A Framework for Data Encryption Based on Joint Employment of Cryptography and Coding

Miodrag Mihaljević
Mathematical Institute
Serbian Academy of Sciences and Arts

SEE Forum on Data Science

Belgrade, 20 June 2016

Roadmap

- Preface & Abstract of the Talk
- A Historical Prospective on Cryptology
- Models of Communication Channels
- The LPN Problems and Two Paradigms for Security Enhancement
- An Illustration of Recent Results:

  - A Framework for Security Enhanced Encryption Based on Channels with Synchronization Errors

  - Information-Theoretic Security Evaluation

  - Computational-Complexity Security Evaluation
- Concluding Notes

I. Preface
&
Abstract of the Talk

what we are talking about

GRAND CHALLENGES FOR ENGINEERING

Imperatives, Prospects, and Priorities

SUMMARY OF A FORUM

Prepared by Steve Olson
for the
NATIONAL ACADEMY OF ENGINEERING

THE NATIONAL ACADEMIES PRESS
Washington, DC

tractable engineering system challenges that must be met in this century for human life as we know it to continue on this planet. The committee received thousands of inputs from around the world to determine its list of Grand Challenges for Engineering, and its report was reviewed by more than 50 subject-matter experts, making it among the most reviewed of Academy studies. The 14 Grand Challenges for Engineering are to

Make solar energy economical
Provide energy from fusion
Develop carbon sequestration
Manage the nitrogen cycle
Provide access to clean water
Improve urban infrastructure
Advance health informatics
Engineer better medicines
Reverse-engineer the brain
Prevent nuclear terror
Secure cyberspace
Enhance virtual reality

Advance personalized learning
Engineer the tools of scientific discovery.

The Grand Challenges were not ranked in importance or likelihood of solution, nor was any strategy proposed for solving them. Rather, they were offered as a way to inspire the profession, young people, and the public at large to seek the solutions.

In 2010 a plan was put forth to prepare engineering students to think about careers devoted to addressing the Grand Challenges. Called the NAE Grand Challenge Scholars Program, it was the first specific action taken toward achieving solutions to the challenges on a global scale.

In 2013 the first Global Grand Challenges Summit was held in London, cosponsored by the Royal Academy of Engineering, the Chinese Academy of Engineering, and the US National Academy of Engineering in their first joint effort. In September 2015 a second Global Grand Challenges Summit was held in Beijing, with more than 800 attendees invited by the three academies. The third Global Grand Challenges Summit, to be hosted by the NAE in the United States in 2017, will be held in conjunction with a new FIRST Robotics international event aimed at engaging the world's youth on

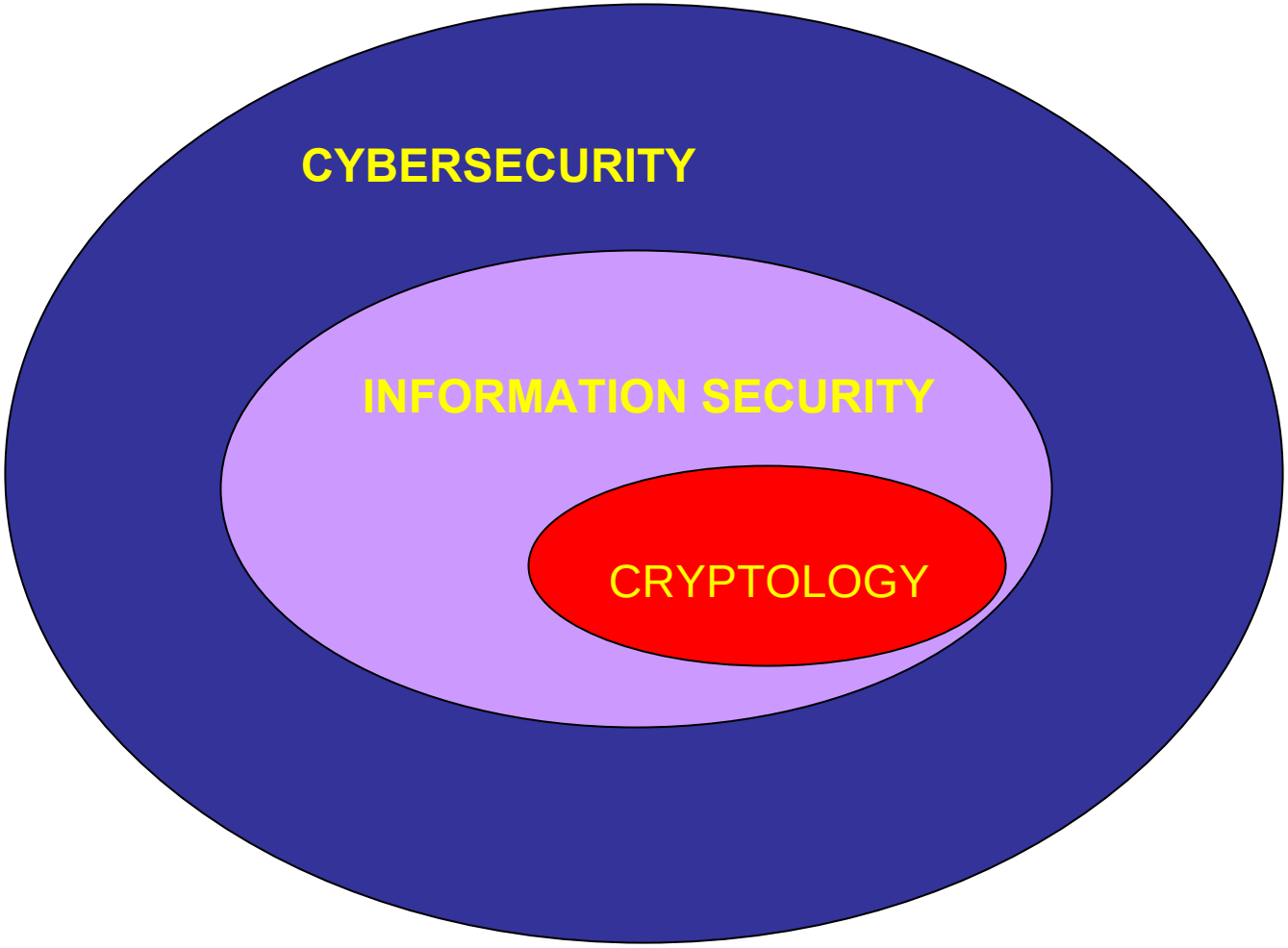
The 14 challenges were selected from hundreds of suggestions from engineers, scientists, policymakers and ordinary people around the world

- Make solar energy affordable.
- Provide energy from fusion
- Develop carbon sequestration methods.
- Manage the nitrogen cyclee.
- Provide access to clean water.
- Restore and improve urban infrastructure.
- Advance health informatics.
- Engineer better medicines.
- Reverse-engineer the brain.
- Prevent nuclear terror.
- Secure cyberspace.
- Enhance virtual reality.
- Advance personalized learning.
- Engineer the tools for scientific discovery.

.

