Volume 3, Issue 4, April 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Speaker Detection of Encrypted Speech Communication Using Talk Patterns

Swornalakshme D
Department of Computer Science and Engineering, K.S.R. College of Engineering, India

Gowtham M.C
Department of Computer Science and Engineering, K.S.R. College of Engineering, India

Kalai Selvan B
Department of Computer Science and Engineering, K.S.R. College of Engineering, India

Latha G
Department of Computer Science and Engineering, K.S.R. College of Engineering India

Prakash M
Department of Computer Science and Engineering, K.S.R. College of Engineering India

Abstract— In this paper, we propose a new class of traffic analysis attacks to encrypted speech communications with the goal of detecting speakers of encrypted speech communications. These attacks are based on packet timing information only and the attacks can detect speakers of speech communications made with different codec's. The detection can be made with the talk patterns. These talk patterns are observed from the encrypted speech packets and this can be analyzed for speaker detection.

Keywords—Codec's, HMM Training, Constant Bit Rate (CBR), Detection Rate, Cross-Codec Detection

I. INTRODUCTION

Now-a-days the threatening voice calls are in increasing rate. In these cases we may not predict the user, as because of the anonymity network [5]. The network on which we can able to make our data roaming without noticing address of the particular system is called as anonymity network. For making the threatening process the anonymity networks such as Tor [4] and JAP can be used. The other common way for making such threatening message is re routing the usual way, by choosing longer router. In this proposed attack, the adversary first collects the traces of encrypted speech communication made by the speaker. Then adversary extracts application level feature of speaker's speech communication's and trains a Hidden Markov Model (HMM) [7]. To check the speaker of one communication, the adversary should collect the traces of communication first. And also calculate the likelihood of the traced speech. On comparing with previous attacks, the proposed attack can detect the speaker of encrypted speech communication with more probability and accuracy. In comparing with previous traditional attacks, the proposed attacks are differentiated by the followings: Packet time information and Packet size information. The packet size information cannot get throw the traffic analysis. Because the voice packets are generated with same size by implementing the Constant Bit Rate (CBR) and also the packets in anonymity network will be of same size, such networks are Tor, and this is because to prevent the traffic analysis attacks.

The summarizations made from the above information are:

- These are passive attacks and these uses the HMM tools. It is a powerful tool to model the temporal data.
- The participants in this communication are at least 20 hops away from each other and end-to-end delay must be 80ms at minimum.
- The encrypted traffic will be in a small amount.
- Here the Intersection attacks also used to improve the effectiveness and efficiency of the attacks.

II. PROBLEM DEFINITION

We focus on detecting speakers of encrypted speech communication by analyzing the talk patterns. The application level patterns can be recovered from the network traffic. The goal of speaker detection is to find the speaker of one specific speech such as a presenter of presentation through audio cast [1] or an instructor of e-learning course. In example, the speaker to be trace is Ram. The adversary collects the traces of Ram's previous speech in advance. The speech may be in the form of encrypted speech communication. By using these traces, one can detect whether Ram is the speaker of the given speech or not. This can be done by comparing the traces of specific speech with previous traces of a single speaker. The traces which are used for speaker detection are collected at different time and in different network. The speech can also make with different voice in various forms.

A. Network Model

The encryption can be done with the secure versions of RTP [2] protocol such as SRTP [3] or ZRTP [2] used in Zfone. These are used to protect confidentiality of speech. For better privacy, Ram routes his encrypted speech through anonymity network. And for voice quality, he use low-latency network such as Tor and JAP. Despite of the existing traffic analysis attacks [8], low-latency anonymity network [10] still effectively protect communication confidentiality and anonymity. The attacks launched by the adversary will not disturb the existing network traffic. Hence the proposed attacks are harder to be detected. The traces which are used for training can be collected by access the links connected to the callers. This is used in the traffic analysis attacks on anonymity network and tracing VoIP [9]. The threat model is defined for traditional privacy-related traffic analysis attacks. This model does not require simultaneous access to links; such links are connected to participants with their communication. Let the adversary collect the speech communication traces in advance. The collections can be used for detect the speaker. This can be done by comparing the traces with current speech. In these speech packets, the packet size information is not available for traffic analysis. Because of the CBR codec's [6] and packet encryption format. The packet encryption will prevent the content in packet from the adversary. Hence the adversary is available with packet timing information only. This is used to launch the privacy attacks.

