Code Based Cryptography

Vishal Saraswat
vishal.saraswat@gmail.com
http://crypta.in/

Indian Statistical Institute, Kolkata
20 May 2016

Outline

1. CRRao AIMSCS
2. Vishal Saraswat
3. Anonymous Signatures
4. Anonymous Identity Based Encryption
5. How to Leak a Secret and Reap the Rewards too
6. Anonymous Proxy Signatures
7. Lightweight Block Ciphers
8. Secret Sharing Schemes
9. Introduction
10. Coding Theory
11. McEliece Cryptosystem
12. Niederreiter Cryptosystem
13. McEliece v/s Niederreiter
14. McEliece PKC: Advantages and Disadvantages

Section Outline

1. CRRao AIMSCS
   - CRRao AIMSCS
   - Current Funding Agencies
   - Current Projects
   - Other Projects

CRRao AIMSCS

- An autonomous institute of higher learning and research
- Mainly funded by projects from the Government of India
- Conduct research in Mathematics, Statistics, Computer Science and convergence of these areas
- Provide consultancy, project services to government and non-government organizations, the industry, the corporate and private sector.

Current Funding Agencies

- Ministry of Statistics and Program Implementation (MoSPI)
- Department of Science and Technology (DST)
- Centre for Mathematical Studies and Analysis (CMSA)
- Defence Research and Development Organization (DRDO)
- Cabinet Secretariat (CSO)

Current Projects

- (CMSA)
  - Design of Proprietary Block Cipher
  - Design of Proprietary Stream Cipher
  - System Engineering to Prevent Side-Channel Cryptanalysis
  - Development of Test Suite for Analysis of Cipher Strength
  - Secret Sharing
- (DRDO-SAG)
  - Post-Quantum cryptology
  - Software Tools for SMT Solvers Based Cryptanalysis
  - SAT Solvers based Cryptanalysis
  - Lattice based Cryptography and Cryptanalysis

Projects in Discussions

- Secret Sharing Schemes
- Discrete-Log Problems
- Digital Signatures using Asymmetric Pairings
- Mechanics of Electromagnetic Radiation (Side Channel Cryptanalysis)
- Design of a 128-bit Block Cipher

Other Projects

- High-Performance Computing Toolkit for Algebraic Cryptanalysis
- Secure Operating System for Distributed Secure Transactions
- Design of Cryptographic Boolean Functions
- Lightweight Secure Cryptosystems based on Finite Automata
- Application of Artificial Neural Networks to Cryptanalysis

Section Outline

2. Vishal Saraswat
   - Professional Experience
   - Education
   - Research Interests
   - Projects at AIMSCS
   - Some other projects

Professional Experience

- **Assistant Professor**, C.R.Rao Advanced Institute of Mathematics, Statistics and Computer Science (AIMSCS), Hyderabad, India
- **External Faculty**, Indian Institute of Technology, Hyderabad (IITH)
- **Honorary Lecturer**, School of Physics, University of Hyderabad
- **Lecturer/TA**, University of Minnesota (UMN), Minneapolis, MN, USA
- **Mentor**, Interdisciplinary Research Experience for Undergraduates, Institute of Mathematics and its Applications (IMA), UMN
- **Research Assistant**, Intelligent Storage Consortium, Digital Technology Center (DTC), UMN
- **Research Visitor**, Center for Discrete Mathematics and Theoretical Computer Science (DIMACS), Rutgers University, Piscataway, NJ
- **Research Fellow**, Minnesota Center for Industrial Mathematics (MCIM), UMN
- **Research Scholar**, Tata Institute of Fundamental Research (TIFR), Bombay

Education

- **Ph.D.**, University of Minnesota, Minneapolis-MN, USA (UMN)
  - Topic: **Anonymity and Privacy in Public Key Cryptography**
  - Advisor: **Professor Andrew M. Odlyzko**
- **M.S. – Computer Science**, UMN
- **M.S. – Mathematics**, UMN
- **P.G.Certificate in Statistical Methods and Applications**, Indian Statistical Institute, Calcutta
- **B.Sc. – Mathematics**, St. Xavier's College, University of Calcutta, Calcutta

Research Interests

- Number Theory
- Information Security
- Cryptography and Cryptanalysis
- Anonymity and Privacy
- Post Quantum Crypto

Research Interests

- Sieve Methods
  - Number Field Sieve
  - Function Field Sieve
- Generic Ring Algorithms
  - To analyze various public key cyptosystems
- Lattices
  - Distribution of invertible polynomials in $\mathbb{Z}_p[X]/g\mathbb{Z}_p[X]$
- Error Correcting Codes
  - Compact keys for McEliece Cryptosystem
- Lattices vs. Error Correcting Codes

Projects at AIMSCS

- Design of an indigenous block cipher
- Post Quantum Crypto
- Secret Sharing
- Discrete Log Computation
- Side Channel Cryptanalysis
- Lattice Based Cryptanalysis
  - using lattice reduction methods to attack various cryptosystems

Some other projects

- Secure and Efficient Long Term Data Management
  - Guarantee confidentiality, integrity and real-time global availability of large amounts data for a long period of time.
- Secure and Efficient Large Scale Key Management
  - Incorporate efficient, scalable and dependable key management into the file system so that it is transparent from the users.
- Applied Remote Cache-timing Attacks against AES [1]
  - To **remotely** recover the AES encryption key of a target server.

