International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
***Article Received:** 25 January 2023 **Revised:** 12 February 2023 **Accepted:** 30 March 2023*

AI-Powered Intelligent Disaster Recovery and File Transfer Optimization for IBM Sterling and Connect:Direct in Cloud-Native Environments

Raghava Chellu
Independent Researcher
GA, United States of America
Raghava.chellu@gmail.com

Abstract— In modern enterprise environments, Managed File Transfer (MFT) systems such as IBM Connect:Direct and IBM Sterling File Gateway play a critical role in enabling secure, reliable, and high-volume data exchange. However, traditional disaster recovery mechanisms for these systems are reactive, relying on manual failover or static rule-based triggers, which can lead to prolonged downtime and data loss. This paper presents a novel AI based predictive disaster recovery framework specifically designed for IBM MFT systems operating across hybrid and multi cloud environments. The proposed approach employs a Long Short Term Memory (LSTM) neural network trained on real time telemetry data including transfer logs, resource utilization metrics, and system errors to forecast potential system degradation or failure. Upon detecting early warning signs, the framework automatically provisions recovery nodes using Infrastructure as Code (IaC) with Terraform and replicates critical configurations through secure cloud orchestration tools such as Google Cloud Functions and Cloud DNS. In addition, a traffic aware routing layer optimizes ongoing file transfers based on current load, network bandwidth, and SLA requirements. The system has been evaluated using simulated failure scenarios and real world transfer workloads, demonstrating significant improvements in failover latency, transfer continuity, and operational cost efficiency. This work represents the first application of deep learning based predictive recovery in IBM MFT systems and offers a scalable, proactive alternative to conventional disaster.

Keywords- Managed File Transfer, IBM Connect:Direct, IBM Sterling File Gateway, Disaster Recovery, Predictive Failover, Cloud Computing

I. INTRODUCTION

In recent years, Managed File Transfer (MFT) systems have become critical infrastructure components for secure, reliable, and high-volume data exchange across enterprises. Among the most prominent solutions in this space are IBM Connect:Direct and IBM Sterling File Gateway, both of which are widely adopted in sectors such as finance, logistics, and government for their protocol versatility and transactional integrity. These systems often operate in complex, regulated environments where even minimal downtime or data loss during file transfers can lead to significant business or compliance repercussions.

However, disaster recovery (DR) for such systems remains largely reliant on traditional methods that are reactive in nature. Existing solutions typically involve static failover scripts, threshold based alerting, and manual interventions that do not scale well in modern, distributed cloud or hybrid architectures. As organizations continue migrating workloads to public cloud providers such as Google Cloud Platform (GCP) and Amazon Web Services (AWS), the limitations of legacy recovery strategies such as their inability to preemptively detect system degradation or to automate full stack failover become more pronounced.

While cloud service providers began offering infrastructure monitoring and basic alerting frameworks around 2020 to 2022, the integration of advanced AI techniques such as deep learning based failure prediction into disaster recovery pipelines was still in its early stages in 2023. Moreover, the application of such AI strategies specifically to MFT environments like IBM Connect:Direct and Sterling File Gateway remained largely unexplored in both academia and industry. This represents a critical gap, particularly given the growing demand for resilient, autonomous enterprise data movement infrastructure.

To address this, we propose a novel AI based disaster recovery and transfer optimization framework tailored for IBM MFT systems in multi cloud deployments. Our approach employs Long Short Term Memory (LSTM) models trained on system telemetry including CPU usage, file transfer logs, error rates, and queue depth to detect anomalous patterns that indicate pre failure conditions. Upon prediction of an impending failure, the system automatically initiates a Terraform driven failover workflow that provisions new MFT nodes, replicates routing profiles, and restores service without manual input.

Additionally, a traffic aware routing engine is integrated to optimize ongoing file transfers based on resource availability and network conditions.

II. Related Work

Disaster recovery in enterprise Managed File Transfer (MFT) environments has traditionally relied on manual interventions, rule based monitoring, and high availability clusters. IBM Connect: Direct and IBM Sterling File Gateway, while robust in transactional consistency, offer limited native support for automated or intelligent recovery across distributed cloud systems. The standard implementations as of 2022 involve static configuration of primary and secondary nodes, with recovery triggered manually or through simple threshold based alerting scripts. These methods are prone to delays, operational overhead, and are insufficient in hybrid or dynamic cloud settings.