Cybersecurity

Cryptology, Information Security & CyberSecurity

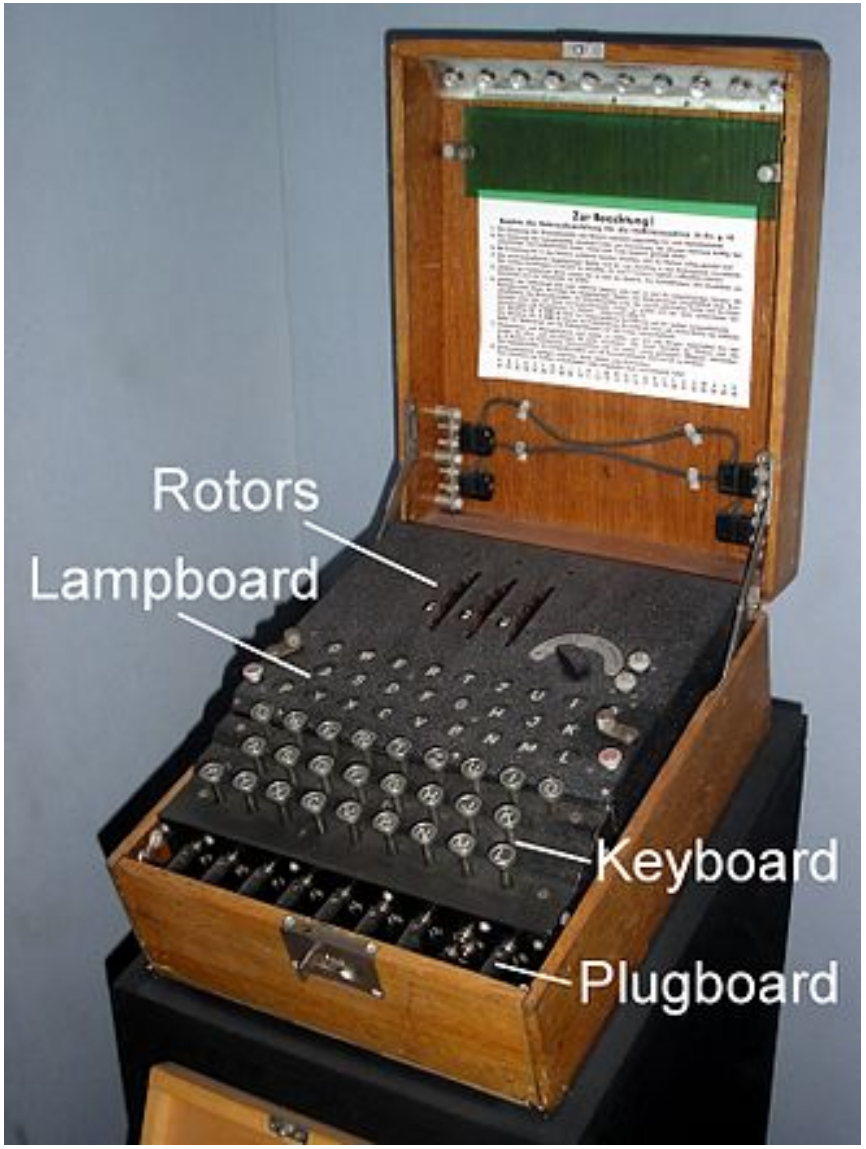


Abstract

- An important topic of Data Science is Data Security where data confidentiality appears as a very important issue. When a heavy employment of encryption is necessary, minimization of the overheads and fit into the implementation constraints are required which preserve cryptographic security as well
- Accordingly, this talk addresses an approach for design of compact encryption which supports minimization of the overheads, fits into the asymmetric implementation constraints and provides certain level of the provable security.
- The addressed approach is based on a combination of traditional encryption and coding in order to provide security enhancement of lightweight encryption algorithms which fits into the implementation constraints.

II. A Historical Prospective

The First Computer
&
From Art of Secrtet Writing to Cryptology


Rotors
Lampboard
Keyboard
Plugboard

Cryptanalysis of Enigma

Known plaintext attacking scenario

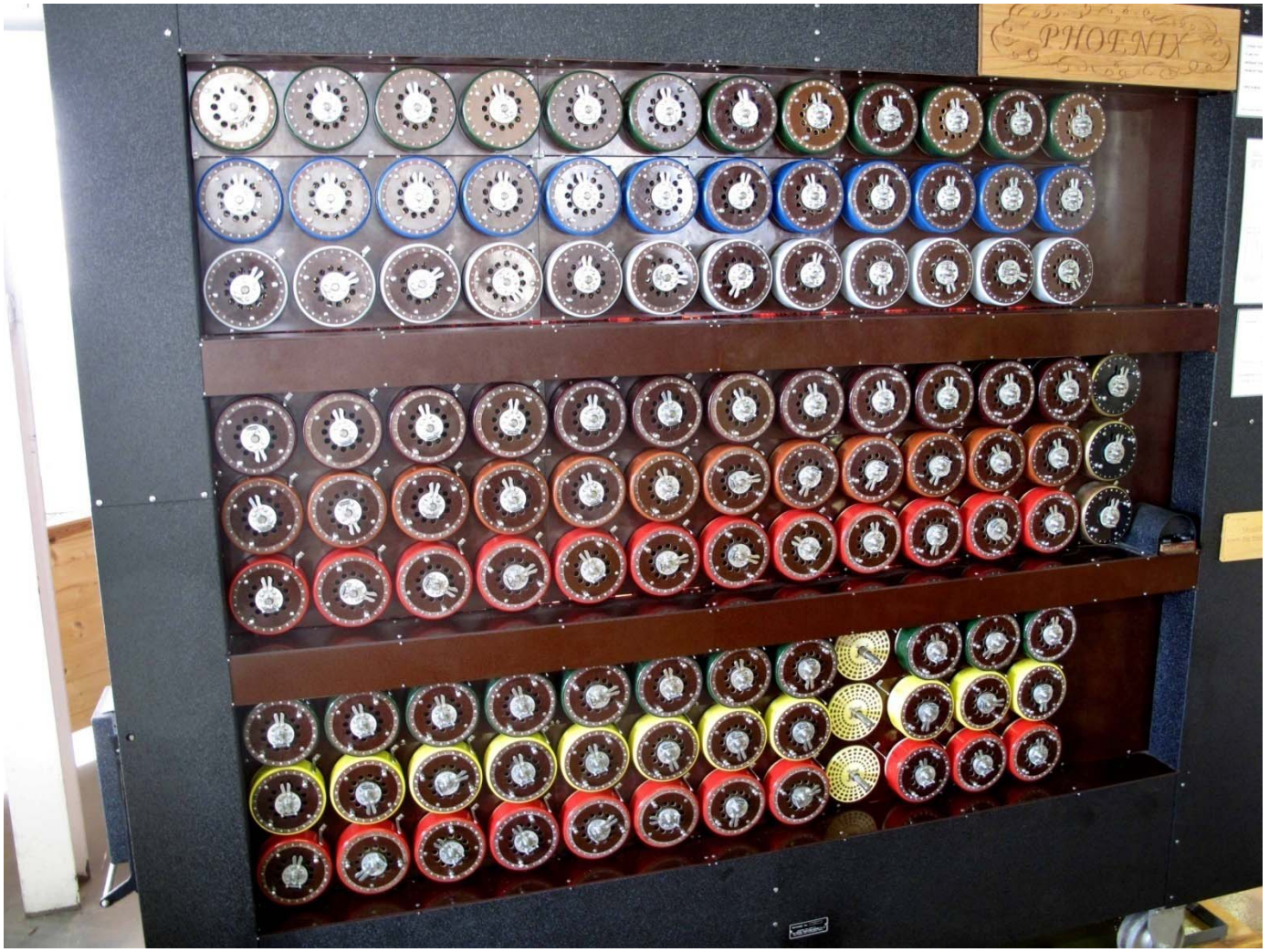
- How to obtain the pairs of corresponding plaintext & ciphertext
- ...

Paradigm of Exhaustive Serach

- How to perform serch over a set of hypothesis
- ...

The working rebuilt Bomb at Bletchley Park museum. Each of the rotating drums simulates the action of an Enigma rotor.

The (electro-mechanical) computer



Establishment of Cryptology and Information Theory

Claude Shannon (1916-2001)

Two Key Papers

- **Information Thory**
- C. E. Shannon, “**A mathematical theory of communication**”. Bell System Technical Journal, vol. 27, pp. 379–423 and 623–656, July and October 1948.

- **Cryptology**
- C. E. Shannon "**Communication Theory of Secrecy Systems**". Bell System Technical Journal, vol. 28 (4), pp. 656–715, 1949.

III. Some Models of Noisy Channels

Additive Noise

&

Synchronization Noise

Noisy Channels

Channels with Additive Noise

- Erasure Channel
- Binary Symmetric Channel
- Gaussian Channel
- ...

Channels with Synchronization Noise

- Channels with insertion
- Channels wit deletions
- Channels with Insertion, deletion and additive noise
- ...

