UNIVERSITÉ DE GENÈVE

Microsoft Research

Security Analysis of Robust Data-Hiding with Geometrically Structured Codebooks

E. Topak[(a)], S. Voloshynovskiy[(a)], O. Koval[(a)], M. K. Mihcak[(b)] and T. Pun[(a)]

[(a)]Stochastic Image Processing (SIP) Group,
University of Geneva, Switzerland
&
[(b)]Microsoft Research, Redmond, USA

Agenda

- **Problem formulation;**
- **Channels with geometrical attacks;**
- **Information theoretic (IT) framework for geometrically robust data-hiding;**
- **Structured codebooks;**
- **Analysis of security leaks and attacking strategies;**
- **Conclusions;**
- **Future research directions.**

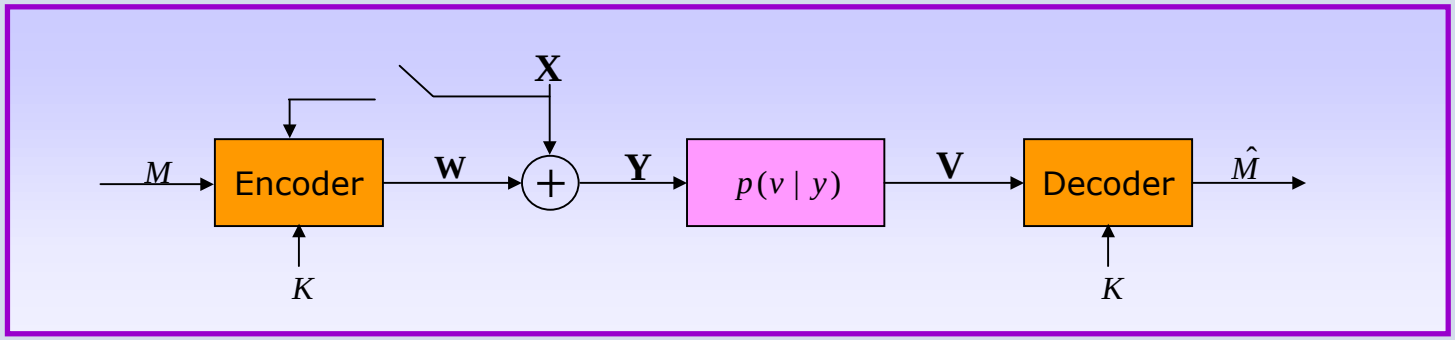
Problem formulation

Objectives:

- **To analyze the conditions of reliable communications in channels with geometrical transformations;**
- **To study capacity achieving geometrically-robust data-hiding codes;**
- **To investigate security leakages of structured codebooks and corresponding attacking strategies.**

Problem formulation

Data-hiding problem: Given $|\mathcal{K}|$ users each with its own key $k \in \{1,2,\cdots|\mathcal{K}|\}$, communicate reliably message $m \in \mathcal{M}$, $\mathcal{M} = \{1,2,\cdots,2^{NR}\}$, embedded in the host image $\mathbf{X} \in \mathcal{X}^N$ through the channel $p(v \mid y)$.



If the host state is taken into account or not in watermark generation:

$\mathbf{W} = \mathbf{W}(M, \mathbf{X}, K) \rightarrow$ **Random binning approach**

$\mathbf{W} = \mathbf{W}(M, K) \rightarrow$ **Random coding approach**

Problem formulation

Performance criterion:

$$P_e^{(N)} = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \Pr\left[\hat{M} \neq m \mid M = m\right]$$