Section Outline

3 Anonymous Signatures
- Problem Scenario

Anonymous Signatures

- Anonymous Auctions
  - A minister wants to bid for an IPL cricket team
    BUT wants to remain anonymous unless he wins the auction.
- Anonymous Paper Review
  - A professor wants to publish the proof of $P = NP$
    BUT wants to remain anonymous unless the paper is accepted.
- Anonymous Authentication Protocols
  - Boyd and Park key transport protocol:
    $A \rightarrow B : PKE_B(ID_A, \sigma, count)$
    $A \leftarrow B : Enc_\sigma(count, r_B)$
    $A \rightarrow B : Sig_A(ID_B, h(count, \sigma, r_B))$
- anonymous credential systems, anonymous transaction systems, bid secrecy and verifiability, e-voting, ...
- We introduce notions of security and anonymize generically any signature scheme. [5]

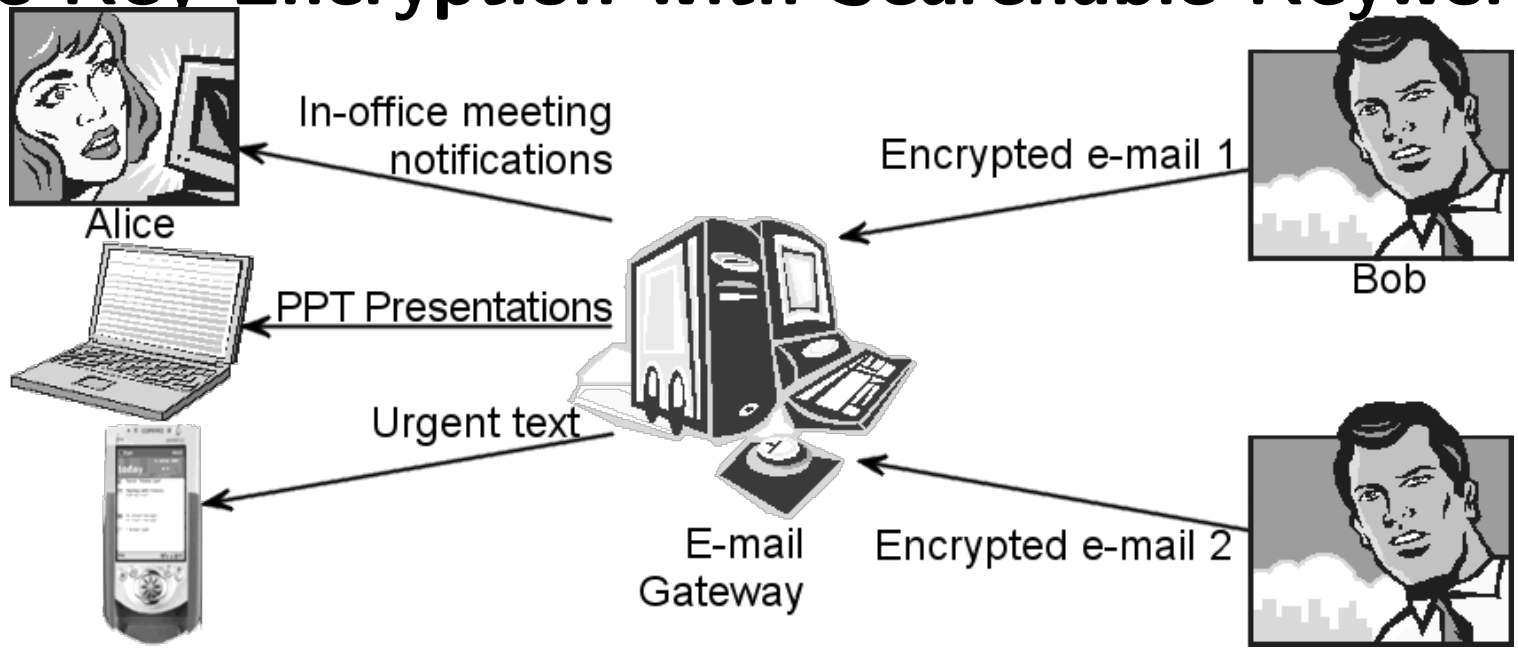
Section Outline

4 Anonymous Identity Based Encryption
- Problem Scenario 1
- Problem Scenario 2

Anonymous Identity Based Encryption

- An Indian agent in Pakistan wants to receive messages from India or other Indian agents.
- All emails entering Pakistan from India are monitored
  - to whom they are going etc.
- Even if the email cannot be read they can trace the route etc.
- Anonymous identity based encryption (aIBE) allows the recipient:
  - to remain anonymous.
  - to communicate with other agents who do not know the full identity but only certain attributes.
- We provide the only existing non-trivial method of converting a non-anonymous IBE to an anonymous IBE. [6]

Public-Key Encryption with Searchable Keywords



- Alice wishes to read her e-mail on 3 devices: desktop, laptop, pager
- E-mail server has to route each e-mail to appropriate device according to e-mail keywords
- If e-mail is encrypted then server cannot read keywords
- Public key encryption with keyword search (PEKS) allows:
  - Server to route e-mail appropriately
  - E-mail to remain private against server
- We solve the open problem of a PEKS not based on elliptic curves. [6]

Section Outline

5 How to Leak a Secret and Reap the Rewards too
- Motivation I
- Motivation II
- Applications

Godfather Scenario

- Suppose the government announces a reward for information leading directly to the apprehension or conviction of the boss B of a criminal gang.
- B is a powerful person feared by all and has many officials in the government secretly working for him.
- A member S of the gang is tempted by the award and wants to reveal some information about B.
  - S needs to make the information public and not tip off a select few officials since S does not know if they might be working for B and might hand him over to B.
  - S cannot reveal his identity to anyone for same reasons.
  - S wants/needs to reveal his identity to claim the reward
    - only after B has been convicted; and
    - only to a designated official V.
  - Finally, V should not be able to prove that S had leaked the information.

Auction Scenario

- A bidder A wants to anonymously participate in an auction without revealing his identity even to the auctioneer.
- The auctioneer B is not willing to accept 'completely anonymous' bids. For example, bids from millionaires or people from certain organization, may be accepted.
- When A wins the auction, A needs to prove his identity to claim the win while preserving his anonymity.

- A could possibly use a ring signature (with a sufficiciently anonymizing ring formed of potential millionaire bidders) but then A will not be able to prove that he indeed placed the winning bid.
- Use our proposed $divrs$ with B as the designated verifier.
  - If A loses the auction, there is no loss of anonymity.
  - If A wins the auction, only B gets to identify A.
  - Note that B cannot prove to others that A bid for the item.
  - The winning bid though can be publicly verified.

How to Leak a Secret and Reap the Rewards too

- Many applications of ring signatures like whistle blowing or anonymity-preserving auctions come with a 'reward' for the signer (whistle blower or auction winner).
- To claim the reward the signer would need to reveal their identity to a designated authority (government official or auctioneer) and prove that they indeed produced the signature.

