UNIVERSITY OF PATRAS
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

DEPARTMENT of electronics and computers (PC)
INTERACTIVE TECHNOLOGIES LABORATORY

Study of a real time voice phishing detection system with machine learning technology

DIPLOMA THESIS

IOANNIS SKARPETIS

SUPERVISOR: FEIDAS CHRISTOS

PATRAS – FEBROUARY 2024

University of Patras, Department of Electrical and Computer Engineering.

Ioannis Skarpetis

CERTIFICATION

It is certified that the Diploma Thesis titled

Study of a real time voice phishing detection system with machine learning technology

of the Department of Electrical and Computer Engineering student

IOANNIS SKARPETIS

Registration Number: 1066539

was presented publicly at the Department of Electrical and Computer Engineering at

……../……../……

and was examined by the following examining committee:

Name Surname, Title, Affiliation (supervisor)

Name Surname, Title, Affiliation (committee member)

Name Surname, Title, Affiliation (committee member)

The Supervisor	The Director of the Division
Feidas Christos Substitute Professor	Name Surname Title

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΕΠΙΒΛΕΠΟΝΤΟΣ
ΕΡΓΑΣΤΗΡΙΟ ΕΠΙΒΛΕΠΟΝΤΟΣ

ΤΙΤΛΟΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Δ Ι Π Λ Ω Μ Α Τ Ι Κ Η Ε Ρ Γ Α Σ Ι Α

ΟΝΟΜΑΤΕΠΩΝΥΜΟ ΦΟΙΤΗΤΗ

ΕΠΙΒΛΕΠΩΝ: ΟΝΟΜΑΤΕΠΩΝΥΜΟ ΕΠΙΒΛΕΠΟΝΤΟΣ

ΠΑΤΡΑ - ΜΗΝΑΣ ΕΤΟΣ

ΠΙΣΤΟΠΟΙΗΣΗ

Πιστοποιείται ότι η Διπλωματική Εργασία με τίτλο

ΤΙΤΛΟΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

του/της φοιτητή/τριας του Τμήματος Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών

ΟΝΟΜΑ ΕΠΩΝΥΜΟ ΤΟΥ **ΠΑΤΡΩΝΥΜΟ**

Αριθμός Μητρώου: XXXXXXX

Παρουσιάστηκε δημόσια στο Τμήμα Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών στις

……../……../…….

και εξετάστηκε από την ακόλουθη εξεταστική επιτροπή:

Όνομα Επώνυμο, Βαθμίδα, Τμήμα (επιβλέπων)

Όνομα Επώνυμο, Βαθμίδα, Τμήμα (μέλος επιτροπής)

Όνομα Επώνυμο, Βαθμίδα, Τμήμα (μέλος επιτροπής)

Ο/Η Επιβλέπων/ουσα	Ο/Η Διευθυντής/τρια του Τομέα
Ονοματεπώνυμο Βαθμίδα	Ονοματεπώνυμο Βαθμίδα

PREFACE

The Preface is optional. Here the author puts any information that is not directly related to the scientific content of the Diploma Thesis, such as acknowledgements, etc.

The structure, form and extent of the preface are at the student's discretion, unless otherwise specified by the supervisor.

ABSTRACT

DIPLOMA THESIS TITLE

STUDENT NAME, SURNAME: **SUPERVISOR NAME, SURNAME:**

The objectives, methods, procedures, experiments and results of the Diploma Thesis are briefly described.

The structure, format and scope of the abstract are at the student's discretion, unless otherwise specified by the supervisor.

ΕΚΤΕΤΑΜΕΝΗ ΕΛΛΗΝΙΚΗ ΠΕΡΙΛΗΨΗ

ΤΙΤΛΟΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

ΟΝΟΜΑΤΕΠΩΝΥΜΟ ΦΟΙΤΗΤΗ: **ΟΝΟΜΑΤΕΠΩΝΥΜΟ ΕΠΙΒΛΕΠΟΝΤΟΣ:**

Σύμφωνα με τον κανονισμό Διπλωματικών Εργασιών, η συγγραφή της Διπλωματικής Εργασίας στην Αγγλική γλώσσα θα συνοδεύεται απαραιτήτως από εκτενή περίληψη στα Ελληνικά, τύπου επιστημονικής εργασίας (paper).

