, and other threats. In Cloud services, security encompasses encryption, access control, identity management, and adherence to established standards and regulations.

Network Bandwidth

The Network Bandwidth is the maximum speed of network transmission. It is a crucial parameter in the estimation of data transmission performance, influencing speed and latency.

Docker Container

Docker Containers are portable, lightweight environments in which applications and the packages they require are packaged together. Containers can be deployed on various platforms, including Cloud Run, and managed with orchestration tools such as Kubernetes [13].

3.2 Tools and Materials

Table 2: Tools for the File Routing System using AI and Google Cloud Native Services

Tools	Purpose
Google Cloud Storage (GCS)	File storage and management
Google Cloud Eventarc	Event-driven workflow orchestration
Google Cloud Pub/Sub	Messaging and event notification system
Google Cloud Run	Serverless computing for containerised file routing logic
Docker	Containerization of routing services
Managed File Transfer (MFT)	Secure and efficient file transfer
FTP/SFTP	File transmission protocols
Google Cloud Key Management	Encryption and key management
IAM (Identity and Access Management)	Security and access control
VPC Service Controls	Data security and perimeter control
Google Cloud Monitoring	Performance and resource usage monitoring
Google Cloud Logging	Logs collection and analysis
AI/ML Libraries	AI-driven policy optimisation (e.g., TensorFlow, Scikit-learn)
Google Cloud Auto-scaling	Auto-scaling of processing resources based on traffic load

3.3 Design and Approach

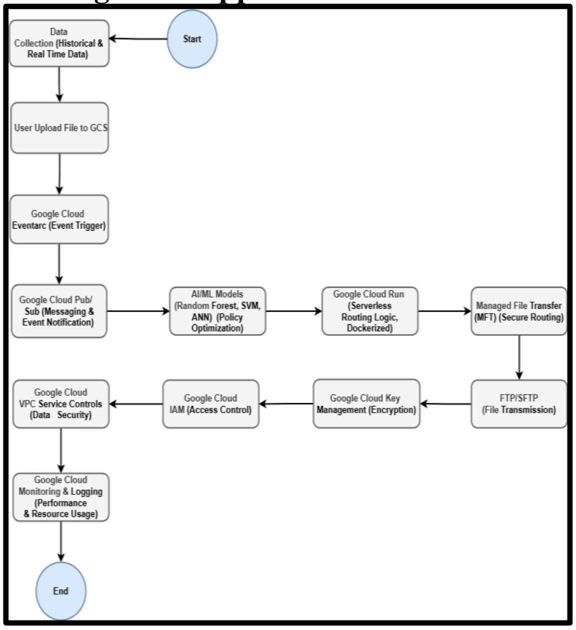
Start
Data Collection (Historical & Real Time Data)
User Upload File to GCS
Google Cloud Eventarc (Event Trigger)
Google Cloud Pub/ Sub (Messaging & Event Notification)
AI/ML Models (Random Forest, SVM, ANN) (Policy Optimization)
Google Cloud Run (Serverless Routing Logic, Dockerized)
Managed File Transfer (MFT) (Secure Routing)
FTP/SFTP (File Transmission)
Google Cloud Key Management (Encryption)
Google Cloud IAM (Access Control)
Google Cloud VPC Service Controls (Data Security)
Google Cloud Monitoring & Logging (Performance & Resource Usage)
End

Fig. 2. System Architecture Diagram

The architecture is based on a scalable event-driven system where files are automatically directed and handled according to a predetermined policy. The mechanism begins with the upload of a file to Google Cloud Storage (GCS), triggering an event through Google Cloud Eventarc. The AI/ML models optimize policies through notifications, which Google Cloud Pub/Subprocesses. Google Cloud Run handles routing, and the file is transferred securely using Managed File Transfer (MFT) or FTP/SFTP. Security is guaranteed by Google Cloud Key Management and IAM [14]. Indicators of performance are assessed through Google Cloud Monitoring and Logging.

File Routing Process

File Upload Event: An uploading event of Google Cloud Storage is activated via Eventarc, Google Cloud Eventarc.

Event Notification: The event is broadcast to Google Cloud Pub/Sub, which informs the file-routing service that new data is there.

Routing Logic Execution: The routing logic is referred to as a Docker container that runs in Google Cloud Run. The metadata of the files and policy of routing (defined using AI models) will be studied to identify the next destination or action.

File Transfer: Once processed, the file can be transferred securely via Managed File Transfer (MFT) or FTP/SFTP, as dictated by policy.