Section Outline

6 Anonymous Proxy Signatures
- Motivation

Anonymous Proxy Signatures

- Proxy signatures
  - You need to go out of town but you are expecting a registered post.
  - Authorize a neighbor to receive the post on your behalf.
- Proxy multi-signature
  - Full family is going out of town but want registered posts to be received.
  - Each member can authorize a neighbor to receive posts on their behalf.
  - All members give one authorizing letter to a neighbor.
- Multi-proxy signature
  - What if you do not want to rely on one single person?
  - You can give authorizing letters to multiple people.
- Anonymous proxy signature
  - In the authorizing letter, we explicitly mention the authorized person.
  - Or we can just give a "bearer" authorizing letter.
- Threshold anonymous proxy multi-signature [3]

Example

Consider a secure multi-party computation setting:

- Parties $\mathcal{O}_1, \ldots, \mathcal{O}_n$ start a process $\mathcal{P}$ after authenticating themselves.
- Once the process $\mathcal{P}$ is started, $\mathcal{O}_1, \ldots, \mathcal{O}_n$ do not need to stay online
- while the process $\mathcal{P}$ may remain active but
- need access to additional resources that require further authentication.
- The parties thus delegate their rights to the process $\mathcal{P}$ and
- the resources allow access to $\mathcal{P}$ as long as the resources can verify
- that $\mathcal{P}$ was indeed authorised by the original parties.
- The resources do not need to know the 'identity' of the process at all
- and $\mathcal{P}$ may remain anonymous to them.
- Most of the times the resources do not even need to know whether it is actually the original parties who were authenticated or their proxy $\mathcal{P}$.
- But in case of a malicious process, the original parties should be able to expose the process and restrict any further activities by it on their behalf.

Section Outline

7 Lightweight Block Ciphers
- Applications

Lightweight Block Ciphers

- Tailor made for devices which have very limited resources
  - cellphones
  - smart cards
  - RFID tags
  - sensor nodes

- These devices use less resources in terms of
  - power
  - memory
  - time
  - space

- The heat generated, the electromagnetic emmisions and similar by-products have a huge impact on the functioning, on the life-span, and on the security of these devices

Section Outline

8 Secret Sharing Schemes
- Problem Scenario
- Naive Solution
- Problems with Naive Solution

Missile Key / Bank Vault Key

- Suppose one single person can fire a missile system.
- One bad fight with his wife
- and the whole village of in-laws goes BOOM.
- So need a mechanism so that one single person cannot fire the missile.

- Suppose the bank vault can be opened by one single key held by the manager.
- One bad day at the racing
- and the bank-vault goes empty
- So need a mechanism so that one single person cannot open the vault.

Naive Solution

- Meet CRRao in AIMSCS at six pm.
  - MEETCRRAOINAIMSCSATSIXPM
- Distribute in two **shares**
  - M E C R O N I S S T I P — MECRONISSTIP
  - E T R A I A M C A S X M — ETRAIAMCASXM
- Distribute in three **shares**
  - M T R I I C T X — MTRIICTX
  - E C A N M S S P — ECANMSSP
  - E R O A S A I M — EROASAIM
- What if you can get **only** two out of three shares?
  - EE CR AO NA MS SA SI PM — EECRAONAMSSASIPM
- Distribute such that two out of three **shares** can recover full secret!
  - ME TC RA IN IM CS TS XP — METCRAINIMCSTSXP
  - M ET RR OI AI SC AT IX M — METRROIAISCATIXM
  - EE CR AO NA MS SA SI PM — EECRAONAMSSASIPM

Problems with Naive Solution

- Each of the n shares gives '$1/n$'-th of the information about the secret.
- And any m of the n shares, when combined, give 'm/n'-th of the information about the secret.
- We use mathematical techniques so that any share should give '0' information about the secret.
- In fact even if $t-1$ shares are obtained there is no extra information obtained.

Section Outline

9 Introduction
- Post-quantum cryptography
- Code-based cryptography

Post-quantum cryptography

- PQC [**?**] is the study of cryptosystems that
  - run on classical computers; and yet
  - are secure against attacks by quantum computers. [**?**, **?**]
- Some existing systems that fall under PQC are
  - Code-based cryptography [**?**, **?**, **?**, **?**]
  - Lattice-based cryptography
  - Multivariate-quadratic-equations cryptography
  - Hash-based cryptography
  - Secret-key cryptography

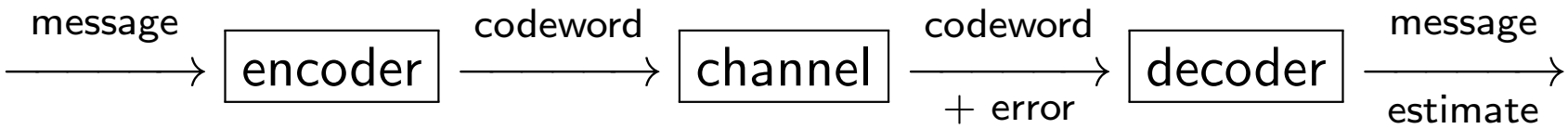
Code-based cryptography

- The McEliece cryptosystem [**?**]
- The Niederreiter cryptosystem [**?**]

Section Outline

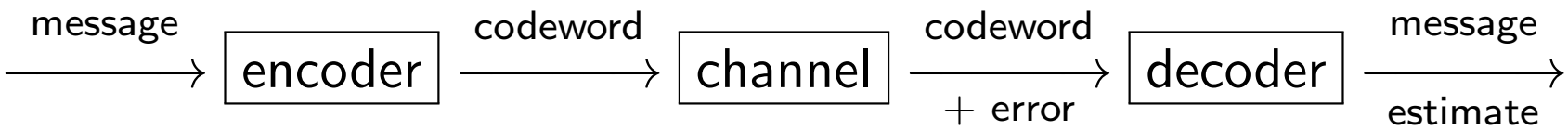
10 Coding Theory
- Codes
- Linear Codes
- Hamming distance
- Minimum distance
- Parity-check matrix
- Decoding
- Cyclic Codes
- BCH Codes
- GRS codes
- Alternant Codes
- Goppa Codes

Code



- Coding theory deals with design of "error detecting/correcting codes" for reliable transmission of information over noisy channels.
- Codes protect against errors that occur in messages transmitted in a noisy channel.
- An encoder transforms a message word into a codeword by adding redundancy.
- The decoder uses a decoding algorithm to detect and/or correct errors which might have occurred during transmission.