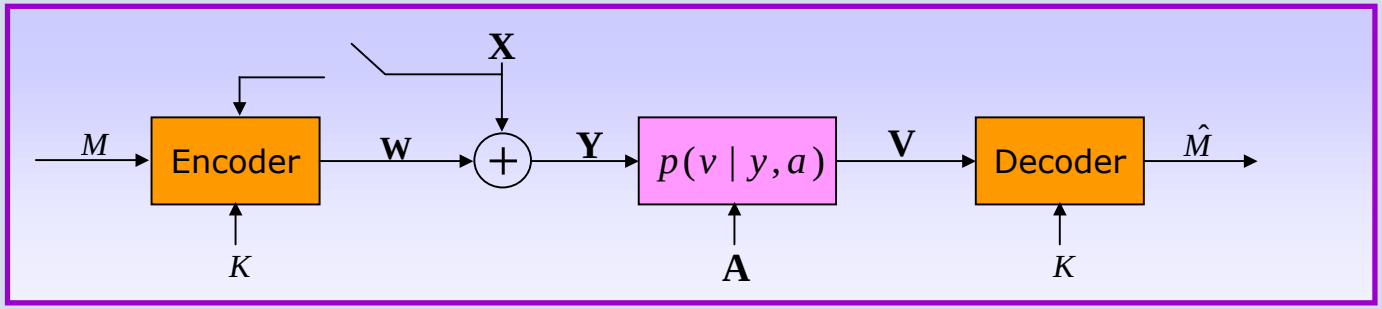
Conditions for reliable communications:

Random coding	$R \leq \frac{1}{N} I(\mathbf{W}; \mathbf{V} \mid K)$
Random binning	$R \leq \frac{1}{N} \left[I(\mathbf{U}; \mathbf{V} \mid K) - I(\mathbf{U}; \mathbf{X} \mid K) \right]$

Theoretical set-up $(N \to \infty)$	$P_e^{(N)} = 0$
Practical set-up $(N < \infty)$	$P_e^{(N)} \neq 0$

Problem formulation

Data-hiding in channels with geometrical attacks



Trade-offs:

Geometrical channels → Synchronization framework

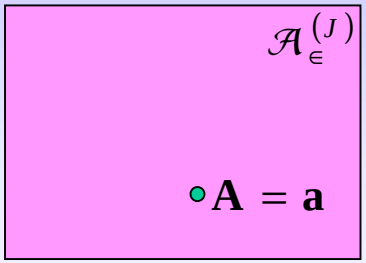
Syncronization framework → Security leakages

Problem: To analyze conditions of reliable communications in the case of geometrical attacks avoiding security leakages

Channels with geometrical attacks

A decoder without a synchronization framework has to perform an exhaustive decoding through all possible geometrical transformations!

<u>Assumption:</u> Applied transformation belongs to the set of typical geometrical transformations:



$$\mathbf{a} = (a_1, \ldots, a_J), \quad a_i \sim p_A(a)$$

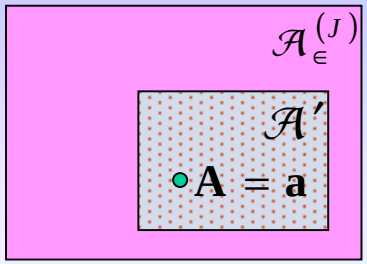
Average probability of error:

$$P_e^{G(N)} = \sum_{\mathbf{a} \in \mathcal{A}_\epsilon^{(J)}} p_{\mathbf{A}}(\mathbf{a}) P_e^{(N)}(\mathbf{a})$$

Theoretical set-up $(N \to \infty)$	$P_e^{(N)}(\mathbf{a}) = 0$	$P_e^{G(N)} \to 0$	**No impact on communications performance in price of increase in decoding complexity**
Practical set-up $(N < \infty)$	$P_e^{(N)}(\mathbf{a}) \neq 0$	$P_e^{G(N)} \to 1$	**Geometrical attacks completely destroy reliable communications**

Channels with geometrical attacks

Data-hider strategy: add synchronization part into the codebook.



Constrained search space: $\left|\mathcal{A}'\right| \leq \left|\mathcal{A}_{\in}^{(J)}\right|$

Average probability of error: $P_e^{G(N)} = \sum_{\mathbf{a} \in \mathcal{A}'} p_{\mathbf{A}}(\mathbf{a}) \tilde{P}_e^{(N)}(\mathbf{a})$

rate loss $\tilde{R}$ **due to synchronization**

Geometrical synchronization based on structured codebooks:

Estimation of the applied geometrical transformation from the attacked data: channel state estimation (CSE)

→

Compensation of the estimate: channel state compensation (CSC)

Channels with geometrical attacks

Structured codebooks

Template-based structured codebooks
(a specially designed template is used to perform CSE and CSC)

Redundant-based structured codebooks
(codewords have special statistics to aid CSE and CSC)