III. FORMAL DEFINITION

The proposed traffic analysis attack is to find the speaker of speech communication from the pool of traces. The communication trace will compared with the candidate traces. The identification of speaker can be evaluated with the rate of detection, false positive rate and false negative rate.

A. Detection Rate

The ratio of the number of successful detections to the number of attempts is called as the Detection Rate. The detection is successful is the trace is available in the traced pools and the term attempt is defined as trial of identification.

B. False Negative Rate

Ram's speech in communications which are detected as speech communication can be made by other speakers.

C. False Positive Rate

Sometimes, the communication of other speakers may detect as Ram's speech by mistake.

IV. DETECTION OF SPEAKER

The packet of proposed attack contains Packet timing information only. Hence the silence suppression technique can used to extract the talk patterns by talk spurts and silence gaps. From the talk patterns which are recovered. We can create a Hidden Markov Model (HMM). Whenever the adversary wants to determine the trace, he can check the talk patterns recovered from candidate trace against the model.

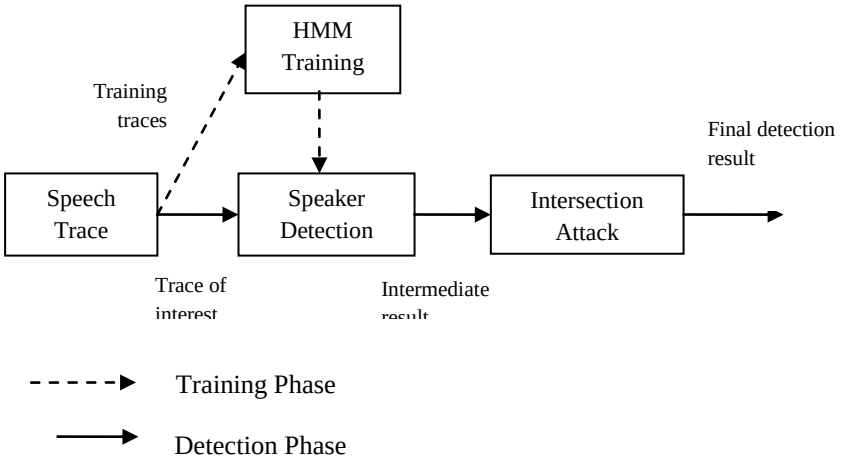

Fig. 1 Steps of proposed attacks

There are two phases in the proposed attacks. They are, Training phase and Detection phase. Each step has their specific features. The Training phase is responsible for HMM training for the previous traces. The detection phase is responsible for searching and comparing the current trace from the pool. This includes the following steps: Speaker detection and intersection attack. Although the intersection step is optional, the adversary can use this step to improve the efficiency and accuracy of the final detection result. The final result will be more accurate than previous attacks.

A. HMM Training

The input for this step is feature vectors and the output is trained HMM. The HMM is a powerful tool with a property called Markov property. This property performs that the transition from the current state to the next state, which depends only on the current state. The output for the model is not visible for outer world, but the states are observed. Each state has a probability distribution over the possible output. Hence, the sequence of outputs which are generated by HMM provides more information about the sequence of states. This tool is used to model the temporal data and also used in temporal pattern recognition. The pattern recognition includes speech recognition, hand-writing recognition and gesture recognition. The adversary considers each talk period including one talk spurt and the following silence gap as an invisible state. The observation which made from the previous output is the length of a talk spurt and the length of the

silence gaps. The trace of speech communication is process through these hidden states, because each state in the model corresponds to a talk period.