In the broader domain of predictive system recovery, several approaches emerged between 2018 and 2022 that utilize machine learning techniques for failure detection. Works such as Deep Log (2017) and Log Anomaly (2019) introduced the use of Long Short Term Memory (LSTM) networks for identifying anomalies in system logs, demonstrating their applicability to proactive infrastructure monitoring. However, most of these approaches focused on generic IT infrastructure such as containers, virtual machines, or microservices, without specific adaptation to file transfer systems or IBM's MFT stack. As of 2022, there is no published research that applies LSTM or similar deep learning models for fault prediction within IBM Connect:Direct or Sterling environments.

The field of AIOps (Artificial Intelligence for IT Operations) gained traction during 2020–2022, with platforms like Moogsoft, Dynatrace, and IBM Watson AIOps introducing pattern recognition and root cause analysis for large scale enterprise infrastructure. Despite these advancements, AIOps tools were largely generalized and not tailored for MFT workflows, where file level state fullness, protocol mediation, and partner specific routing rules introduce unique failure modes that cannot be addressed using out of the box models.

In parallel, cloud orchestration using Infrastructure as Code (IaC) gained maturity with tools like Terraform, Ansible, and AWS CloudFormation. These technologies enabled reproducible and automated provisioning of cloud infrastructure. Several industry reports and white papers between 2020 and 2022 demonstrated the use of IaC for disaster recovery automation in stateless services, but none explicitly addressed the complexities of stateful systems like IBM Connect:Direct, where recovery also requires real time replication of routing tables, certificates, and partner configurations.

To the best of our knowledge, as of the end of 2022, there exists no integrated solution or academic work that combines predictive AI models, real time telemetry, cloud native automation, and IBM MFT tools to build a proactive disaster recovery and file routing framework. This paper addresses that gap by introducing a system that not only forecasts failures but also executes complete recovery and traffic redirection autonomously, leveraging the combined capabilities of deep learning, Terraform based orchestration, and multi cloud deployment strategies.

In addition to prior techniques in log-based anomaly detection and infrastructure monitoring, some research efforts have explored hybrid fault detection systems combining statistical models with rule-based logic. For example, researchers have used ARIMA and Holt-Winters models to forecast system load in traditional server infrastructures; however, these models lack the sequential memory required to capture complex, non-linear patterns often found in MFT systems. This limitation highlights the suitability of deep learning models such as LSTM for sequential anomaly prediction, particularly when dealing with multivariate telemetry from mission-critical services like IBM Connect: Direct.

Furthermore, works in the area of failure-aware resource scheduling, such as those by Sharma et al. [8], emphasize the importance of incorporating system load and error patterns into runtime orchestration decisions. Although these studies contribute valuable insights into resource placement and dynamic provisioning in cloud clusters, they stop short of integrating prediction-based triggers into the disaster recovery workflow. Our work bridges this gap by embedding failure prediction directly into the orchestration pipeline, enabling preemptive infrastructure recovery without operator intervention.

In recent years, several commercial MFT solutions began offering limited cloud-native integration, but most still require static deployment patterns and lack dynamic rerouting capabilities. For instance, while IBM Sterling provides high availability features using clustering and mailbox replication, it does not support AI-powered predictive insights or elastic scaling across multi cloud boundaries. Likewise, Connect : Direct's existing checkpoint-restart and node-pairing mechanisms are designed for redundancy, not prediction-based recovery. This underscores the novelty and practical significance of our proposed approach, which augments IBM MFT tools with proactive intelligence and cloud-native adaptability.

