International Journal of Emerging Trends in Computer Science and Information Technology
ISSN: 3050-9246 | https://doi.org/10.63282/3050-9246.IJETCSIT-V5I2P123
Eureka Vision Publication | Volume 5, Issue 2, 215-223, 2024

Original Article

Explainable AI anomaly Detection in Banking Microservices

Lalitha Reddy Badam
Independent Researcher, University of New Haven, USA.

Abstract - In a move to offer real time, resilient and scalable banking services, banks are also resorting to distributed microservice-based banking applications. The dynamics of interactions among microservices however create heterogeneous logs, metrics, traces, and transaction events, and it is challenging to detect and interpret anomalous behavior in a timely manner. Traditional approaches to anomaly detection are useful at pointing out anomalies in system behavior, but tend to give few attempts to explain why they made their predictions, diminishing its value to security analysts and banking operations staff teams. The present paper will present an explainable AI (XAI) architecture of banking microservices based anomaly detection that will possess such modules as the behavioral monitoring module, multi-source telemetry analysis module, anomaly detection by machine learning and explanation module. The framework will associate service-level and transaction-level metrics with each other to differentiate regular service variation with operations and instances of suspicious or abnormal behavior. Mechanisms of explainability would provide evidence of the characteristics and service engagement and patterns of behavior that have resulted in each anomaly decision that are interpretable. The solution will be to enhance the quality of anomaly detection, transparency of diagnosis, and analyst confidence and aid in timely responding to incidents in diverse banking settings. The systematic basis of reliable AI-assisted monitoring of distributed financial services would be offered through its scalability, interpretability, and applicability to operations as also brought out in the framework.

Keywords - Explainable Artificial Intelligence, Anomaly Detection, Banking Microservices, Machine Learning, Distributed Systems, Financial Cybersecurity, Service Monitoring.

1. Introduction

The result of the rapid pace of the digital banking forms, real-time payment solution, mobile applications, and the distributed financial ecosystem is that modern banking information systems have become highly sophisticated. Microservice architectures have gained new significance since large banking applications are broken down into independently deployable microservices that deal with authentication, customer management, payment processing, transaction validation, fraud analysis and account services. Whereas this architecture lets organizations enhance scalability and flexibility as well as the availability of services, it will institute very dynamic interactions between distributed components. Abnormal behavior can therefore manifest itself as an individual-service or transaction or inter-service level and it might be hard to detect where the analysis of system activity believes the use of individual monitoring indicators. Anomaly detection is thus a valuable financial system feature in the sense that abnormal behavior may signal a failure of systems, fraud, and intrusions, as well as variances of how transactions should work [2]. ScienceDirect

The growing amount and diversity of banking telemetry also add to the burden of anomaly detection. An environment based on microservices can produce sustained stream of transaction records, service traces, application logs, service traces, authentication traces, network traces, and behavioral traces. Such data sources can include varying levels of granularity and different periods of time and it is therefore not easy to get an interpretation of what is abnormal. Traditional anomaly detectors can be used to find exceptions in behavior using statistical, machine learning, clustering or reconstruction-based detectors, but can not reveal why an observation is an exception. This is particularly important in the banking environment where the security analysts and operational departments must be presented with intelligible evidence before starting in the investigation or response of an alert. Studies have equally put the research problems of financial anomaly detection on the tangibility and representativeness of real-life financial data, which might influence the reliability and feasibility of detection schemes [2].

Deep learning and machine learning have provided viable processes of identification in advanced and previously unknowingly formed designs in massive-dimensional data. Namely, it is possible to train autoencoder-based models to learn the properties of normal behaviour and to discover those that will lead to abnormally large reconstruction errors. Interpretable architecture Autoencoders have demonstrated that anomaly detection can be combined with a system that produces the decisions made by the system more readable by domain experts [1]. However, one issue has been brought about by the growth in complexity of machine learning models: the high level of detection might not guaranteed translate into the high level of interpretability. A model can rightly indicate a suspect transaction or service interaction but it can give inadequate details to show the particular characteristics of behavior that informed the decision.

Explainable artificial intelligence (XAI) will alleviate this disadvantage by offering ways in which machine learning decisions can be explained and exchanged with human participants. The concept of transparency, interpretability, accountability, and user trust are now a relevant concern when applying AI systems in high sensitivity areas of applications and this is what has made XAI an imperative research topic [4]. Explainability in banking microservices can enable analysts to go further than simply provide an anomaly score and explain the evidence. Examples of such evidence can involve anomalous transaction properties, anomalous service behaviour, resource utilisation, anomalous authentication behaviour, or connections between various telemetry properties. As a result, explainability can encourage a more knowledgeable investigation and decrease the complexity of interpretation of the warnings produced by complicated detection frameworks.

