# Towards Universal Weakly-Secure Codes for Data Exchange and Storage

DIMACS workshop

Newark, NJ

April 2, 2015

Alex Sprintson

spalex@tamu.edu

Joint work with Swanand Kadhe, Muxi Yan, and Igor Zelenko

# Weakly Secure Coding

Set of files to be stored: $S = \{S_1, S_2, \ldots, S_{B_s}\}$

Set of coded files observed by Eve: $E$

- Perfectly secure scheme: $I\left(S; E\right) = 0$
- Weakly secure scheme: $I\left(S_i; E\right) = 0$
- $g$-weakly secure scheme

$$\boxed{I(S_{\mathcal{G}}; E) = 0 \quad \forall \mathcal{G} : |\mathcal{G}| \leq g}$$

# Weakly Secure Coding

Weakly secure against $g$ guesses

$$\boxed{I(S_{\mathcal{G}}; E) = 0 \quad \forall \mathcal{G} : |\mathcal{G}| \leq g}$$

- Equivalent to maximizing the minimum Hamming weight of any vector in the span of the codewords
- Requires that no meaningful information is exposed to Eve
- Example

$$S_1 + S_2 + S_3 + S_4$$
$$S_1 + 5S_2 + 12S_3 + 8S_4$$

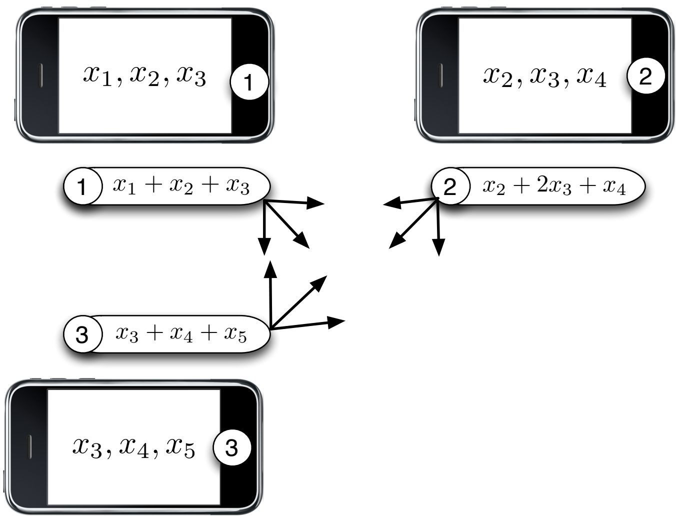
# Cooperative Data Exchange Problem

Clients need to share their local packets with other clients

Clients use a lossless broadcast channel

One packet or function of packet is broadcasted at each time slot.

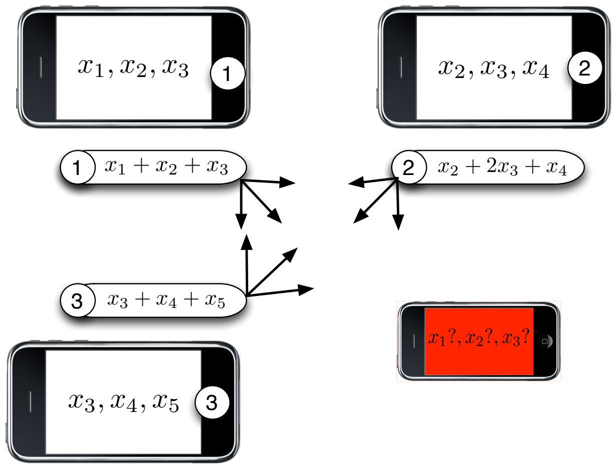
Related to the **key distribution** and **omniscience** problems