Code



- A *code* C is a finite subset of some mathematical structure.
- Members of this subset are called *codewords*.
- The codewords are chosen in such a way that even if some "error" is added to any codeword, it can be "identified" (with good probability).
- *Encoder* is a function that transforms a message to a codeword.
- *Decoder* is a function that transforms a received codeword+error to a message estimate.

Example — Repetition Code

- $\mathcal{M} = \{0, 1\}$.
- $C = \{000, 111\} = \langle 111 \rangle$.
- **Encoder:** $0 \to 000$; $1 \to 111$.
- **Decoder:**
  - If the codeword has more 0s, decode it to 0;
  - If the codeword has more 1s, decode it to 1.
  - Maximum likelihood (ML) Decoding.
- This is $[3, 1, 3]$ binary linear code.
- $G_{1\times 3} = [I_1|A] = [111]$.
- $H_{2\times 3} = [A^T|I_2] = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$.

Example — Even Parity Code

- $\mathcal{M} = \{0,1\}^3 = \{000, 001, 010, 011, 100, 101, 110, 111\}$.
- $C = \langle 100011, 010101, 001110 \rangle$
  $= \{000000, 001110, 010101, 011011, 100011, 101101, 110110, 111000\}$.
- **Equations:** $x_2 + x_3 + x_4 = 0$, $x_1 + x_3 + x_5 = 0$, $x_1 + x_2 + x_6 = 0$.
- This is $[6, 3, 3]$ binary linear code.
- $G_{3\times 6} = [I_3|A] = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$.
- $H_{3\times 6} = [A^T|I_3] = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$.

Linear Codes

- If the encoding function E is linear in the message set, C is called a *linear code*.
- If the messages are binary strings of length k, i.e., $\mathcal{M} = \mathbb{F}_2^k$, and $E : \mathbb{F}_2^k \to \mathbb{F}_2^n$ is linear, then C is called a *binary linear code*.
- A (n, k) *q-ary linear code* C of length n and dimension k is a k-dimensional subspace of $\mathbb{F}_q^n$.
- $r = k/n$ is called the *code rate*.

Linear Codes

- Any linear map $E : \mathbb{F}_2^k \to \mathbb{F}_2^n$ can be represented using a $k \times n$ matrix G over $\mathbb{F}$ such that $E(x) = xG$.
- G is called a *generator matrix* for the code C and

$$C = \text{linear span of } G = \{mG : m \in \mathbb{F}_2^k\}.$$

- The matrix G forms a basis of C and corresponds to the encoding map sending k-bit strings to n-bit strings.

Generator matrix of a linear code

- For example

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

  generates a linear code of length $n = 7$ and dimension $k = 4$ over $\mathbb{F}_2$.

- The *message*

$$m = (0011)$$

  is then *encoded* to the *codeword*

$$c = (0011)G = (0011001).$$

Hamming distance

- The *Hamming weight* of a word is the number of non-zero coordinates.

$$hw(x) = \sum_{i=1}^{k} x_i \quad \text{where } x = (x_1, \ldots, x_n)\,.$$

- The *Hamming distance* between two words in F_2^n is the number of coordinates where they differ.

$$d(x, y) = hw(x \oplus y)\,.$$

- Hamming weight of a word is its Hamming distance from the zero.

$$hw(x) = hw(x \oplus 0) = d(x, 0)\,.$$

Minimum distance

- The *minimum distance* of a linear code C is the smallest Hamming weight of a non-zero codeword in C.

$$d = \min\{hw(x) | x \in C \setminus \{0\}\}.$$

- C is then called a (n, k, d) *code*.
- The example code is a $(7, 4, 3)$ binary Hamming code with minimum distance 3.
- The example codeword (0011001) has the minimum weight 3.
- *Singleton bound*: $d \leq n - k + 1$.
- **Theorem** (*Error correction*)
  A linear code C with minimum distance $d \geq 2t + 1$ can correct any t errors.

Example — Single Error Correcting Code

- $m > 1$.
- $n = 2^m - 1$.
- $k = 2^m - 1 - m$.
- $d = 3$.
- Hamming Code $\mathcal{H}_m$.

Parity-check matrix

- A *parity-check matrix* of linear code C is a $(n-k) \times n$ matrix H such that

$$cH^T = 0$$

  for all codewords $c \in C$.
- In particular, for any generating matrix G,

$$G \cdot H^T = 0\,.$$

- The code C can be defined using the *parity check* $H : \mathbb{F}_2^n \to \mathbb{F}_2^{n-k}$.

$$C = \text{span}(G) = \text{null space}(H) = \{x \in \mathbb{F}_2^n | xH^T = 0\}\,.$$

Parity-check matrix

- Relation between G and H:

$$G_{k\times n} = [I_k | A_{k\times(n-k)}] \iff H_{(n-k)\times n} = [A^T_{k\times(n-k)} | I_{n-k}]$$

- Given a generating matrix G, a parity-check matrix H can be computed using the Gaussian elimination to compute the $(n-k)\times n$ kernel of G.
- **Theorem** (*Independence of the H matrix*)
  A linear code C has minimum distance d if and only if any set of $d-1$ columns of H are linearly independent.

$\mathcal{H}_3$ — $[7, 4, 3]$ **Hamming Code**

- Generating Matrix

$$G = G_{4\times 7} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

- Parity check Matrix

$$H = H_{3\times 7} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

Decoding problem

- *Decoding problem:* Given $y \in \mathbb{F}_2^n$, find the codeword $c \in C$ "closest" to y, assuming that there is a unique closest codeword.
- *Generic decoding is hard* for a binary linear code with no obvious structure.
- *Berlekamp, McEliece, van Tilborg [?]* showed that the general decoding problem for linear codes over $\mathbb{F}_2$ is NP-complete.

Good News

- There are code families with fast decoding algorithms
  - Hamming codes
  - BCH codes
  - Reed-Solomon codes
  - Reed-Muller codes
  - Gabidulin codes
  - alternant codes
  - Algebraic-geometric codes
  - ...
  - Goppa codes

Syndrome decoding

- Given $y \in F_2^n$, determine an *error vector* e of a given weight t such that

$$c = y - e$$

  is a codeword.
- Given $y \in F_2^n$, compute the *syndrome*

$$s = s(y) = yH^T = (c + e)H^T = eH^T .$$

- Find a weight-t word e such that $s = s(y) = eH^T$.
  - Tricky part for the decoder.
  - Here code structure helps build efficient algorithms
- Find $c = r - e$ and then use linear algebra to recover x.

