Tardos符号の検出に関する研究

セキュリティ情報学　山口研究室　研究紹介

デジタルコンテンツの保護

近年，インターネット技術の発達，普及することによって，音楽などのデジタルコンテンツが様々な分野において利用されています．しかし，デジタルコンテンツはコピー，編集が容易であるため，著作権問題が深刻化しています．この問題に対し，利用者の識別情報 (ID) をコンテンツに埋め込む電子指紋方式 (Digital Fingerprinting あるいは Transaction Tracking) が提案されています．電子指紋方式は埋め込む ID 情報を特別な検出器を使用しないと認識できないように ID 情報を埋め込む方式です．

悪意のある利用者集団「結託攻撃」

電子指紋方式で埋め込まれた ID はコンテンツのデータからでは認識できないが，複数の利用者のデータを比較すると，元々のコンテンツのデータは同一のものであるため，データの違いは ID 部分と特定でき，単純な編集によって，電子指紋方式を無効化することができます．この複数の悪意のある利用者が互いのコンテンツデータを比較して，改竄する攻撃を「結託攻撃」といいます．

改竄された ID から結託者を特定「結託耐性符号」

結託攻撃は攻撃に利用したコンテンツデータと非常に強い関係性を持つ．この関係（相関）を利用してある人数以下の結託者をある誤る率で検出可能な符号を「結託耐性符号」(collusion-secure code) という．結託耐性符号は結託攻撃に対して非常に耐性があるが，一方で他の符号に比べて，符号長が冗長になってしまう．そのため結託耐性符号の符号長を短縮する研究が行われています．

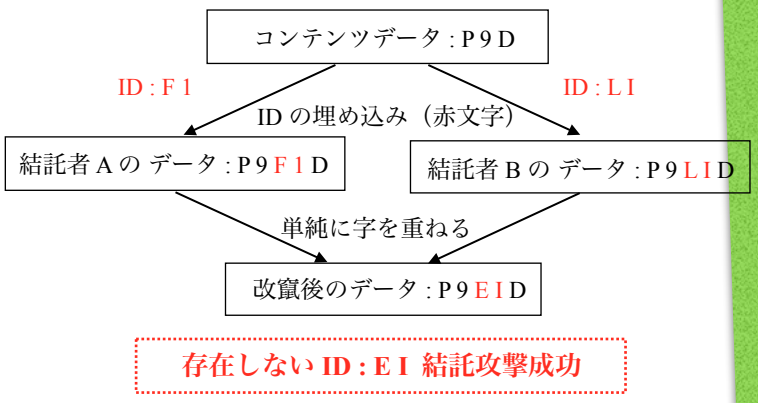

図 1 : 結託攻撃の一例

表 1 : 結託耐性符号の符号長

ユーザ数 100　誤り率 0.001 の場合	
Boneh-Shaw 符号	24,168,000
c-secure CRT 符号	241,000
Tardos 符号	115,000

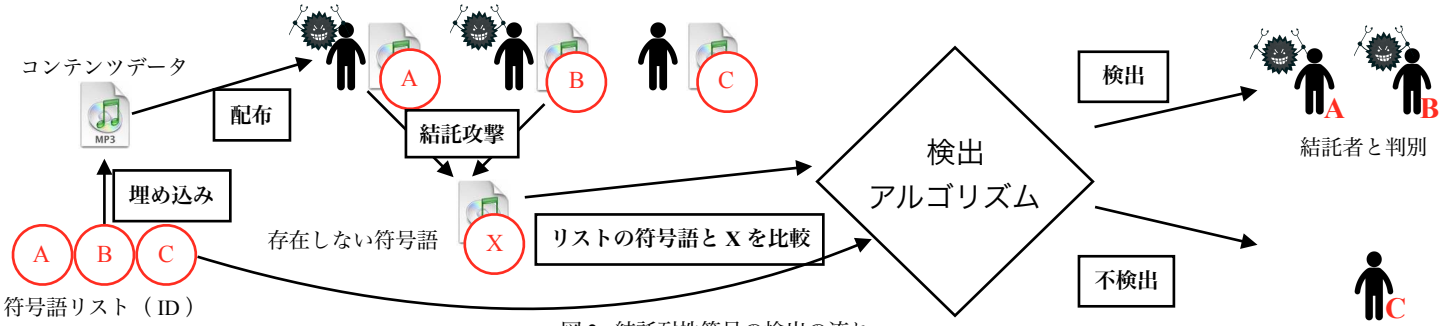

図 2 : 結託耐性符号の検出の流れ

Tardos符号

Tardos符号は結託耐性符号の中で改竄に強い符号と期待されていて，その符号はビット列で構成されています．符号語の各要素ごとに独立に連続型確率分布 P に従った確率 $\Pr[x=1]=p$ で生成し，この確率 p と改竄によって生成された語を利用して，全ての要素の疑わしさの合計をスコアとして計算し，このスコアが閾値を超えた場合，そのユーザを結託攻撃者として検出します．

Tardos符号のスコア分布特性

Tardos符号は一般ユーザのスコアが低く分布し，結託攻撃者のスコアは高く分布をする．この一般ユーザの分布と結託攻撃者の分布が重なるとき，一般ユーザを結託攻撃者として告発するfalse positiveが発生し，また，結託攻撃者のスコアが高い分布を取らないとき，結託攻撃者の告発ができないfalse negativeが発生します．本研究はスコアの符号語の要素数分の合計値によって求められることに着目し，スコアの分布が符号長と密接に関係があると考え，その関係をシミュレーションにより求めることにより，スコア分布を推測し，それから閾値を一般ユーザの分布ギリギリに設定する手法を提案しました．

今後の研究

提案した手法は検出率に関してはTardos符号を上回っていますが，その誤り率に関しては，理論的な評価ができていません．したがって，誤り率の評価を考慮した新たな閾値の設定方法を考察しようと考えています．

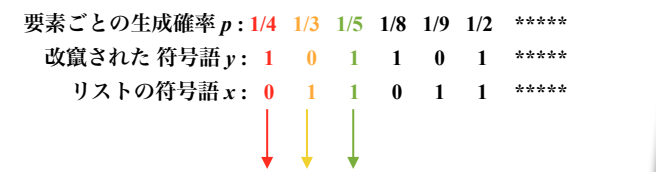

図 3 : Tardos符号のスコア計算

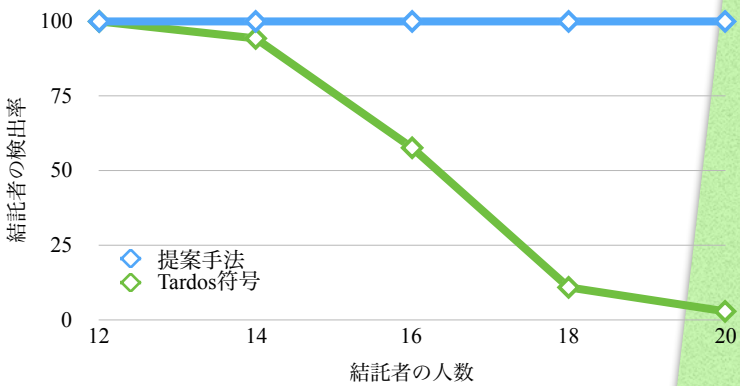

図 4 : 10人まで対応可能な条件での検出性能比較