A Collaborative Privacy-Enhanced Alibi Phone

Hsien-Ting Cheng[1], Ching-Lun Lin[2], and Hao-hua Chuinst[1]

[1] Department of Computer Science and Information Engineering,
Graduate Institute of Networking and Multimedia,
National Taiwan University, Taipei, Taiwan
{r92006, hchu}@csie.ntu.edu.tw
[2] Department of Computer Science,
Columbia University,
New York, NY 10027, USA
cl2399@columbia.edu

Abstract. This paper presents a collaborative privacy protection approach that not only filters context information and reduces its granularity, but also intelligently replaces the filtered-out context with an artificial context considered appropriate by its user. The benefit of this approach is that individuals accessing the filtered context cannot detect the presence of filtering, namely, filtering becomes imperceptible. This new approach is used as a basis for designing, implementing and evaluating a collaborative privacy-enhanced alibi phone, allowing user to imperceptibly conceal surrounding ambient sound from callers, while leaving callers unaware of this filtering.

1 Introduction

Many individuals feel that they reveal more (context) information to others regarding their daily life than is necessary. For example, when receiving a voice call or participating in a video conferencing session, individuals not only disclose to callers our voice and facial expression, which is the only information they actually want them to hear/see. Microphones and cameras also capture and transmit ambient sound and background scenery to callers, reveal additional context information regarding our current location and activities. In some situations, such additional information can cause unnecessary embarrassment and misunderstanding to the callees. To avoid these situations, many callees often refuse to communicate with callers when they consider it inconvenient. This work uses the following two scenarios to illustrate situations like those described above:

1. Joe has told his girlfriend Jane that he was going to play basketball with his male friends. Unfortunately, some of his male friends did not show up, leading to cancellation of the basketball game. Instead, Joe decided to meet some female friends in a coffee shop for a friendly chat. At the coffee shop, Joe then receives a phone call from Jane. Joe was hesitant to answer this call,

Y.-C. Chung and J.E. Moreira (Eds.): GPC 2006, LNCS 3947, pp. 405–414, 2006.

because he was concerned that Jane would hear that he was with female friends, possibly causing an unnecessary misunderstanding.

2. Joe noticed a video phone call from his supervisor while he was entertaining an unexpected client at a local jazz bar. Since this was an unexpected visit, Joe had not informed his supervisor Jill about it. Again, Joe was hesitant to pick up the video phone call because he did not want Jill to see or hear the bar environment and loud Jazz music and thus conclude that Joe was slacking off from work.

A simple solution to the above dilemmas would be to filter out the ambient sound and background scene [1]. However, this simple solution is insufficient in situations in which the callers are expecting certain types of ambient sound or background scenes from the callees. In the 1st scenario, Jane would expect Joe to be in the basketball court and expect to hear the sound of basketball being played. In the 2nd scenario, Jill would expect Joe to be working in a busy office and expect see sights and hear sounds confirming this. The ideal solution should produce the expected ambient noise and background scene V the basketball court or the busy office. Notably, filtering alone can lead to a noticeable absence of ambient sound and background scenes, particularly when callers are expecting callees to be in certain places with distinctive ambient sounds and scenes. Filtering may create the undesirable impression that callees are intentionally and explicitly hiding certain information from callers. Therefore, a new approach to privacy protection is required that does not simply protect the context information of callees, but simultaneously can make such filtering imperceptible to callers.

This paper proposes a new privacy protection approach that not only filters out context information, but also intelligently substitutes the filtered-out context information with artificial context information considered appropriate by its user, thus creating the appearance of imperceptible filtering to callers. This approach is collaborative in the sense that other peers on the network who may have access to such artificial context information can help by contributing them. Based on this new approach, this work has designed an audio-based privacy protection system for use with a mobile phone. To achieve imperceptible filtering, it is designed to do the following: (1) filter out background ambient sound from the voice of the callee, (2) find an appropriate ambient sound source expected by the caller over a peer network, and (3) mix the selected ambient sound source with the voice of the callee. Consider the 1st scenario described previously. When his privacy-enhanced cell phone detects that Joe is not currently at the expected location on the basketball court, it can filter out the background chatter of Joes female friends, find an ambient sound source on a basketball court and mix this ambient sound with Joes voice. Consequently, Jane will hear ambient sound resembling the location where she expects Joe to be, namely a basketball court. Joe thus can feel comfortable picking up phone calls anytime anywhere regardless of his current ambient environments.