Οι υπόλοιπες λεπτομέρειες σχετικά με τη δομή και τη μορφή της Εκτεταμένης Ελληνικής Περίληψης είναι στη διακριτική ευχέρεια του φοιτητή, εκτός αν προδιαγράψει διαφορετικά ο επιβλέπων.

Table Of Contents

Table Of Figures

1. Introduction

Since the dawn of the 21st century, humanity has embarked on a remarkable journey of technological progression. As our society becomes ever more reliant on technology for its daily functions, we've seen a corresponding escalation in the complexity of cybercriminal activities. The advancement of technology has not only streamlined our daily tasks but has simultaneously opened numerous vulnerabilities. These gaps are frequently exploited by cybercriminals, targeting those who struggle to stay abreast of the latest technological developments.

Voice phishing, or 'vishing', is a notable vulnerability within the cybersecurity landscape, manifesting as a sophisticated form of phishing conducted via phone calls. While phishing as a broader concept is not novel, having roots that extend well before the internet age, vishing represents its adaptation to the telecommunication realm. Phishing, at its core, is about manipulating individuals to reveal confidential information under misleading pretenses, a strategy that has seen various incarnations over many decades. The term "phishing" gained widespread recognition during the 1990s alongside the internet's growth. Originally, it denoted email-based frauds wherein culprits masqueraded as legitimate entities to coax out sensitive details like login credentials and financial data from unsuspecting victims.

Vishing exploits the same fundamental principle of deception but leverages the immediacy and personal touch of voice communication. Here, fraudsters make use of phone calls to impersonate legitimate organizations or authorities, aiming to instill a sense of urgency or fear in victims. This tactic often involves compelling narratives designed to prompt quick action, leading individuals to inadvertently disclose personal or financial information.

Efforts to bolster security against such attacks have been ongoing, yet the advent of caller ID spoofing technology has notably heightened the efficacy of vishing attacks. This technology enables attackers to disguise their actual identity, presenting a significant obstacle in identifying and preventing these fraudulent calls. Consequently, the dynamic nature and sophistication of vishing have established it as a significant and continually evolving threat within the cybersecurity domain. This scenario highlights the critical necessity for advanced and adaptable security measures capable of effectively neutralizing the ever-changing tactics of cybercriminals. Consequently, the focus has shifted towards innovative methods for detecting vishing attacks, particularly by analyzing the conversation content between the two parties in real-time. This approach represents a significant step forward in proactive cybersecurity, aiming to identify and mitigate vishing threats as they occur.

2. Analysis of the problem under research

Before delving into the analysis of the research problem addressed in this dissertation, it is essential to first define the central focus, which is Phone Call Scam Detection. This foundational concept forms the basis of our investigation and discussion, setting the stage

for a thorough exploration of the strategies and technologies employed in identifying and mitigating fraudulent phone communications.