The significance of the explanations of mechanisms has been shown to be important in the detection of anomalies of our detecting view as a consequence of a recent study. Autoencoders have been used to identify anomalies, whose causes such as the presence of features that are related to high reconstruction error have been explained using SHAP-based methods [3]. This is especially applicable when dealing with a banking context since such approaches can offer an explanation at an instance level as opposed to giving an explanation about the general behavior of a detection model. Likewise rule-extraction strategies to find unsupervised anomalies have shown that explainable rules may be obtained out of typically interpretationally inexplicable detection models, and that their quality to clarify importance is gauged by measures like understandability, representativeness, stability, and variation [5].

In spite of the developments, explainable anomaly detection in the case of distributed microservices in banking is hard to apply. Recent tools tend to focus on a data set, a single framework on anomaly detection, or general purpose explainability instead of a unified analysis of mixed microservice monitoring and banking-related conduct indicators. This is why the need of a structured framework in which multi-source monitoring, machine learning-driven anomaly detection, and the production of insightful explanation will be practiced in a distributed banking context exists. Such a structure should not just be able to point out cases of abnormal behavior, but also provide quality information about the services, features and even trends of behavior that could have led to the event of occurrence of the abnormality.

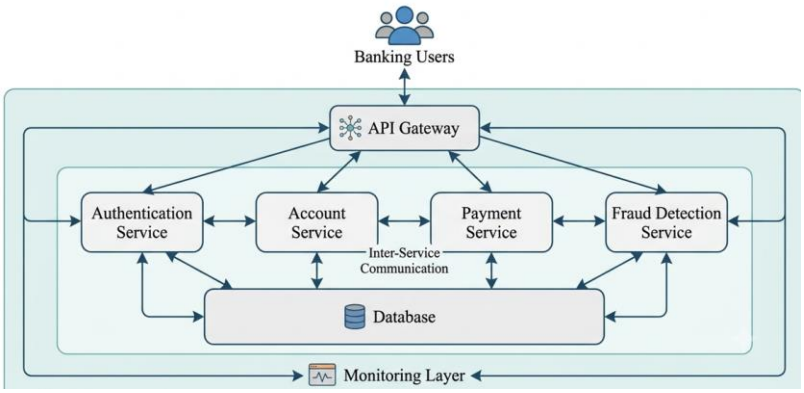


Figure 1. Banking Microservice Anomaly Detection Ecosystem

In this paper, it is important to note that they respond to this requirement and present an explainable AI framework to detect anomalies in banking microservices. The framework is a combination of behavioral monitoring, heterogeneous telemetry analysis and machine learning-based anomaly detection and explanation mechanisms to give transparent interpretation on anomalies detected. The offered approach will make the process of detection, human understanding, and trust towards AI-facilitated surveillance more visible by tying the outcomes of anomaly decisions to behavioral evidence, which can be explained to humans. The framework offers a platform of reliable and scalable anomaly analysis within complex banking setting, where quick detection and significant interpretation of abnormal behavior are also vital.

2. Related Work

Anomaly detection has emerged as an important topic of study in the identification of abnormal, suspicious or unexpected behavior in complicated digital systems. The issue is especially vexing in financial settings, since patterns of transactions are very dynamic, data distributions are frequently skewed, and anomalies can be few in number than regular transactions. Ahmed et al. gave a general overview of the methods of anomaly detection in financial situation and discussed statistical, classification, clustering and other machine learning techniques of detecting abnormal financial situation [1]. Their work clarified that various detection methods need to be selected with respect to the traits of financial information and highlighted the point that it is challenging since the representative real-life databases might hardly exist. The observations shape the need of adaptive anomaly detection controls to aid heterogeneous and ceaselessly changing financial environments.

With the development of the deep learning, a representation-learning approach has gained popularity in learning anomalies. Aguilar et al. proposed an interpretable auto encoder that tries to add to the interpretability of the learned auto encoder a mechanism that runs on decision-trees and autoencoder learning [2]. Their work is particularly timely as using conventional autoencoders can perhaps be effective in executing a successful detection with providing limited information why an observation has been identified as an anomaly. The technique attempts to bridging that void between human perception and detection limits by incorporating interpretable framework. The technique is an excellent place to start with explainable anomaly detection in banking systems which theoretically, an operator ought to understand behavioral characteristics that resulted in an alert.