Syndrome decoding

- Corresponding to any error vector of weight upto t there is a unique syndrome.
- Syndrome decoding for errors of weight upto t, $2t + 1 \leq d \leq n - k + 1$.
- Syndrome table for $\mathcal{H}_3$:

Syndrome	Error
000	0000000
001	0000001
010	0000010
011	1000000
100	0000100
101	0100000
110	0010000
111	0001000

Syndrome decoding for $\mathcal{H}_3$

- Let $y = (0011011)$.
- Then

$$yH^T = (0011011)\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = (010)$$

- From Syndrome table , $(010) \leftrightarrow e = (0000010)$.
- Therefore, $xG = y + e = (0011011) + (0000010) = (0011001)$.

Standard Array Decoding

- The *coset* of a code C corresponding to a vector $a \in \mathbb{F}_2^n$ is

$$C_a = C + a = \{x + a | x \in C\}.$$

- A vector $x_0 + a \in C + a$ is called a *coset leader* of $C + a$ if

$$hw(x_0 + a) = \min\{hw(y) | y \in C + a\}.$$

- The syndrome of any element $y \in C_a$ is same.

Standard Array Decoding

- Let $C = \langle 0111, 0101 \rangle = \{0000, 0111, 0101, 0010\}$
- Then the cosets of C are
  - $C_{0000} = \{0000, 0111, 0101, 0010\}$
  - $C_{1000} = \{1000, 1111, 1101, 1010\}$
  - $C_{0100} = \{0100, 0011, 0001, 0110\}$
  - $C_{1001} = \{1001, 1110, 1101, 1011\}$
  - (Blue colour vectors represents the coset leaders.)
- The syndrome of any element $y \in C_a$ is same.
- Corresponding to any error vector of weight upto t there is a unique syndrome.

Cyclic Codes

- C is a *cyclic code* if

$$(c_0, c_1, \ldots, c_{n-1}) \in C \Rightarrow (c_{n-1}, c_0, c_1, \ldots, c_{n-2}) \in C$$

- The code is *invariant* under cyclic shifts.
- Using the polynomial representation $\mathbb{F}_2^n \simeq \mathbb{F}_2[X]/\langle X^n - 1\rangle$, C is a cyclic code if

$$c(X) \in C \Rightarrow Xc(X) \in C$$

Cyclic Codes

- For any (n, k) cyclic code C, we can compute a $(n-k)$ degree monic polynomial $g(X)$, called a *generator polynomial*, such that

$$C = \{m(X)g(X) | m(X) \in \mathbb{F}_2[X], \deg(m(X)) < k\}.$$

- Generator polynomial of n-length cyclic codes divide $X^n - 1$.
- *Parity check polynomial* of C is a polynomial $h(X)$ such that

$$(X^n + 1) = h(X)g(X)$$

  for some generating polynomial $g(X)$ of C.
- **Encoding:** $m(X) \to m(X)g(X) \mod X^n - 1$ where $m(X) \longleftrightarrow (m_0, m_1, \ldots, m_{k-1})$.

Cyclic Codes — Generating Matrix

- Let $g(X)$ be the generating polynomial then

$$G = \begin{pmatrix} g(X) \\ Xg(X) \\ \vdots \\ X^{n-r-1}g(X) \end{pmatrix}.$$

- If $g(X) = g_0 + g_1X + g_2X^2 + \cdots + g_{r-1}X_{r-1} + g_rX^r$ then

$$G = \begin{pmatrix} g_0 & g_1 & \cdots & g_r & 0 & \cdots & 0 \\ 0 & g_0 & \cdots & g_{r-1} & g_r & \cdots & 0 \\ & & & \cdots & & & \\ 0 & \cdots & 0 & g_0 & g_1 & \cdots & g_r \end{pmatrix}.$$

Cyclic Codes — Parity Check Matrix

- Let $h(X)$ be the parity check polynomial then parity check

$$H = \begin{pmatrix} \hat{h}(X) \\ X\hat{h}(X) \\ \vdots \\ X^{n-r-1}\hat{h}(X) \end{pmatrix} \text{ where } \hat{h} = \sum_{i=0}^{k-1} h_{k-1-i}X^i \,.$$

- If $h(X) = h_0 + h_1X + h_2X^2 + \cdots + h_{k-1}X_{k-1} + h_kX^k$ then

$$H = \begin{pmatrix} 0 & \cdots & 0 & h_k & \cdots & h_1 & h_0 \\ 0 & \cdots & h_k & h_{k-1} & \cdots & h_0 & 0 \\ & & & \cdots & & & \\ h_k & h_{k-1} & \cdots & h_0 & 0 & \cdots & 0 \end{pmatrix} .$$

Bose-Chaudhari-Hocquenghem codes

- Let α be the n-th root of unity in $\mathbb{F}_{q^m}$ for a given m.
- A (narrow-sense) *BCH code* with design distance $2t+1$ and length n over $\mathbb{F}_q$ has generator polynomial

$$g_{BCH}(X) = LCM(minpoly_q(\alpha) minpoly_q(\alpha^2) \dots minpoly_q(\alpha^{2t}))$$

  where $minpoly_q(\alpha^i)$ is the minimum degree polynomial with coefficients from $\mathbb{F}_q$ with α^i as a root.
- $minpoly_q(\alpha^i)$ divides $x^n - 1$ for all i.

BCH codes — Parity Check matrix

- Thus, any for any codeword $c(X)$,

$$(c(\alpha), c(\alpha^2), \ldots, c(\alpha 2t)) = 0.$$

- In other words, the parity check matrix

$$H_{BCH} = \begin{pmatrix} 1 & \alpha & \alpha^2 & \ldots & \alpha^{(n-1)} \\ 1 & \alpha^2 & \alpha^4 & \ldots & \alpha^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha^{2t} & \alpha^{4t} & \ldots & \alpha^{2t(n-1)} \end{pmatrix}$$

- $BCH_q(n, 2t) = NullSpace(H_{BCH})$ in $\mathbb{F}_q^n$.
- Any set of $2t$ columns from H_{BCH} is linearly independent over $\mathbb{F}_q$. Therefore BCH code with design distance $2t+1$ can correct any t errors.