Binary Erasure Channel

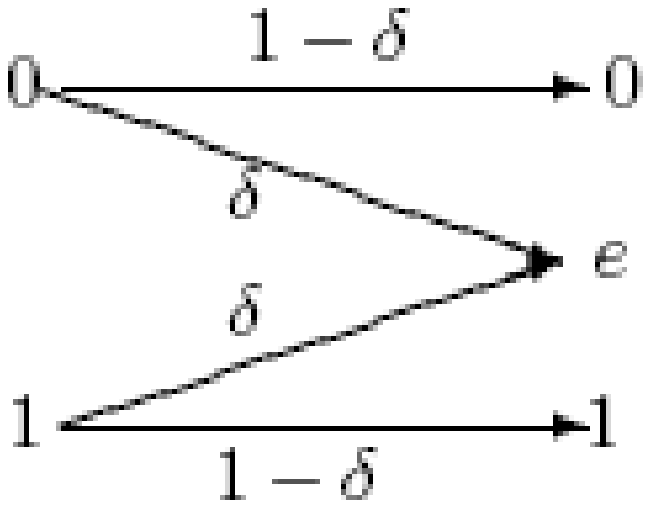
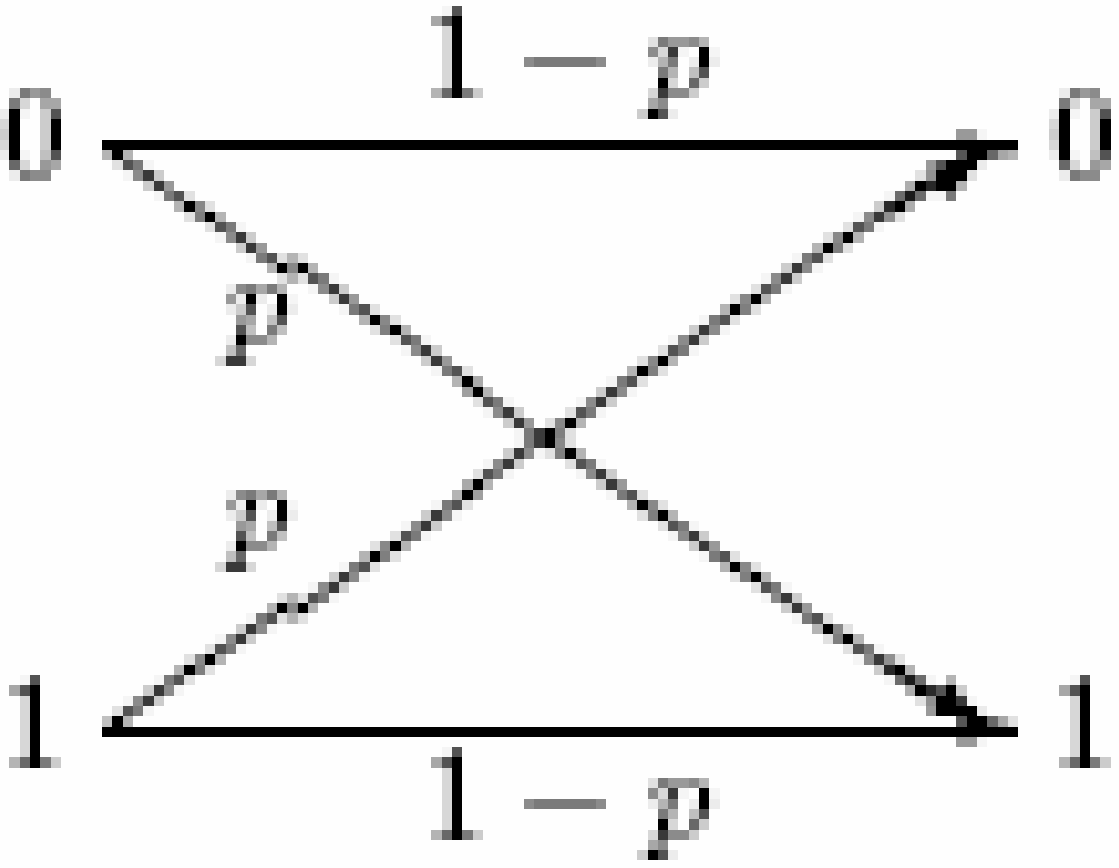
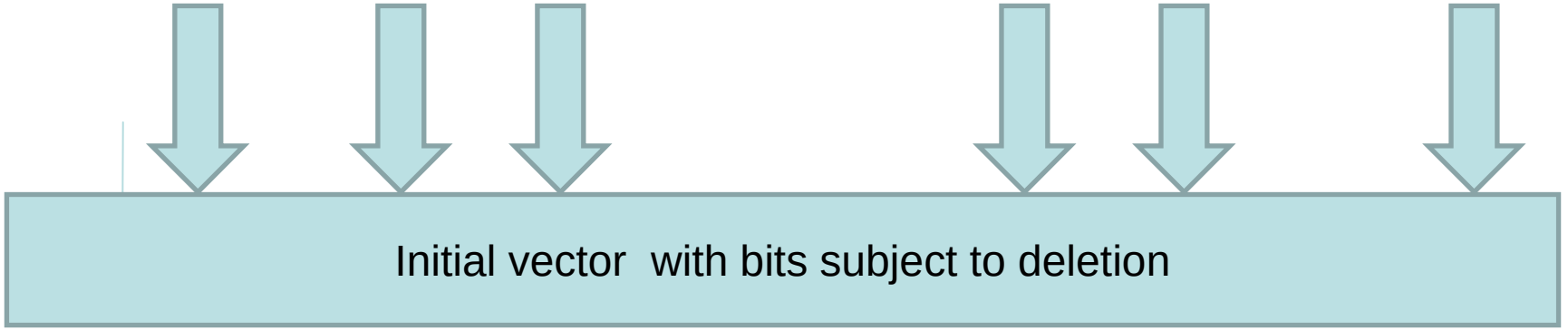


Figure 1: Binary erasure channel (BEC) with erasure probability δ.

Binary Summetric Channel



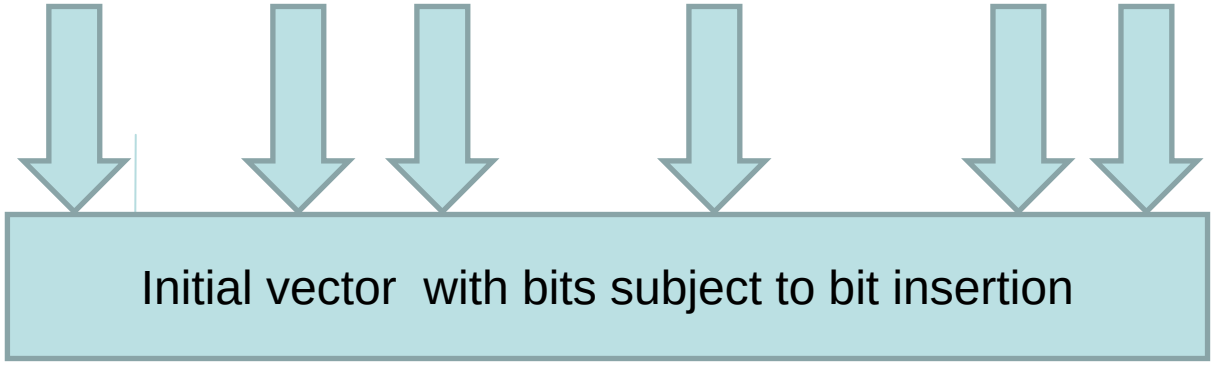
Binary Channel with Random Bit Deletion



Shrinked vector after the chanel with random bits deletion

Deletion of bits is RANDOM – Positions of deleted bits are UNKNOWN

Binary Channel with Random Bit Insetion



Expanded vector after the chanel with random bits insertion

Insertion of bits is RANDOM – Positions of insereted bits are UNKNOWN

IV. The LPN Problems and Two Cryptograhic Paradigms

Security of Encryption and Implementation Complexity

- Mainly based on heuristic assumptions
- Particularly when the encryption is based on employment of finite state machines
- Lightweight encryption implies additional challenges

...

- Security enhancement appears as an interesting approach

...

- Asymmetric implementation complexity of encryption and decryption also appears as an interesting issue

The LPN Problems (Learning Parity in Noise)