Another important direction is the post-hoc explanation methods of the complex models of anomaly detection. Antwarg et al. investigated the use of Shapley Additive Explanations (SHAP) for explaining anomalies detected by autoencoders [3]. Their approach accepts the attributes which most successfully breed reconstruction errors and by consequence an elucidation of individual aberrant observations. The importance of this work is that it is capable of gaining an explanation at instance level without introducing significant changes to the model with which a detection is to be obtained. The parallel description in the framework of banking microservice may prove to be generally useful to help analysts understand what attributes of a transaction, service-level metrics, or behaviors attributes lead to an unusual event.

Arrieta et al. [4] were able to explore a systematic broadening of the establishments of explainable artificial intelligence through a discussive method. They have made important ideas, taxonomies, and concerns regarding explainability, interpretability, transparency, and responsible AI through their work. The analysis showed that the generation of explanations can be based on various methods depending on the intrinsic interpretability of the model or when the post-hoc explanation mechanism is needed. These thoughts find more applicability in the monetary sphere of cybersecurity as computerized anomaly detectors can aid in decision making regarding fraud detection and security oversight and response to an incident. Consequently, the explainability is not a mere visualization aspect but an important part of the dependable monitoring with the help of AI.

The rule extraction was studied in another research carried out by Barbado et al. to improve the explainability of a model that was created to detect anomalies by the One-Class SVM as an application setting [5]. They were interested in the rules to depict the behavior that a generally hard to understand model of detecting anomaly was trained on. Other properties that assessed the explanations were categorized under comprehensibility, representativeness as well as stability and diversity. The approach has weaknesses in terms of operational specifics: It is not always easy to interpret a concise decision rule instead of complex output of numerical models. However, it is often difficult to maintain explanations made in terms of rules because system behaviour changes with rapid rates or aberrant behaviour due to interactions between a large number of microservices.

Carletti et al. [6] proposed an interpretable feature importance extraction methodology to obtain an anomaly detector, based on an Isolation Forest, named DiFFI. The approach aims at isolating the value of single attributes towards anomaly decisions, thus enhancing the meaning of the tree-based detection models. This paper demonstrates that interpretable feature-level data can be fed into the execution of anomaly identification algorithms rather than as something to be viewed as an auxiliary post-processing activity only. These feature-level explanations might be applicable to banking applications where analysts might want to know whether volume of transactions, response time, resource usage, authentication, or other signs was part of an alert.

Boehmer and Rainerle-Ma explored the issue of anomaly detection in a process-based viewpoint through the mining of association rules based on dynamic process run time behavior and exploration with this information to explain potential root causes [7]. This is particularly applicable to the field of distributed systems because they may be unable to explain the occurrence of abnormal behavior with respect to a single observation and instead may event to the unexpected turn of events or association. Another framework of anomaly detecting and explaining convolutional autoencoders based on data storage systems was also suggested by Assaf et al. [8]. They have demonstrated that their work can be employed in the infrastructure monitoring context to implement deep learning on an anomaly detection context with explanation mechanisms.

The reasons why neural-network-based anomaly detection needed to be explained to the users via explainable deep neural networks have been described in Amarasinghe et al. papers [9]. The first to offer deep-learning anomaly detection to system logs with DeepLog and demonstrate how sequential system log-data can be modeled to detect and diagnose abnormal system behavior is Du et al. [10]. These works are extremely applicable to the microservice architectures as logs and consecutive service events will represent useful evidence on the basis of deviant interactions. Yet, the log-centric analysis might not be sufficient to sufficiently capture the anomalies that can engage the same transactions, metrics, traces and service dependencies.

The paper by Kauffmann et al. discusses how one-class models can be described using deep Taylor decomposition [11]. Through their work, they were able to show how feature contributions could be estimated as to why an observation was said to be anomalous. Another significant issue that Li et al. solved is the suggestion of a multivariate time-series anomaly detection

and interpretive strategy that used hierarchical inter-metric and time-based relationships [12]. This is particularly notable to banking microservices, where a deviation in behavior can be identified over time based on time dependencies between a collection of service-level metrics rather than through an independent deviation modeled by a single metric.

Lipton discussed the conceptual problem of defining interpretability and stressed that interpretability is multi-dimensional and it should not be treated as a monolithic one universally defined [13]. Lundberg and Lee introduced SHAP as providing a unified language to interpret the findings of the model predictions [14], which provides a strong basis upon which a feature-attribution-based explanation can be established. Hwang and Lee also illustrated the use of explainability to anomaly detection by using their E-SFD framework in sensor fault detection in industrial control systems [15]. Their input shows that the process of detection of anomalies and explainability can be integrated to offer support on more understandable decisions in safety- and security-sensitive environments.