# Eavesdropper

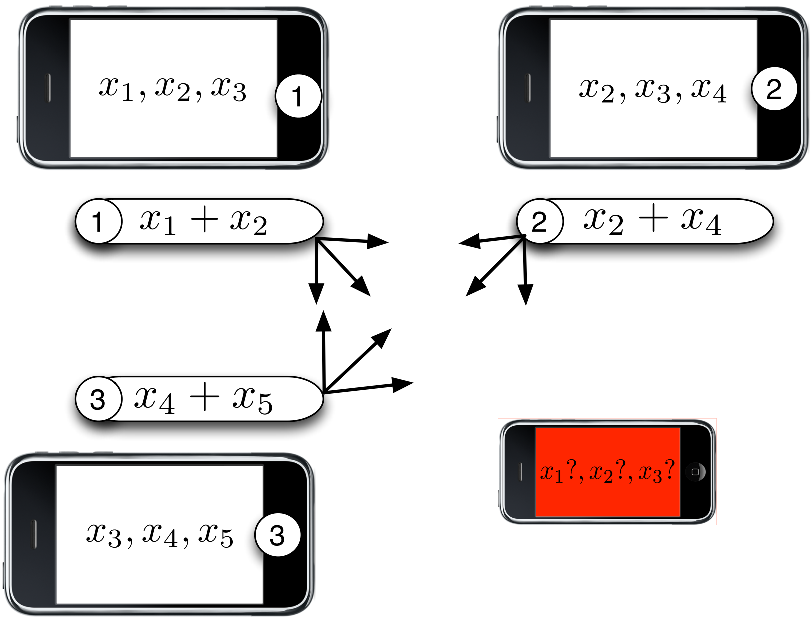
Wants to obtain information about packets held by the clients

Has access to any data transmitted over the broadcast channel



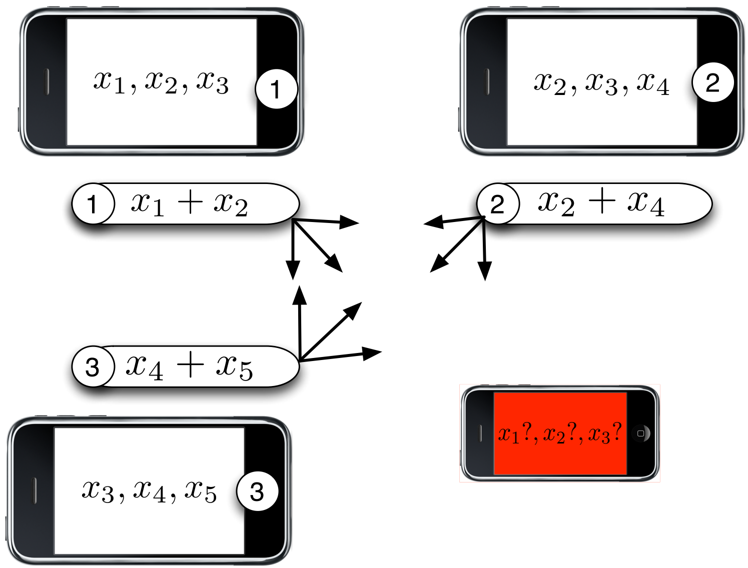
# $g$-weak Security

For each subset $S_G$ of $X$ of size $g$ or less it holds that $I(S_G; P) = 0$



# Example

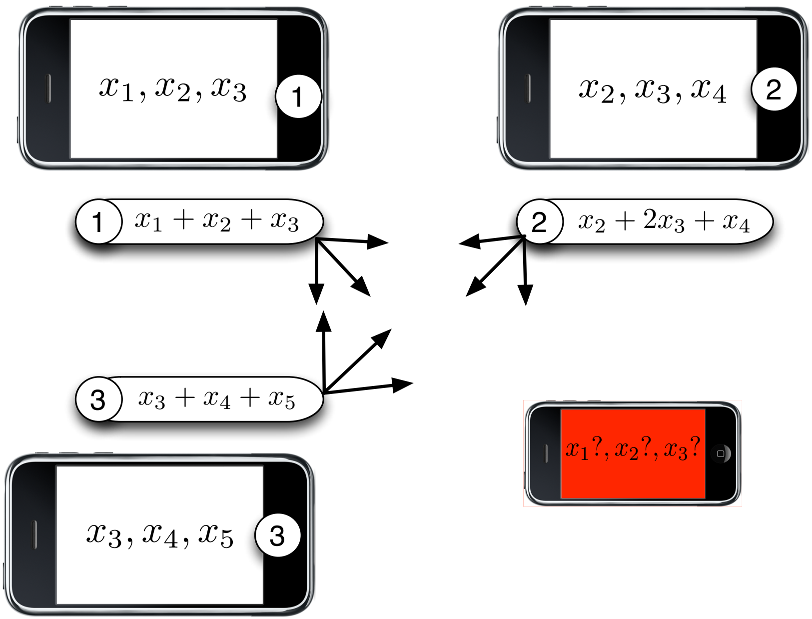
Eavesdropper can only get value of $x_1 + x_2$, $x_2 + x_4$, and $x_4 + x_5$,
– cannot get value of the original packets $x_1, \cdots, x_4$
– this solution is 1-weakly secure



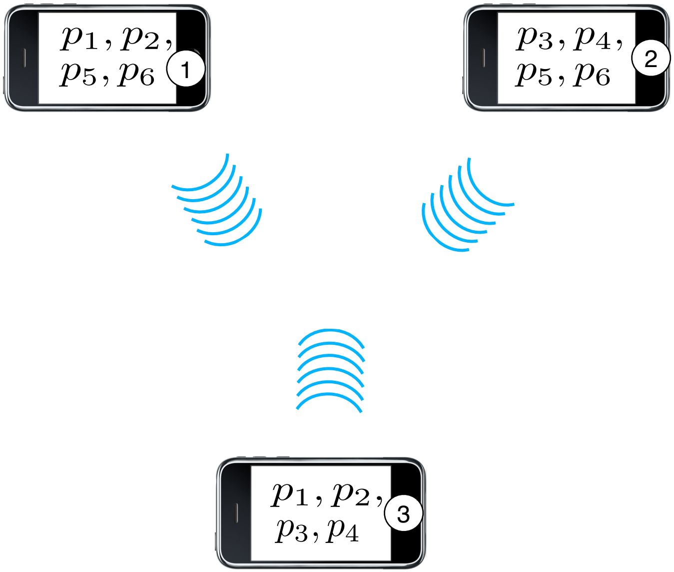
# Example (cont.)

Eavesdropper cannot obtain a combination of any two original packets

This solution is 2-weakly secure



# Constrained Matrix Completion Problem




| | $p_1$ | $p_2$ | $p_3$ | $p_4$ | $p_5$ | $p_6$ |
|---|---|---|---|---|---|---|
| 1 | x | x | 0 | 0 | x | x |
| 2 | 0 | 0 | x | x | x | x |
| 3 | x | x | x | x | 0 | 0 |

# Matrix completion problem

When is it possible to complete the matrix so it will satisfy the MDS condition?

– When it does not contain an all zero submatrix of size $a \times b$, such that $a + b \geq OPT + 1$

$$\begin{array}{cc} & n \\ \mathrm{OPT} & \begin{array}{|c|c|c|c|c|c|}\hline \mathrm{x} & \mathrm{x} & 0 & 0 & \mathrm{x} & \mathrm{x} \\ \hline 0 & 0 & \mathrm{x} & \mathrm{x} & \mathrm{x} & \mathrm{x} \\ \hline \mathrm{x} & \mathrm{x} & \mathrm{x} & \mathrm{x} & 0 & 0 \\ \hline \end{array} \end{array}$$

Fragouli, Soljanin, '06
Halbawi, Ho, Yao, Duursma '14
Dau, Song, Yuen '14

# Matrix completion problem

Our case: constraints on the code construction

– Due to the side information available at the clients

Random code works with high probability

– Hard to check since finding a minimum distance is an NP-hard problem

$$\begin{array}{c|c|c|c|c|c|c|}
 & \multicolumn{6}{c}{\longleftarrow \quad n \quad \longrightarrow} \\
\cline{2-7}
 & \mathrm{x} & \mathrm{x} & 0 & 0 & \mathrm{x} & \mathrm{x} \\
\cline{2-7}
\mathrm{OPT} \updownarrow & 0 & 0 & \mathrm{x} & \mathrm{x} & \mathrm{x} & \mathrm{x} \\
\cline{2-7}
 & \mathrm{x} & \mathrm{x} & \mathrm{x} & \mathrm{x} & 0 & 0 \\
\cline{2-7}
\end{array}$$

# Theorem

Can achieve the distance

$$n - OPT + 1$$

– with high probability at least $1 - \binom{n}{OPT} \frac{OPT}{q}$
– requires field size $\binom{q>n}{OPT} OPT$