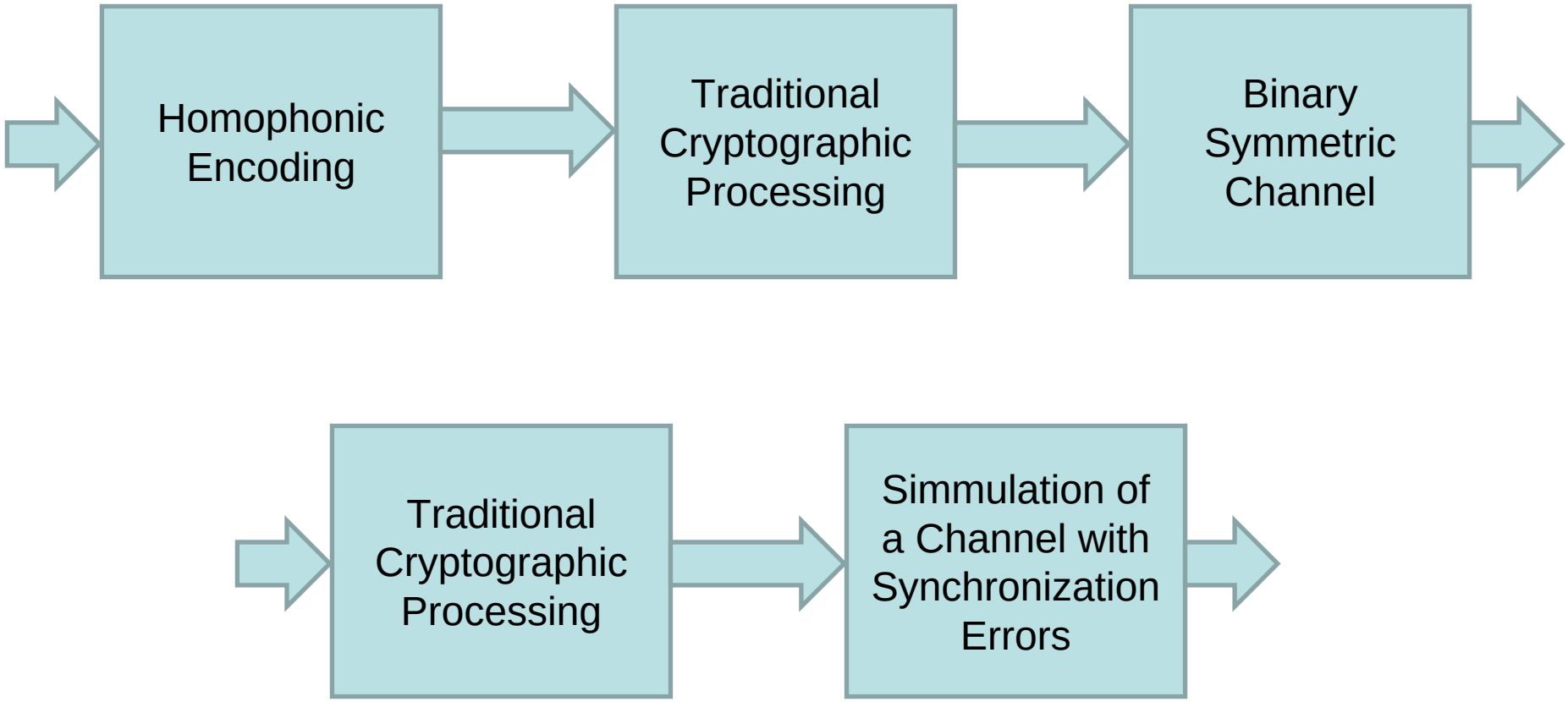
Basic

- Informally: Solving a system of linear equations with “the right-hand sides” visible through a binary symmetric channel.

Generalized

- Informally: Solving a system of linear equations with “the right-hand side” visible through a channel with synchronization errors .

Two Paradigms for Security Enhancement



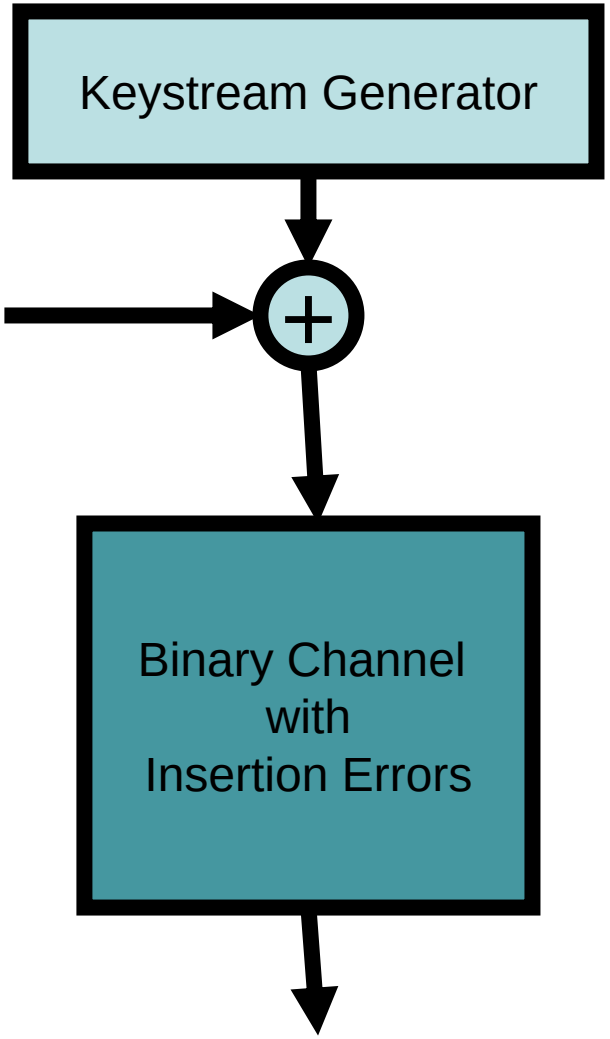
Illustrative References

- M.J. Mihaljevic, A. Kavcic and K. Matsuura, "An Encryption Technique for Provably Secure Transmission from a High Performance Computing Entity to a Tiny One," Mathematical Problems in Engineering, vol. 2016, Article ID 7920495, 10 pages, May 2016. doi:10.1155/2016/7920495.
- S. Tomovic, M.J. Mihaljevicv, A. Perovic and Z. Ognjanovic, "A Protocol for Provably Secure Authentication of a Tiny Entity to a High Performance Computing One," Mathematical Problems in Engineering, vol. 2016, Article ID 9289050, 9 pages, May 2016. doi:10.1155/2016/9289050.
- A. Kavcic, M.J. Mihaljevic and K. Matsuura, "Light-Weight Secrecy System Using Channels with Insertion Errors: Cryptographic Implications", IEEE Information Theory Workshop 2015, Jeju Island, Korea, 11-15 Oct. 2015, Proceedings, pp. 257-261, 2015.
- F. Oggier and M.J. Mihaljevic, "An Information-Theoretic Security Evaluation of a Class of Randomized Encryption Schemes", IEEE Transactions on Information Forensics and Security, vol. 9, no. 2, pp. 158-168, Feb. 2014.
- M.J. Mihaljevic, S. Gangopadhyay, G. Paul and H. Imai, ``Internal State Recovery of Keystream Generator LILI-128 Based on a Novel Weakness of the Employed Boolean Function", Information Processing Letters, vol. 112, no. 21, pp. 805-810, November 2012.
- M.J. Mihaljevic, S. Gangopadhyay, G. Paul and H. Imai, ``State Recovery of Grain-v1 Employing Normality Order of the Filter Function", IET Information Security, vol. 6, no. 2, pp. 55-64, June 2012.

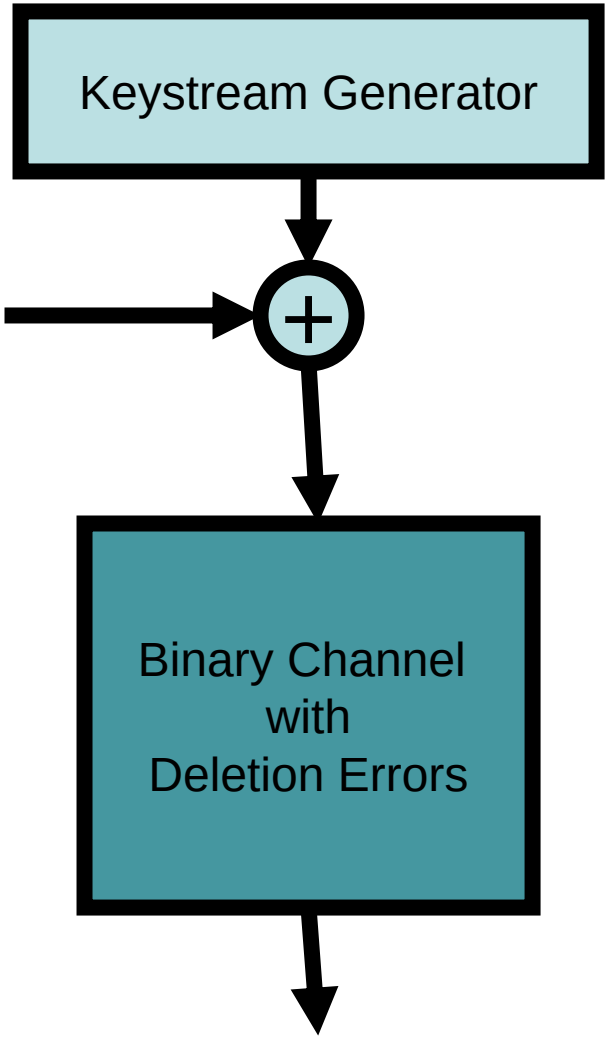
V. A Framework for Security Enhancement Based on the Channels with Synchronization Errors