2 Related Work

Previous works on protecting context information were focused primarily on information filtering and granularities. Project Aura [2] proposed an access control mechanism for filtering out fine-grained information from raw context data, so that the provided context information would match the access privileges given to the request. For example, location information can be determined based on an image captured by a camera. If a user is only granted access to the location information, Aura will filter out and remove the image, and return only a text-based location description. In comparison, the proposed system not only filters information granularity, but also intelligently substitutes the filtered-out context information with the artificial context information. As a result, the proposed system can create the appearance of imperceptible filtering to people accessing the context information of call recipients.

Several commercial products are available that can eliminate ambient noises. The Boom Noise Canceling Headset [2] enables users to communicate clearly in loud noisy environments. This headset can be plugged into most cellular phones. The headset is fitted with two microphones. The mouthpiece microphone collects the voice of the user together with some of the ambient noises surrounding the user. Meanwhile, the noise microphone picks up all the ambient noise but little of the users voice. The handset subtracts the ambient noise gathered by the noise microphone from the audio signals gathered by the mouthpiece microphone. The net result of the subtraction is a pure recording of the speakers voice.

Some cellular phone service providers, such as TransAsia Telecom in Taiwan [3], are currently offering services that enable users to mix background music into their phone calls locally and centrally, respectively. To use these services, users must establish a schedule for mixing the music or sound effect, such as noise of traffic jam, a circus parade, a thunderstorm, a ringing phone, into the phone call. Additionally, users can decide what and how to mix the music or sound effect based on the identity of the callee/caller. Nevertheless, the proposed system differs from these services in several respects. First, both of the previous system lack the ambient noise filtering feature, which will be crucial in a location with extremely noise background, simply mixing is unable to replace the original ambient noise. Second, the previous service provided by TransAsia Telecom [3] is designed to make phone calls more entertaining rather than being designed to provide privacy protection. Moreover, the selection of background music for mixing with voice in previous systems is static and limited, only some prerecorded sound effect are available. Our proposed system can search through a peer network of collaborative users and the available background sound at their current locations, to find the desired ambient and mix it in real time, which actually meets the requirements of users.

A group of cell phone users have formed an alibi and excuse club [4], to provide more substantial excuses, club member has to pick up a phone to let the boss know of a buddys tardiness and make his wife believe her husband has an important meeting when he is really at bar. Such club uses a real manual conversation to hide context information, however, even the founder of the club admits there existing moral problems from an integrity standpoint.

3 Challenges and Approach

This study identifies the following technical challenges in realizing this privacy-enhanced phone.

- Detecting whether the user is at the expected location. Detecting this requires the user to maintain a schedule of whether he/she is expected to be at different times. By consulting the schedule of a user and comparing his/her expected location with his/her current location, the system can determine whether the user is at the expected location. Currently, the most popular locating system is GPS. However, GPS does not work indoors. To overcome this problem, the proposed system pre-defines some location profiles, such as office, transportation, countryside, and so on. User can then manually change current location profile. Additionally, audio recognition techniques can be applied to ambient environment sound to automatically infer the current location profile of the user.
- Quickly locate an appropriate ambient sound source at the expected location. The amount of time required to locate an appropriate ambient sound source must not exceed a few phone rings, which is the amount of time the caller is willing to wait for the callee to answer the call. If no appropriate ambient sound can be identified sufficiently quickly, the caller might abandon the call, as well as the callee may miss the phone call.
- Filter out background ambient sound and mix the selected ambient sound in real time. Audio filtering and mixing have been active areas of research on speech processing. When selecting speech processing techniques for the proposed system, the limited processing power on mobile devices must be considered, and the real time constraints of voice calls.
- Security attacks: numerous attacks can be made on the proposed system to check if a user is using an alibi background sound. Consider the following attack. To find out whether Joe is at his expected location, an attacker (or his/her friend) can make a phone call to Joe from Joes expected location. If Joe is currently not at the expected location, his cell phone will geocast a message over the peer network that requests the ambient sound source at the expected location. The attacker (or his/her friend) will then receive this request message from Joe immediately after calling Joe. This means that the attacker can tell whether Joe is at the expected place or not, depending on whether the attacker receives an ambient sound request at the expected place or not. In the 2nd attack, the attacker can monitor the data packet containing requests for ambient sound sources. The attacker can then extract the IP address in the data packet and map it to the likely physical location.
- Reliability: an active ambient sound source can sometimes fail during a call. Failures can result from wireless network disconnection, the source device running out of battery, and so on. Since failure of ambient sound source can make filtering visible to the callers, the proposed system has to be reliable under all these unexpected conditions.
- Peer-to-peer architecture vs. centralized architecture: two possible methods exist for building a peer network of collaborative users: peer-to-peer vs.

