Mind Your Voice: Security Risks and Recommendations for Audio-centric Social Media Platforms

Technical Brief
With an estimated growth rate of about half-million users per day, or 6 million users per month, ClubHouse is probably the most popular audio-only social network at the moment, albeit not the only one: Riffr, Listen, Audlist, and HearMeOut are some of the alternatives, while we've been waiting for the upcoming Twitter Spaces. There's also Discord, which we didn't consider in our analysis because, despite having audio features, it's not centered around audio.

This technical brief aims to provide users, app creators, and social network service providers practical security recommendations about the security risks brought forth by audio-only social networks.

In parallel and independent from the research published by the Stanford Internet Observatory (SIO), we analyzed the apps (primarily ClubHouse but also including Riffr, Listen, Audlist, and HearMeOut). Among other attacks, we also found that, under some circumstances, an attacker can passively collect sensitive information about ongoing conversations, even "locked" ones; including participants, their identifiers, names, photos, and so on, without the need to join that room. This poses a clear privacy risk. Some security risks that we highlight here are unique to the ephemeral nature of audio-only social networks, while some are shared with other modern social network platforms (audio-centered or not).

This research has been conducted in February 8-11 of this year. At the time of publication, some of the issues described in this document might have been or are currently being fixed by the app vendors.

We acknowledge that ClubHouse has rapidly responded to the concerns raised by the SIO and other researchers. Also, we have independently obtained and analyzed the software tools used to allegedly “spill audio from ClubHouse,” an episode that was highlighted in the press and promptly blocked by ClubHouse. We want to underline that this wasn’t a security breach. What happened is that a developer has created a mirror website that allowed others to listen in, using the developer’s only account instead of their personal account. While this certainly breaks the terms of service, by no means any specific security weakness has been used and, most importantly, the mirror website was not doing any recording: the audio was still being streamed from ClubHouse servers to the requesting client, and was never going through the mirror website. In other words, that website was nothing more than a client, based on JavaScript as opposed to iOS. Although this type of service abuse can be made more difficult, no web service or social network is immune to them, because there’s no technical way to reliably block abuses without impacting the availability to legitimate users.

Audio-only or Audio-centric Social Networks

Audio-only social networks are a growing technology that allows users to communicate almost exclusively via audio. Private or group textual chats are substituted by voice chats where two or more users can join private or public rooms as they wish and start listening immediately.
Figure 1. The main entities and data and interactions among them in a typical audio-only social network.

In these platforms, like in other online social networks, users can organize in groups (sometimes called "clubs") and schedule events. Some apps like Riffr or Audlist allow permanent recording of uploaded content, providing a micro-podcasting experience.

ClubHouse has taken a different approach. Any conversation on ClubHouse is ephemeral, for streaming only, and exclusive: one can only follow it by joining the so-called room. The terms of services prevent users from recording the audio without the participants' written consent, but clearly, this doesn't prevent a malicious listener from doing that — and we have found out that this is happening.
Figure 2. From left to right: User interfaces of Riffr, Audlit, and ClubHouse.
Security and Human Risks of Audio-centric Social Networks

This section provides samples of attacks that can be perpetrated via audio or a combination of audio and social network features implemented in audio-centered social network services and apps.

By analyzing the apps and their communication protocols, we confirmed that a malicious actor could automate different kinds of attacks, including massive data collection, user monitoring, and so on. For the end-users, this means that their entire social circle (e.g., names and pictures of friends and friends of friends) can be obtained by an attacker without them noticing; and sometimes, even through automated means.

Not only did we find that this was possible, but in late December 2020 a group going by the name of "Reverse and Code - Reverse Engineering and Bot Coding" has advertised the development of a ClubHouse Bot, mentioning that their goal is to offer a "Fully web-based Clubhouse SaaS bot to schedule all daily tasks to run in the background on autopilot 24/7/365."

![Figure 3. An online group advertising the development of a ClubHouse bot to automate user activity.](image)

Given the fast growth of some of these services, the high competition for visibility among users, and the "exclusivity" of certain clubs, we understand why many discussions on these networks revolve around ways to monetize them. Some users are already talking about "paid clubs" on ClubHouse, to name one. This and other examples may create more and more opportunities for illicit businesses as well.

Here are some of the security risks in audio-centered platforms. Most of these can be automated as well.
1. Network Traffic Interception and Wiretapping

While HTTPS is used to secure all non-audio traffic, except for the download of recorded audio files on those networks that allow it, live audio calls are normally transported over dedicated protocols, often based on UDP. Real-Time Communication (RTC), or WebRTC, is the most common family of protocols used for live audio-video calls. If an attacker can intercept network traffic, they can partially or completely tap into unencrypted audio streams and tune in to what other users are listening to.