Desired Model of Encryption an Attacker Should Face

Encryption at Party I

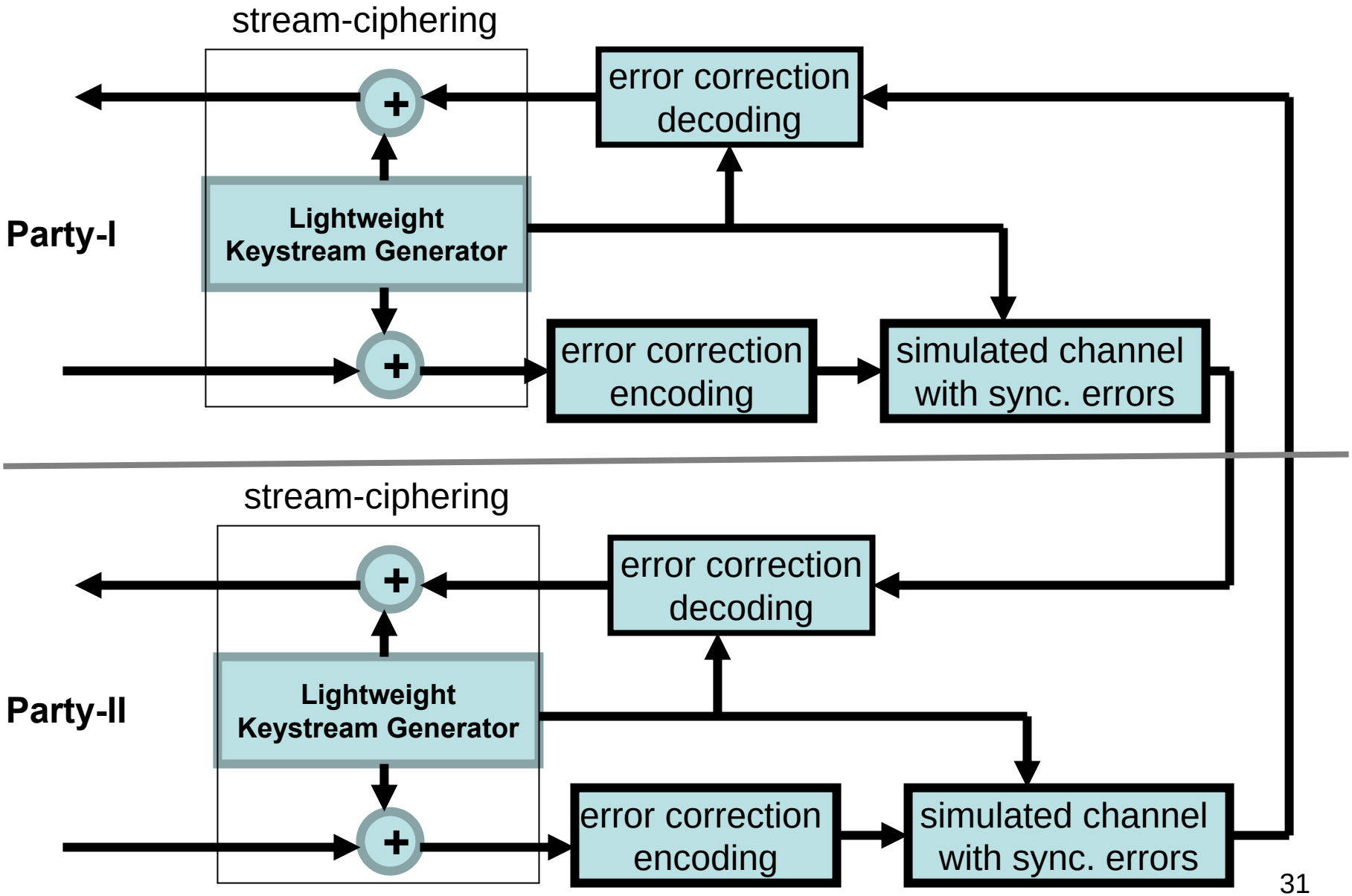


Encryption at Party II

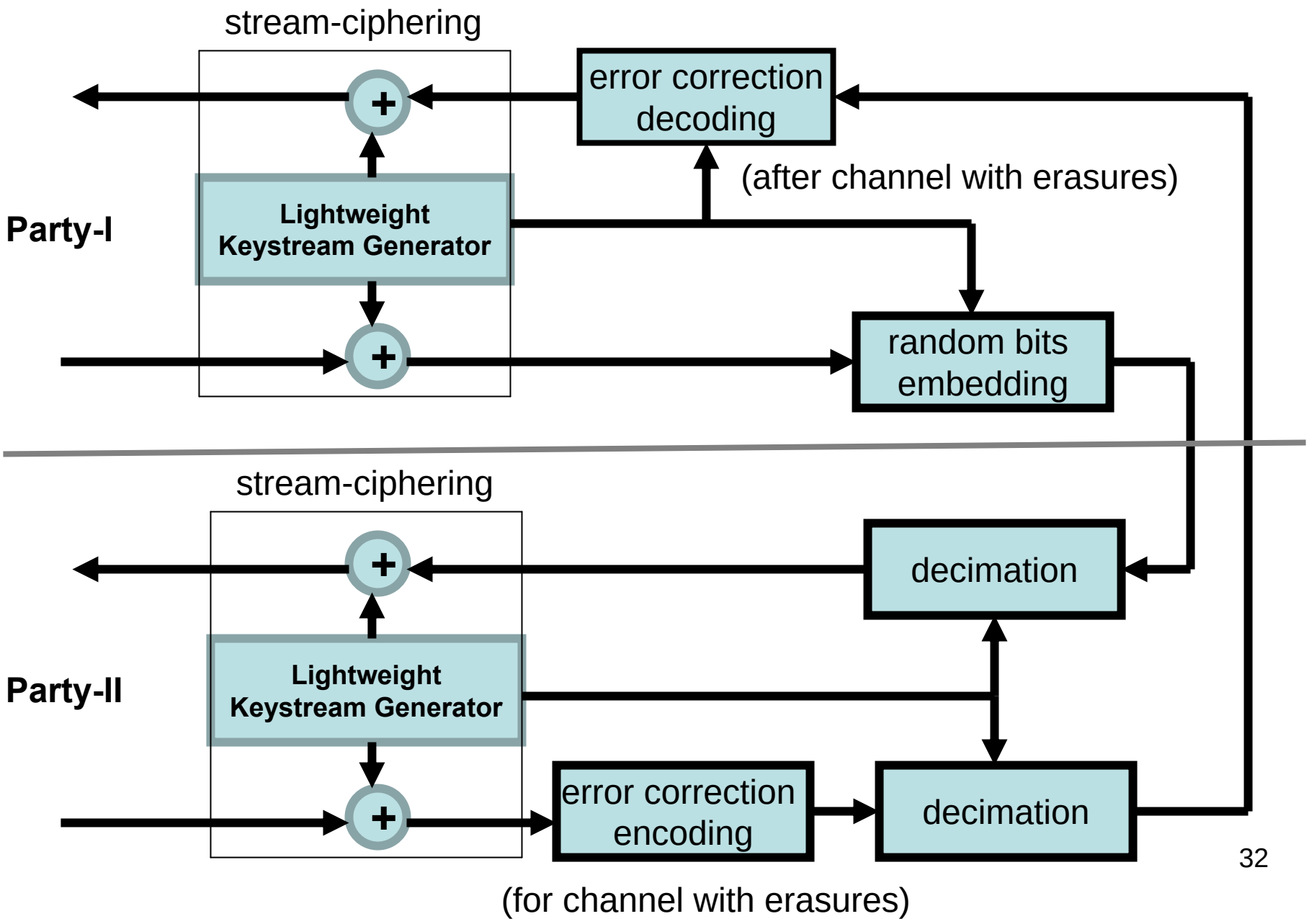


Attacker Side

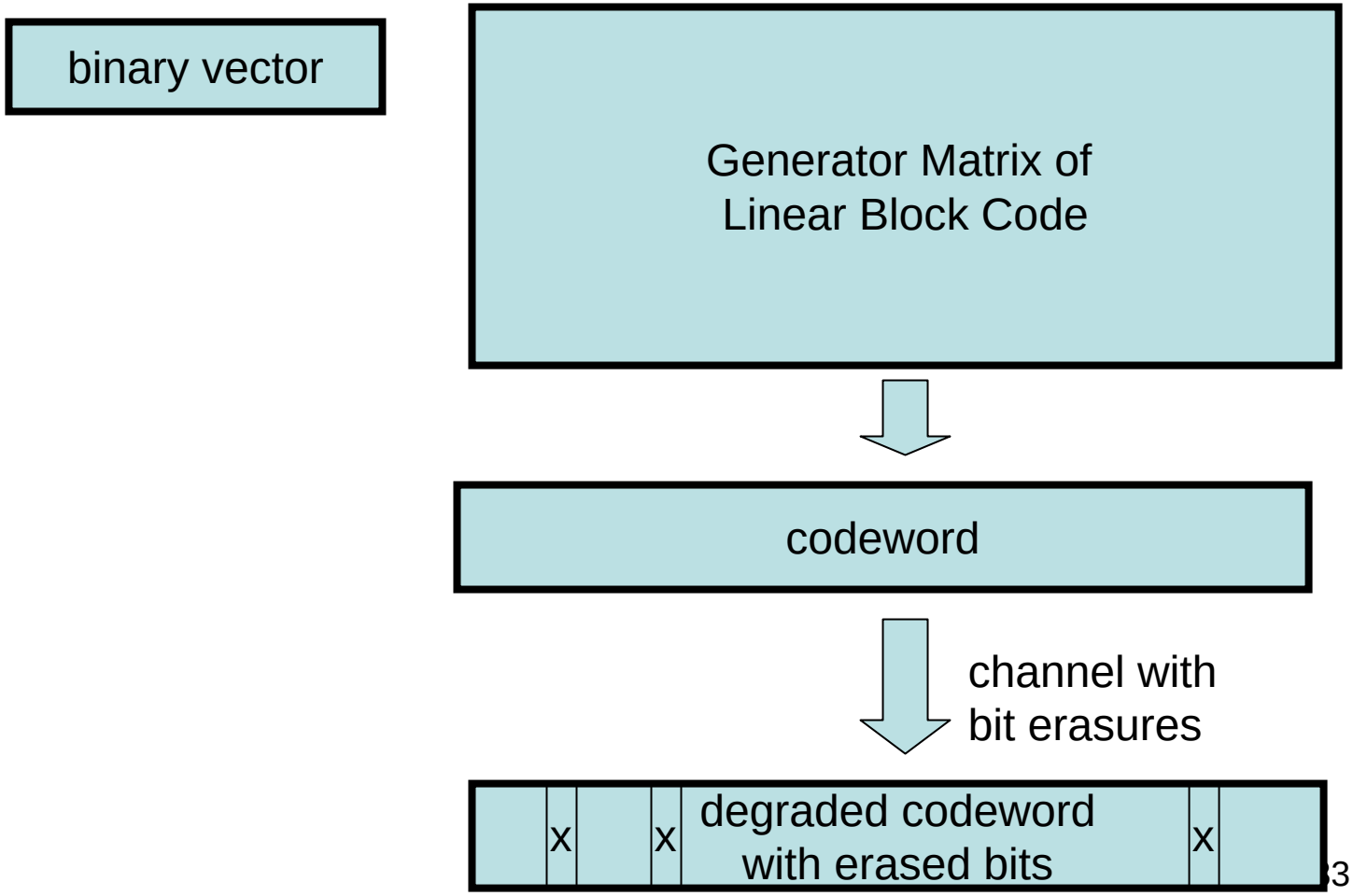
A Framework for Encryption Based on Simulated Channels with Synchronization Errors



A Framework for Encryption with Asymmetric Implementation Complexity

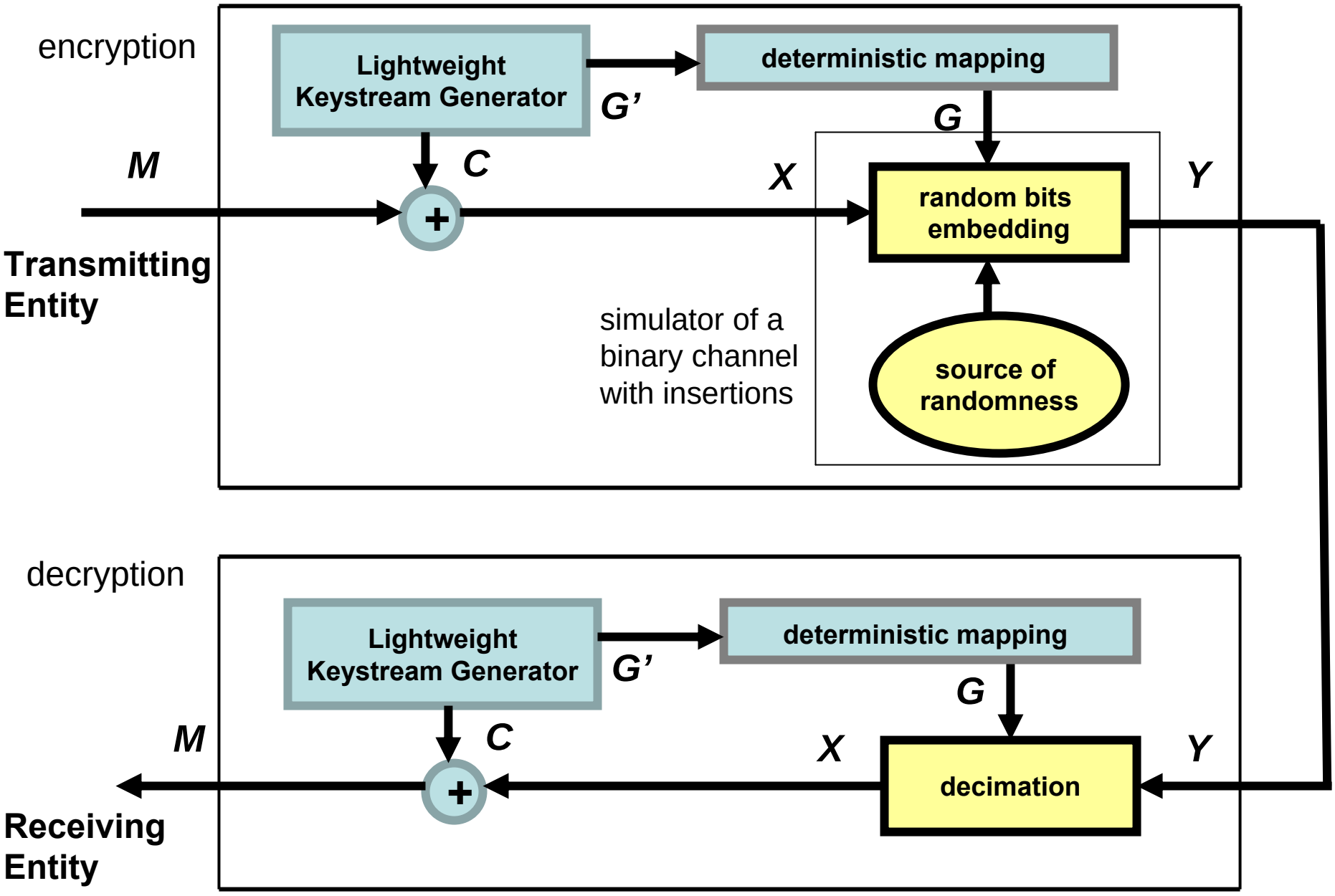


A Linear Binary Block Code Encoding Paradigm



V.1 Particular Instantiation Under Security Evaluation

A Framework for Encryption and Decryption with Asymmetric Implementation Complexity



Two Approaches for Security Evaluation

- **Information Theoretic Security Evaluation**
- **Computational Complexity Security Evaluation**

V.2 Information-Theoretic Security Evaluation

Preliminaries

Eve (the eavesdropper) and Bob (the intended receiver) both receive the string $Y^{(n)}$ containing the randomly inserted symbols. The eavesdropper, not having access to the shared source of randomness G^n, cannot easily parse the string $Y^{(n)}$ to recover X^n. The intended receiver, on the other hand, has access to G^n, and since G_k represents the length of the inserted string between any two symbols X_k and X_{k+1}, the intended receiver (Bob) can easily remove the inserted symbols $\underline{B_k}$ from $Y^{(n)}$ (i.e., decimate $Y^{(n)}$) to recover X^n. In other words, by sharing the source of randomness G^n.

The sequence C^n is a pseudo-random sequence, but for the purpose of computing information-theoretic quantities, we assume that C^n is modeled to be statistically indistinguishable from a sequence of iid Bernoulli-$\frac{1}{2}$ random variables.

