Bandwidth Optimal Steganography Secure Against Adaptive Chosen Stegotext Attacks

Tri Van Le[1] and Kaoru Kurosawa[2]

[1] Department of Computer Science
Florida State University
Tallahassee, Florida 32306-4530, USA
levan@cs.fsu.edu
[2] Department of Computer and Information Sciences
Ibaraki University 4-12-1
Nakanarusawa, Hitachi, Ibaraki 316-8511, Japan
kurosawa@cis.ibaraki.ac.jp

Abstract. We provide construction of steganographic schemes secure against adaptive chosen stegotext attacks. Our constructions achieve embedding rate equals to the Shannon entropy bound on steganographic channel capacity. Further the covertext distribution can be given as either an integrable probability function or as a random covertext sampler. We also introduce steganographic codes that are of interests in constructing other steganographic protocols such as steganographic secret sharing or steganographic distributed computations.

Keywords: bandwidth, information hiding, steganography, adaptive chosen stegotext attack.

1 Introduction

Definition. The *Prisoner's Problem* introduced by G.J. Simmons [14] and generalized by R. Anderson [1] can be stated informally as follows: Two prisoners, Alice and Bob, want to communicate to each other their secret escape plan under the surveillance of a warden, Wendy. In order to pass Wendy's censorship, Alice and Bob have to keep their communications as innocent as possible so that they will not be banned by Wendy.

Motivation. A fundamental question to steganography is what are the limits of provably secure steganography? We answer this question constructively and positively by constructing provably secure schemes with extremely low overhead. We prove that our schemes are secure and essentially optimal. For covertext distributions that support high bandwidth (e.g. thousands of bits per cover), our schemes achieve this bandwidth (Section 5) and are several orders of magnitude better than all previously known secure schemes.

Our schemes are very flexible in that they can work with either an integrable probability function or a random covertext sampler. Their security can be chosen

J. Camenisch et al. (Eds.): IH 2006, LNCS 4437, pp. 297–313, 2007.

in the information theoretic setting or in the computational complexity theoretic setting and are proved in the corresponding setting. In the information theoretic setting, we show matching bounds for both cases of probability function and covertext sampler. In the computational complexity theoretic setting, matching bound is proved only for the most general case of random covertext sampler. Our results show that a probability model of the covertext distribution is sufficient for practical secure steganography, regardless of the security setting.

Discussion. We solve the steganographic problem in a novel way. At the heart of our solution are uniquely decodable variable length coding schemes Γ, called $\mathcal{P}$-codes, with source alphabet Σ and destination alphabet C such that: if $x \in \Sigma^\infty$ is chosen uniformly randomly then $\Gamma(x) \in C^\infty$ distributes according to $\mathcal{P}$, where $\mathcal{P}$ is a given distribution over sequences of covertexts.

Note that such a coding scheme is quite related to homophonic coding schemes [7], which are uniquely decodable variable length coding scheme Γ' with source alphabet C and destination alphabet Σ such that: if $c \in C^*$ is chosen randomly according to distribution $\mathcal{P}$ then $\Gamma'(c) \in \Sigma^*$ is a sequence of independent and uniformly random bits.

Of course, one can hope that such a homophonic coding scheme Γ' will give rise to a uniquely decodable $\mathcal{P}$-code Γ. However, this is not necessarily true because Γ' can map one-to-many, as in the case of [7]. Therefore by exchanging the encoding and decoding operations in Γ', we will obtain a non-uniquely decodable $\mathcal{P}$-coding scheme Γ'', which is not what we need.

To construct these $\mathcal{P}$-codes, we generalize an idea of Ross Anderson [1] where one can use a perfect compression scheme on the covertexts to obtain a perfectly secure steganographic scheme. Nevertheless, in practice one can never obtain a perfect encryption scheme, so we have to build our $\mathcal{P}$-coding schemes based on the idea of arithmetic coding. The result is a coding scheme that has near optimal information rate, no decoding error and provable security.

Related work. Previously, the Prisoner's Problem was considered in the secret key setting by: Cachin [3], Mittelholzer [11], Moulin and Sullivan [12], Zollner et.al. [16] in the unconditional security model; and Katzenbeisser and Petitcolas [10], Hopper et.al. [8], Reyzin and Russell [13] in the conditional security model. In this article, we consider the problem in the *public key* setting. In this setting, Craver [4] and Anderson[1] proposed several general directions to solve the problem. Katzenbeisser and Petitcolas [10] gave a formal model. Hopper and Ahn [9] constructed proven secure schemes, and then modified it in [15] to remove the dependence on unbiased functions [3]. Michael Backes and Christian Cachin [2] have been able to improve efficiency of Hopper and Ahn's scheme by some factor. Nevertheless all the approaches outlined above have very high overhead and extremely low bit rate [3,8,9,13,2]. In some cases, the bit rate is less than a hundredth of a bit per cover.

Organization. The paper is organized as follows: we describe the model in Section 2, our new primitive $\mathcal{P}$-codes in Section 3, show constructions of public