Computers and Electrical Engineering 37 (2011) 559–569

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Crypt analysis of two time pads in case of compressed speech ☆

L.A. Khan [a,*], M.S. Baig [b], Ali Hassan [c]

[a] College of Signals, National University of Sciences and Technology (NUST), Islamabad, Pakistan
[b] Centre for Advanced Studies in Engineering, Islamabad, Pakistan
[c] Centre for Cyber Technology and Spectrum Management, NUST, Islamabad, Pakistan

ARTICLE INFO

Article history:
Received 13 February 2009
Received in revised form 29 March 2011
Accepted 4 April 2011
Available online 6 May 2011

ABSTRACT

Keystream reuse, also known as the two time pad problem, is a well known weakness in stream ciphers. The implementers of the cryptographic algorithms are still underestimating this threat. The keystream reuse exploitation techniques presented so far assume the underlying plaintext to be textual data and all the heuristics presented previously are based on the language characteristics of the underlying text based data, which fail when compression is applied on the plaintext before encryption. This paper presents exploitation techniques for two time pads in case of stream ciphered digitized and compressed speech signals. In this paper we show that how an adversary can automatically recover the digitized speech signals encrypted under the same keystream provided the language (e.g. English) and digital encoding/compression scheme details of the underlying speech signals are known. Our technique of cryptanalysis is based on the modeling of the speech parameters by exploiting the inter frame correlations between each components of the speech vectors in different frames and then using these models to decode the two speech signals in the keystream reuse scenario. The technique is flexible enough to incorporate all modern speech coding schemes based on parameter or hybrid encoding and compression techniques. The simulation experiments have showed promising results for most of the present day speech digitization and compression techniques.

1. Introduction

The key reuse problem in stream cipher has been the focus of cryptanalysts for quite some time and has also been mentioned in the literature as the "two time pad" problem [1]. It can simply be described as: if two different plaintexts p and q are encrypted with the same keystream k in an additive stream cipher then their ciphertexts $c_1(p \oplus k)$ and $c_2(q \oplus k)$, once bitwise XORed, neutralize the effect of the keystream and produce $p \oplus q$. The vulnerability of keystream reuse exists with many practical systems which are still in use. Some practical systems which are vulnerable to such type of attacks include Microsoft Office [1–3]; 802.11 Wired Equivalent Privacy (WEP) [4–6]; WinZip [7], the point to point tunneling protocol (PPTP) [8], encrypted data storage applications [9], etc. This problem has not been only underestimated till the recent past but is predicted to remain there for quite some time in the future also because of the endorsement of Advanced Encryption Standard (AES) counter mode by National Institute of Standards and Technology (NIST) [1]. Moreover, an effective and secure key management has still been an uphill task for the crypto designers and we still have a hard time getting the key management right [10].

The keystream reuse problem discussed previously is with respect to the textual data and all the exploitation techniques presented so far, starting from the famous (rather notorious) *VENONA* Project of NSA [11,12] followed by Rubin [13], Dawson and Neilson [14], Mason et al. [1] and Ekera and Terelius [15], assume the underlying plaintext to be uncompressed text-based

☆ Reviews processed and proposed for publication by Associate Editor Dr. Ferat Sahin.
* Corresponding author. Address: D 50/10 E-9/1 PAF Complex, Islamabad, Pakistan. Tel.: +92 51 9506965; fax: +92 51 9506909.
E-mail address: liaquatalikhan@gmail.com (L.A. Khan).

0045-7906/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compeleceng.2011.04.008

data digitally encoded with the standard coding schemes such as ASCII coding. We have tried to address this problem from the speech signals point of view. The proposed work is the first ever exploitation technique for keystream reuse in case the underlying data is compressed digitized speech signal. All the previous works [1,11–15] consider the underlying data as text files whereas, our previous work [16] deal with uncompressed speech. In all the techniques, prior to our work, the properties of text based data with ASCII coding are used to extract $p(text)$ and $q(text)$ from $p \oplus q$ (random data). In this paper, we present techniques to extract p(compressed speech) and q(compressed speech) from $p \oplus q$ (random data). The approach presented in this paper is based on language modeling, hidden Markov models (HMMs), and speech synthesis. The use of HMMs, although having a very rich mathematical structure [17,18], is a relatively new concept in cryptography and cryptanalysis. HMMs have recently been used for several problems in these areas. For example in [19], standard Markov modeling techniques are used to reduce the search space of human memorable passwords; in [20,21], HMMs are used to model key stroke timings to deduce key sequences and password lengths in Secure Shell (SSH) protocol; in [22], HMMs are used to decipher simple substitution ciphers in the presence of noise; in [23], keyboard acoustic emanations are modeled using HMMs to recognize the typed characters from the sounds of the keys; in [24], randomized counter measures against side channel cryptanalysis are modeled as HMMs to recover the keys in case of hardware based implementation of cryptographic algorithms; in [25], the packet lengths in case of encrypted voice over IP communication are modeled as HMMs to deduce the identity of the speaker. In our case, HMMs are employed to model the inter frame correlation of different parameters in the Code Exited Linear Prediction (CELP) coded speech signals. These models are then applied in a special case of stream enciphered speech where two different speech files are enciphered by the same keystream. We have used hidden Markov models in a completely different context from the previous works. The contributions of our work can be enumerated as under:

- The introduction of new exploitation techniques for two time pad for compressed speech signals based on hidden Markov models for which no exploitation techniques existed previously to the best of our knowledge.
- Application of language modeling techniques for modeling the inter-frame correlations between different parameters in a CELP coded speech signal.

The rest of the paper is organized as follows: In Section 2, we discuss the background information with relevance to our work. Section 3 presents details about the proposed approach. In Section 4 we present the implementation details for the proposed approach along with experimental results and complexity analysis. Section 5 concludes the paper and gives directions for future work.

2. Background information

In this section, we present some background information relevant to our model of attack on the two time pads in case of compressed speech signals.

2.1. Speech encoding and compression

In speech coding, the most important criterion is preservation of intelligibility and quality of speech, with a constrained amount of transmitted data. Mainly two types of speech encoding techniques exist, namely waveform encoding and parameter encoding. A third type is a combination of the previous two types of encoding mechanisms, known as the Hybrid encoding. In the waveform encoding the speech signal waveform is sampled, quantized (and compressed) and then digitally encoded. The A-law and μ-law algorithms [26] used in traditional Pulse Coded Modulation (PCM) digital telephony can be seen as early precursors of speech digitization based on waveform encoding. In case of parameter encoding speech is represented as a source filter model in which the parameters of the model along with excitation information are used for digital representation of speech. The most common coding technique in this domain is CELP, which is the most widely used speech coding algorithm. It is for this reason that we have selected a particular implementation of CELP for our attack on parameter/hybrid encoded speech. Different standard implementations of the CELP algorithms are available and are in use in most of the modern speech processing applications.

2.2. Language models and smoothing

Although our use of the language models will be in a total different context as compared to the Natural Language Processing (NLP) community, yet the basic concepts can still be borrowed from them and adapted to our situation. A language model is a statistical model which assigns a probability to each word/alphabet of the language through a probability distribution. Language models are used in many natural language processing applications such as speech recognition, machine translation of speech, parts of speech tagging, information retrieval etc. Language models developed from even very large corpora of text based data do not incorporate all possibilities as certain sequence may not be observed during the training process. In order to solve this inherent problem with language models, they are often approximated using smoothed *n-gram* models. Smoothing is a technique used to estimate better probabilities when data available is not sufficient to estimate accurate probabilities [27]. Out of the many smoothing techniques available we used additive and Witten Bell smoothing [27] techniques during