The information-theoretic quantity of interest is the *iud information rate* defined as the information rate between X^n and $Y^{(n)}$ when the symbols X_k are independent and uniformly distributed (iud)

$$\mathcal{I}_{\text{iud}}(X;Y) \triangleq \lim_{n\to\infty} \frac{1}{n} I\left(X^n; Y^{(n)}\right)\Big|_{p(x^n)=2^{-n}}.$$

The information rate $\mathcal{I}_{\text{iud}}(X;Y)$ represents the amount of information that the eavesdropper can "*learn*", on average, about X after observing Y. The information rate $\mathcal{I}_{\text{iud}}(X;Y)$ is not computable in closed-form, but is attainable using Monde-Carlo techniques.

The information rate $\mathcal{I}_{\text{iud}}(X;Y)$ is not computable in closed-form, but is attainable using Monde-Carlo techniques. For example, known bounds are

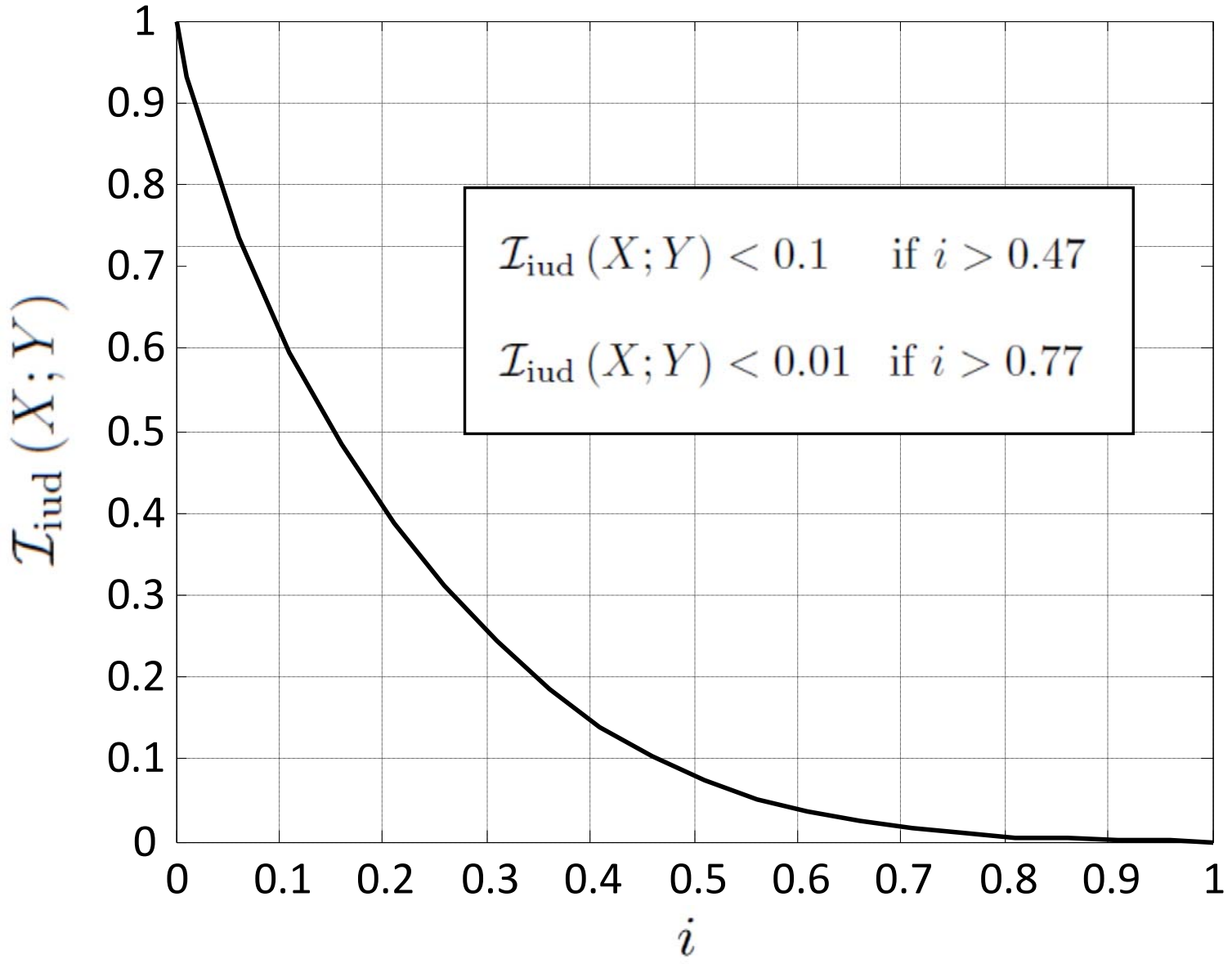
$$\mathcal{I}_{\text{iud}}(X;Y) \geq \frac{1}{n} I\left(X^n; Y^{(n)}\right)\Big|_{p(x^n)=2^{-n}} - \frac{1}{n} H\left(\mathcal{L}\left(Y^{(n)}\right)\right) \quad (1)$$

$$\mathcal{I}_{\text{iud}}(X;Y) \leq \frac{1}{n} I\left(X^n; Y^{(n)}\right)\Big|_{p(x^n)=2^{-n}}. \quad (2)$$

For large n, the correction term $\frac{1}{n} H\left(\mathcal{L}\left(Y^{(n)}\right)\right)$ equals

$$\frac{1}{n} H\left(\mathcal{L}\left(Y^{(n)}\right)\right) = \frac{1}{2n} \log_2\left(\frac{2\pi \mathrm{e} \cdot i \cdot n}{(1-i)^2}\right) + O\left(n^{-2}\right). \quad (3)$$

Illustrative Numerical Example



We already established that learning X after observing Y is extremely unfavorable for the eavesdropper because the information rate $\mathcal{I}_{\text{iud}}(X;Y)$ is low for large insertion probabilities i. However, the eavesdropper may adopt a strategy in which she first attempts to *learn* the sequence G^n, and then attempt to crack X^n. To study the effects of this strategy, let us define the following quantities:

$$\mathcal{I}_{\text{iud}}(G;Y) \triangleq \lim_{n\to\infty} \frac{1}{n} I\left(G^n; Y^{(n)}\right)\Big|_{p(x^n)=2^{-n}}$$

$$\mathcal{I}_{\text{iud}}(X,G;Y) \triangleq \lim_{n\to\infty} \frac{1}{n} I\left(X^n, G^n; Y^{(n)}\right)\Big|_{p(x^n)=2^{-n}}$$

$$\mathcal{I}_{\text{iud}}(X;Y|G) \triangleq \lim_{n\to\infty} \frac{1}{n} I\left(X^n; Y^{(n)}\Big| G^n\right)\Big|_{p(x^n)=2^{-n}}$$

$$\mathcal{I}_{\text{iud}}(G;Y|X) \triangleq \lim_{n\to\infty} \frac{1}{n} I\left(G^n; Y^{(n)}\Big| X^n\right)\Big|_{p(x^n)=2^{-n}}$$

Proposition 1:

$$\begin{aligned}
\mathcal{I}_{\mathsf{iud}}(G;Y) &= 0 && (1)\\
\mathcal{I}_{\mathsf{iud}}(X;Y|\,G) &= 1 && (2)\\
\mathcal{I}_{\mathsf{iud}}(X,G;Y) &= 1 && (3)\\
\mathcal{I}_{\mathsf{iud}}(G;Y|\,X) &= 1-\mathcal{I}_{\mathsf{iud}}(X;Y)\,. && (4)
\end{aligned}$$

V.3 Computational Complexity Security Evaluation

Definition 1: The Adversarial Indistinguishability Experiment consists of the following steps:

1. The adversary $\mathcal{A}$ chooses a pair of messages $(\mathbf{m}_0; \mathbf{m}_1)$ of the same length n, and passes them on to the encryption system for encrypting.

2. A bit $b \in \{0,1\}$ is chosen uniformly at random, and only one of the two messages $(\mathbf{m}_0; \mathbf{m}_1)$, precisely $\mathbf{m}_b$, is encrypted into ciphertext $\text{Enc}(\mathbf{m}_b)$ and returned to $\mathcal{A}$;

3. Upon observing $\text{Enc}(\mathbf{m}_b)$, and without knowledge of b, the adversary $\mathcal{A}$ outputs a bit b_0;

4. The experiment output is defined to be 1 if $b_0 = b$, and 0 otherwise; if the experiment output is 1, denoted shortly as the event $(\mathcal{A} \to 1)$, we say that $\mathcal{A}$ has succeeded.