Without going into details, an attacker can use different methods to intercept network traffic: DNS poisoning ARP spoofing in unsecured or shared Wi-Fi networks, rogue LTE stations, or rogue APs are just some examples. More in general, a rogue network operator (e.g., a compromised or untrustworthy ISP) could access call information. For these reasons, audio-video messaging apps use additional layers of encryption that mitigate the information disclosure risk.

The version of the ClubHouse app that we have examined, which is based on the Agora RTC\(^9\) (based on Real-time Transport Protocol or RTP) framework, does not enable audio stream encryption, while the official technical documentation\(^10\) recommends doing so. The lack of encryption allows an attacker to intercept the communication data and know who is talking to whom, including in private rooms. From the analysis of the app, we can say that ClubHouse is using an older version of the Agora library, which implements encryption methods that are no longer up to date.

For example, with minimal effort, an attacker can know who is talking to whom by simply analyzing network traffic and looking for RTC-related packets. In the following screenshots, we can see how an attacker can automate this procedure and intercept the RTC control packets to obtain sensitive information about a private chat created with two users in it. We also conducted and recorded a demonstration of this.\(^11\)

![Figure 4. Automating network analysis and RTC packet search](image-url)
AgoraRTC adopts a mix of slightly proprietary protocols that make audio decoding more complicated but certainly feasible given enough time. We wanted to make the users aware of this risk by quickly providing an example of what can be done with almost zero effort by an attacker. We're confident that a complete reverse engineering of the protocols — currently ongoing — would allow an attacker to also listen to ongoing conversations. Meanwhile, we recommend that users avoid using this app on shared Wi-Fi networks or untrustworthy networks until a patch is released.

By examining the Agora SDK library for iOS (the Android version was also analyzed by McAfee in a separate report) used by ClubHouse, we found out that the encryption primitives of that specific version of the library have currently been deprecated by Agora, according to their technical documentation. However, those (outdated) encryption primitives are not used by ClubHouse, which Stanford Internet Observatory's report also independently confirmed. That version of the Agora SDK library also contains three hardcoded IP addresses, some of which are contacted by the app. Those IP addresses belong to the software-defined network service provider that enables Agora to scale worldwide.

2. User Impersonation and Deepfake Voice

A malicious user could impersonate a public persona and, by cloning their voice, make them say things they never would, with consequences on their reputation. An attacker could also clone the voice and create a fake profile of a famous trader, attract users into joining a room, and endorse a certain financial strategy.

User impersonation on audio-only social networks is easier than on video-centric social networks such as YouTube. The feasibility of user impersonation depends on two factors: the content-forging capability of the attacker and the presence of account-verification countermeasures. If the service provider does not implement a strict user-verification program like Facebook or Twitter does, forging accounts or buying them from underground marketplaces is relatively easy.

![ClubHouse invites for sale on underground forums](image)

Figure 5. ClubHouse invites for sale on underground forums
We're not aware of any user-verification program in audio-only social networks. For example, by searching on ClubHouse for the G20 leaders, we found multiple accounts for most of them, as well as accounts of other public figures. Fortunately, ClubHouse allows users to connect other social media accounts to their profiles, which helps "validate" the authenticity of the real profile.

Forging high-quality, natural audio or video content for user impersonation is only slightly more difficult than producing text or other static content (e.g., images), but deepfake technology has now become accessible, making this a more tangible threat.

Figure 7. A free real-time voice cloning toolbox

*Image Source: Corentin Jemine Youtube Channel*
Replicating someone's voice is certainly easier and can be done almost in real-time, compared to preparing an offline deepfake video, for which we already have quite advanced tools in the market.

Many companies offer legitimate voice-cloning services; in 2019 a fraudster successfully cloned the voice of an unnamed UK-based energy firm's executive and used it to order the CEO to transfer approximately $250,000 to its account. We verified that voice-cloning services are also on sale on dark web marketplaces.

We recently published a report that offers a deeper explanation of how AI technology can be abused on audio and video content.
3. Opportunistic Recording

Opportunistic recording can ease the implementation of existing attacks (mainly extortion-based), making them more powerful or scalable, especially if combined with user impersonation. For example, if someone is speaking naturally in a room with many users, there’s no way to tell if a malicious listener is recording the speech. Later on, the malicious listener can clone the account, automatically follow all of the account's contacts to make it look more authentic, join any other room, and use the sample of the speech to make the "cloned" voice say other potentially embarrassing phrases.

The content of most audio-only social networks is meant to be ephemeral and “for participants only,” but some can actually take recordings, as the following Twitter post alleges.

![Figure 10. A conversation on Twitter where users discuss recording ClubHouse rooms](image)

*Image Source: Twitter*

Major audio-only social networks such as ClubHouse clearly state that users must not record the audio stream of any room without the participants' consent. ClubHouse’s Terms of Service states that users should "agree to not use the Service to: [...] record any portion of a conversation without the express written consent of all of the speakers involved."