BCH Codes — Decoding

- $r(X) = c(X) + e(X)$, $hw(e) \leq t$.
  - find syndrome,
  - find error,
  - find information symbols.
- For any α^i, $i = 1, 2, \ldots, 2t$ we have

$$r(\alpha^i) = c(\alpha^i) + e(\alpha^i) = e(\alpha^i) = \sum_{j=0}^{n-1} e_j (\alpha^i)^j \, .$$

BCH Codes — Decoding

- Suppose e has errors in ν locations for some $\nu \leq t$.
- Let those locations be $j_1, j_2, \ldots, j_\nu$.
- Then,

$$r(\alpha^i) = \sum_{l=1}^{\nu} e_{j_l} (\alpha^i)^{j_l} \,,\, \; i = 1, 2, \ldots, 2t \,.$$

BCH Codes — Decoding

- Let $X_l = \alpha^{j_l}$ and $S_i = r(\alpha^i)$.
- Therefore we have the set of equations

$$\begin{aligned} S_1 &= e_{j1}X_1 + e_{j2}X_2 + \cdots + e_{j\nu}X_\nu\,; \\ S_2 &= e_{j1}X_1^2 + e_{j2}X_2^2 + \cdots + e_{j\nu}X_\nu^2\,; \\ &\cdots \\ S_{2t} &= e_{j1}X_1^{2t} + e_{j2}X_2^{2t} + \cdots + e_{j\nu}X_\nu^{2t}\,. \end{aligned}$$

- Note that $X_l = \alpha^{j_l}$ indicates the location of the l th error (i.e, j_l) while e_{j_l} is the error value at that position.
- We want to get both X_l and the e_{j_l} in that order.
- Direct solving for X_l involve nonlinear equations.
- So we use another trick.

BCH Codes — Decoding

- *Error Locator Polynomial*: A polynomial whose roots are X_l^{-1}, $l = 1, \ldots, \nu$.

$$\Gamma(x) = \prod_{i=1}^{\nu}(1 - X_l x) = 1 + \Gamma_1 x + \Gamma_2 x^2 + \cdots + \Gamma_\nu x^\nu \, .$$

- If we have the coefficients Γ_i, then getting the roots of $\Gamma(x)$ is equivalent to finding error locations
  - can be done by evaluations of $\Gamma(x)$.
- If we have the error locations, we can use the equations in the previous slide to get the error values.
- Coefficients Γ_i and the syndromes are related by Newton's identities.

BCH Codes — Decoding

- *Newton's identities*:

$$\begin{pmatrix} S_1 & S_2 & \dots & S_\nu \\ S_2 & S_3 & \dots & S_{\nu+1} \\ \vdots & \vdots & \ddots & \vdots \\ S_\nu & S_{\nu+1} & \dots & S_{2\nu-1} \end{pmatrix} \begin{pmatrix} \Gamma_\nu \\ \Gamma_{\nu-1} \\ \vdots \\ \Gamma_1 \end{pmatrix} = \begin{pmatrix} S_{\nu+1} \\ S_{\nu+2} \\ \vdots \\ S_{2\nu} \end{pmatrix}$$

- Above equation is well defined for $\nu \leq t$.

BCH Codes — Decoding

- Set $\nu = t$.
- Form M_ν and find $\det(M_\nu)$.
- If $\det(M_\nu) = 0$ then set $\nu \leftarrow \nu - 1$ and repeat the previous step.
- If M_ν is invertible, solve for coefficients Γ_i, $i = 1, 2, \ldots, \nu$.
- Finally solve for the error values.

Reed Solomon codes

- RS Code is a BCH Code with $n = q^m - 1$ over $\mathbb{F}_q^m$. Thus,

$$g_{RS}(X) = (X - \alpha)(X - \alpha^2) \dots (X - \alpha^{2t})$$

- Another way to encode RS code: For any $m(X)$ (upto degree $k-1$), the codeword is

$$(m(1), m(\alpha), \dots, m(\alpha^{n-1})$$

- min distance: $d = n - k + 1$.

Generalized Reed Solomon codes

- GRS Codes also have max distance $d = n - k + 1$.
- $v = (v_1, v_2, \ldots, v_n)$: non-zero elements in $\mathbb{F}_q^m$.
- $\beta = (\beta_1, \beta_2, \ldots, \beta_n)$: distinct elements in $\mathbb{F}_q^m$.
- The $GRS(\beta, v)$ is the set of all vectors of the form $(v_1 m(\beta_1), v_2 m(\beta_2), \ldots, v_n m(\beta_n))$, where $m(X)$ is any polynomial of degree $< k$.

GRS codes

- The parity matrix H of a GRS Code takes the form

$$H_{GRS} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \beta_1 & \beta_2 & \dots & \beta_n \\ \beta_1^2 & \beta_2^2 & \dots & \beta_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ \beta_1^{n-k-1} & \beta_2^{n-k-1} & \dots & \beta_n^{n-k-1} \end{pmatrix} \begin{pmatrix} y_1 & \dots & & \\ & y_2 & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ & \dots & & y_n \end{pmatrix} = XY$$

  where $y = (y_1, \dots, y_n)$ is some vector (with non-zero y_i) such that H_{GRS} is an appropriate H matrix to $GRS(\beta, v)$.
- $GRS(\beta, v) = NullSpace(H_{GRS})$ in $\mathbb{F}_q^m$.

Alternant Codes

- Long BCH codes are not good
  - rate (k/n) and error correction (d/n) don't keep growing together.
- Rectified by Alternant codes.
- Subcodes of GRS codes.
- *Alternant Code*: For β consisting of n distinct values from $\mathbb{F}_{q^m}$, and y being non-zero values from $\mathbb{F}_{q^m}$,

$$A(\beta, y) = NullSpace(H_{GRS}) \text{ in } \mathbb{F}_q\,.$$

Goppa Codes

- Fix
  - a prime power q,
  - a positive integer m,
  - a positive integer $n \leq q^m$,
  - an integer $t < \frac{n}{m}$,
  - distinct $\alpha_1, \ldots, \alpha_n \in \mathbb{F}_{q^m}$, and
  - a polynomial $g(x) \in \mathbb{F}_{q^m}[x]$ of degree t such that for all i, $g(\alpha_i) \neq 0$.
- The *Goppa code* $\Gamma(\alpha_1, \ldots, \alpha_n, g)$ consists of all words $c = (c_1, \ldots, c_n)$ in $\mathbb{F}_q^n$ such that

$$\sum_{i=1}^{n} \frac{c_i}{x - \alpha_i} \equiv 0 \pmod{g(x)}.$$

- $g(x)$ is called the *Goppa polynomial*.