centralized. In section 5, advantages and disadvantages of these two approaches are compared, and the peer-to-peer method is chosen for implementation.

4 Design

Figure 1 shows the design of our privacy-enhanced Phone. The design comprises four components: Context Agent, Location Scheduler, Ambient Sound Locator, and Voice Processor. The executing flow is described through the following five steps:

1. Receive a ring-tone on a mobile.
2. Context Agent determines the current location of the callee via GPS or by checking the pre-defined location of the user.
3. Location Scheduler compares the current location of the callee with his/her expected location schedule. If the callee is not at the expected location, they are prompted to see if they need ambient sound at the expected location.
4. Ambient Sound Locator identifies several ambient sound sources and selects one as the active sound source.
5. Voice Processor filters out the original ambient sound and mixes in the ambient sound source.

This work has implemented the phone on HP iPAQ running the Microsoft Windows CE Operating System. This study has developed and deployed a voice processor capable of filtering and mixing ambient sounds. The ambient sound mixer is implemented by adding two waveforms and then adjusting the coefficients a and b to yield the optimum performance.

$$S(t_k) = aS_1(t_k) + bS_2(t_k) \tag{1}$$

S_1, S_2: voice signal; t_k: time index

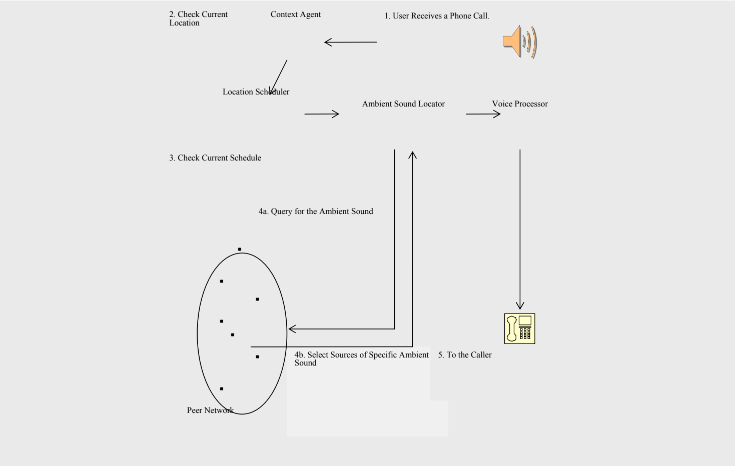
2. Check Current Location
Context Agent
1. User Receives a Phone Call.
Location Scheduler
Ambient Sound Locator
Voice Processor
3. Check Current Schedule
4a. Query for the Ambient Sound
4b. Select Sources of Specific Ambient Sound
5. To the Caller
Peer Network

Fig. 1. The executing flow of the privacy-enhanced phone

However, designing a smooth and effective ambient sound filter is more complex than mixing. Developing such a filter is equivalent to the problem of noise reduction in the field of speech processing. This study examines two general approaches for noise reduction. The 1st approach employs the whole voice sequence as an input, and then calculates a global optimal noise signal. This approach requires advance knowledge of the entire voice sequence. The 2nd approach is a frame-by-frame method based on the Wiener Filter [5] that takes two frames at a time, and then calculates noise signals locally. This approach has the advantage property of online frame-by-frame processing, which is applicable to the targeted voice conferencing application. In addition, the 2nd approach is fast enough to run in real time. Therefore, it is used.

