Asymptotic channel capacity of collusion resistant watermarking for non-binary alphabets

ir. Dion Boesten, dr. Boris Škorić

Introduction
Forensic watermarking is a technique to combat the unauthorized distribution of digital content. A unique and imperceptible watermark is embedded into each copy. The main challenge in this area is to create codes that are resistant to collusion attacks. In a collusion attack a group of users (pirates) compare their copies and try to create an untraceable copy. Information theory can provide a lower bound on the required code length of a reliable collusion resistant code.

Bias based codes
- Bias vectors $\mathbf{p}^{(j)}$ are drawn from a distribution $F$
- Codeword symbol $X_{ij}$ is generated stochastically using bias $\mathbf{p}^{(j)}$: $\Pr[X_{ij} = a] = p_{ij}^{(j)}$

Restricted digit model
- Pirates choose one of the symbols from the ones they have seen in each segment
- By the Marking Assumption they cannot modify a segment in which they all received the same symbol

Accusation
An accusation algorithm compares the detected watermark with the matrix $X$ to produce candidate coalitions. This is most often done by assigning scores to individual users and then applying a threshold.

Attack channel
- counter variable $\mathbf{z}$
- pirate strategy $\theta$
- output symbol $Y$
- Optimal attack is segment independent
- Attack can be seen as noise on a communication channel
- $\Sigma_\alpha$ counts the number of $\alpha$'s that the pirates received
- Channel output $Y$ is generated using strategy $\theta$ that is allowed to be non-deterministic: $\theta_{y|\mathbf{z}} \triangleq \Pr[Y = y|\mathbf{z} = \mathbf{z}]$

Channel capacity
- The mutual information $I(Y; \mathbf{z} | F = \mathbf{p})$ measures how much information $Y$ reveals about the identity of the pirates (equivalent with $\mathbf{z}$):
  \[ H(\mathbf{z}) I(Y; \mathbf{z} | F = \mathbf{p}) H(Y) \]
- The channel capacity $C_q$ is derived as the optimal value of a max-min game:
  \[ C_q = \max_{\theta} \min_F \int F(\mathbf{p}) I(Y; \mathbf{z} | F = \mathbf{p}) d\mathbf{p} \]

Asymptotics ($c \to \infty$)
- Binary case was solved by Huang and Moulin[2010]:
  \[ \theta^*_{y|\mathbf{z}} = \frac{\sigma_Y}{c} \quad F^*(\mathbf{p}) = \frac{1}{\pi \sqrt{p_0p_1}} \]
- Non-binary case was solved by us:
  \[ C_q \sim \frac{q - 1}{2c^2 \ln q} \]

Discussion
The fingerprinting capacity $C_q$ is an increasing function of $q$. Hence there is an advantage to switch to higher alphabets if the embedding scheme allows this.