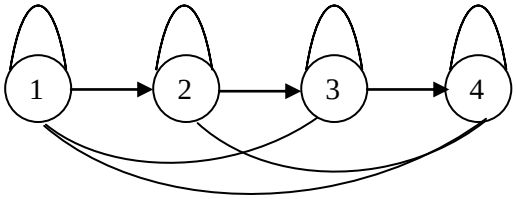

Fig. 2 HMM

The traffic analysis attack uses the modified left-right HMM as shown in the above fig. These are based on the left-right models because of the nonergodic nature of the speech signals. The attributes of the signal whose properties are change over time. The speaker specification model can be obtained by training the HMM with the traces. These can be used in the speaker detection phase.

B. Speaker Detection

The input given to this step is Ram's HMM trained in the previous step and the output from this step is the intermediate detection result. There are two phases in the detection step: The feature vector is calculated with the HMM trained and the trace with the highest match, which will declare as the trace of Ram. Sometimes, the intersection step will be used. If it so, then the result from intersection will be treated as the final detection result. It is optional and used to improve the accuracy of result. In this step, the intermediate results can be fed into the optional intersection attack.

C. Intersection Attack

The detection accuracy can be improved by this step. The input for this step is the intermediate result from the speaker detection and the output taken from this step is the final detection result. The main idea as: we can improve the accuracy with number of trials and final result can be determined by combining the results from all the trials, instead of making decision from a single trial.

In other way, the adversary can make the detection by using the cross-codec detection method.

1) *Cross-Codec Detection:* In this method, the traces and the training traces from HMM are generated with some codec's. This detection is more trustable because, (i) by implementing practically, the traces and trained traces will be generated on different time by different speakers. Hence they must be in different codec's. (ii) Since the speech packets are encrypted and padded to a fixed length, the adversary cannot able to differentiate the speech communication made with the different codec.

V. ARCHITECTURE OVERVIEW

Here we can evaluate the effectiveness of the proposed traffic analysis attacks.

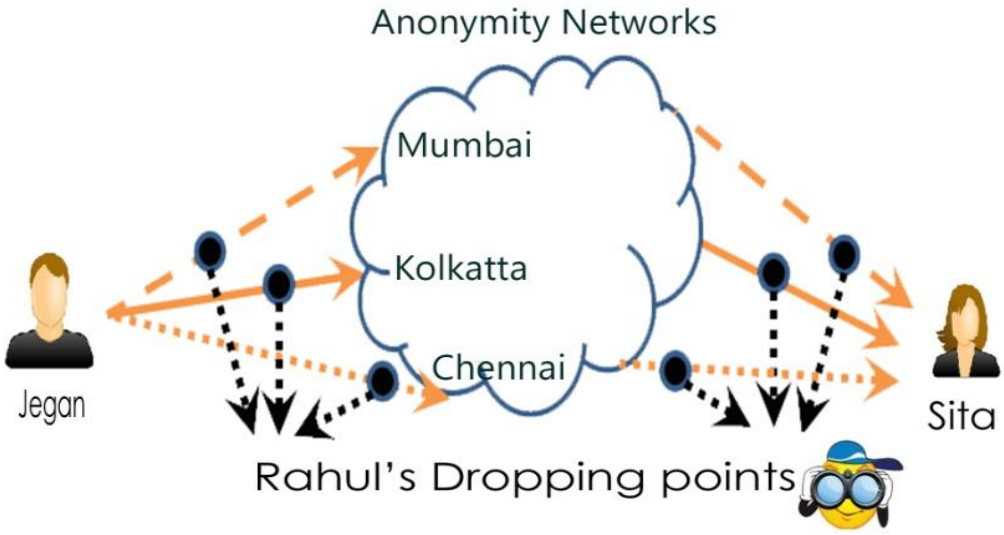

Fig. 3. Architecture of proposed attack.

Here the speech packets are directed to the anonymity network. This can be done by using the website findnot.com before reaching the receiver side or by using the rerouting concept. This concept usually routes the packet in a randomly selected and usually longer path instead of shortest path. Speaker may use the commercial anonymous communication service which is provided by the findnot.com because it is possible to select every point in the anonymous network by the speaker. In the architecture given above, speech packets are directed from the cities like Mumbai, Kolkata and Chennai. These communications are made through anonymous network. Hence the end-to-end delay is 80ms at least. And also the participants in the communication are at least 20 hops away from each other. A small part of the speech communication

is made through the campus network. So the traces of speech communication are collected from different type of networks which are available.