$n$

OPT

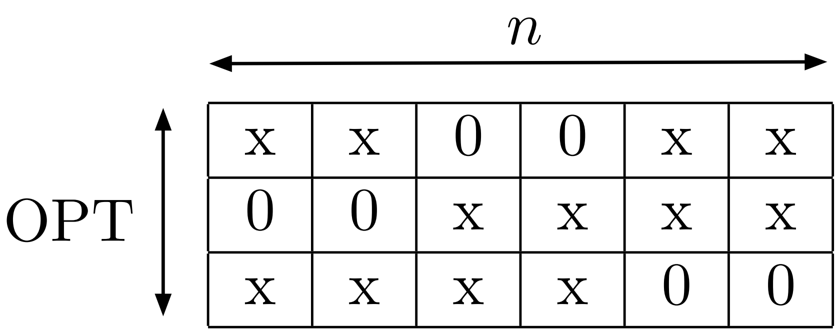
| | | | | | |
|---|---|---|---|---|---|
| x | x | 0 | 0 | x | x |
| 0 | 0 | x | x | x | x |
| x | x | x | x | 0 | 0 |

# Deterministic algorithm

Use matrix completion

– Fill $i^{th}$ entry of the matrix with a value if $GF(2^i) \subset GF(2^{i-1})$

– Determinant of any $OPT \times OPT$ matrix is guaranteed to be full rank



# Structured Codes

Can we use standard codes, e.g., Reed-Solomon

Then, perform a linear transformation to complete the matrix?

Generalized Reed-Solomon code

$$G = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \alpha_1 & \alpha_2 & \dots & \alpha_n \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_1^{\mu-1} & \alpha_2^{\mu-1} & \dots & \alpha_n^{\mu-1} \end{bmatrix}.$$

# Structured Codes

Can we use standard codes, e.g., Reed-Solomon

Then, perform a linear transformation to complete the matrix?

$$\begin{bmatrix} X & X & X & X & 0 & 0 \\ X & X & 0 & 0 & X & X \\ 0 & 0 & X & X & X & X \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 & \alpha_6 \\ \alpha_1^2 & \alpha_2^2 & \alpha_3^2 & \alpha_4^2 & \alpha_5^2 & \alpha_6^2 \end{bmatrix}$$

Unfortunately, the transformation matrix is not guaranteed to be full-rank

# Negative example

A negative example:

$$\begin{bmatrix} X & X & X & X & 0 & 0 \\ X & X & 0 & 0 & X & X \\ 0 & 0 & X & X & X & X \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 & \alpha_6 \\ \alpha_1^2 & \alpha_2^2 & \alpha_3^2 & \alpha_4^2 & \alpha_5^2 & \alpha_6^2 \end{bmatrix}$$

$\alpha$: primitive element of $GF(8)$ with primitive polynomial $x^3 + x + 1$

# Conjecture

If the configuration matrix can be completed to MDS,

– i.e., it does not contain a zero submatrix of dimension $a \times b$ such that $a + b \geq OPT + 1$

Then the determinant of $T$ is not identically equal to zero

$$\begin{bmatrix} X & X & X & X & 0 & 0 \\ X & X & 0 & 0 & X & X \\ 0 & 0 & X & X & X & X \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 & \alpha_6 \\ \alpha_1^2 & \alpha_2^2 & \alpha_3^2 & \alpha_4^2 & \alpha_5^2 & \alpha_6^2 \end{bmatrix}$$

# Reformulation of the problem

Let $N_1, \ldots N_\mu$ be subsets of $[n]$ such that $|N_i| = \mu - 1$

Define the collection of $\mu$ polynomials $P_1, \ldots, P_\mu$ in $\mathbb{F}[\alpha_1, \ldots \alpha_2][x]$:

$$P_i = \prod_{j \in N_i} (x - \alpha_j).$$

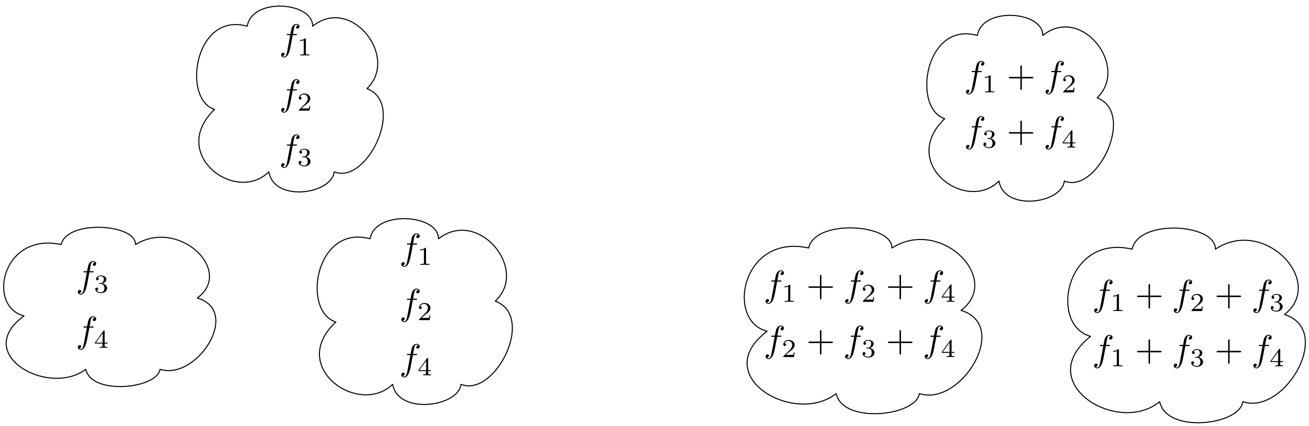
Question: Under what condition on the collection of sets $\{N_i\}_{i=1}^{\mu}$ the polynomials $\{P_i\}_{i=1}^{\mu}$ are linearly dependent over the ring $\mathbb{F}[\alpha_1, \ldots, \alpha_n]$?

# Security for Storage: Motivation

There are numerous service providers

Some of these cloud networks can be compromised

Any of the storage nodes in a compromised network can be eavesdropped

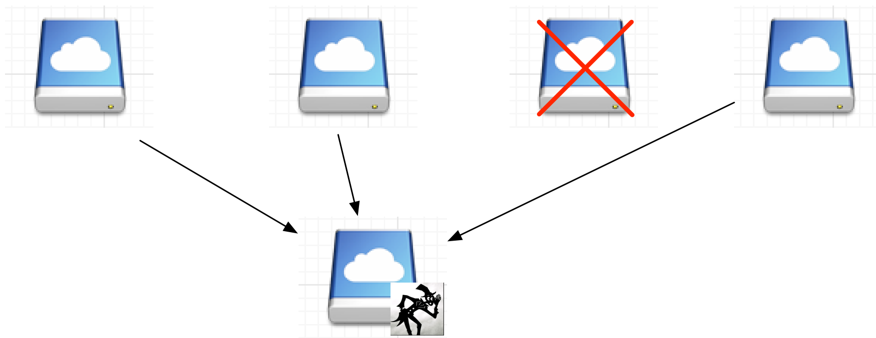


# Security for Storage: Challenges

Storage system is a dynamic system with nodes continually failing and being replaced

At a particular node location, eavesdropper can keep on observing the data downloaded during multiple repairs

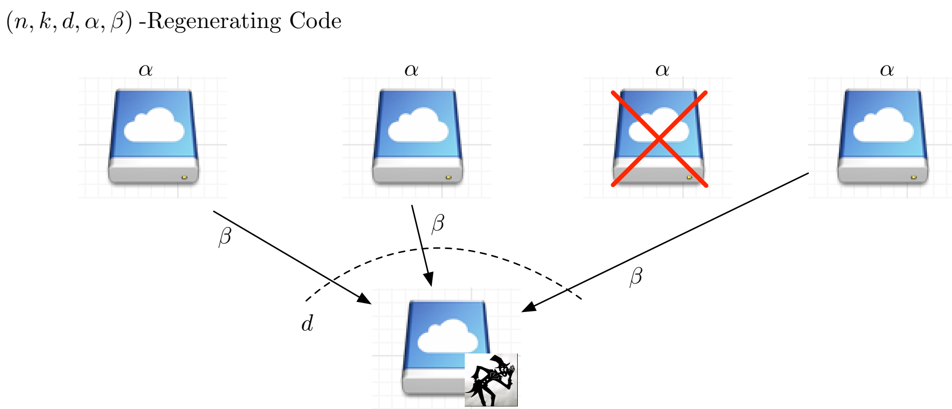
– Random coding is not helpful



# Regenerating Codes

A special class of erasure codes that optimally trade-off storage space for repair bandwidth