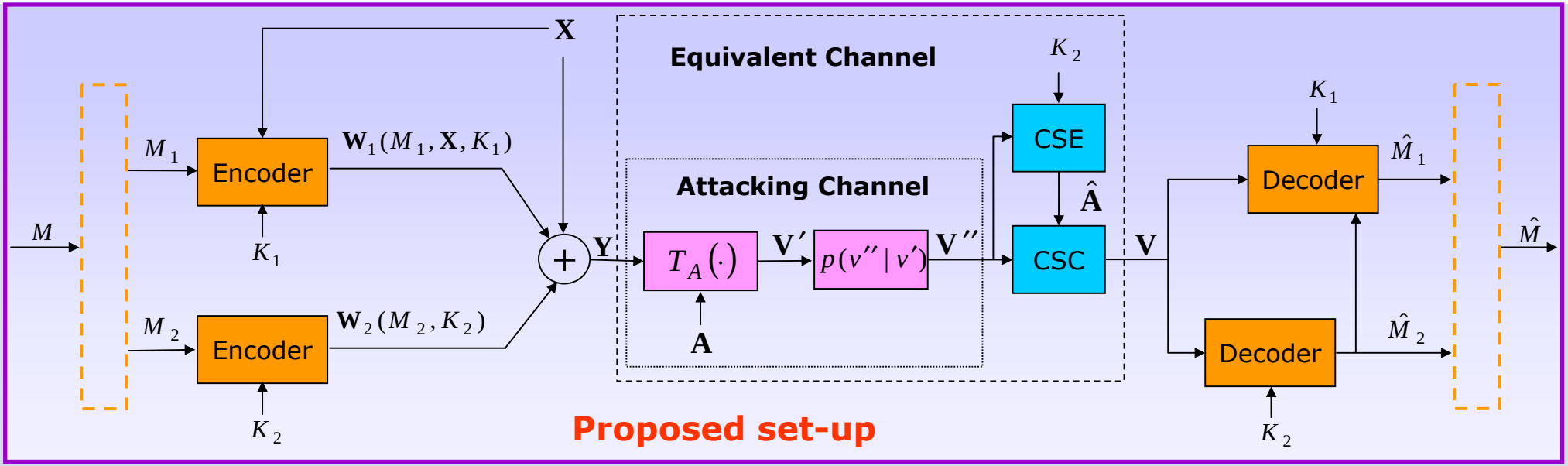
Our objectives

capacity achieving data-hiding
host interference problem to be solved based on random binning dependent on host data

robustness to geometrical attacks
codewords with synchronization features to be generated according to statistics that are independent from those of the host data

Problem: How to combine these conflicting requirements?

IT framework for geometrically robust data-hiding



$\mathbf{W}_1 \rightarrow$ **carries only information about** M_1**.**

$\mathbf{W}_2 \rightarrow$ **has synchronization features using:**

- **redundant-based design,**
- **template-based design.**

Practical implementation principles:

- **CDMA/SDMA signalling**
- **Genie-aided decoding**

(Multistage decoder)

IT framework for geometrically robust data-hiding

A $\left(2^{NR_1}, 2^{NR_2}, N\right)$ code for MAC consists of:

Index sets: $\mathcal{M}_1 = \left\{1,2,\cdots,2^{NR_1}\right\}, \mathcal{M}_2 = \left\{1,2,\cdots,2^{NR_2}\right\};$

Encoding functions:

$$f_1 : \left\{1,2,\cdots,2^{NR_1}\right\} \times \left\{1,2,\cdots,|\mathcal{K}_1|\right\} \times \mathcal{X}^N \rightarrow \mathcal{W}_1^N;$$
$$f_2 : \left\{1,2,\cdots,2^{NR_2}\right\} \times \left\{1,2,\cdots,|\mathcal{K}_2|\right\} \rightarrow \mathcal{W}_2^N;$$

Decoding function:

$$g : \mathcal{V}^N \times \left\{1,2,\cdots,|\mathcal{K}_1|\right\} \times \left\{1,2,\cdots,|\mathcal{K}_2|\right\} \rightarrow \left\{1,2,\cdots,2^{NR_1}\right\} \times \left\{1,2,\cdots,2^{NR_2}\right\}.$$

Average probability of error for $\left(2^{NR_1}, 2^{NR_2}, N\right)$ code:

$$P_e^{(N)} = \frac{1}{2^{N(R_1+R_2)}} \sum_{(m_1,m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} \Pr\left[g(\mathbf{V}, K_1, K_2) \neq (m_1, m_2) \,|\, (M_1 = m_1, M_2 = m_2)\right].$$