Sometimes, the adversary may need to trace the audio signals form video. The audio signal extracted from video's hosted on the Research Channels. These audio signals which are extracted can be downloaded from the corresponding sites. The traces which are used for both training and detection should be 14.7 minutes long on average, but it is not specified. For fair comparisons, traces used should contain the same number of talk periods. Feature vectors generated from these traces should be of the same size. Because, the different in length of the talk periods in different traces, traces used in experiments are of different length in minutes. At least three speeches are traced for a single speaker and the speech packet must send through at least four different network entry points. The campus entry point is one of the choices. All the traces which are used for speaker detection can be collected on the link of Ram's computer. The traces on detection phase can be taken from the link in the path from sending side to receiver side. The timing of packets which is collected from the receiver side will be noisiest in nature. This is because the accumulated randomness of the network delays. The detection rate for candidate traces collected from the sending end is comparable with the detection rate for candidate trace collected from receiving side. This is followed because of the technique called Filtering. This filtering technique is mainly used in the silence test can largely filter out noise caused by the random network delay at the receiving end. For better training, the training traces should be longer than test tracing.

VI. CONCLUSIONS

The above passive traffic analysis attacks are proposed to compromise the privacy of speech communications. These attacks are based on application-level and feature extracted from the traces of speech communication. The proposed attacks are evaluated by performing the extensive experiments over different networks. The networks may be a commercial anonymity network or may be a campus network. It shows that the proposed attacks can detect the speakers of the encrypted speech communications with high detection rates based on speech communication traces of 15 minutes long on average.

REFERENCES

[1] S. Casner and S. Deering, *"First ietf Internet Audiocast," SIGCOMM Computer Comm. Rev.*, vol. 22, pp. 92-97, http://doi.acm.org/10.1145142267.142338, July 1992.

[2] P. Zimmermann, A. Johnston, and J. Callas, *"Zrtp: Media Path Key Agreement for Secure rtp Draft-Zimmermann-Avt-Zrtp-11,"* RFC, United States, 2008.

[3] M. Baugher, D. McGrew, M. Naslund, E. Carrara, and K. Norrman, "*The Secure Real-Time Transport Protocol (srtp),*" 2004.

[4] R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The Second-Generation Onion Router," *Proc. 13th USENIX Security Symp.*, pp. 303-320, Aug. 2004.

[5] O. Berthold, H. Federrath, and S. Köpsell, *"Web MIXes: A System for Anonymous and Unobservable Internet Access," Proc. Designing Privacy Enhancing Technologies: Workshop Design Issues in Anonymity and Unobservability*, H. Federrath, ed., pp. 115-129, July 2000.

[6] J.M. Valin, *"Speex: A Free Codec for Free Speech," Proc. Australian Nat'l Linux Conf.*, 2006.

[7] C. Rathinavelu and L. Deng, *"Hmm-Based Speech Recognition Using State-Dependent, Linear Transforms on Mel-Warped dft Features," Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing (ICASSP '96)*, pp. 9-12, 1996.

[8] Y. Zhu, X. Fu, B. Graham, R. Bettati, and W. Zhao, *"Correlation-Based Traffic Analysis Attacks on Anonymity Networks," IEEE Trans. Parallel and Distributed Systems*, vol. 21, no. 7,pp. 954 -967, July 2010.

[9] X. Wang, S. Chen, and S. Jajodia, *"Tracking Anonymous Peer-to-Peer Voip Calls on the Internet," Proc. ACM Conf. Computer and Comm. Security*, pp. 81-91, Nov. 2005.

[10] B.N. Levine, M.K. Reiter, C. Wang, and M.K. Wright, *"Timing Attacks in Low-Latency Mix-Based Systems," Proc. Eighth Int'l Financial Cryptography (FC '04) Conf.*, pp. 251-265, Feb. 2004.