- $(n, k)$-MDS property: any $k$ nodes are sufficient for data reconstruction
- Minimize the repair bandwidth $d\beta$



# Product-Matrix (PM) Codes

We focus on a special class of regenerating codes,
– Product-Matrix framework based Minimum Bandwidth Regenerating (PM-MBR) Codes

Explicit codes, unlike random coding

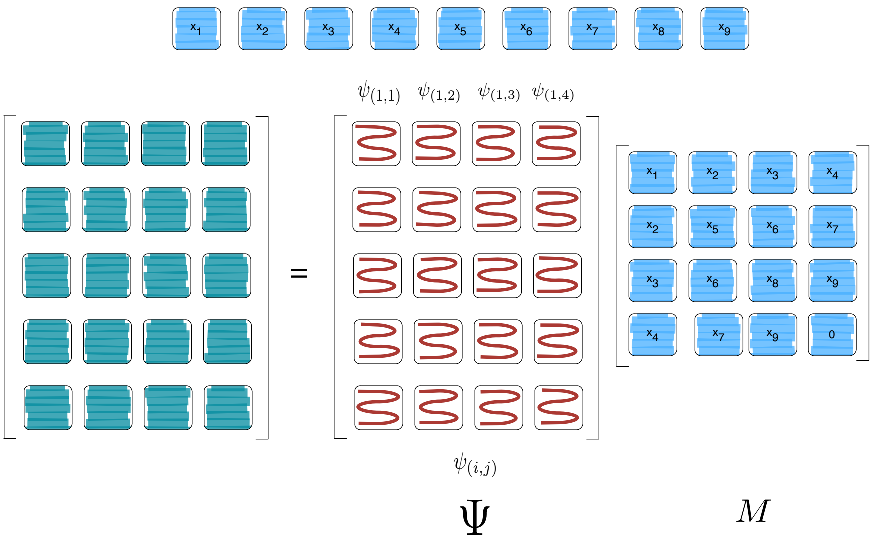
Designed for exact regeneration
– Repaired node is an exact replica of the failed node

Construction for all values of $(n, k, d)$
– Efficient in terms of field size – Very practical!

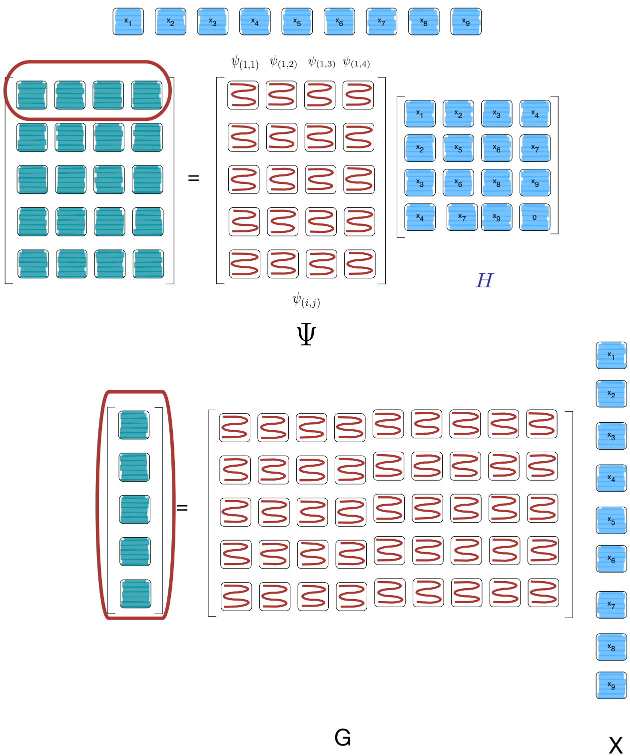
# Product-Matrix (PM) Codes

PM code is obtained by taking a product of encoding matrix $\Psi$ and message matrix $M$

- Both $\Psi$ and $M$ have have specific structures
- Choosing $\Psi$ as a Vandermonde or a Cauchy matrix works