Altogether, the analyzed papers suggest that the process of explainable anomaly detection, feature attribution, interpretable models, log-based diagnosis, and understanding of multivariate behaviors are now significantly improved. However, what is clearly lacking is how to bring these abilities into the banking microservice design in which heterogeneous transaction data, transaction logs, metrics, traces, authentication events and inter-service dependencies are to be analysed together. The available literature usually discusses individual-data sources, or discrete anomaly detection algorithms, or individual domains of application. As such, a unified XAI system is necessary to identify anomalies that exist across distributed banking services as well as the ability to offer demonstrable evidence on the nature of features, interactions among services and patterns in behavior that caused each anomaly to be identified. Such an integration can not only improve the anomaly detection, but also the transparency of the diagnostic and can give an analyst confidence in AI-enhanced banking cybersecurity.

3. Explainable Behavioral Intelligence Framework for Banking Microservice Anomaly Detection

The framework taken into consideration will provide a logical method of anomalous behavior detection, interpretation, and prioritization within microservices of a distributed banking. Unlike more traditional anomaly detection systems, in which results are primarily anomaly scores, or binary alerts, the proposed system comprises heterogeneous telemetry analysis with machine learning-based detection and interpretability systems. It would provide banking environment that would allow integration of transaction services, authentication services, account management, payment gateways, fraud monitoring services, and supporting infrastructure into a network of interdependent microservices using the framework. Within its overall architecture, it has oriented itself towards four mutually supporting objectives i.e. multi-source behavioral visibility, accurate detection of anomaly, generable understandable decisions and actionable security insight.

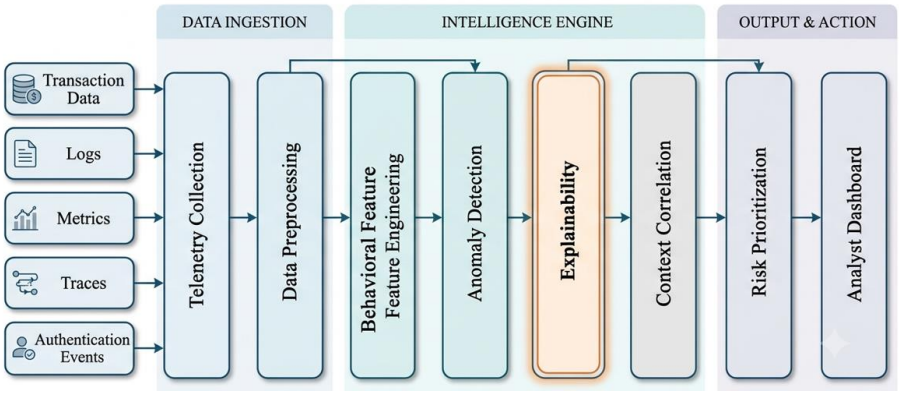


Figure 2. Explainable Behavioral Intelligence Framework

3.1. Framework Architecture

Its design is based on a layered system that consists of data gathering, telemetry preprocessing, behavioral feature engineering, anomaly detection, explainability, and analyst-friendly decision support layers. Distributed banking microservices send data in an iterative manner to the surveillance structure where the heterogeneous observations are described using one-representation of system and transaction behavior. With the structure, individual elements will be logically independent and will correlate information across different services in case of anomaly detection.

The framework is the first layer that collects the heterogeneous telemetry availed by the banking microservices. Application logs, transaction logs, service metrics, distributed traces, authentication events, API requests, response times, resource consumption indicators, and the inter-service communication data are sources of input. Transaction-level data might contain frequency of transaction, after reason data, transaction status, geographical or time-based results both geographical or time-based variations and authorization-based data. Observations on the service level can be CPU utilization, memory

consumption, request rates, error rates, latency, failed requests and service availability. This make-up of a group of sources expands the spectrum of behavioral visibility compared to streams of monitors that rely on a single stream.

3.2. Multi-Source Telemetry Processing

The telemetry gathered is sent through a normalization and preprocessing component. Banking microservices can generate the information in different formats and with different frequencies, as such, just by integrating the raw observations crudely, one can result in inconsistency and redundancy of the information. The preprocess will involve performing a clean and normalization, time stamp alignment, removal of duplicates, missing values and category encoding. Temporal data is applied to coordinate the events the etiology of which belong to multiple services in order to analyze related activities as a sequence of behaviors.

Observations are then correlated by means of a correlation mechanism with the relevant services, transactions, sessions and request chains. Of special concern in microservice landscapes is that, when an abnormal event happens in one service it can be caused or spread to another service. For example, an unusual spike of the payment-service latency may be attributed to the break down of authentication processing, slowness of the responses of the database, or even the failure of API traffic. With these relationships, the framework can construct more context on the behavior, and the decision on anomaly will then be constructed.