Goppa Code as an Alternant Code

- $G(z)$ a polynomial with coefficients from $\mathbb{F}_{q^m}$.
- $\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ are n elements such that

$$G(\alpha_i) \neq 0, \text{ for all } i = 1, 2, \ldots, n\,.$$

- Let $y = (G(\alpha_1)^{-1}, G(\alpha_2)^{-1}, \ldots, G(\alpha_n)^{-1})$.
- Then $Goppa(\alpha, G(z)) = A(\alpha, y)$
- If α is set of all non-zeros of $G(z)$ then the Goppa code is completely determined by $G(z)$.
- Has an optimised decoding algorithm because of its further structure.

Properties of Goppa codes

- $\Gamma(\alpha_1, \ldots, \alpha_n, g)$ has length n and dimension $k \geq n - mt$.
- The minimum distance is at least $\deg g + 1 = t + 1$.
  - For $q = 2$, the minimum distance is $2t + 1$.
- Patterson decoding efficiently decodes t errors in the binary case; otherwise only $t/2$ errors can be corrected.

Section Outline

11 McEliece Cryptosystem
- Introduction
- Setup
- Key Generation
- Encryption
- Decryption

McEliece Cryptosystem [?]

- McEliece proposed a public-key cryptosystem based on error-correcting codes in 1978.
- Secret key is a linear error-correcting code with an efficient decoding algorithm.
- Public key is a transformation of the secret inner code which is hard to decode.

Setup

- The (public) system parameters are three positive integers n, k, t with $n \geq n - kt \geq 2t + 1$.
- The public key is a random-looking $k \times n$ matrix G_{pub} with entries in $\mathbb{F}_2$.
- A message is a vector in $\mathbb{F}_2^k$.
- A ciphertext $c \in \mathbb{F}_2^n$ is a codeword with an added error of weight t.

Key Generation

- Let G be a generator matrix of a Goppa code Γ
  - of length n and dimension k and error-correcting cappability t;
  - and which allows for "efficient" decoding.
- Let S be a random $k \times k$ invertible matrix (called *scrambler*).
- Let P be a random $n \times n$ permutation matrix.
- The secret key is the triple (G, S, P).
- The public key is $G_{pub} = SGP$.
- Finding G given G_{pub} seems to be harder than generic decoding.

Encryption

- A message $m \in \mathbb{F}_2^k$ is encrypted to a ciphertext $c \in \mathbb{F}_2^n$ as

$$c = mG_{pub} + e$$

  where $e \in \mathbb{F}_2^n$ is a random error vector of weight t.
- Assuming G_{pub} is indistinguishable from a random code and c is inditinguishable from a random vector in $\mathbb{F}_2^n$, the hardness of the generic decoding guarantees the security.

Decryption

- The legitimate receiver knows G, S, P with $G_{pub} = SGP$ and a decoding algorithm for Γ.
- Computes $yP^{-1} = mSG + eP^{-1}$.
  - Since P is a permutation, eP^{-1} has exactly the same values as in e with their positions permuted so eP^{-1} is just another error vector of weight t.
- Applies the decoding algorithm of to recover mSG which is a codeword in Γ.
- Recovers mS and then obtains $m = (mS)S^{-1}$.

Section Outline

12 Niederreiter Cryptosystem
- Introduction
- Setup
- Key Generation
- Encryption
- Decryption
- Systematic Matrices

Niederreiter Cryptosystem [?]

- A variation of the McEliece cryptosystem.
- Instead of the generator matrix, the parity check matrix is used for the encryption.
- Similar security
- More efficient implementation

Setup

- The (public) system parameters are three positive integers n, k, t with $n \geq n - kt \geq 2t + 1$.
- The public key is a random-looking $(n - k) \times n$ matrix H_{pub} with entries in $\mathbb{F}_2$.
- A message is an "error" vector in $\mathbb{F}_2^n$.
- A ciphertext is a syndrome in $\mathbb{F}_2^{n-k}$.

Key Generation

- Let H be a parity-check matrix of a Goppa code Γ
  - of length n and dimension k and error-correcting cappability t;
  - and which allows for "efficient" decoding.
- Let S be a random $(n-k) \times (n-k)$ invertible matrix (called *scrambler*).
- Let P be a random $n \times n$ permutation matrix.
- The secret key is the triple (H, S, P).
- The public key is $H_{pub} = SHP$.

Encryption

- A message $m \in \mathbb{F}_2^l$, $l \approx \lfloor \log_2 \binom{n}{t} \rfloor$, is encrypted to a ciphertext $c \in \mathbb{F}_2^{n-k}$ as follows:
- First, m is encoded as a binary string e of length n and weight t.
- Set $c = H_{pub} e^T$.

Decryption

- The legitimate receiver knows H, S, P with $H_{pub} = SHP$ and a decoding algorithm for Γ.
- Compute $\hat{c} = S^{-1}c$.
- Obtain $\hat{e}$ from $\hat{c}$ using the decoding algorithm.
- Compute $e = P^{-1}\hat{e}$ of length n and weight t.
- The message m can now be recovered by inverting the encoding.

Key Generation for Systemetic Parity Matrix

- Let Γ be a Goppa code
  - of length n and dimension k and error-correcting cappability t;
  - and which allows for "efficient" decoding.
- Select a random Goppa polynomial $g(z)$ of degree t over $\mathbb{F}_q$.
- Randomly choose n elements of $\mathbb{F}_q$ that are not roots of $g(z)$ as the support L.
- Compute the $(n-k) \times n$ parity check matrix H according to L and $g(z)$.
- Bring H to systematic form using Gauss-Jordan elimination: $H_{pub} = SH$.
- The secret key is the triple $(L, g(z), S)$.
- The public key is H_{pub}.