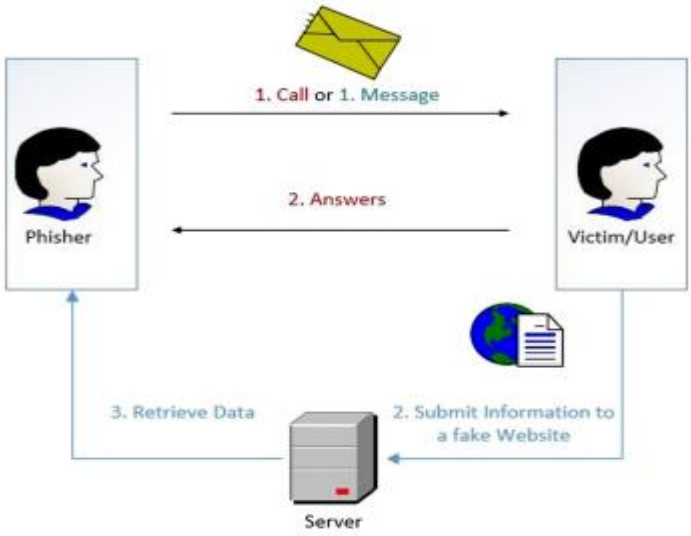

Fig 1: Mobile Phone Phishing Attack

2.1 Vishing attacks and methods

To effectively develop security measures against vishing attacks, it is crucial to first comprehend the vast array of techniques employed by Vishers. Understanding their strategies and tactics provides the foundational knowledge necessary to design and implement robust countermeasures. The techniques implemented by Vishers can be broadly categorized into two main phases [4]:

Before the vishing attack:

- **Automated information gathering:** Vishers frequently use automated tools to 'scrape' large amounts of user data. This bulk collection can provide a broad base from which specific targets are identified and proceed to be evaluated as potential victims.

- **Victim Evaluation:** Vishing scams often begin with phishing attacks via automated emails, texts, or social media messages, posing as legitimate entities to gather contact information. Those who respond to these initial contacts are more likely to fall for subsequent vishing calls. Some Vishers skip initial contact and directly call numerous potential victims using automated services, aiming to reach susceptible individuals.

- **Data accumulation via research**: A key element of preparation for Vishers involves extensive research and collection of personal and private information about their potential victims. This phase often involves the use of social media platforms to gather details such as phone numbers, employment data, and location information, which, while not overly sensitive, can lay the groundwork for a more targeted approach.

- **Direct Methods:** In some instances, Vishers may resort to direct, personal methods of information gathering, like sifting through an individual's trash also known as

Dumpster Diving. This approach is more physical and hands-on where the perpetrators physically search through garbage to find discarded documents, letters, bills, and other materials that contain personal or sensitive information. Even seemingly harmless details such as phone lists, calendars, or organizational charts can be instrumental for attackers when planning strategy. [5]

- **Caller ID Spoofing:** An essential element in the success of vishing attacks is the ability to avoid detection by the victim. To achieve this, Vishers use a tactic known as "Caller ID Spoofing." This technique involves altering the phone number that appears on the recipient's caller ID. It proves highly effective because many individuals depend on the caller ID to verify the legitimacy of the caller. [6]

During the vishing attack:

- **Impersonation and Authority:** Vishers often pretend to be from reputable organizations, such as banks, government agencies, tech support, law enforcement agencies or members of a financial institution. They use authoritative tones and language to gain the victim's trust and compliance. [8]

- **Confirmation Bias:** They use known information about the victim (gathered during the preparation phase) to build trust and credibility. For example, referencing a recent transaction or a known issue to make the call seem legitimate.

- **Creating urgency:** Vishers frequently employ the tactic of instilling a sense of immediate urgency. They may assert that urgent action is required to address issues like a supposed security breach in the victim's bank account or a pressing financial matter. The aim of the attacker is to prompt the victim to act quickly and impulsively, bypassing the usual steps of call verification and thoughtful consideration.

- **Emotional Manipulation:** Vishers adeptly exploit the emotions of their victims, wielding tactics that evoke fear, sympathy, or excitement to steer their responses. They might threaten legal action or warn of imminent financial loss to instill fear and a sense of urgency. Conversely, they can appeal to the victim's emotions through sympathetic stories or scenarios, creating a false sense of connection or trust. By manipulating emotions in these ways, Vishers aim to cloud the victim's judgment, making it easier to extract sensitive information or coerce them into specific actions that further the scam. [7]

- **Information Overload:** When individuals are quickly presented with an excessive amount of information, it can lead to sensory overload, impairing their ability to logically process and evaluate the arguments being made. In such scenarios, people are more inclined to accept the statements presented to them, as their capacity for critical thinking is overwhelmed by the sheer volume of information. This tactic increases the likelihood of the person acquiescing to the Visher's demands or assertions. [8]

- **Diversion Tactics**: Victims are often redirected to alternate communication channels or actions. This could involve being instructed to call a different number, visit a fraudulent website, or mail sensitive documents to a specific address. These tactics

serve to deepen the victim's involvement in the scam, making the deception appear more legitimate and complicating the victim's ability to discern the fraud.