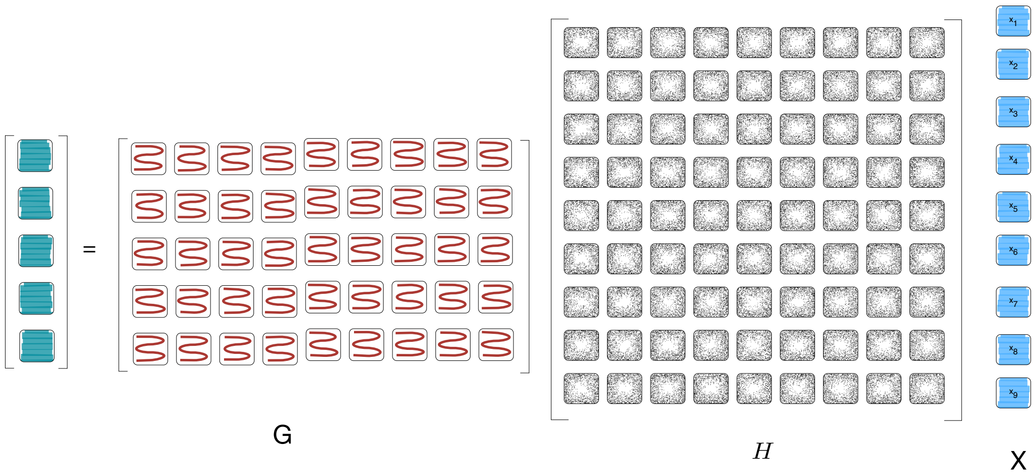


# Eavesdropping a PM-MBR Code



# Coset Coding Based Outer Codes

Can we utilize the elegant structure of Product Matrix codes to explicitly design $H$ that satisfies the condition above?

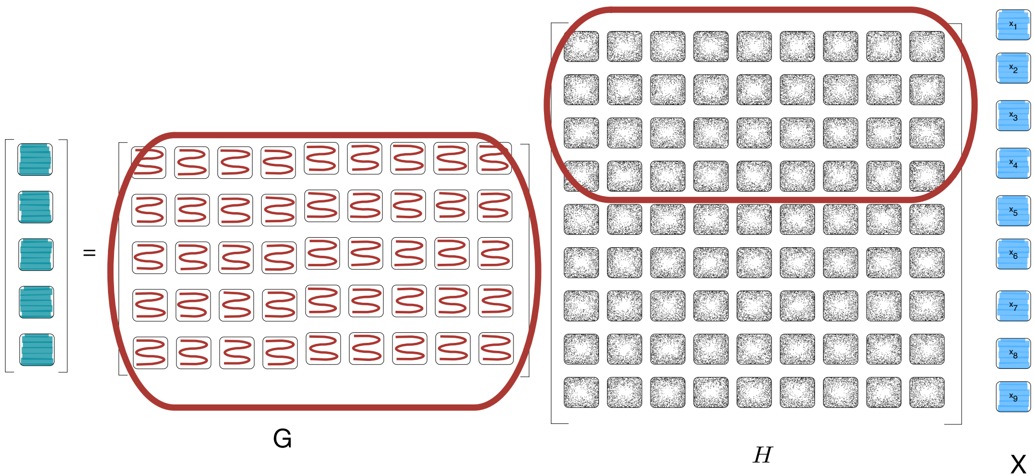


# Outer Code Design

How to design $H$ that satisfies this condition?

$$rank \begin{bmatrix} H_{\mathcal{G}'} \\ G_E \end{bmatrix} = rank(H_{\mathcal{G}'}) + rank(G_E),$$

where $H_{\mathcal{G}'}$ is any $(g+1) \times B$ sub-matrix of $H$

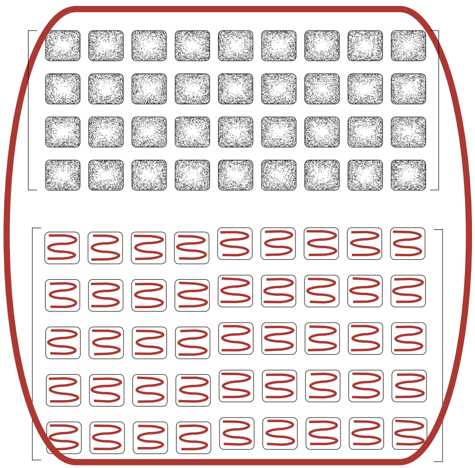


# Outer Code Design

How to design $H$ that satisfies this condition?