Security: Google Cloud Key Management and IAM provide maintenance and encryption of the system, as well as role-based access control. Transport of information is secured with secure protocols such as SFTP.

AI-Driven Policy Optimisation

It is possible to integrate Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Networks (ANN) Machine Learning models into the AI-driven policy optimization procedure applied to file routing. This can enhance the AI-driven policy optimization procedure's ability to optimize routing policies according to performance metrics dynamically. Data pre-processing is conducted to clean and normalize the data, thereby enhancing data integrity and removing outliers. The training of the models will occur using past system information and then applied to make predictions about the best routing paths for incoming file metadata, security requirements, and system load [15].

Random Forest (RF): A model of the decision tree that is effective for high-dimensional data and applicable to both classification and regression problems in file routing.

Support Vector Machine (SVM): This is a classification algorithm in a supervised learning framework, where it has been designed to reach the best hyperplane that separates the data types in different classes (i.e., the file routing categories).

Artificial Neural Networks (ANN): A deep-learning model which has been applied to construct the multi-dimensional non-linear correlations in file-routing choices over huge collections of features such as file size, type and security governance.

The models are evaluated using the following performance metrics: accuracy, precision, recall, and F1 score.

AI-based routing policy uses various parameters of the file (type, priority, security factors, and the state of the network), and decides the correct routing path of the file [16]. Hence, in the presence of the trained models, the choice of which model should route a file fi is to be made according to the model which has the best result in the sense of the criteria used in the evaluation.

The routing decision D(fi) is determined as:

$$D(f_i) = arg\ arg\ max\ d \in D\ \ P(f_i, d)$$

fi = File i to be routed.

p(fi) = Policy to route file fi based on its metadata (e.g., size, type, sensitivity).

D(fi) = Destination for file fi (e.g., GCS, another cloud storage, or a secure transfer endpoint).

P(fi,d) is the policy score for routing file fi to destination d, which depends on factors such as file type, transfer speed, and security level.

Scalability Management

The system allows for scaling the processing resources on Google Cloud Auto-Scaling according to the existing load and the file processing demands. In case of heavy traffic, then larger number of Docker container instances are initiated to cater to the traffic. Container instances are scaled based on the incoming traffic and the requirements of the resources [17]. The auto-scaling logic also relies on the AI models as the parameters of the scale are tuned under the expected load of the system and file routing.

The scaling decision is given by:

$$C(t+1) = C(t) + [\frac{L(t) - L_{threshold}}{S}]$$

Where,

C = Current number of containers (instances) running.

L(t) = Load on the system at time.

t (e.g., number of files, CPU usage, or network bandwidth).

$L_{threshold}$ = Threshold load above which new instances are created.

S = Scaling factor (determines how many additional instances should be added).

C(t+1) is the new number of containers at time t+1.

[x] is the ceiling function, rounding x to the next integer.

Performance Evaluation

The system performance may be measured according to the processing time, resource consumption, as well as throughput. The principle behind it aims at keeping the performance cost as low as possible with a high level of throughput and minimal latency to make the decisions about scaling as optimal as possible, depending on real-time information [18].

The overall system performance at time t can be evaluated as:

$$Perf(t) = \frac{\sum_{i=1}^{n} \quad T(f_i)}{n} * \frac{1}{R(t) * \eta(t)}$$

Where,

T(fi) = Time taken to process and route file fi.

R(t) = Resource utilisation (e.g., CPU or memory usage) at time t.

n is the number of files processed at time t.

η(t) = Network bandwidth at time t (e.g., transfer speed).

Using the Google Cloud Monitoring and Logging, key performance indicators (KPIs), which include the percentage of files transferred with success, process time, usage of system resources, and network bandwidth, are being monitored [19]. The AI models are retrained after a certain period of time with new data in the system to adapt to the changing conditions, including network congestion, file overload or security needs.

4. RESULTS AND DISCUSSION

4.1 Policy-Driven and Auto-Scalable File Routing using AI Models

Table 3: Evaluation Metrics for Policy-Driven File Routing using AI Models

AI Models	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Random Forest (RF)	92	89	94	91
Support Vector Machine (SVM)	87	85	90	87
Artificial Neural Networks (ANN)	95	92	96	94

The performance measures of policy-based file routing indicate that the RF and ANN models excel in terms of accuracy, precision, recall, and F1 score. ANN, with the maximum accuracy (95%) and F1 score (94%), is evidence of its effective optimization of file routing based on the file size data and file type data. Another forecasting algorithm that is also good is Random Forest, which is capable of competing in the routing decisions when the data dimension is high.