- **Follow-Up and Persistence:** In vishing scams, Vishers often employ persistent follow-up calls, especially if the victim shows initial doubt or reluctance. This relentless approach aims to wear down the victim's defenses, increasing the likelihood of succumbing to the scam. Such persistence is a key tactic in gradually eroding skepticism and reinforcing the scam's credibility.

2.1.1 Typical Vishing Attack Examples

Vishers select from a diverse range of scenarios to deceive their victims. The subsequent examples will elaborate on the most frequently utilized themes in vishing scams. [2]

- **IRS and Immigration Scams:** Scammers impersonate IRS officials or immigration officers, threatening arrest or deportation if alleged debts aren't paid.

- **Romance Scams:** Fraudsters establish romantic connections, often through dating apps or phone calls, claiming to be past lovers in need of emergency funds.

- **Tech Support Scams**: Scammers pose as tech support, alleging urgent computer issues like viruses, gaining remote control of computers to access sensitive information or install malware.

- **Debt Relief and Credit Repair Scams:** Fraudsters offer debt relief or credit repair services for a fee, often worsening the victim's financial situation.

- **Business and Investment Scams:** Scammers pretend to be financial experts, persuading victims to invest money in fraudulent schemes.

- **Charity Scams:** Perpetrators pose as charity workers soliciting donations for fake causes, with funds going directly to the scammers.

- **Auto Warranty Scams:** Scammers make fake calls about car warranties, seeking personal information or payments for nonexistent warranty renewals.

- **Parcel Scams:** Aimed at immigrants, scammers claim a parcel is linked to a financial crime, posing as couriers and police to coerce money for a fake investigation.

- **Kidnapping Scams:** Scammers claim to have kidnapped a relative, using research or social engineering to extract ransom payments.

2.2 Current Approaches and Emerging Frontiers in Vishing Detection

It's evident that vishers possess an extensive arsenal of options, tactics, and scenarios, making the challenge of effectively detecting vishing attacks a complex one. Historically, methods like real-time detection of caller ID spoofing, with software like STIR/SHAKEN [2] and iVisher [6] and blacklisting phone numbers, represented a critical line of defense against such attacks. However, the integration of machine learning into cybersecurity presents an evolving frontier. These new tools offer enhanced capabilities to security professionals, equipping them to more effectively counteract the sophisticated strategies employed by

vishing scammers. This evolution in anti-vishing technology reflects the dynamic nature of the cybersecurity field, where constant innovation is essential to stay ahead of threats.

Efforts to integrate machine learning into vishing detection have primarily focused on analyzing metadata, both statistical and non-statistical, such as CALLING_NUMBER, CALLED_LOCATION, and CALL_DURATION [10]. However, the area of real-time detection that involves processing the actual speech content within a scam call, post-answer by the victim, remains a largely unexplored frontier in the field. This gap suggests potential for future advancements in vishing detection methodologies that could more directly address the nuances of in-call scam tactics. In this dissertation, we shift our focus to developing a system capable of real-time vishing detection. This system will leverage the actual conversational content of calls, utilizing machine learning algorithms. This approach represents a new direction in vishing detection, aiming to directly analyze the dynamics of scam calls as they occur.

2.3 Why Real-Time In-Call Detection Is Essential

Previous efforts to mitigate vishing attacks have predominantly focused on the pre-answer phase of the phone call. Pioneering research in this domain, such as the work of Jian Xing [10], [6], and Manh-Hung Tran [12], has been directed towards proactive strategies against vishing. The concept of real-time detection during a call is relatively novel in this field. Minyoung Lee and Eunil Park's [11] research was among the first to explore real-time vishing detection. Utilizing machine learning techniques, their study specifically addressed vishing attacks in the Korean language and yielded some promising results.