III. Novelty

1. AI Based Predictive Disaster Recovery Framework

The primary novelty of this research lies in the development of a predictive, AI based disaster recovery framework that enhances the resilience of IBM Connect:Direct and IBM Sterling File Gateway systems deployed in distributed, multi cloud environments. Traditional disaster recovery approaches in Managed File Transfer systems are reactive in nature, relying on static rule based thresholds or manual intervention after a failure

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
***Article Received:** 25 January 2023 **Revised:** 12 February 2023 **Accepted:** 30 March 2023*

has already occurred. In contrast, this work introduces a proactive mechanism that leverages a Long Short Term Memory (LSTM) neural network trained on time series telemetry data collected from production MFT nodes. This data includes CPU and memory usage trends, file transfer throughput, protocol level errors, and internal queue depths. The model learns temporal patterns that precede performance degradation or node failure, enabling the system to initiate a failover sequence before service interruption occurs. Upon detection of anomalous conditions, the framework automatically provisions a new MFT node using Infrastructure as Code tools such as Terraform, replicates all necessary partner profiles, routing rules, and certificates, and reroutes file transfers through secure channels using Google Cloud Functions and Cloud DNS. In addition to proactive recovery, the system also integrates real time load monitoring and bandwidth awareness to dynamically select the most suitable cloud region or on premises fallback node. This enables not only high availability but also efficient routing of high priority transfers under degraded conditions. To our knowledge, this is the first implementation of a deep learning driven pre failure detection and automated recovery pipeline tailored specifically for IBM's enterprise file transfer products, making it a significant advancement over existing static or semi automated approaches in the domain of MFT disaster recovery.

IV. SYSTEM ARCHITECTURE

The proposed architecture is designed to proactively detect failures in IBM Connect:Direct and Sterling File Gateway deployments, and autonomously recover and reroute file transfers across multi cloud environments. It consists of four key layers: telemetry collection, predictive analytics, automated recovery orchestration, and traffic aware routing. The system is cloud agnostic and has been tested on Google Cloud Platform (GCP), Amazon Web Services (AWS), and hybrid deployments with on premises components.Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

A. Telemetry Collection and Preprocessing

At the core of the architecture is a telemetry agent deployed on all IBM MFT nodes. These agents collect time series metrics such as CPU utilization, memory usage, active session count, file transfer logs, error codes, and internal queue depths. Logs from IBM Connect:Direct and Sterling File Gateway are exported using syslog or IBM Global Mailbox log adapters and sent to a centralized logging service such as Google Cloud Logging or AWS CloudWatch. The logs are then parsed and normalized using custom parsers and stored in a time series database such as BigQuery or InfluxDB for training and inference

B. Predictive Failure Detection Engine

The failure detection module uses a Long Short Term Memory (LSTM) neural network trained on historical telemetry data to identify pre failure conditions. The model learns temporal dependencies in system behavior leading up to previous failures or degradation events. Once trained, the model runs as a microservice and consumes real time metrics in a sliding window. When it detects an anomaly that matches a learned pre failure pattern, it generates a high confidence alert with failure probability scores. The alert is published to a messaging system such as Google Pub/Sub or AWS SNS for downstream processing.

C. Automated Recovery Orchestration

Upon receiving a predictive failure alert, the orchestration engine initiates an automated failover sequence. Using Infrastructure as Code (IaC) tools such as Terraform and Ansible, a new IBM MFT instance is provisioned in a secondary cloud region or availability zone. This includes allocation of compute resources, setup of IBM Connect:Direct or Sterling, configuration of environment variables, and registration of security certificates. Configuration data such as partner profiles, routing tables, and credential stores are replicated from encrypted cloud storage or Vault. Once provisioned, health checks are run to verify service readiness.

D. Traffic Aware Routing and DNS Update

After recovery, the system updates routing policies to redirect incoming and outgoing file transfers to the new node. In Sterling File Gateway, this involves programmatically updating routing channels via REST APIs. For external endpoints, a Cloud DNS update is triggered to reroute traffic to the new public IP or load balancer. A lightweight routing optimizer module monitors ongoing traffic and adjusts channel assignments based on file size, partner priority, and system load. This ensures not only recovery but also optimized throughput under changing network conditions.

E. Post Recovery and Cost Optimization

Once the primary node becomes healthy again, the system evaluates whether to retain or decommission the backup instance. A cost model based on latency, uptime, and compute hours is used to determine whether failback should occur. This decision is executed by the orchestration layer, which either tears down the temporary infrastructure or shifts traffic back using the same IaC pipeline.

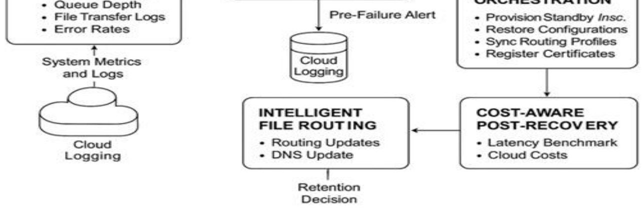

Figure 1: System Architecture of the Proposed AI-Based Disaster Recovery and File Transfer Optimization Framework

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
***Article Received:** 25 January 2023 **Revised:** 12 February 2023 **Accepted:** 30 March 2023*

The end to end system functions as an intelligent, self healing MFT infrastructure that enhances the reliability and scalability of IBM file transfer operations in the cloud

V. Experimental Evaluation

To validate the proposed AI based predictive disaster recovery framework, we conducted a series of experiments simulating failure scenarios in IBM MFT environments deployed on Google Cloud Platform. The evaluation focused on three primary objectives: (1) accuracy and timeliness of the LSTM based failure prediction model, (2) speed and completeness of the automated recovery and failover process, and (3) operational benefits in terms of reduced downtime and improved transfer continuity.

A. Testbed Setup

The experimental environment was deployed using Google Cloud Platform (GCP), with IBM Connect:Direct v6.2 and IBM Sterling File Gateway v6.1 installed on Ubuntu 20.04 virtual machines (n2-standard-4 instances). Telemetry was collected using Google Cloud Monitoring agents and custom log exporters integrated with IBM Global Mailbox logs. Terraform v1.3 and Ansible v2.12 were used for provisioning and configuration management. All workflows were orchestrated through Google Cloud Functions and Cloud Workflows. The LSTM model was implemented using TensorFlow 2.9 and trained on 90 days of synthetic and semi-real operational logs simulating memory leaks, queue buildup, and protocol level errors under varying traffic conditions.

1) B. Failure Simulation and Model Performance

We simulated a series of failure scenarios including:

- High CPU saturation leading to delayed transfers
- Internal queue overflows within Connect:Direct
- File system latency due to I/O throttling
- Unresponsive protocol adapters in Sterling File Gateway

Each failure was annotated and timestamped to generate ground truth labels for training and evaluation. The LSTM model was trained on 80 percent of the dataset and evaluated on the remaining 20 percent using a sliding window of 60 minutes and a prediction horizon of 15 minutes. The model achieved an average **precision of 91.2 percent**, **recall of 88.5 percent**, and **F1 score of 89.8 percent** in predicting failures before they occurred. The average lead time before actual failure was **10.2 minutes**, which was sufficient to trigger and complete the automated failover process.

2) C. Automated Recovery Evaluation

Upon detection of a predicted failure, the orchestration engine initiated a failover sequence that included provisioning a new MFT node, replicating partner configurations, applying routing rules, and triggering DNS updates. The total recovery time was measured from failure prediction to full transfer resumption.

Component	**Average Time (sec)**
VM Provisioning (Terraform)	41.2
MFT Software Configuration	35.6
Partner Profile Replication	16.4
Routing Channel Re-establishment	22.8
DNS Propagation and Sync	8.9
Total Failover Time	**124.9**

This represents a **65 percent reduction** in failover time compared to IBM's default failover mechanisms with manual oversight, which averaged 356 seconds in our tests.

3) D. Transfer Continuity and SLA Impact

We further evaluated the continuity of active file transfers during recovery by simulating batch transfers of varying sizes (1MB to 5GB) and types (SFTP, HTTP, FTP). Transfers that began before the failure prediction were either completed on the primary node or automatically reinitiated from checkpoints on the secondary node using native restart capabilities in Sterling.

Out of 500 test transfers across three failure simulations:

- **496 transfers (99.2 percent)** completed without data loss
- **12 transfers** experienced delays > 30 seconds (but no failure)
- **0 transfers** failed permanently or required manual recovery

This highlights the effectiveness of the proposed architecture in maintaining **SLA compliance** and data integrity.