Decryption for Systemetic Parity Matrix

- Compute $\hat{c} = S^{-1}c$.
- Obtain e from $\hat{c}$ using the decoding algorithm.
- The message m can now be recovered by inverting the encoding.
- No need for the permutation.

Section Outline

13 McEliece v/s Niederreiter
- Equivalence
- Comparison

Security Equivalence of McEliece and Niederreiter

- Let C be an $(n, k, 2t + 1)$ linear code.
- Then given a generator matrix G of C, a parity check matrix H of C can be found using linear algebra with complexity $O(n^3)$.
- Also given a parity check matrix H of C, a generator matrix G of C can be found using linear algebra with complexity $O(n^3)$.
- For instance, by converting the generator matrix into the systematic form $G = [I_k A]$, the parity check matrix is simply given by $H = -[A^T I_{n-k}]$.

McEliece $\Longleftarrow$ Niederreiter

- Let $c = mG_{pub} + e$ be a ciphertext in the McEliece cryptosystem with public key G_{pub}.
- Obtain a parity check matrix H from G_{pub} and let $z = cH^T$.
- Then, $z = cH^T = mG_{pub}H^T + eH^T = eH^T$.
- So, if Niederreiter cryptosystem is broken, then given the "syndrome" z, the "error" e can be found.
- Then, $mG_{pub} = c - e$ is a codeword and m can be obtained using linear algebra, breaking McEliece cryptosystem in complexity $O(n^3)$.

McEliece $\Longrightarrow$ Niederreiter

- Let $z = yH_{pub}^T$ be a ciphertext in the Niederreiter cryptosystem with public key H_{pub}.
- Obtain a generator matrix G from H_{pub}.
- A solution $c \in \mathbb{F}_2^n$ such that $z = cH_{pub}^T$ of weight larger than or equal to t can be obtained using linear algebra with complexity $O(n^3)$.
- Then, $0 = cH_{pub}^T - z = cH_{pub}^T - yH_{pub}^T = (c-y)H_{pub}^T$.
- So $c - y$ is a codeword in C and there exists some $m \in \mathbb{F}_2^k$ such that $c - y = mG$ or $c = mG + y$.
- So, if McEliece cryptosystem is broken, then given c, m and hence e can be found, breaking Niederreiter cryptosystem in complexity $O(n^3)$.

Size Comparison

Size	McEliece	Niederreiter	Remarks
Public Key	$(n-k) \times k$	$(n-k) \times k$	same
Secret Key	$(n-k)^2 + tk + n \log n$	$(n-k)^2 + tk + nt$	almost same when $k \approx n/2$
PlainText	k	$\approx \lfloor \log_2 \binom{n}{t} \rfloor$	almost same when $k \approx n/2$
CipherText	n	$n-k$	smaller in Niedereiter so better information rate

Efficiency Comparison

Operation	McEliece	Niederreiter	Remarks
Encryption	1 matrix multiplication	1 (simpler) matrix multiplication + encoding	similar
Decryption	1 matrix multiplication + decoding + permutation	1 matrix multiplication + decoding + message decoding	Niederreiter is more expensive

Summing up the Comparison

- Niederreiter has smaller ciphertext blocks.
- Results in higher information rate.
- Decryption is more expensive. Some workload may be moved to the encryption, which still remains a fast operation.

Section Outline

14 McEliece PKC: Advantages and Disadvantages

McEliece and Niederreiter PKC

- McEliece PKC - proposed in 1978, almost as old as Public-Key Cryptography.
  - Based on binary Goppa codes
  - The hard problem: GDP
- Niederreiter PKC - proposed in 1986
  - Based on GRS codes, but replaced with Goppa after Sidelnikov-Shestakov attack, 1992
  - The hard problem: SDP

Advantages of Goppa Codes

- have a very efficient decoding algorithm
- share many characteristics with random codes
- does not disclose any visible structure that an attacker could exploit
- there are sufficiently large number of codes with given parameters

Description: McEliece PKC

Private-key

- Generator matrix G of a t-error correcting code C
- An invertible matrix $S_{k\times k}$.
- A permutation matrix $P_{n\times n}$.

Public-key

$G'_{k\times n} = S_{k\times k}G_{k\times n}P_{n\times n}$.

Encryption

$c \to y = cG' + e$

Decryption

- Compute $x = yP^{-1} = cSG + eP^{-1}$.
- Decode x with the secret algorithm for C. This gives cSG.
- Compute c from cSG.

Description: Niederreiter PKC

Private-key

- Parity-check matrix H of a t-error correcting code C
- An invertible matrix $S_{(n-k)\times(n-k)}$.
- A permutation matrix $P_{n\times n}$.

Public-key

$H'_{(n-k)\times n} = S_{(n-k)\times(n-k)} H_{(n-k)\times n} P_{n\times n}$.

Encryption

$c \to s' = cH'^T$, where $w(c) = t$.

Decryption

- Compute $s = s' S^{-1^T} = cP^T H^T = c' H^T$, where $c' = cP^T$.
- Decode s with the secret algorithm for C. This gives c'.
- Compute c from cP^T.

Advantages and Disadvantages of the schemes

Advantages

- No structural attacks: that reveal the structure of hidden code
- No significant decoding attacks: that recover an information set of symbols without errors, that try to decode a random code without structure
- very fast encryption and decryption
  - **Encryption** (in bo/ib): 514(McEliece), 50(Niederreiter) and 2402(RSA)
  - **Decryption** (in bo/ib): 5140(McEliece), 7863(Niederreiter) and 738112(RSA)

Disadvantages

- Low Information Rate
- Huge Key-size
- Key Size (in bytes): 67072(McEliece), 32750(Niederreiter) and 256(RSA)

Bibliography


Secure and Efficient Scheme for Delegation of Signing Rights,
ICICS, LNCS, Springer, 2014


How to Leak a Secret and Reap the Rewards too,
LatinCrypt, LNCS 8895, Springer, 2014


A Secure Anonymous Proxy Multi-signature Scheme,
SeCrypt, SciTePress, 2014


Remote Cache-timing Attacks Against AES,
CS2, HiPEAC, ACM, 2014


Anonymous Signatures Revisited,
ProvSec, LNCS 5848, pp.140-153, Springer, 2009


Public-Key Encryption with Searchable Keywords based on Jacobi Symbols,
IndoCrypt, LNCS 4859, pp.282-296, Springer, 2007