$$rank \begin{bmatrix} H_{\mathcal{G}'} \\ G_E \end{bmatrix} = rank(H_{\mathcal{G}'}) + rank(G_E),$$

where $H_{\mathcal{G}'}$ is any $(g+1) \times B$ sub-matrix of $H$



# Explicit Outer Code Construction

Observation: generator matrix for any node $e$ has the same *structure*

$$G_e = \begin{bmatrix} \Psi(e,1) & \Psi(e,2) & \Psi(e,3) & \Psi(e,4) & 0 & 0 & 0 & 0 & 0 \\ 0 & \Psi(e,1) & 0 & 0 & \Psi(e,2) & \Psi(e,3) & \Psi(e,4) & 0 & 0 \\ 0 & 0 & \Psi(e,1) & 0 & 0 & \Psi(e,2) & 0 & \Psi(e,3) & \Psi(e,4) \\ 0 & 0 & 0 & \Psi(e,1) & 0 & 0 & \Psi(e,2) & 0 & \Psi(e,3) \end{bmatrix}$$

Notion of type

- A length-$B$ encoding vector $h^{(i)}$ is of type $i$ if it has form as the $i$-th row of $G_e$
- Essentially, the type specifies the locations of the non-zero coefficients

# Explicit Outer Code Construction

Design $H$ such that each row belongs to one of the $d$ types

It is sufficient to specify the number of rows of each type and the values of the non-zero coefficients

Let $\theta_i$ denote the number of rows of type $i$ that are present in $H$

– We call $\theta_i$ as the type cardinality of type $i$

$$\theta_i = \begin{cases} 0 & \text{if} \quad i = 1, \\ d - k + j & \text{if} \quad 2 \leq i \leq k-1, \\ d - 1 & \text{if} \quad i = k, \\ 1 & \text{if} \quad k+1 \leq i \leq d. \end{cases}$$

# Explicit Outer Code Construction

Example : $(n = 5, k = 3, d = 4)$ PM-MBR Code, $B = 9$, $B_s = 7$

$$H = \begin{bmatrix} 0 & \hat{\Psi}(1,1) & 0 & 0 & \hat{\Psi}(1,2) & \hat{\Psi}(1,3) & \hat{\Psi}(1,4) & 0 & 0 \\ 0 & \hat{\Psi}(2,1) & 0 & 0 & \hat{\Psi}(2,2) & \hat{\Psi}(2,3) & \hat{\Psi}(2,4) & 0 & 0 \\ 0 & \hat{\Psi}(3,1) & 0 & 0 & \hat{\Psi}(3,2) & \hat{\Psi}(3,3) & \hat{\Psi}(3,4) & 0 & 0 \\ -- & -- & -- & -- & -- & -- & -- & -- & -- \\ 0 & 0 & \hat{\Psi}(1,1) & 0 & 0 & \hat{\Psi}(1,2) & 0 & \hat{\Psi}(1,3) & \hat{\Psi}(1,4) \\ 0 & 0 & \hat{\Psi}(2,1) & 0 & 0 & \hat{\Psi}(2,2) & 0 & \hat{\Psi}(2,3) & \hat{\Psi}(2,4) \\ 0 & 0 & \hat{\Psi}(3,1) & 0 & 0 & \hat{\Psi}(3,2) & 0 & \hat{\Psi}(3,3) & \hat{\Psi}(3,4) \\ -- & -- & -- & -- & -- & -- & -- & -- & -- \\ 0 & 0 & 0 & \hat{\Psi}(1,1) & 0 & 0 & \hat{\Psi}(1,2) & 0 & \hat{\Psi}(1,3) \end{bmatrix}$$

First three rows are of type 2

Next three rows are of type 3

Last row is of type 4

# Theorem

Proposed outer code that results in a $g$-weakly secure code for $g = d + k - 3$

The secure storage capacity of the proposed construction is $B_s = B - 2$

- Improvement over uncoded security level of $k - 1$ guesses
- Roughly twofold enhancement in the security level
  * Still far from maximum possible level of security
  * $g_{max} = B - d - 1 = \mathcal{O}\left(k^2\right)$
  * Does not require an increase in the field size

# Conclusions

- A promising way to provide reliability and security
- Light-weight alternatives to cryptographic primitives
- In many cases, reliability and security can be provided at no or little additional cost
- Many exciting research problems