3.3. Behavioral Feature Representation

The framework processes the Raw telemetry after which behavioral attribute is produced and that can be utilized in machine learning based analysis. The feature generation occurs at various levels that include transaction level, service level, temporal, interaction level and system level features. The transaction-level attributes describe the frequency of transaction anomalies, distribution of amount, attributes of authentication and result of a transaction. Service-level acknowledgements comprise the latency, error rate, volumes, resource usage and the availability. Temporal features characterise time-related dynamics of the behaviour, and features at the interaction level characterise how it relies on and how patterns of communication between microservices.

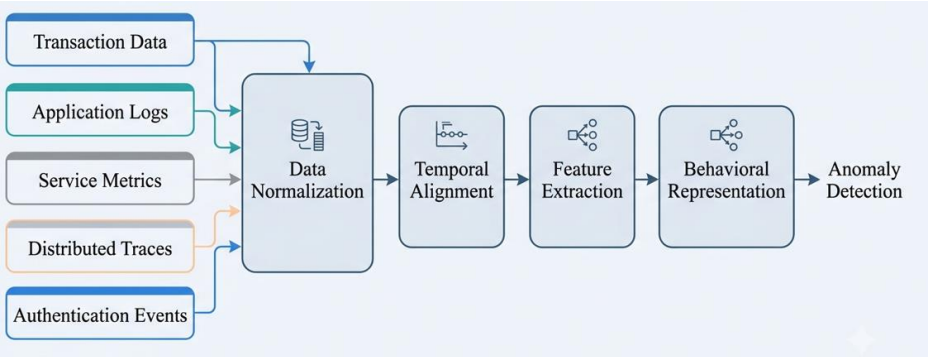


Figure 3. Multi-Source Telemetry and Behavioral Correlation Pipeline

The framework contains the individual and aggregate behavioural representations. The detection mechanism requires information about individual representations that allow the detection of particular deviantiations, and aggregated representations, which provide information about system conditions in general. This multi-level representation is significant since an observation that looks abnormal in one instance might be normal at a given circumstance in operations. Provision of contextual information about behavioral patterns therefore helps to reduce the number of false warnings as a result of valid workload variations.

3.4. Machine Learning-Based Anomaly Identification

The processed representation of the behavioral representation is fed into an anomaly detector layer. This aspect can have machine learning that is supervised, semi-supervised or unsupervised depending on labeled banking data and its availability and quality. It will be model and algorithm agnostic so the algorithms such as autoencoders, Isolation Forest, One-Class SVM, and clustering algorithms or any other suitable anomaly detection models can be added.

The detection layer makes an image of the expected behaviour and matches the received observations to the learnt behavioural patterns. The score on the measurement of each observation of the way it is behaving towards it is an anomaly. Rather than when this score is the final answer, the framework forwards the outcome of the detection and the feature information to explainability layer. The decoupling allows the anomaly detector model to focus on the performance in detector and a specialized explanation system, such as interpretable evidence, of the model complex behaviour.

3.5. Explainability and Evidence Generation

The key element of the offered framework is the explainability layer. It aims to find out the reason behind an observation being categorized as anomalous and the factors which that had the greatest influence over the decision. Depending on the model used to detect anomalies, feature-attribution procedures like SHAP, local explanation procedures, feature-importance mechanisms, rule-based methods, or model-specific explanation procedures can be combined to accomplish this.

All the anomalies located via the framework will result in an account that features the most important behavioral characteristics and their applications in making a decision regarding the anomaly. As an example, an alert related to a payment microservice can be a high rate of transactions, an abnormal delay in responding, failed authentication, or an unfamiliar interaction between the service as the reasons of notable extent. The interpretation then converts an anomaly score which is generated by a machine into evidence to be examined by a security analyst.

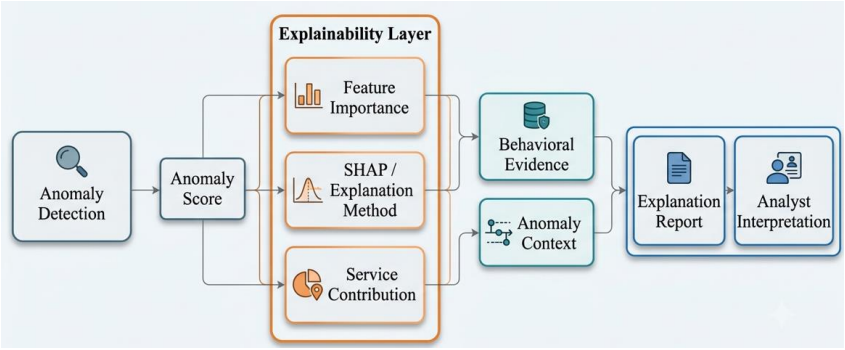


Figure 4. Explainable Anomaly Analysis and Evidence Generation

The structure further distinctly divides the feature-level and service-level descriptions. The feature level explanations figure out the specific attributes in relation to the anomaly, whereas very often service level explanations figure out the microservices or service interactions involved in the abnormal behavior. This difference comes in handy especially in distributed banking systems since the root cause of an anomaly might not be in the service that triggers the alert.