However, the streaming and lightweight nature of audio content makes it very easy for a malicious user to record an ongoing conversation and use it later on, possibly recombined in a way to make the victim "say" arbitrary content — either through deepfake technology or with simple remixing.

Needless to say, the recordings of some of the most exclusive rooms (e.g., with VIPs among the speakers) have already been shared on other services, while hundreds of thousands of recordings and instructions on how to take these can be found on Google using keywords such as "clubhouse recording."
4. Harassment and Blackmailing

Even without recording, a malicious user can harass or blackmail users. An attacker that follows their victim will get notified when that person joins a public room. Upon receiving a notification, the attacker could join that room, ask the moderator to speak, and use this opportunity to say something or stream pre-recorded audio to blackmail the victim. We verified that all of this could be easily scripted to run automatically.

Fortunately, major providers such as ClubHouse allow users to block or report other users (including their introducers), which would trigger an investigation that could rely on encrypted samples of recorded content (if within a given time window).

![Alpha Exploration Co.]

*Last Updated Nov 2, 2020*

Thank you for reviewing the updated Terms of Service! One important change we’ve made is the introduction of temporary and encrypted audio recording, solely for Trust & Safety purposes. By using Clubhouse, you now consent to having your audio temporarily recorded in case a Trust & Safety incident is reported. We’ve also updated our Privacy Policy and our Community Guidelines to reflect this.

Figure 11. Highlight from ClubHouse's Terms of Service

However, a user could maliciously report another user about an ongoing conversation to have their account suspended. Only the social network operators can validate that report by opening the case and listening to the conversation, which puts a lot of burden on the operators' side. It's currently unclear how this could scale.

5. Underground Content Promotion Services

Whether they're legitimate or not, content-promotion services are commonly used and abused in any social network to inflate content or user popularity artificially. While such services do not directly represent a security risk, engaging with the actors that provide illicit services can be dangerous or even illegal, not to mention that unaware purchasers may get re-targeted in the future.

Right after its launch, we found active discussions about ClubHouse on the surface Web. Some users are already discussing purchasing followers, with some alleged developers promising to reverse engineer the API to create a bot in exchange for an invite — something that we also verified to be feasible.
Figure 12. Forum showing the post of a user who's willing to buy fake followers for ClubHouse

Figure 13. Forum users discussing making and/or buying a bot service for marketing purposes
We cannot exclude that similar opportunities could arise after engaging with entities that sell invites to exclusive social networks such as ClubHouse for up to US$125.

An Invitation to ClubHouse (The app) Clubhouse App Invite

<table>
<thead>
<tr>
<th>Condition:  --</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantity: 1</td>
</tr>
<tr>
<td>Price: <strong>US $125.00</strong></td>
</tr>
</tbody>
</table>

*No Interest if paid in full in 6 mo on $99+*

Figure 14. A ClubHouse invite for sale on eBay

For example, by posting a generic message on his public Twitter feed asking if anyone can provide ClubHouse invites, the author of this technical brief post was received direct messages by two sellers in less than a day. Similarly, we found so-called "silent rooms" where joiners are immediately allowed to sit in the speakers’ lineup at the top, which gets them visibility and followers, as in the example below:

Figure 15. Direct messages from sellers and speakers’ lineup in silent rooms

Although it’s not clear what strategy lies behind these rooms, it’s another sign of an intensely competitive environment.
6. Audio Covert Channels

Threat actors see any new social network as another opportunity to create covert channels for command & control (C&C) or hiding or transmitting information using steganography. If audio-centric social networks alternative channel: for example, an attacker can create multiple rooms and have bots joining them to dispatch commands without leaving any trace (except in the encrypted recordings, if any).

In 2016, we analyzed how common chat APIs (e.g., Slack, Discord, Telegram, Hipchat, Mattermost, Twitter, Facebook) could be used to implement C&C channels, uncovering known and future abuses for botnet operations. The upcoming Twitter Spaces aside, none of the current audio-only social networks support an official API, but, given the fast growth of ClubHouse, it’s reasonable to assume that there will be one in the near future.

We foresee similar malicious use cases with audio-only social networks, with the uniqueness offered by the use of sound as a covert channel. Audio steganography, which has been proven effective, can be used to implement cover channels. Ultrasound also have been demonstrated to offer a reliable covert channel over audio; for example, they’ve been used by the advertising industry to implement cross-device tracking.

While most of the aforementioned covert channels could be detected deterministically and filtered, we predict that future threat actors could use language-based coding techniques to blend encoded information in meaningful conversations. For example, specific sequences of known words or the delay between words could be used to represent an arbitrary alphabet, on top of which any information could be transmitted at slow rates.