IT framework for geometrically robust data-hiding

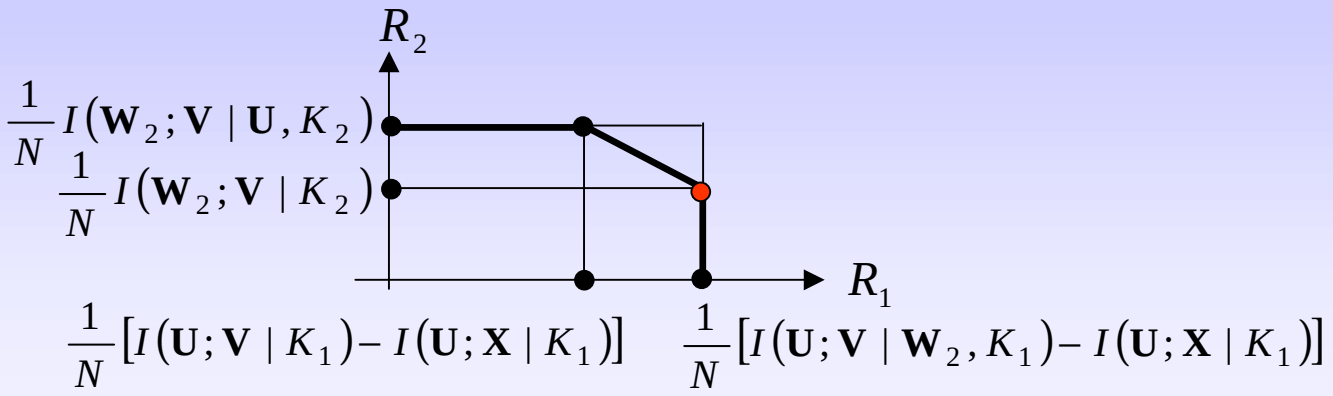
The achievable rates:

$$R_1 \le \frac{1}{N}\left[I(\mathbf{U};\mathbf{V} \mid \mathbf{W}_2, K_1) - I(\mathbf{U};\mathbf{X} \mid K_1)\right]$$

$$R_2 \le \frac{1}{N} I(\mathbf{W}_2;\mathbf{V} \mid \mathbf{U}, K_2)$$

$$R_1 + R_2 \le \frac{1}{N}\left[I(\mathbf{W}_2, \mathbf{U};\mathbf{V} \mid K_1, K_2) - I(\mathbf{U};\mathbf{X} \mid K_1)\right]$$

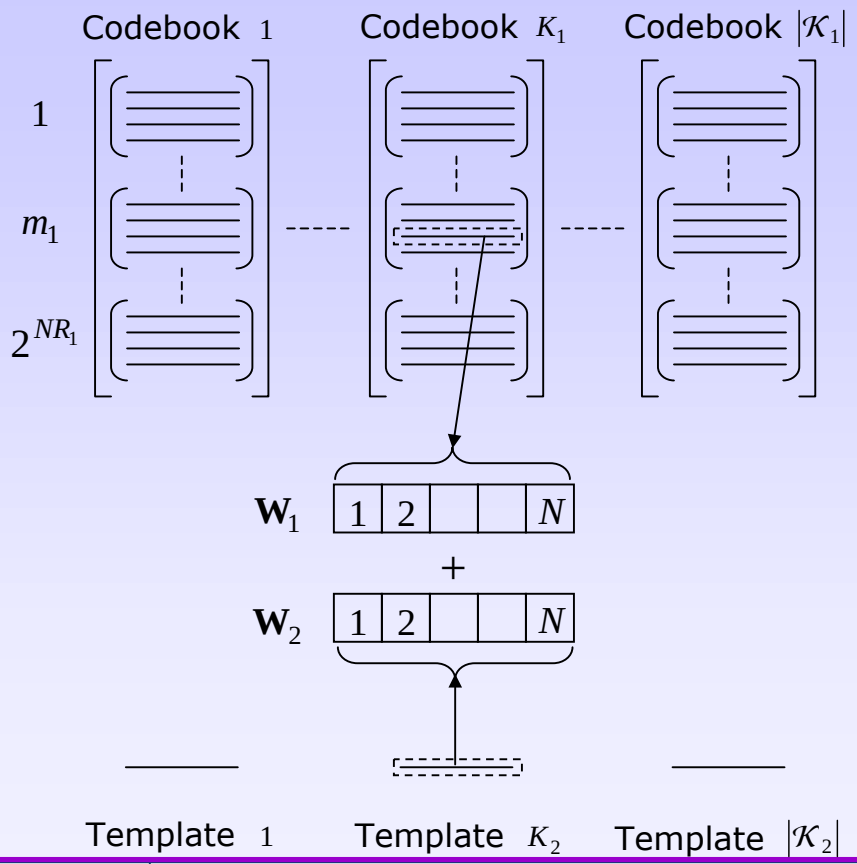
The capacity region:



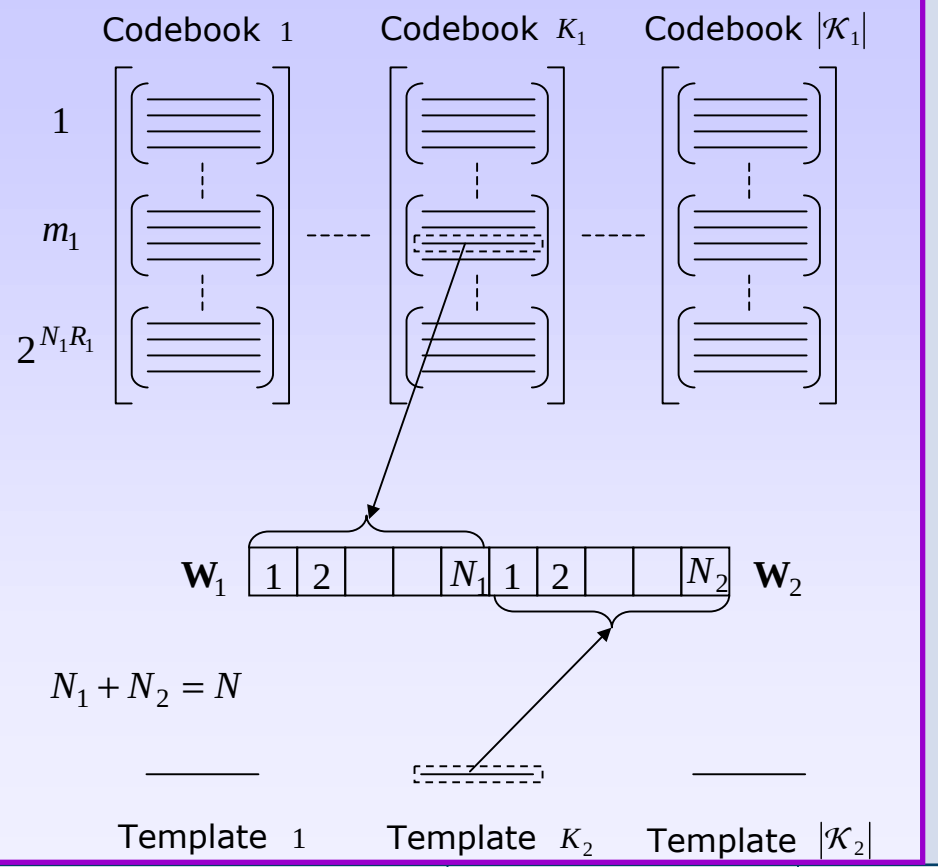
Structured codebooks

Template-based structured codebooks

CDMA signalling:



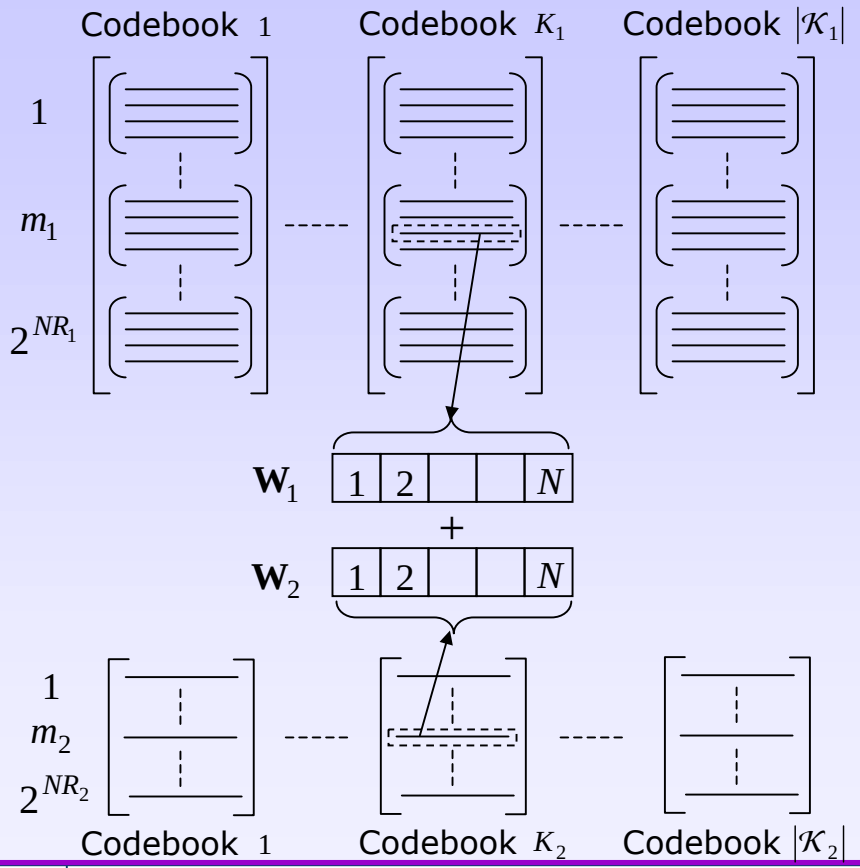
SDMA signalling:



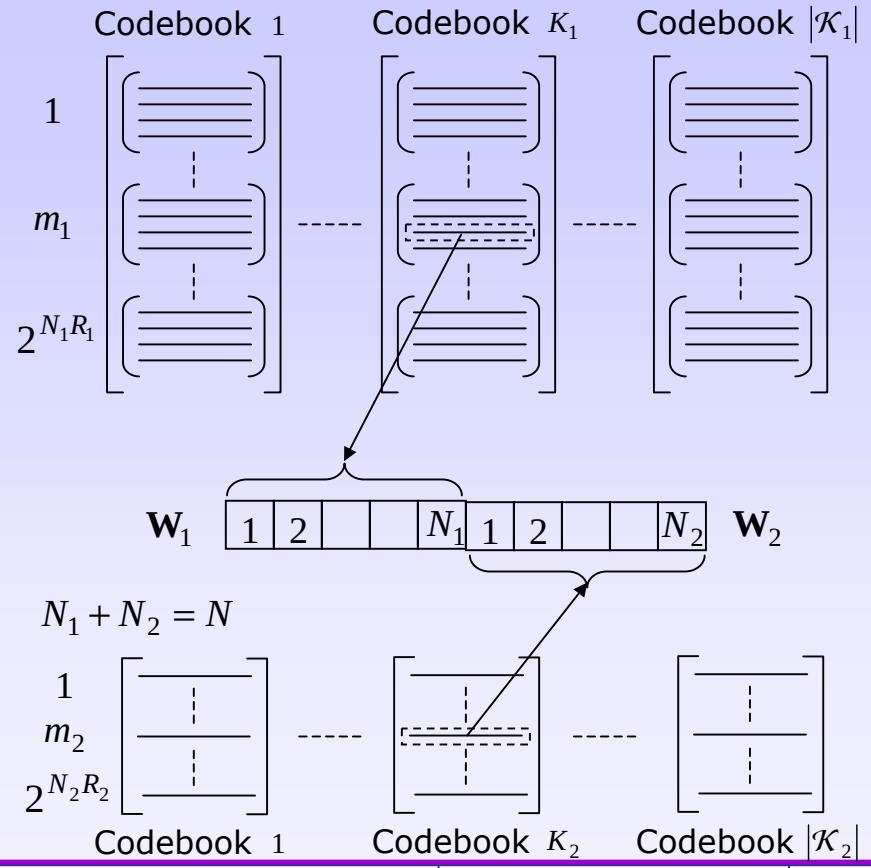
Structured codebooks

Redundant-based structured codebooks

CDMA signalling:



SDMA signalling:



Analysis of security leaks and attacking strategies

Attacker's objective: **To destroy reliable communications.**

Attacker's approach: **To exploit all available prior information and all security leakages.**

Assumptions based on Kerckhoff principle:

Attacker has access to:

- **encoding and decoding algorithms,**
- **codebooks.**

Attacker does not know:

- **secret keys K_1 and K_2,**
- **indexes M_1 and M_2,**
- **the original host image $\mathbf{X}$.**

Analysis of security leaks and attacking strategies

Attack Type	Goal	Attacking Strategy	Required Priors
Statistical signal processing attacks	**To decrease the rate of reliable communications**	**Subtracting an estimate of the watermark sequence from the stego data and adding noise to avoid the attack inversion**	**Host and watermark statistics**
Geometrical attacks	**To increase the decoding complexity on the data-hider side**	**Signal desynchronization**	**-**
Key space search attacks	**To destroy reliable communications completely**	**Exhaustive search in codebooks for the communicated watermark in order to subtract it from the stego data**	**Codebook construction, host and watermark statistics**

Analysis of security leaks and attacking strategies

Key space search attacks

Attacks against template-based structured codebooks

Security consideration:

Template $\mathbf{w}_2$ is only key-dependent and unique for a particular key $K_2 = k$.

Attacks against redundant-based structured codebooks

Security consideration:

By observing stego data, the attacker could estimate the statistics of $\mathbf{w}_2$ even when K_2 is not available.

Attacks against template-based structured codebooks

Particular scenario	Attack complexity
$K_1 = K_2 = K$, and there is a one-to-one correspondence between the codebooks of $\mathbf{w}_1$ and $\mathbf{w}_2$ for a given K	$\lvert\mathcal{K}_2\rvert + 2^{N[R_1+R']}$
$K_1 \neq K_2$, and there is no relationship between the codebooks of $\mathbf{w}_1$ and $\mathbf{w}_2$	$\lvert\mathcal{K}_2\rvert + \lvert\mathcal{K}_1\rvert 2^{N[R_1+R']}$
$K_1 \neq K_2$, but K_2 is fixed and is the same for all users	$1 + \lvert\mathcal{K}_1\rvert 2^{N[R_1+R']}$

Attacks against redundant-based structured codebooks

Particular scenario	Attack complexity
The statistics of $\mathbf{W}_2$ are the same for all codebooks	$\lvert\mathcal{K}_2\rvert 2^{NR_2} + \lvert\mathcal{K}_1\rvert 2^{N[R_1+R']}$
The statistics of $\mathbf{W}_2$ are different for all user codebooks and there is a one-to-one relationship between the codebooks of $\mathbf{W}_1$ and $\mathbf{W}_2$	$\lvert\mathcal{K}_2\rvert 2^{NR_2} + 2^{N[R_1+R']}$

Analysis of security leaks and attacking strategies

Assumption: Generate codebooks in the way that each one contains unique codewords and every possible codeword is included in a unique codebook.

Trial efforts without security leakage analysis:

Random coding	$2^{H(\mathbf{W})} = \lvert\mathcal{K}\rvert 2^{NR}$
Random binning	$2^{H(\mathbf{U})} = \lvert\mathcal{K}\rvert 2^{N[R+R']}$

Trial efforts with security leakage analysis:

Random coding	$2^{H(\mathbf{W}\lvert\mathbf{Y})} = 2^{H(\mathbf{W})-I(\mathbf{W};\mathbf{Y})}$
Random binning	$2^{H(\mathbf{U}\lvert\mathbf{Y})} \le 2^{H(\mathbf{U})-[I(\mathbf{U};\mathbf{Y})-I(\mathbf{U};\mathbf{X})]}$

Conclusions

- **The conditions of reliable communications based on structured codebooks in channels with geometrical transformations are analyzed from an information-theoretic point of view;**
- **The MAC framework is developed to design capacity achieving geometrically robust data-hiding;**
- **The analysis of security leakages for each codebook structure is performed.**

Future directions

- **Consideration of collusion attacks;**
- **Emphasizing the impact of host data statistics on the security;**
- **Extension of the proposed set-up to practical scenarios, with $N < \infty$;**
- **Particular low-complexity search algorithms reducing the complexity of the attacker search based on the security leakages.**