5 Centralized vs. P2P Architecture

Two possible approaches exist for realizing the peer network of collaborative users who can help others by acting as ambient sound sources: centralized vs. peer-to-peer (P2P). In a centralized architecture, the system maintains a directory server of client locations, which can be accomplished by server periodically polling the locations of clients or alternatively client pushing the location information to the server. The location directory server responds to a client request for the ambient sound sources at a specified location or a location profile. The centralized architecture can also deploy powerful, stationary voice processing servers for running audio filtering and mixing software, and thus can alleviate the problem of limited processing and power for mobile devices. In the

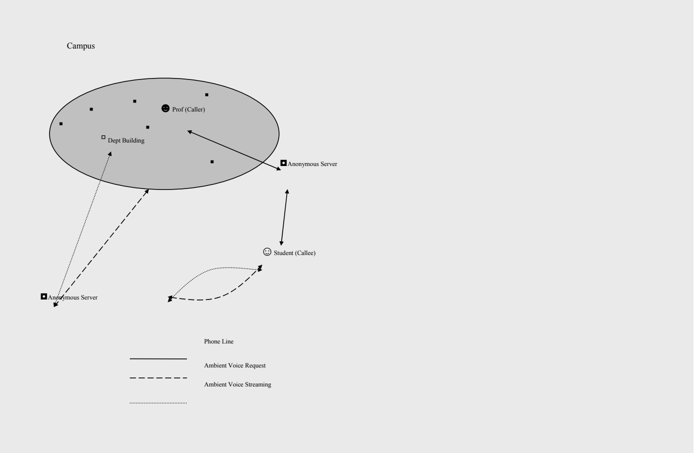

Fig. 2. A professor wants to check if a student is currently working in the department building or is just fooling around. If the student is not at his position, the student can ask for an ambient sound source provided by other students on campus through anonymous server.

centralized architecture, all voice and data communications must go through some centralized servers that are responsible for the privacy and security protection of users making requests or providing ambient sound sources.

However, the centralized architecture requires infrastructure support for deployment. Therefore, this work favors the P2P architecture; namely, direct peer node communications between the requestors and providers of ambient sound sources [6]. This work favors the P2P architecture because the application presented in this work fundamentally has a P2P flavor V in which the participating users act as service providers sharing their ambient sounds, and also act as service consumers when they require ambient sounds.

To address the security attacks described in section 2, this work proposes using anonymous redirection servers. The security attacks involve an attacker who can monitor the data packet containing requests for ambient sound sources, and then infer physical location using the IP address in the data packet. Using anonymous servers stops this attack because the requests are redirected through anonymous servers, and these anonymous servers can then remove the IP address from the request. This is shown in Fig 2.

6 Evalution

This work evaluated the performance of the privacy enhanced phone both objectively and subjectively. In objective part, the evaluation metrics include ambient sound filtering quality and communication delay; user study was conducted subjectively to evaluate the audible sound quality and the psychological state of user.

6.1 Objective Evaluation

- Ambient sound filtering: To measure the noise reduction performance of the Weiner Filter, this work employs segmental SNR (signal-to-noise ratio) improvement which is calculated by:

  $$SNR_{improve} = segSNR_{out} - segSNR_{in} \tag{2}$$

  Three types of background noise (speech, jazz, rock music) are artificially added to a sample of clean speech. Each noise has different SNR: -5dB and 5dB. Table 1 lists the noise reduction results.
- Communication delay: the communication delay time was measured, including the processing times for audio filtering and mixing, and network delay. The results listed in Table 2 exhibited good performance in terms of voice processing time and network delay time. Experiments were performed under three different conditions: without filtering and mixing, with filtering only, and with both filtering and mixing.

6.2 Subjective Evaluation: User Study

The system design should consider actual audible sound quality and user psychological state. Current systems are all unable to transmit and mix ambient

Table 1. The improvement of the noise reduction using Weiner Filter for three different noise types: speech, jazz, and rock music. And each type with two inputs SNR (measured in dB).

Noise type	Speech		Jazz		Rock Music	
Input SNR	5	−5	5	−5	5	−5
Improvement	7.3	11.6	8.9	13.3	8.2	12.7

Table 2. Delay time (process + transmit): evaluated with none of filter and mixer, filter only, and both filter and mix respectively

	None	Filter Only	Both Filter and Mix
Delay time (sec)	0.64	1.53	1.71

sound in real time. Thus, current systems cannot provide users with full privacy protection. This study describes the actual experiences of 32 participants who evaluated the system from different perspectives.

