Providing Warnings For Fake Incoming Calls

Victor Carbune
Matthew Sharifi

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation
Carbune, Victor and Sharifi, Matthew, "Providing Warnings For Fake Incoming Calls", Technical Disclosure Commons, (July 09, 2020)
https://www.tdcommons.org/dpubs_series/3412

This work is licensed under a Creative Commons Attribution 4.0 License.
This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.
Providing Warnings For Fake Incoming Calls

ABSTRACT

Phishing attempts using artificially synthesized audio (“deep fakes”) are increasingly used to deceive users that receive voice calls via telephone networks or calling apps. Voice cloning technology can be used to conduct a fake phone call in which the caller audio is faked to imitate someone else’s identity.

The techniques described in this disclosure analyze an incoming call and assist a user in the verification of caller identity. Fake voice calls are detected automatically and alerts are provided to the user. The detection is based on user-permitted information such as caller phone number, profile information, voice profiles of callers, and other contextual factors. Fake calls can be detected prior to the user answering the call or during an ongoing call. The described techniques are implemented on-device and with specific user permission.

KEYWORDS

- Fake call
- Caller identification
- Deepfake speech
- Synthesized speech
- Dialer interface
- Contextual factors
- Voice profile
- Phishing attempt
BACKGROUND

Phishing attempts using artificially synthesized audio (“deep fakes”) are increasingly used to deceive users. For example, malicious actors can use voice cloning technology to imitate a caller that is known to a user, for example, to imitate an executive’s voice [1] where the executive is a known caller to a call recipient.

The consequences of such imitation can be significant because the voice cloning technology can be used to perpetrate fraud on the user receiving the call. For example, the cloned executive’s voice can be used to transmit instructions for the call recipient to transfer funds to an improper account.

DESCRIPTION

Fake phone calls that are placed using voice cloning technology to mimic the caller’s voice are hard to detect with conventional techniques. The techniques described in this disclosure analyze an incoming call and assist a user in the verification of caller identity. The described techniques are implemented on-device and with specific user permission.

As used herein, a fake phone call refers to a call from a fraudulent caller who is trying to imitate the identity of a legitimate caller who is known to the user. The techniques enable caller identity verification and fake call detection by using user-permitted information such as caller phone number, caller app, profile information, voice profiles from prior calls from the caller, and other contextual factors.

For example, upon detecting that a call is likely fake, either upon receipt or during the call, the user is warned the call is likely a fake with prompts on a user device that is used to answer the call, e.g., a smartphone, smart speaker, wearable device, computer, etc. Examples of prompts include vibrating the device and/or providing an alert warning the user of the fake call.
The alert can be provided in the app used to participate in the call, e.g., a smartphone dialer application or other application. In another example, an in-call audio warning can be provided to the user indicating that the call is a likely phishing attempt. In some cases, the prompt can be delivered even before the call starts.

Upon receipt of an incoming call, the calling number or caller phone app profile (e.g., via a calling service such as a messaging platform, voice/video calling platform, etc.) is matched with on-device and/or off-device records of previous calls from the caller to extract a variety of factors that are used as input to a trained machine learning model that outputs the likelihood of the call being fake. Such user-permitted factors can include one or more of: prior calls with the caller, prior non-call interactions with the caller, metadata about caller activities, contextual factors such as time of day, etc.

- **Speaker ID:** With user permission, prior calls with the caller are utilized with an on-device model to train a speaker identifier for the particular caller. The model can indicate whether the voice of the caller matches the speaker ID that was trained from prior calls.

- **Prior non-call interactions:** Other interactions with the caller, e.g., emails or chat messages exchanged with the caller, can also be analyzed and used for detection of fake calls. These messages can be prioritized based on factors such as recency. With user permission, the messages are run through a pre-trained machine learning model to generate embeddings that can be provided as context to the fake call detection model.

- **Metadata:** Metadata about the known caller’s activities can include, with user permission, such as meetings scheduled on a shared calendar (e.g., when the recipient has a view of the caller’s calendar). Such metadata may be available to the recipient via their connection with the caller, e.g., the caller and the recipient working at the same company,
being friends on a social network, sharing calendars with each other, etc. For example, if the shared calendar indicates the known caller is currently in a meeting, it is less likely the known caller is also initiating or participating in a phone call. Other types of user-permitted metadata can also be used similarly. Such metadata can be passed through a pre-trained embedding model.

