Low Bandwidth Dynamic Traitor Tracing Schemes

Tamir Tassa
Division of Computer Science,
The Open University,
Raanana, Israel
tamirta@openu.ac.il

Communicated by Dan Boneh
Received 7 June 2002 and revised 20 June 2004
Online publication 21 October 2004

Abstract. Dynamic traitor tracing schemes were introduced by Fiat and Tassa in order to combat piracy in active broadcast scenarios. In such settings the data provider supplies access control keys to its legal customers on a periodical basis. A number of users may collude in order to publish those keys via the Internet or any other network. Dynamic traitor tracing schemes rely on the feedback from the pirate network in order to modify their key allocation until they are able either to incriminate and disconnect all traitors or force them to stop their illegal activity. Those schemes are deterministic in the sense that incrimination is always certain. As such deterministic schemes must multiply the critical data by at least $p + 1$, where $p$ is the number of traitors, they may impose a too large toll on bandwidth. We suggest here probabilistic schemes that enable one to trace all traitors with almost certainty, where the critical data is multiplied by two, regardless of the number of traitors. These techniques are obtained by combining dynamic traitor tracing schemes with binary fingerprinting techniques, such as those proposed by Boneh and Shaw.

Key words. Broadcast encryption, Traitor tracing, Fingerprinting, Watermarking, Binary codes, Pay-TV, On-line algorithms.

1. Introduction

The concept of Dynamic Traitor Tracing was introduced in [7] in order to combat piracy in dynamic data distribution systems. This was a natural extension of the concepts of Traitor Tracing, Watermarking and Fingerprinting that were relevant for static data distribution systems. A data distribution system is any setting in which a data provider (the center) is providing data to a large group of paying users. Piracy occurs when one or few legal users redistribute the data for which they paid to illegal customers, thus rendering financial losses to the legitimate data provider. Hence, tracing techniques are required in order to find the source of information leakage, disconnect it and press charges against those traitorous users.
In order to distinguish between the static and dynamic models, we refer henceforth to the legal data distribution, from the center to the paying users, as the primary data distribution, while the illegal data distribution from the traitors to the illegal customers is referred to as the secondary data distribution.

A static data distribution system is one in which the primary data distribution completely precedes in time the secondary data distribution. Namely, primary and secondary data distributions are perceived as actions that have no duration: the first occurs at some time $t_1$ and the second occurs at a later time $t_2 > t_1$. Assume that the data provider captures such secondary distributed material. Obviously, if all legal users get the very same data, there is no way of tracing back one of the traitors, given the pirate material. Hence, it is necessary to personalize the data prior to primary distribution in such a way that any captured pirate material would enable the tracing of at least one of the colluding traitors. Traitor tracing schemes, watermarking, fingerprinting, traceability schemes and parent-identifying codes are different terms that were given to techniques that were designed to answer this need. One of the reasons for the plurality of terms is the diverse scenarios that give rise to this problem. In the context of decoders for Pay-TV, one usually thinks in terms of “traitor tracing”. On the other hand, schemes that are designed to protect visual material are usually called “watermarking” or “fingerprinting”. “Traceability schemes” or “parent-identifying codes” are the mathematical tools that such schemes use.

In view of the above, we feel that it is imperative to set a general framework that encompasses all scenarios of data distribution and all counter-piracy methods.

The Traitor Tracing Framework. Let $U = \{u_1, \ldots, u_n\}$ denote the set of users and $T = \{t_1, \ldots, t_p\} \subset U$ be the subset of traitorous users. $T$ is called the pirate while its members are referred to as traitors. An algorithm that aims at locating the traitors is called a traitor tracing scheme. Such a scheme consists of the following ingredients:

- A marking alphabet, $\Sigma = \{\sigma_1, \ldots, \sigma_r\}$.
- A codebook $\Gamma$ that is used for personalizing copies of the data prior to its distribution. $\Gamma$ is an $(r, \ell, n)$-code, meaning that $\Gamma = \{w^1, \ldots, w^n\} \subset \Sigma^\ell$, $|\Sigma| = r$.
- A personalization scheme $P: U \rightarrow \Gamma$ that determines how to mark the data that is provided to a particular user with a codeword from the codebook.
- A generation assumption: let $P(T) = \{P(t_1), \ldots, P(t_p)\}$ denote the set of codewords in the copies that are owned by the traitors. We let $\langle P(T) \rangle$ denote the set of codewords that could be generated by the pirate and be placed in the pirate copy. $\langle P(T) \rangle \subset (\Sigma \cup \{?\})^\ell$, where “?” denotes an unreadable mark. Different assumptions were made in different studies about the strength of the generation operation $\langle \cdot \rangle$, depending on the underlying application.
- A tracing algorithm that, given a pirate copy, aims at tracing back (at least) one traitor that collaborated in producing that copy. This algorithm may be therefore viewed as a function $A: \langle P(T) \rangle \rightarrow U$. Obviously, a desired property is that whenever $w \in \langle P(T) \rangle$, $A(w) \in T$. However, this is not always the case, as indicated by the next ingredient.
- An upper bound on the probability of false incrimination, $\varepsilon \geq 0$. In other words, if $w \in \langle P(T) \rangle$, then $\text{Prob}(A(w) \in U \setminus T) \leq \varepsilon$. Traitor tracing or watermarking schemes in which $\varepsilon = 0$ are called deterministic while those in which false incrimination may occur are called probabilistic.