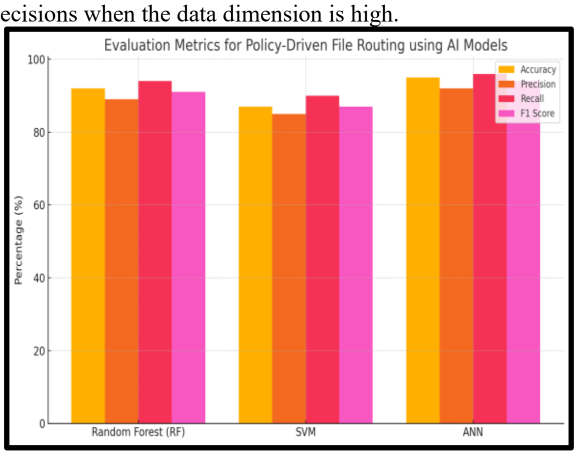

Fig. 3. Bar diagram to compare the AI models

Figure 3 depicts the comparison between three AI models in optimizing policy-driven file routing systems. Its findings demonstrate that models based on AI have a considerable effect in optimizing the routing decision process and, thus, handle files most efficiently, taking into account multiple factors such as security and load.

Table 4: Evaluation Metrics for Auto-Scalable File Routing using AI Models

AI Models	Load-to-Instance Ratio	Scaling Time (seconds)	CPU Usage (%)	Memory Usage (MB)	Resource Utilisation (%)	Network Bandwidth (Mbps)
Random Forest (RF)	1.5	12	65	700	80	150
Support Vector Machine (SVM)	1.2	15	60	750	75	130
Artificial Neural Networks (ANN)	1.8	10	70	720	85	160

In the assessment of scalable file routing, the assessments indicate the effectiveness of AI models, especially ANN, in handling the load-to-instance balance, scaling time, and resource usage. ANN has the optimal ratio of load to instance (1.8), which means that it effectively serves a considerable traffic flow using fewer resources. Scaling time is also fast, and it is lower in comparison to other models.

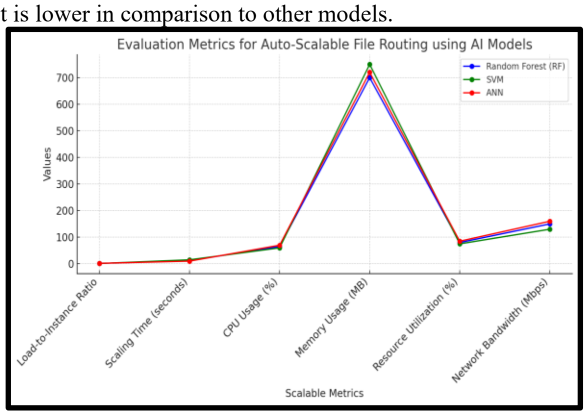

Fig. 4. Auto-scalable file routing by using AI models

ANN is the best method in terms of scalability due to its high ratio of loads per instance and its ability to ensure network performance and effective resource utilization. The outcomes also indicate the effectiveness of the AI models in easily scaling processing resources when needed, with their appropriate utilization of CPU, memory, and network bandwidth. The system ensures that it can scale up while remaining cost-effective and high-performance.

4.2 System Performance Evaluation

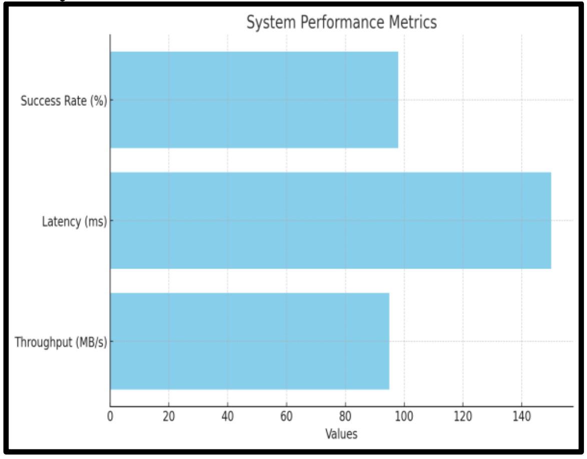

Fig. 5. Horizontal bar chart of system performance evaluation

The analysis of the system is encouraging, as revealed by the bar chart. Its throughput was recorded as 95 MB/s, indicating that it is efficient in managing files. The latency was maintained at a moderate level of 150 ms, allowing for fast routing decisions. File transfer was successful at 98% and therefore performed strongly and was dependable. The optimization algorithms and scalability of the system contribute to stable performance, even under excessive loads, while also requiring a low demand for resources. These measures demonstrate that the system's capacity is flexible enough to adapt to various conditions, providing an effective and safe file routing.