3.6. Anomaly Contextualization and Risk Prioritization

Once the generation of explanations, the framework runs an analysis of the context to find out the operational importance of each anomaly. This could be ascribed to numerous aspects that involve severity of anomalies, service impacted, sensitivity of the transaction, numerous incidences, and the similar margin of the service and the margin of deviation. Alerts can then be slotted into various levels of priorities to enable the analysts concentrate on the events which could be of significance.

The framework also links the related anomalies to reduce the alert fragmentation. Named products can be a collection of lower-level notification issued during authentication, payment, and account services that is one bigger behavior incident. Correlation therefore enables the system to develop a unifying context of anomalies, rather than displaying event taking place in isolation. In banking scenarios, where the patterns of distributed attacks or failure may spread a number of interconnected services, it is important.

3.7. Analyst-Oriented Output and Feedback

The final level displays the detected anomaly, the level of the anomaly, the set of services, the features, the behaviour and description of the anomaly to the security or operations analysts. The framework does not assign a specific score on anomaly, but provides a logical presentation of evidences shown in the detection. The analysts are therefore in a position to be aware of the cause of the alert and use the evidence to facilitate investigations and response.

Analyst decisions, such as confirmation, rejection or reclassifying of alerts can be optional and documented as part of the feedback mechanism. This feedback can then be optimised to an optimal detection threshold or feature representation, or model parameters. This forms a cycle of continuous improvement where the knowledge about operations can be added to engineer out improvement in the seamless nature of anomaly in an increasing manner.

3.8. Overall Framework Workflow

The entire workflow will commence with continuous collection of mixed banking telemetry and then proceed to preprocess it, time matching, behavioral feature soaking, identifying anomalies, generating explanations, context, and correlation, and prioritizing risk. The final product is a statement of interpretation of the anomaly that consists of the behavior identified, the context of service or transaction affecting it, the characteristics considered to cause it, and the level of impact it has had.

The primary input of the framework therefore is in the integration of detection and explanation of different dimensions of behaviour of banking microservices. The given architecture places interpretability at the stage of anomaly-analysis, in contrast to the visualization of explainability as a layer. This architecture offers a systematic basis to credible AI-aided supervision in distributed banking settings, where suspicious activity cannot be correctly recognized merely, but analysts additionally need interpretable evidence to scrutinize and confirm suspicious activity.

4. Empirical Validation and Evidence-Based Assessment of the Proposed Framework

The presented explainable behavioral intelligence framework can be assessed through an empirical mindset in the areas of detection capability, explainability, behavior coverage as well as utility of its functions. Rather than evaluating another common experimental comparison of algorithms, the evaluation attempts to deal with the analysis of whether the framework provides a rational process of the identification of abnormal behavior and subsequent material conclusions over the diversified banking microservices. Evaluation is considering the correlation of the identified anomalies, contributing characteristics of behavior, services affected and resulting evidence towards the analyst.

4.1. Empirical Assessment Dimensions

The framework is tested on the four complementary dimensions, i.e., anomaly detection, quality of the explanation, behavioural coverage and operational relevance. Anomaly detection pays attention to whether the framework is able to differentiate between deviations and the normal patterns of behavior. The quality of explanation measures the extent to which observed anomalies can be attributed to explicable contributive factors, and interactions among services. Behavioral coverage determines how well the framework is able to assimilate the information of transaction, service, time, and inter-service. Relevance in operations are to do with the fact whether the obtained information will assist the analysts in investigating and prioritizing the valuable events.

The importance of such multidimensional assessment is that the conceptual framework that has been tailored to explainable detection of bank anomalies cannot be assessed on the basis of the outcomes of detection. A detector mechanism is only of practical little use when a very precise one can be brought to bear, and the security and operations people are unable to interpret or place decisions in any kind of context.

4.2. Anomaly Detection Capability

The empirical assessment considers the ability of the framework in identifying deviations when there is a range of behavior. The observations of interest on the transaction level would involve atypical transaction frequency, atypical transaction properties, recurring failures and atypical pattern of authentication. Some of the indicators at the service level that can help in detecting the anomalies include increased latency, increase in error rates, greater volume of request compared with normal and abnormal use of resources.