Why though, is real-time in-call detection essential? Real-time in-call vishing detection is mandatory for several reasons, which can be effectively communicated through bullet points:

- Limited Individual Defense: Once the phone call is answered, individuals primarily rely on their personal knowledge, emotional stability, and understanding to defend themselves against vishing attacks. This reliance is often insufficient as scammers are typically well-prepared, extensively researched, and skilled in manipulative tactics.

- Emotional Manipulation: As mentioned before, Vishers exploit emotional vulnerabilities, using tactics like fear, urgency, and sympathy, which can overwhelm the victim's rational decision-making. Real-time detection can provide a safeguard during these high-pressure situations, where individuals might not be able to think clearly.

- Information Discrepancy: Scammers often have more information about the victim than the victim has about the caller, creating an imbalance that can be exploited. Real-time detection systems can help level this playing field by providing additional context or warnings about the call.

- Sophistication of Scams: Vishing scams have evolved to be highly sophisticated, often bypassing traditional pre-answer detection methods. Real-time detection can adapt to these evolving tactics, offering a dynamic line of defense as the call progresses.

- Support for vulnerable populations: Certain groups, like the elderly or less tech-savvy individuals, may be more susceptible to vishing. Real-time detection can provide an additional layer of protection for these vulnerable populations.

Hence, it becomes clear that there is a pressing need for an enhanced defensive system. Such a system would bolster individual protection during the post-answer phase of phone calls, particularly in situations where proactive vishing detection systems fall short.

2.4 Addressing the Identified Problem

The objective of this dissertation is to explore the feasibility and effectiveness of a system that processes phone call conversations in real time for vishing detection. This proposed system will integrate technologies like speech-to-text conversion, machine learning algorithms, and streaming techniques to accomplish its goals. The research will focus on key questions:

- Is it possible to train a machine learning algorithm to detect vishing within a conversation as it happens?

- Can such an algorithm provide real-time predictions with a sufficiently low response time to effectively protect the victim?

This dissertation is structured to methodically address the research questions posed. Chapter 3 delves into the PySpark framework, which is pivotal for data preprocessing, model training, and enabling the streaming functionality of the system. Chapter 4 discusses the implementation of the Speech to Text module, with a particular focus on the integration of the Google Speech to Text API.

Chapters 5 and 6 are dedicated to dataset creation and preprocessing, detailing the methodologies and approaches used in preparing the data for analysis. Chapter 7 outlines the architecture of the machine learning algorithms deployed, covering aspects of their training and hyperparameter tuning.

Chapter 8 provides an overview of the system’s functionality as a cohesive unit. The testing procedures and their corresponding results are the focus of Chapter 9, offering insights into the system's performance and efficacy.

The dissertation concludes with Chapters 10, which reflect on the challenges encountered during system development and propose potential future enhancements. This final chapter also presents overall deductions and conclusions drawn from the research.

3. PySpark

Before discussing PySpark's role in our system, it's crucial to address Apache Spark. As the foundational framework for PySpark, Spark's advanced data processing capabilities and efficiency in managing large datasets are central to our research. Its powerful architecture, which enables fast and scalable data handling, sets the stage for understanding how PySpark enhances these functionalities to suit our specific application needs.

3.1 Apache Spark Fundamentals

Apache Spark is a powerful, open-source unified analytics engine designed for large-scale data processing and analytics. Its power lies in the ability to perform both batch and real-time data processing, making it a versatile tool in the field of data science and machine learning.