As a side note, with the proliferation of smart speakers, audio-only social networks are yet another opportunity to trigger unwanted commands, as we’ve explained in our previous work.
Permissions Used and Data Collected by the Apps

We briefly highlight some technical findings that we believe can help explain the potential impact on user privacy. It's important to underline that most of the apps nowadays integrate various services that cover specific functionalities. For example, many apps, such as ClubHouse, use Amplitude for analytics and Instabug for bug and crash reporting, and integration to other social network services such as Twitter, Instagram, or Facebook. So, even when using a single app or service, different data types are shared with various third-party services. This includes, for example, the following data:

- Language (as set on the phone)
- Country
- Device manufacturer (e.g., Apple)
- Carrier
- Device model
- Operating System version

Most of the users we talked to were concerned with audio-only social network apps requesting many permissions, so we think we should clarify which app uses which permissions. To this end, we used our Mobile App Reputation Service (MARS), which offers a free-scan plan, to analyze the five available apps (four Android apps plus one iOS app).

All of the apps are classified by MARS as safe, showing no significant indicators of suspicious behavior potentially affecting the users' privacy. All of the apps legitimately request access to the following permissions in order to implement their functionalities:

- **Record audio or microphone access.** Obviously, there's no way to avoid granting this permission in order to use the app.

- **Read contacts in the address book.** Users can deny this request if they prefer not to share the address book — in most instances, it doesn't affect the apps' normal operation. We analyzed the apps and verified that the most popular one, ClubHouse, only uploads the phone numbers from the user's contact list. But, when sending an invite, the full name (as stored in the address book) is uploaded alongside the number. So, for instance, if the user has "Duffy Duck" in his contact list and tries to invite this contact, the social network service provider will know Duffy Duck's number.

- **Camera.** Although not directly needed by the apps, some of them use this permission to allow taking photos for user profiles.

- **Read and write external storage (Android only).** Some apps require this permission to access files (e.g., pictures) on the external storage.

- **Access network state.** This may sound suspicious at first because one may think that such an app should not require local network access. We verified that no sensitive information about the local network gets uploaded. There's a clear technical reason behind this request as explained in the case of ClubHouse. As stated in Agora's Dev Center, "Agora RTC SDK for iOS with the version earlier than v3.1.2 detects the connection quality between the client and the user's local router, and reports the round-trip delay between the client and the user's local router by using the gatewayRtt parameter of the reportRtcStats callback. The iOS system determines the connectivity detection as a search for local network devices. Thus,
although the app does not connect to any devices on the user’s local network, the user sees a prompt to find local network devices when launching an iOS app for the first time."

Figure 16. Screenshot of the ClubHouse app requesting permission to access the local network. This request is not used to collect any sensitive information.

We leave a detailed privacy-policy analysis to experts on legal matters. We understand this is already being done in Europe, where ClubHouse is being scrutinized about General Data Protection Regulation (GDPR) compliance.²⁶,³⁷
Security recommendations

To ensure the secure use of audio-only apps, here are a few recommendations for users of audio-only social networks:

- **Join public rooms and speak as if in public.** Users should only say things that they are comfortable sharing with the public, as there is a possibility that someone in the virtual room is recording (even if recording without written consent is against the Terms of Service of most, if not all, of these apps).

- **Do not trust someone by their name alone.** These apps currently have no account-verification processes implemented; always double-check that the bio, username, and linked social media contacts are authentic.

- **Only grant the necessary permissions and share the needed data.** For example, if users don’t want the apps to collect all data from their address book, they can deny the permission requested.

Based on our technical analysis of the apps and communication protocols, we recommend that current and future service providers consider implementing the following features unless they have done so already:

- **Do not store secrets (such as credentials and API keys) in the app.** We have found cases of apps embedding credentials in plain text right in the app manifest, which would allow any malicious actor to impersonate them on third-party services.

- **Offer encrypted private calls.** While there are certainly some trade-offs between performance and encryption, state-of-the-art messaging apps support encrypted group conversations; their use case is different, but we believe that future audio-only social networks should offer a privacy level on par with their text-based equivalent. For example, Secure Realtime Transport Protocol (SRTP) should be used instead of RTP.

- **User account verification.** None of the audio-only social networks currently support verified accounts like Twitter, Facebook or Instagram do, and we have already seen fake accounts appearing on some of them. While waiting for account-verification features, we recommend users to manually check whether the account they’re interacting with is genuine (e.g., check the number of followers or connected social network accounts).

- **Real-time content analysis.** All of the content-moderation challenges that traditional social networks face are harder on audio- or video-only social networks because it’s intrinsically harder to analyze audio (or video) than text (i.e., speech-to-text takes resources). On the one hand, there’s a clear privacy challenge that arises if these services implement content inspection (because it means that they have a way to tap into the audio streams). However, content inspection offers some benefits, for instance, in prioritizing incidents.
References:


