**ProtecTalk: Real-Time Scam Call Detection and Prevention**

**1. Introduction**

**1.1. Purpose, Key Benefits, and Scope**

This document outlines the product requirements for ProtecTalk, a real-time scam call detection and prevention application. Its primary goal is to identify and mitigate fraudulent calls by providing immediate visual and haptic alerts. In cases of suspected fraud, the system can also alert a designated trusted contact to assist the user. By employing a layered security approach, ProtecTalk enhances protection, mainly for individuals who may be more vulnerable to scams, while ensuring confidence in daily communications.

**1.2 User Types**

* **Protegee:** The primary user who receives protection from both the application’s fraud detection methods and the trusted contact.
* **Trusted Contact:** A designated individual entrusted by the protegee to take action in case of a suspected fraud attempt.

**1.3. Overview of the Alert System**

* **Before Answering**: ProtecTalk displays a warning based on existing information, such as known scam numbers, when an unknown call is received.
* **During the Call**: The alert level updates in real time based on ongoing analysis, using visual and haptic cues. A trusted contact may be notified if needed.

**2. System Overview**

**2.1. Application Requirements**

2.1.1. The application is designed for Android smartphones.

2.1.2. It will be available for free download on the Google Play Store.

2.1.3. Initially the app will support spoken English calls.

2.1.4. An active internet connection is required for the application to function.

2.1.5. The application will require user permissions for file system access, microphone usage, contacts, and location.

**2.2. High-Level Description and Principles**

2.2.1. The application is implemented using a client-server architecture.

2.2.2. It features a simple and accessible interface designed for users of all age groups.

2.2.3. The application runs continuously in the background, eliminating the need for manual activation.

2.2.4. It is optimized for low resource consumption, ensuring minimal impact on device performance.

2.2.5. Privacy-first design: All data processing is designed to protect user privacy. Sensitive information is processed locally whenever possible, reducing the need for data transmission. In cases where remote processing is required, data is anonymized and encrypted to ensure the user’s privacy is maintained.

2.2.6. Secure HTTP communication is used as the protocol for data exchange between the client and server, with JSON as the communication format.

**3. Core Functionality**

**3.1. Application Activation**

3.1.1. The application remains idle when no calls are in progress.

3.1.2. Upon receiving a call, the application verifies whether the number is saved in the user’s contacts.

3.1.3. If the incoming call is from an unknown number, the application activates automatically.

**3.2. Unknown Number Validation**

3.2.1. Before answering, the application checks a remote database containing phone number history and analysis to determine if the caller is potentially fraudulent.

3.2.2. If the database identifies a high likelihood of fraud, a visual alert will be displayed to the protegee at this stage.

**3.3. Real-Time Scam Detection**

3.3.1. During the call, the application transcribes speech into text.

3.3.2. A local AI model analyzes the transcription, identifying scam-related keywords in real-time while filtering out sensitive information.

3.3.3. The application sends the processed text to a remote AI model to calculate the probability of the call being a scam.

3.3.4. Based on the calculated probability, the application visually presents an alert using a traffic light system. Optionally, the protegee may receive an in-call voice warning.

3.3.5. If the protegee has one or more trusted contacts, an alert will be sent to them.

3.3.6. In cases of high scam risk, the call transcript will be stored locally on the user's device.

**3.4. Alerts and Notifications**

3.4.1. The protegee will receive real-time indications of the potential scam risk of a call through vibrations and a visual traffic light system with the following meanings:

* Green – No scam indicators have been detected by the remote model so far.
* Yellow – Some risk has been identified in the call.
* Red – A high risk of fraud is detected.

3.4.2. Based on the probability calculated by the remote model, the application will send an alert to the trusted contact, notifying them of the ongoing call.

3.4.3. If the application is deleted from the protegee’s device, their trusted contacts will receive a notification about its removal.

**3.5. Feedback for Learning and Improvement**

**3.5.1. After a call ends:**

* The protegee will have the option to send the full call transcription to improve the AI model’s detection accuracy.
* The trusted contact will have the option to provide feedback on whether the call was indeed a scam, further refining the model’s ability to identify fraudulent activity.

**4. User Interaction Flow**

**4.1. User Registration**

4.1.1. User identification occurs after installation by manually entering a phone number and receiving a randomly generated 6-digit verification code.

4.1.2. The user must accept the terms of use and privacy policy.

4.1.3. The user will be prompted to grant the necessary permissions.

4.1.4. The user will enter their name and date of birth.

4.1.5. A skippable screen will prompt the user to either request protection from a trusted contact or offer protection to a specific protegee.

4.1.6. Based on the user’s choice, the application will redirect them to the appropriate screen to establish a protection link between a trusted contact and a protegee.

**4.2. User Login**

4.2.1. User identification will occur after app installation by manually entering their phone number and receiving a randomly generated 6-digit verification code.

4.2.2. Upon successful verification, the user will be redirected to the application’s main screen.

**4.3. Establishing Protection**

4.3.1. Protection requests follow a two-way approval process.

4.3.2. Both the protegee and the trusted contact can initiate a protection request through a dedicated screen.

4.3.3. The requesting user must provide the other user’s phone number, name, and relationship.

4.3.4. The application server will verify the request and proceed as follows:

* If the other user is registered, the application will confirm that the request was sent successfully.
* If the other user is not registered, the requesting user will be notified accordingly.

4.3.5. Upon receiving the request, the other user can choose to approve or reject it.

4.3.6. Based on the decision, the requesting user will receive a correspondingnotification.

**4.4. Removing Protection**

4.4.1. Both the protegee and the trusted contact can remove a protection relationship through the protection management screen.

4.4.2. Each contact entry will display relevant details, allowing the user to select which protection to remove.

4.4.3. Upon removal, the application server will update the system and notify the other user (either the protegee or the trusted contact) accordingly.

**4.5. Notifications and Alerts Center**

4.5.1. The user can view the history of scam alerts.

4.5.2. The user can access past notifications, including pending protection requests.

**4.6. Configurability and Customizations**

4.6.1. The user can choose to disable the "Unknown Number Validation" and "Real-Time Scam Detection" functionalities.

4.6.2. If the user has trusted contacts, they will receive an alert notifying them that these options have been disabled.

**5. Data Management**

**5.1. User Data** – The system will store the following information for each user:

5.1.1. User’s Phone Number – The unique phone number associated with the user’s account.

5.1.2. User’s Name – The full name of the user for identification purposes.

5.1.3. User Type (Protegee/Trusted Contact) – Defines the user’s role in the system.

5.1.4. Linked Contact/User Details – Stores information about other linked users (e.g., protegees or trusted contacts).

5.1.5. Old Linked Contacts – Stores information about previous linked contacts that were removed.

**5.2. Detection Data**

5.2.1. Scam Patterns – A database of known fraud patterns and severity levels.

5.2.2. Fraud Scripts – Stores common scam scripts and phrases used by fraudsters.

5.2.3. Keywords – A list of important words and contextual weights used for scam detection.