3.1.1 Spark Architecture

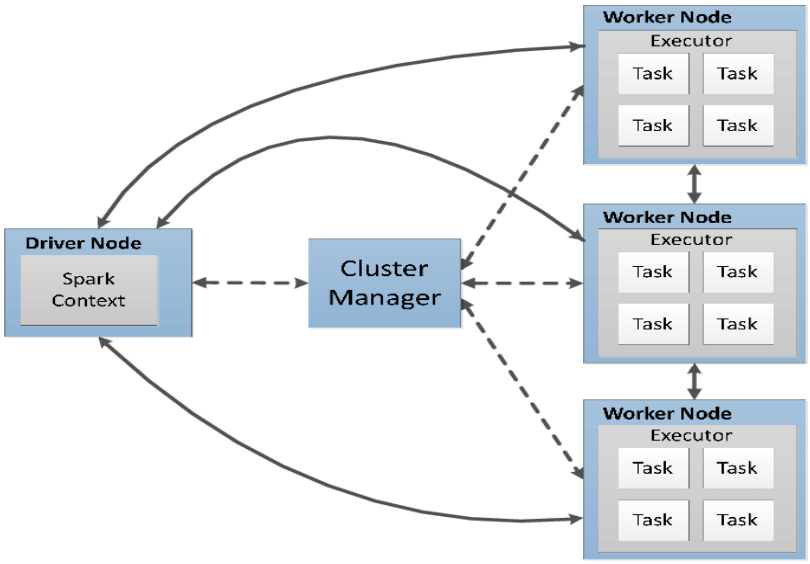

Fig 2: Spark Cluster Architecture

Spark operates as a distributed computing system; in this section we explore its architecture. The architecture is depicted in Figure 3.1 as a **Spark Cluster**, showcasing its ability to coordinate complex processing tasks across multiple machines for efficient data management and computation. At the center of the architecture lies the **Driver Node**, in which the Spark Context is housed. This central coordinator initiates and manages the lifecycle of various Spark jobs. It converts the user application into tasks that are then distributed across the cluster for execution. The **Cluster Manager** is the resource negotiator responsible for allocating resources among the various applications running on the cluster. It ensures optimal utilization of resources and manages the distribution of tasks to the worker nodes. The **Worker Nodes** represent the distributed computational units of the cluster. Each worker node has Executors, which are processes responsible for executing tasks assigned to them by the driver node. Executors run the tasks in multiple threads, which allows Spark to execute tasks in parallel, significantly improving performance. **Tasks** are the smallest units of work in Spark, and they carry out computation and data processing. They are executed by the executors to process the data. The distribution of tasks across executors allows Spark to handle large datasets efficiently [13].

3.1.2 Spark Components

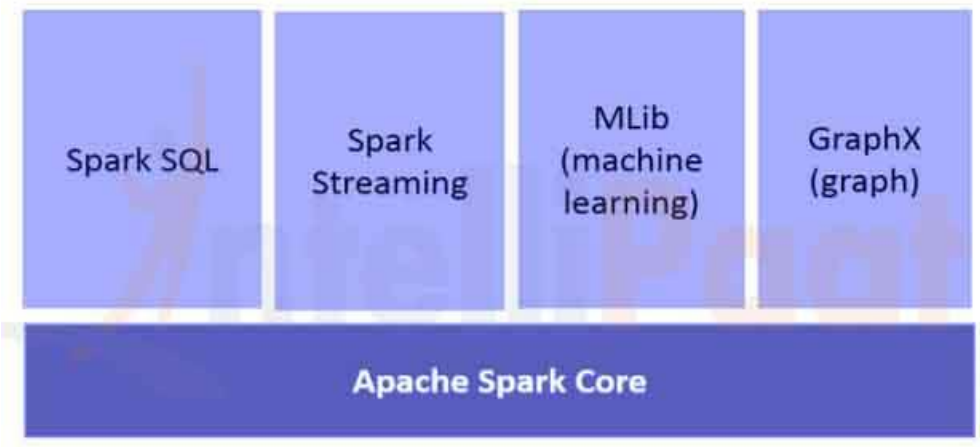

Fig 3: Spark Main Components

The Apache Spark system is composed of several core components, which include the Spark Core and a range of high-level libraries. These libraries cater to specific needs, such as MLlib for machine learning applications, GraphX for graph analytics, Spark Streaming for real-time data processing, and Spark SQL for handling structured data.

3.1.3 Spark Core

Spark Core is the foundational part of Spark, upon which all other functionalities are built. Spark Core provides a variety of core features for distributed task dispatching, scheduling, and basic I/O functionalities, which form the heart of the Spark engine.

To further enhance our comprehension of Spark Core's functionality, and consequently the overall operation of Spark, our examination will pivot to the Resilient Distributed Dataset (RDD).