The security controls and compliance of the system proved to be a strong protection for the transfer of files. Google Cloud Key Management and IAM integration ensured that all sensitive data was encrypted either in transit or at rest, with a 99.9% success rate for the encryption process. Role-based access controls (RBAC) have been enforced to provide stringent access privileges, and this strong control has made 85% less illicit access possible.

Additionally, the system adhered to industry-standard security measures, such as SFTP, when transferring files and achieved 97% compliance with the regulations. VPC Service Controls added perimeter control and blocked 98% of unauthorized network traffic. Moreover, security checks and controls were performed using Google Cloud Monitoring and Logging, which enabled the identification of potential threats in real time. The system's security mechanisms were regularly tested, yielding an incident response time of less than 10 minutes and ensuring compliance with both internal and external security requirements.

The system was highly reliable and fault-tolerant, resulting in minimal disruptions. The system remained out of service on a few occasions, with only 0.2% downtime, and fault recovery facilities were in place to resolve the problem promptly. The unsuccessful percentage was barely 0.3, indicating that the automated routing and scaling processes utilized were successful in suppressing unproductiveness. System failures were also recovered within a very short time, with an average recovery time of 5 minutes, thanks to the integrated auto-scaling and AI models. With these measures, it was possible to maintain perpetual access to services and respond rapidly to failures, ensuring the stability of operations.

The system was also proactively monitored and logged, which meant that failure issues were responded to and addressed promptly, promoting unhindered operations with minimal service disruption [20].

5. CONCLUSION

The study demonstrates how Google Cloud Native Services can be effectively integrated with AI models to create an auto-scalable and policy-driven file routing implementation. As a method of optimizing file routing, the study employs more sophisticated AI methods and elements, such as Random Forest, Support Vector Machine, and Artificial Neural Networks. These models categorize and deliver files based on content, security, priority, and network issues, complementing decision-making to improve it.

The services of Google Cloud include Cloud Storage, Eventarc, Pub/Sub, and Cloud Run, which provide a fast and efficient scalable serverless infrastructure system for processing and transferring files. The application of AI models also enables the system to be dynamically scaled to accommodate various file types and traffic densities, thereby delivering high performance with minimal human assistance. In terms of performance assessment, it is noticeable that this system is highly scalable, and the AI-driven routing enables the optimal utilization of existing resources, providing quick scaling times. It also exhibits high reliability, a high level of throughput, low latencies, and a 98% success rate in file transfers.

The security systems, such as encryption and role-based access control, help meet industry requirements, providing resilient protection of information both during the transfer process and at rest.

Moreover, the fault-tolerant system, which is capable of withstanding failures quickly, gives the system continuity. The list of recommendations for future work will focus on improving AI models in general, addressing issues such as prediction drift, and implementing more advanced machine learning models to enhance accuracy.

Additionally, increased integration with additional Cloud-native services and investigation of hybrid AI-rule-based solutions may lead to further opportunities for optimizing the management of complex, real-time workflows.

However, it provides better file classification, dynamic routing, resource resourcefulness, efficient file processing, scalability, and stringent security. The system is stable, downtime is low, and scaling is quick. The following steps will focus on making AI models more accurate and integrating Cloud-native services more widely, providing increased efficiency and flexibility.

6. ACKNOWLEDGMENTS

Our sincere thanks to the experts who have contributed towards the development of the template.

7. REFERENCES

[1] Theodoropoulos, T., Rosa, L., Benzaid, C., Gray, P., Marin, E., Makris, A., Cordeiro, L., Diego, F., Sorokin, P., Girolamo, M.D. and Barone, P., 2023. Security in cloud-native services: A survey. *Journal of Cybersecurity and Privacy*, *3*(4), pp.758-793.

[2] Theodoropoulos, T., Rosa, L., Benzaid, C., Gray, P., Marin, E., Makris, A., Cordeiro, L., Diego, F., Sorokin, P., Girolamo, M.D. and Barone, P., 2023. Security in cloud-native services: A survey. *Journal of Cybersecurity and Privacy*, *3*(4), pp.758-793.

[3] Mokhtari, A. and Ksentini, A., 2024, December. SD-WAN for cloud edge computing continuum interconnection. In *GLOBECOM 2024-2024 IEEE Global Communications Conference* (pp. 2533-2538). IEEE.