The multi-source representation in the framework enables investigation of the anomalies with multi-behavioral evidence, but not individual indicators. This can be particularly helpful in the context of microservices because a single change in metrics can be a well-purpose workload change, but when multiple indicators are correlated, it can be further arguably that something abnormal is being experienced. Consequently, the framework supports the anomaly detection basing on the situation in comparison to relying on the single threshold violation.

4.3. Explainability and Evidence Quality

One cardinal point of the empirical evaluation is the utility of explanations at anomalies that are identified. Each anomaly is associated with a set of powerful features and where the anomaly is applicable a set of affected microservices or service interactions. The explanation mechanism is assessed assessing the clarity, relevance, consistency and traceability.

Clarity questions the interpretability of the explanation of the model of machine learning by the analysts without references to the internal mechanics of the machine learning model. Relevance is what determines relevance of identified features to the detected anomaly or not. Consistency examines the variations of some causes of behavior in terms of their similarities. Traceability assesses the ability to relate an explanation to an original transaction, service or evidence of telemetry.

Through this evidence-based method the products of assessing the simple anomaly score are converted into an analytical record which can be read. As an example, an issue with a payment service can be justified by logs that there is increased request rate, a longer latency, multiple authentication failures and an unusual interaction with a different service.

4.4. Behavioral Coverage and Cross-Service Analysis

The framework is also assessed according to its capacity to portray heterogeneous and interrelated behavior in the banking. Coverage is dealt in transactional, operational, temporal and inter-service. The transactional coverage is about customer and payment behavior as compared to operational coverage about service health and infrastructure indicators.

Temporal coverage takes into account the change in behavior over time and inter-service coverage takes into account the interrelationships between the distributed components.

This test is especially significant as banking aberrations might not be caused by one of the elements. A series of authentication, account, payment, and notification services can be considered a suspicious event. Outlining the connections among observations against the backdrop of all these parts aids the framework in providing a broader envelope of action and reduce the likelihood of analyzing single cases.

4.5. Operational Utility

Final evaluation element deals with the value of structures output to security and operations teams. The framework can be considered successful in the event that the analyst can then move to evidence-based interpretation without complete understanding of the machine learning model where the analyst must initially recognize that there was an anomaly in the machine learning model. Formatted reports that have indicators of the severity of identified anomalies, the services that have been affected, the features that contribute to an anomaly, the time context, and events are useful to perform a faster investigation and put the information in a higher priority.

5. Discussion

The given explainable behavioral intelligence framework is the reaction to a great shortcoming of the traditional anomaly detection in the banking microservice environment failure to give a sense of meaningful explanation of anomalies that are detected. Conventional methods usually have a tendency to detect occurrences of anomalies in normal behavior but in the case of banking activities, an analyst must comprehend the underlying causes correlating with an alarm. The suggested framework incorporates heterogeneous telemetry, the representation of behavioral features, anomaly detection and explanation mechanisms, which provide a more detailed perspective on anomalous behavior.

One of the strengths of the framework is that it has a multi-source behavioral perspective. Banking micro services generate transactions, logs and metrics, traces, authentication and records of interservice communications in real time. The combination of these sources may give more relevant contextual evidence than analysis of single indicators. This is particularly the case with distributed anomalies whereby abnormal behaviour is not caused by one instance but rather as a consequence of interaction of two or more services.

The explainability layer continues to enhance pragmatic essence of the framework. Explanations at the feature-level and at the service-level can aid analysts to comprehend the behavioral features that may have led to an anomaly and what services might have played a role. This can reduce the interpretational prices of complex machine learning models, and can keep the exploration of incidents more prudent. This approach may also be matched with the research on XAI that is already available and aims at transparency and understanding of a human being in making decisions with the help of AI.

Yet, there are still a number of challenges. The quality of the explanation depends on the selection of detecting model, representation of features, excellence of the telemetry, and the technique applied. The alteration of legitimate behavioral patterns that could be akin to anomalies may also occur under extremely dynamic bank situations that may increase risks of false alerts. In addition, the descriptions must be short enough and always available to be useful during real-time security. Privacy, data management, scaling and integration within the existing banking monitoring infrastructure should be also taken into consideration.

Altogether, the framework is a systematic basis of stable anomaly analysis of microservices in banking. The most important of its inputs will not be that it will do a better job in detecting the anomaly but that it will be able to couple the detection with contextualization and explanation in one analytical step. Analyses of the quality of explaine domain specific explain quality and analyst feedback are further analysed in the future, as well as adaptive behavioural baselines and massive scale validation on realistic banking telemetry.

6. Conclusion

One trend that appears to be growing in banking is the adoption of microservice architectures that poses new challenges to monitor, detect anomalies, and analyze cybersecurity. The scattered and highly networked nature of banking services generates diverse and riddled accounts of transactions, logs, measurements, traces, authentication events and connections on services, which are difficult to detect abnormal behavior using solitary methodology of monitoring. In addition, traditional machine learning-based anomaly detection is able to offer effective detection ability and a little information about why each individual decision has been made. That disadvantage may lower the degree of trust that the analyst has and made the discovery of potentially valuable events a problem.