- **Contextual factors:** Other contextual factors such as time of day for the caller and the recipient, or other factors available locally on the recipient device can also be used in the detection of fake calls.

For each incoming call, the above factors are extracted efficiently, on-device, and in a manner permitted by the user such that the factors are available and can be analyzed by a fake call detection model prior to the user answering the call, or immediately upon the start of a call. At the time the user answers the call, the factors are used to determine the likelihood of the call being fake. Further, as the call progresses, the content of the call (e.g., caller voice) can be utilized with user permission, to further determine whether the call is fake. This can help detect fake callers even when the speaker ID that is determined from prior calls does not immediately match the current caller voice.

The machine learning model that is used for fake call detection can take the various input factors into account to make a determination of whether the call is likely a fake caller. For example, the machine-learning model can be a multi-layered feed forward neural network or can work with variable-length inputs through recurrent neural network layers (e.g., for incorporating both audio and text input as the phone call develops). The generated output of the model can include a likelihood score (e.g., that indicates a likelihood that the caller is fake) and certainty score (e.g., that indicates the confidence of the fake call detection).
The machine learning model can be tuned for high precision to avoid false alerts of fake calls. Training can be done with simulated data and phone calls (using generative models for generating fake audio). This training can also be done using end-to-end generative adversarial networks (GANs) where the generator and discriminator factors are conditioned on the described factors. Ultimately, the discriminator model from the GAN can be used for fake call detection.

The fake call detection techniques described herein can also be used generally to calls received from businesses, e.g., a user pretending to be a representative of a particular business, such as a bank. In this general case, the speaker id profile may be assigned a lower weight (e.g., since the caller may not be associated with prior calls) than the other factors.

The described techniques can be implemented as part of a default telephony app on a smartphone (e.g., a dialer or phone app) to help users identify and handle fake incoming calls. The techniques can also be incorporated into any app that provides voice calling features. The described techniques can identify a likely fake call, e.g., if the caller is calling improperly using a known caller’s phone or app profile (e.g., a messaging or calling app) but the caller voice does not match the known speaker identity.
Fig. 1: Verification of caller and providing a warning of fake phone calls

Fig. 1 shows an operational implementation of the described techniques. A caller (102) initiates a phone call with voice content (106), to a recipient using a recipient device (104) with a call app (112). With user permission, the incoming call is analyzed using a fake call detector (108) running on the user device (104) that implements the techniques described herein.

Upon receipt of and during the call, the fake call detector (108) utilizes input contextual factors (110), including whether the caller’s voice matches a stored speaker identifier. If the call is detected as a fake call, a warning message (114) is provided to the recipient. As seen in Fig. 1, the warning message is highlighted, e.g., by using bold and/or different font colors. The warning message makes it easy for the user to spot a detected fake call and take appropriate verification action. While Fig. 1 illustrates the use of a warning message, any suitable mechanism can be
used in combination or separately from the message, for example, a vibration of the user device, and with respect to user interface elements on the call app, e.g., the “End Call” button, a change in shape and color, any type of highlighting, e.g., a blinking transition, a color-change transition, a change in size, etc. can be used as appropriate.

The threshold values used to provide an indication of a likely fake call can be set by the developers or determined dynamically at runtime. The default thresholds may be set relatively high to avoid incorrect alerts to the recipient.

Further to the descriptions above, a user is provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable use of user information (e.g., information about a user’s known callers, app profiles of callers, speaker identifier determined on device, contextual factors, social network data, calendar, messages), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user’s identity may be treated so that no personally identifiable information can be determined for the user. Thus, the user has control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

The techniques described in this disclosure analyze an incoming call and assist a user in the verification of caller identity. Fake voice calls are detected automatically and alerts are provided to the user. The detection is based on user-permitted information such as caller phone number, profile information, voice profiles of callers, and other contextual factors. Fake calls can
be detected prior to the user answering the call or during an ongoing call. The described techniques are implemented on-device and with specific user permission.

REFERENCES

1. Thieves are now using AI deepfakes to trick companies into sending them money.