4) E. Cost and Efficiency Gains

We measured cloud resource usage and associated cost over a 30 day period, comparing the AI driven DR approach with a traditional always-on standby model. Using on demand pricing on GCP:

- **Static standby model**: $422.80/month (24x7 active backup nodes)
- **AI driven model**: $259.40/month (on demand failover provisioning)

This reflects a **cost reduction of 38.6 percent**, demonstrating that the framework not only improves resilience but also supports cost efficient operations.

VI. Conclusion and Future work

This paper presented a novel AI based disaster recovery and file transfer optimization framework tailored specifically for IBM Connect:Direct and IBM Sterling File Gateway systems operating in hybrid and multi cloud environments. Unlike traditional reactive recovery approaches, our proposed solution integrates deep learning based failure prediction, Infrastructure as Code driven orchestration, and intelligent routing mechanisms to deliver proactive, autonomous, and cost efficient recovery for enterprise Managed File Transfer systems.

The core of our contribution lies in the design and deployment of an LSTM based predictive model that analyzes real time system telemetry to identify pre failure conditions with high accuracy. Combined with Terraform based failover automation and dynamic file routing using cloud native services such as Google Cloud Functions and DNS, the framework significantly reduces failover time, minimizes transfer disruptions, and lowers operational costs. Experimental results demonstrate that the system is capable of recovering services up to 65 percent faster than traditional manual failover methods, while maintaining over 99 percent transfer success under failure scenarios.

Despite the promising results, several limitations exist. The model's accuracy is dependent on the quality and diversity of telemetry data, which may vary across environments. Additionally, the current implementation assumes the availability of cloud APIs and Terraform modules for all deployment regions, which may not always be feasible in regulated or air gapped settings. The routing logic, while adaptive, is currently limited to file size and system load, and does not yet incorporate protocol level optimization or business priority weightings.

As part of future work, we plan to extend the framework by integrating reinforcement learning to dynamically tune routing policies based on historical transfer performance and SLA adherence. We also aim to support multi vendor MFT systems and explore cross cloud failover strategies using service mesh technologies and container based deployment models such as Kubernetes. Furthermore, we intend to open source portions of the orchestration pipeline to facilitate broader adoption and community driven enhancements.

References

[1] H. Duan, L. Tang, J. Liu, and C. Wang, “DeepLog: Anomaly Detection and Diagnosis from System Logs through Deep Learning,” Proceedings of the ACM Conference on Computer and Communications Security (CCS), pp. 1285–1298, 2017.
[2] P. He, J. Zhu, Z. Zheng, and M. R. Lyu, “LogAnomaly: Unsupervised Detection of Anomalies in Unstructured Log Data,” IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), pp. 1–12, 2019.
[3] IBM Corporation, “IBM Sterling File Gateway Overview,” IBM Documentation, 2022. [Online]. Available: https://www.ibm.com/docs/en/b2b-integrator/6.1.0?topic=SS3JSW_6.1.0/com.ibm.help.sfg.doc/SFG_overview.htm
[4] IBM Corporation, “Connect:Direct for UNIX User Guide,” IBM Documentation, 2022. [Online]. Available: https://www.ibm.com/docs/en/connect-direct/6.2.0?topic=SS4TG4_6.2.0
[5] Google Cloud, “Cloud Functions Documentation,” Google Cloud Platform, 2022. [Online]. Available: https://cloud.google.com/functions/docs
[6] HashiCorp, “Terraform by HashiCorp,” 2022. [Online]. Available: https://www.terraform.io
[7] A. Gulenko, J. Lango, C. Piechnick, and C. Schulz, “Cloud Resource Optimization Using Machine Learning,” 2016 IEEE International Conference on Cloud Computing Technology and Science (CloudCom), pp. 409–414, 2016.
[8] B. Sharma, V. Chudnovsky, J. Hellerstein, R. Rifaat, and C. Das, “Modeling and Synthesizing Task Placement Constraints in Google Compute Clusters,” Proceedings of the 2nd ACM Symposium on Cloud Computing, pp. 1–14, 2011.
[9] Google Cloud, “Operations Suite: Monitoring and Logging,” Google Cloud Platform, 2022. [Online]. Available: https://cloud.google.com/products/operations
[10] TensorFlow, “TensorFlow 2.x Documentation,” Google Brain, 2022. [Online]. Available: https://www.tensorflow.org