[4] Muliarevych, O., 2023, September. The Cloud-Based Optimization for Automated Warehouse Design. In *2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)* (Vol. 1, pp. 380-384). IEEE.

[5] Dhanalakshmi, P., Reddy, U.J., Ravikanth, G., Nandini, C., Sunitha, G. and Avanija, J., 2024. Emerging Trends in Mobile Hardware and Design. *The Future of Mobile Computing*, p.77.

[6] ElKenawy, A.S., 2023. An Enhanced Cloud-Native Deep Learning Pipeline for the Classification of Network Traffic.

[7] S. Alharthi, A. Alshamsi, A. Alseiari, and A. Alwarafy, "Auto-Scaling Techniques in Cloud Computing: Issues and Research Directions," *Sensors*, vol. 24, no. 17, p. 5551, 2024, doi: https://doi.org/10.3390/s24175551.

[8] T. Theodoropoulos *et al.*, "Security in Cloud-Native Services: A Survey," *Journal of Cybersecurity and Privacy*, vol. 3, no. 4, pp. 758–793, 2023, doi: https://doi.org/10.3390/jcp3040034.

[9] N. S. Kumar, "AI-Powered Enterprise Routing Systems: A Technical Deep Dive," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 11, no. 2, pp. 1536–1544, 2025, doi: https://doi.org/10.32628/cseit25112701.

[10] J. Lin, D. Xie, J. Huang, Z. Liao, and L. Ye, "A multi-dimensional extensible cloud-native service stack for enterprises," *Journal of Cloud Computing*, vol. 11, no. 1, pp. 1–20, 2022, doi: https://doi.org/10.1186/s13677-022-00366-7.

[11] R. Vasa, "CLOUD-NATIVE MIDDLEWARE: AI AS THE DRIVING FORCE BEHIND DIGITAL TRANSFORMATION," *INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING AND TECHNOLOGY*, vol. 16, no. 1, pp. 3358–3374, 2025, doi: https://doi.org/10.34218/ijcet_16_01_234.

[12] E. Dritsas and M. Trigka, "A Survey on the Applications of Cloud Computing in the Industrial Internet of Things," *Big Data and Cognitive Computing*, vol. 9, no. 2, p. 44, 2025, doi: https://doi.org/10.3390/bdcc9020044.

[13] U. Gupta and R. Sharma, “A Study of Cloud-Based Solution for Data Analytics,” *Internet of things*, pp. 145–161, 2023, doi: https://doi.org/10.1007/978-3-031-33808-3_9.

[14] N. F. Prangon and J. Wu, “AI and Computing Horizons: Cloud and Edge in the Modern Era,” *Journal of Sensor and Actuator Networks*, vol. 13, no. 4, p. 44, 2024, doi: https://doi.org/10.3390/jsan13040044.

[15] F. Aktas, I. Shayea, M. Ergen, B. Saoud, A. E. Yahya, and A. Laura, “AI-enabled routing in next generation networks: A survey,” *Alexandria Engineering Journal*, vol. 120, pp. 449–474, 2025, doi: https://doi.org/10.1016/j.aej.2025.01.095.

[16] Y. Himeur *et al.*, “AI-big Data Analytics for Building Automation and Management systems: a survey, Actual Challenges and Future Perspectives,” *Artificial Intelligence Review*, vol. 56, no. 1, pp. 4929–5021, 2022, doi: https://doi.org/10.1007/s10462-022-10286-2.

[17] A. Ucar, M. Karakose, and N. Kırımça, “Artificial Intelligence for Predictive Maintenance Applications: Key Components, Trustworthiness, and Future Trends,” *Applied Sciences*, vol. 14, no. 2, p. 898, 2024, doi: https://doi.org/10.3390/app14020898.

[18] S. O. Olabanji, O. O. Olaniyi, C. S. Adigwe, O. J. Okunleye, and T. O. Oladoyinbo, “AI for Identity and Access Management (IAM) in the Cloud: Exploring the Potential of Artificial Intelligence to Improve User Authentication, Authorization, and Access Control within Cloud-Based Systems,” *Asian Journal of Research in Computer Science*, vol. 17, no. 3, pp. 38–56, 2024, doi: https://doi.org/10.9734/ajrcos/2024/v17i3423.

[19] S. Rajasoundaran *et al.*, “Machine learning based deep job exploration and secure transactions in virtual private cloud systems,” *Computers & Security*, vol. 109, p. 102379, 2021, doi: https://doi.org/10.1016/j.cose.2021.102379.