3.1.3.1 Resilient Distributed Dataset

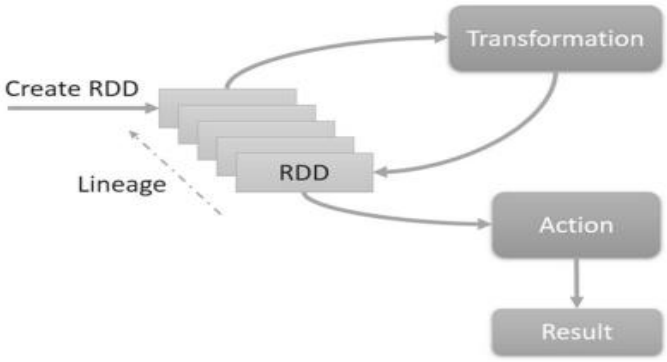

Fig 4: RDD lineage graph

The RDD is the foundational element of Spark, providing the framework for fault-tolerant data distribution and parallel processing that underpins the entire system. RDDs can be generated from external data sources or transformed from existing RDDs. As a key

characteristic, RDDs maintain a lineage graph that allows for the efficient recompilation of data in case of node failures, as shown in Figure 4, rather than relying on data replication. This facilitates robust in-memory data processing and supports a variety of computational models. Moreover, RDDs enable complex operations through transformations, which are operations that create a new RDD from an existing one (e.g. map, filter), and actions, which are operations that trigger computation and produce non-RDD values (e.g. reduce, collect). Both transformations and actions are executed across the cluster only when necessary, optimizing resource utilization and performance [13].

3.2 Spark SQL

4. Speech-To-Text module

5. Dataset Creation

6. Dataset Preprocessing

7. Model Architecture – Training – Tuning

8. Overall System Functionality

9. Experimental Results

10. Deductions, Limitations and Future Research

References

[1] Inside the intricate world of voice phishing (joins.com)

[2] Voice phishing - Wikipedia

[3] What Is a Vishing Attack? | Fortinet

[4] What is Vishing (Voice Phishing)? Examples You Need to Know (softwarelab.org)

[5] What is dumpster diving? (techtarget.com)

[6] Song, J., Kim, H., & Gkelias, A. (2014). iVisher: Real-Time Detection of Caller ID Spoofing. *ETRI Journal*, *36*(5), 865-875.

[7] Norris, G., & Brookes, A. (2021). Personality, emotion and individual differences in response to online fraud. *Personality and Individual Differences*, *169*, 109847.

[8] Jones, K. S., Armstrong, M. E., Tornblad, M. K., & Siami Namin, A. (2021). How social engineers use persuasion principles during vishing attacks. *Information & Computer Security*, *29*(2), 314-331.

[9] What Is a Vishing Attack | Examples & Prevention | Imperva

[10] Xing, J., Yu, M., Wang, S., Zhang, Y., & Ding, Y. (2020). Automated fraudulent phone call recognition through deep learning. *Wireless Communications and Mobile Computing*, *2020*, 1-9.

[11] Lee, M., & Park, E. (2023). Real-time Korean voice phishing detection based on machine learning approaches. *Journal of Ambient Intelligence and Humanized Computing*, *14*(7), 8173-8184.

[12] Tran, M. H., Hoai, T. H. L., & Choo, H. (2020). A third-party intelligent system for preventing call phishing and message scams. In *Future Data and Security Engineering. Big Data, Security and Privacy, Smart City and Industry 4.0 Applications: 7th International Conference, FDSE 2020, Quy Nhon, Vietnam, November 25–27, 2020, Proceedings 7* (pp. 486-492). Springer Singapore.

[13] Salloum, S., Dautov, R., Chen, X., Peng, P. X., & Huang, J. Z. (2016). Big data analytics on Apache Spark. *International Journal of Data Science and Analytics*, *1*, 145-164.

[14] Assefi, M., Behravesh, E., Liu, G., & Tafti, A. P. (2017, December). Big data machine learning using apache spark MLlib. In *2017 ieee international conference on big data (big data)* (pp. 3492-3498). IEEE.