This User Study aims to understand the services offered by the proposed system and whether its performance can meet expectations. A comparison is made with overall system efficiency to see whether the proposed system achieves better efficiency. Additionally, this study surveys user satisfaction with the proposed system. The following outlines complete process and results.

Independent Variables: The calling procedure of individual users.

Dependent Variables: Actual subjective sound heard, call quality transmission, privacy protection, subjective satisfaction ranked based on overall calling experience, user perception of call quality, and system user friendliness.

Participants: Thirty-two participants aged 20-40 years old were selected for the survey. The subjects were frequent PDA and web phone users who possessed their own PDAs. Few of the participants had any experience with ambient sound filtering or mixing.

Procedures: Participants were briefed on the goals and procedures of the user study. The participants were provided a demonstration of the procedures for dialing using the PDA. The entire evaluation comprised two stages. In stage one, each participant was asked to answer three phone calls made to them. The first call involved asking the participants to answer an unfiltered call. The second call involved filtering of ambient sound. The third call used the proposed system for filtering and mixing various ambient sounds as required. In stage two, each participant was asked to rate the overall quality of the proposed system. The score is ranging from 1 to 5 where higher value indicates better performance. In addition, each participant completed the survey form with their background details and personal experience.

Results: The result shows 62.5

- The subject is the boss: Employees feel uncomfortable about the boss being aware of their location.

- The subject is the boyfriend or girlfriend: Relationships with the opposite sex are sometimes complicated, and sometimes when a person in a relationship ends up unexpectedly in a place not previously reported to their partner, even when there is nothing wrong, they may feel concerned about potential misunderstandings should their partner discover their true whereabouts.
- The subjects are parents: Students/adolescents sometimes frequent places that parents disapprove of, for example pubs, KTVs, billiard halls, and video game parlors. When the parents call and find that the children are in such places, the children are likely to feel that their privacy has been invaded. Additionally, children can also feel apologetic at causing their parents unnecessary worry by letting them know they are frequenting such places.

Almost all collected feedbacks are positive. The average score rated by all participants is 4.37. There are 93.75% of participants willing to use the proposed system to protect their privacy. Most users have a high regard for the privacy protection offered by the proposed system. These users consider that the system provides better protection than the previous ambient sound filter telephone system. Furthermore, the proposed system locates and mixes in the desired ambient sounds nearly in real-time, fully satisfying user privacy considerations. However, some users feel that the system efficiency and voice quality can still be improved.

7 Future Work

As shown in Table 1, ambient sound filtering quality for speech is lower than when the filtering is applied to other noise. We believe that this is due to the fact that there is a small difference between the voice of the user and background speech; therefore they are more difficult to distinguish. Future studies can attempt to enhance the filter to enable it to fully detect and remove background speech.

The current system is designed and implemented using the P2P architecture. Future studies should seek to improve this P2P application in the areas of scalability, security and privacy, quality of services, performance, fault tolerance, and so on. We believe that this new system for privacy protection is easily applicable to video and can help users avoid sharing sensitive background scenes withy callers. Future studies can develop video filtering and mixing methods that can substitute existing backgrounds with other scene without making callers aware of the deception.

References

1. The Boom Noise Canceling Headset.
http://www.thetravelinsider.info/roadwarriorcontent/boomheadset.htm (2003)
2. U. Hengartner, P. Steenkiste: Access Control to Information in Pervasive Computing Environments. HotOS. (2003)

3. TransAsis Telecommunicatios, Taiwan.
http://www.hank.net.tw/channel/TL/BGM/service.htm (2002)
4. Elisa Batista: Phone become alibi for liars.
http://www.wired.com/news/wireless/0,1382,63439,00.html (2004)
5. Speech Processing, Transmission and Quality Aspects (STQ); Distributed speech recognition; Advanced front-end feature extraction algorithm; Compression algorithms. ETSI ES 202 050 v.1.1.3 5.1 Noise Reduction. (2003)
6. Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker: Search and replication in unstructured peer-to-peer networks. Proc. of the 16th ACM Intl Conf. on Supercomputing. (2002)