The present paper has given an elucidable model of behavioral intelligence explanatory behavior of detecting anomalies in banking micro services. The architecture integrates multi-source telemetry collection, pre-processing, modeling of behavioral

characteristics, machine learning to identify anomalies, accountability, correlation and hierarchy of risks in context. Its chief objective is to correlate the anomaly detection and evidence that can be interpreted to enable the analysts to not only recognize the fact that an anomaly occurred in an observation but the behavioral features and nature of service relationship that resulted in the decision.

The practical evaluation shows the theoretical and practical usefulness of unification of heterogeneous behavioral messages and explicable anomalies analysis. Temporal level, inter service indicators, the transaction level, and service level provide various perspectives to determining the deviation to the expected behaviour. The combination of the service-level and feature-level explanations also provide assistance in the clear explanation of the observed anomalies and can guide the security and operation teams to investigate the alerts in a more productive way.

The trustworthy AI-assisted monitoring within distributed financial setting can also be based on the proposed framework. It is worth noting, though, that to put it into practice it must consider the quality of data, differences in behavior, how to treat false alerts, the consistency of explanations, privacy, scalability and compatibility with an existing banking infrastructure. The framework can be extended to future research in adaptive behavioral foundations, analyst-feedback schemes, state of the art causal explications, real-time cross-service association, and validated by large-scale empirical banking telemetry. Overall, the proposed solution offers a logical means of carrying out clear-cut, situational, and analystic-driven anomaly identification, which complements the aim of explainable AI to safeguard complex banking microservice infrastructures.

References

[1] Aguilar, D. L., Medina Perez, M. A., Loyola-Gonzalez, O., Choo, K.-K. R., & Bucheli-Susarrey, E. (2023). Towards an interpretable autoencoder: A decision tree-based autoencoder and its application in anomaly detection. *IEEE Transactions on Dependable and Secure Computing, 20*(2), 1048–1049.

[2] Ahmed, M., Mahmood, A. N., & Islam, M. R. (2016). A survey of anomaly detection techniques in financial domain. *Future Generation Computer Systems, 55*, 278–288.

[3] Antwarg, L., Miller, R. M., Shapira, B., & Rokach, L. (2021). Explaining anomalies detected by autoencoders using Shapley additive explanations. *Expert Systems with Applications, 186*, 115736.

[4] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Raja, C., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115.

[5] Barbado, A., Corcho, Ó., & Benjamins, R. (2022). Rule extraction in unsupervised anomaly detection for model explainability: Application to One-Class SVM. *Expert Systems with Applications, 189*, 116100.

[6] Carletti, M., Terzi, M., & Susto, G. A. (2023). Interpretable anomaly detection with DiFFI: Depth-based feature importance of isolation forest. *Engineering Applications of Artificial Intelligence, 119*, 105730.

[7] Böhmer, K., & Rinderle-Ma, S. (2020). Mining association rules for anomaly detection in dynamic process runtime behavior and explaining the root cause to users. *Information Systems, 90*, 101438.

[8] Assaf, R., Giurgiu, I., Pfefferle, J., Monney, S., Pozidis, H., & Schumann, A. (2021). An anomaly detection and explainability framework using convolutional autoencoders for data storage systems. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI)* (pp. 5228–5230).

[9] Amarasinghe, K., Kenney, K., & Manic, M. (2018). Toward explainable deep neural network based anomaly detection. In *Proceedings of the 11th International Conference on Human System Interaction (HSI)* (pp. 311–317).

[10] Du, M., Li, F., Zheng, G., & Srikumar, V. (2017). DeepLog: Anomaly detection and diagnosis from system logs through deep learning. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS)* (pp. 1285–1298).

[11] Kauffmann, J., Müller, K.-R., & Montavon, G. (2020). Towards explaining anomalies: A deep Taylor decomposition of one-class models. *Pattern Recognition, 101*, 107198.

[12] Li, Z., Zhao, Y., Han, J., Su, Y., Jiao, R., Wen, X., & Pei, D. (2021). Multivariate time series anomaly detection and interpretation using hierarchical inter-metric and temporal embedding. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 3220–3230).

[13] Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *ACM Queue, 16*(3), 31–57.

[14] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (pp. 4768–4777).

[15] Hwang, C. H., & Lee, T. (2021). E-SFD: Explainable sensor fault detection in the ICS anomaly detection system. *IEEE Access, 9*, 140470–140486.