Definition 2. An encryption scheme provides indistinguishable encryptions in the presence of an eavesdropper, if for all probabilistic polynomial-time adversaries $\mathcal{A}$

$$\Pr[\mathcal{A} \to 1 | \mathrm{Enc}(\mathbf{m}_b)] \leq \frac{1}{2} + \epsilon \,,$$

where $\epsilon = \mathrm{negl}(n)$ is a negligibly small function.

Proposition 2: Let the encrypted mapping of M^n into X^n be such that $\frac{1}{2}+\epsilon$ equals the advantage of the adversary $\mathcal{A}$ (specified by Definition 2) to win the indistinguishability game (specified by Definition 1), and let the mutual information $\mathcal{I}_{iud}(X;Y)$ be known. Under these assumptions, for large n,

$$\Pr[\mathcal{A} \to 1 | Y^{(n)} = \mathbf{y}] = \frac{1}{2} + \epsilon \cdot \delta, \quad \text{where} \quad (1)$$

$$\delta \triangleq \Pr(X^n = \mathbf{x}_b | Y^{(n)} = \mathbf{y}) < \frac{1}{n} + \frac{1}{n} I\left(X^n, Y^{(n)}\right)\Big|_{p(x^n)=2^{-n}}. \quad (2)$$

Theorem 1: Let the encrypted mapping of M^n into X^n be such that $\frac{1}{2}+\epsilon$ equals the advantage of the adversary $\mathcal{A}$ (specified by Definition 2) to win the indistinguishability game (specified by Definition 1), and let the mutual information $\mathcal{I}_{iud}(X;Y)$ be known. Under these assumptions, for large n,

$$\Pr[\mathcal{A} \to 1 | Y^{(n)} = \mathbf{y}] = \frac{1}{2} + \epsilon \cdot \delta, \quad \text{where} \quad (1)$$

$$\delta < \mathcal{I}_{iud}(X;Y) + \frac{\log_2 \left[\frac{8\pi \mathrm{e} \cdot i \cdot n}{(1-i)^2} \right]}{2n} + O\left(n^{-2}\right). \quad (2)$$

Y. Liron and M. Langberg, “A Characterization of the Number of Subsequences Obtained via the Deletion Channel”, *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2300-2312, May 2015.

Let $D_t(\mathbf{Z})$ be a set of subsequences of $\mathbf{Z}$ that can be obtained from $\mathbf{Z}$ after t deletions. A family of strings, named unbalanced strings has been defined. A string is called unbalanced, if all of the runs of symbols in the string are of length 1, except for one run. Let $U_{\ell,r}^{(i)}$ be a binary string of length ℓ with r runs, in which all runs are of length 1, except for the i-th run which is of length $\ell - r + 1$. Due to symmetry $|D_t(U_{\ell,r}^{(1)})| = |D_t(U_{\ell,r}^{(r)})|$, and consequently define

$$u(\ell, r, t) = |D_t(U_{\ell,r}^{(1)})| = |D_t(U_{\ell,r}^{(r)})| \,. \qquad (1)$$

It has been shown that these extreme cases have the least number of subsequences among the unbalanced strings, as well as that they have the least amount of subsequences among all strings.

Y. Liron and M. Langberg, “A Characterization of the Number of Subsequences Obtained via the Deletion Channel”, *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2300-2312, May 2015.

Theorem (Closed-Form Formula for $u(\ell, r, t)$): For all $t < \ell$, $2 < r \leq \ell$,

(i) when $r > t$:

$$u(\ell, r, t) = d(r, t) + \sum_{i=t+r-\ell-1}^{t-2} d(r-2, i) \ , \quad (1)$$

(ii) when $r \leq t$:

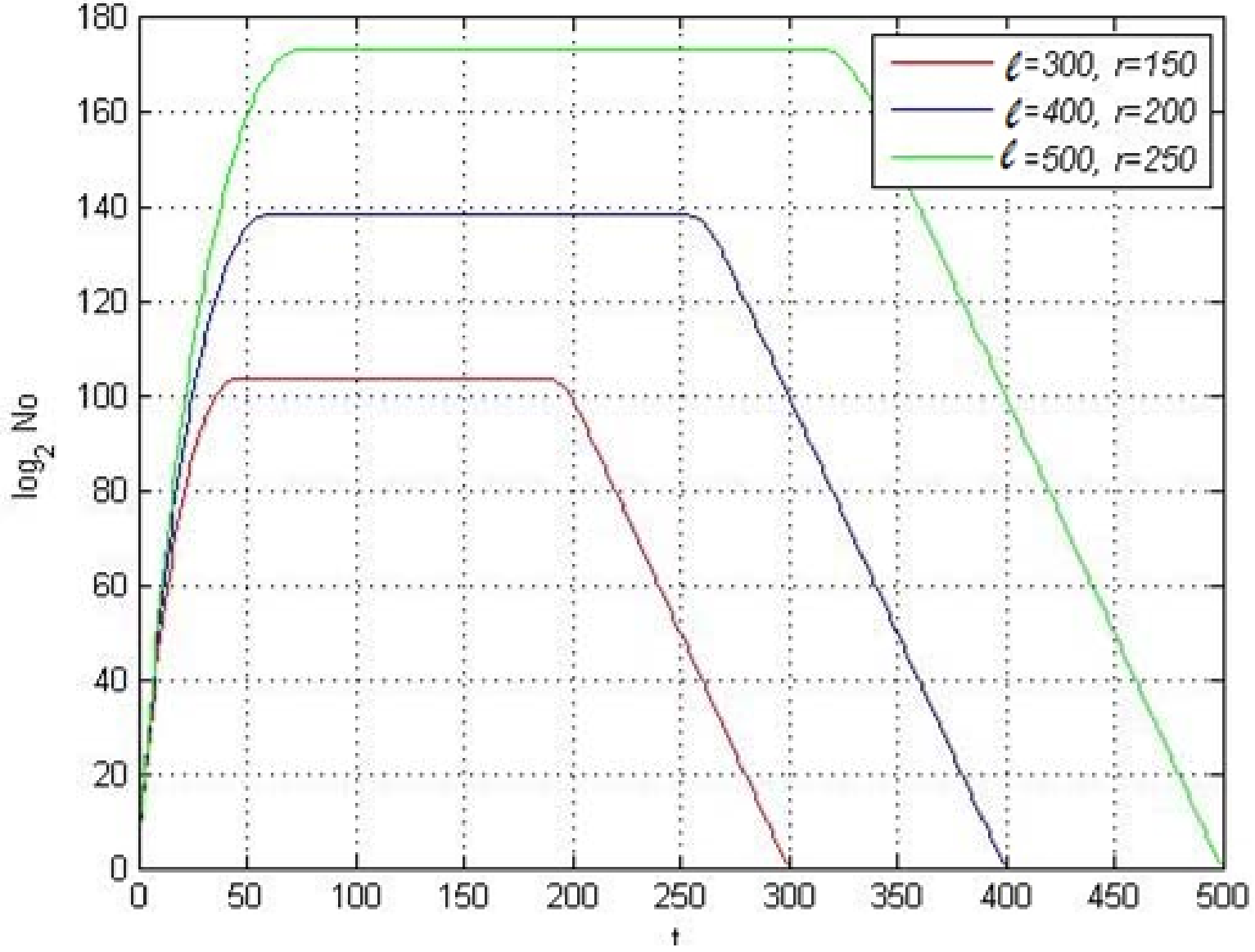
$$u(\ell, r, t) = 2 + \sum_{i=t+r-\ell-1}^{r-3} d(r-2, i) \ , \quad (2)$$

where

$$d(r, i) = |D_i(\mathbf{Z}_r^C)| = \sum_{j=0}^{i} \binom{r-i}{j} \quad (3)$$

assuming that $d(r, 0) = 1$, and for $i < 0$, $d(r, i) = 0$, and that the following conventions are employed:

$$\sum_{i=j}^{k} a_i = 0 \ \text{ when } \ j > k \ , \quad (4)$$


ℓ=300, r=150
ℓ=400, r=200
ℓ=500, r=250
log2 No
t
0
20
40
60
80
100
120
140
160
180
0
50
100
150
200
250
300
350
400
450
500

Theorem 3. Assuming that the employed keystream generator is such that the following is valid:

$$I(\mathbf{M};\mathbf{C}) = 0 \ , \ I(\mathbf{M};\mathbf{G}) = 0 \ , \ I(\mathbf{C};\mathbf{G}) = 0 \ , \tag{1}$$

and

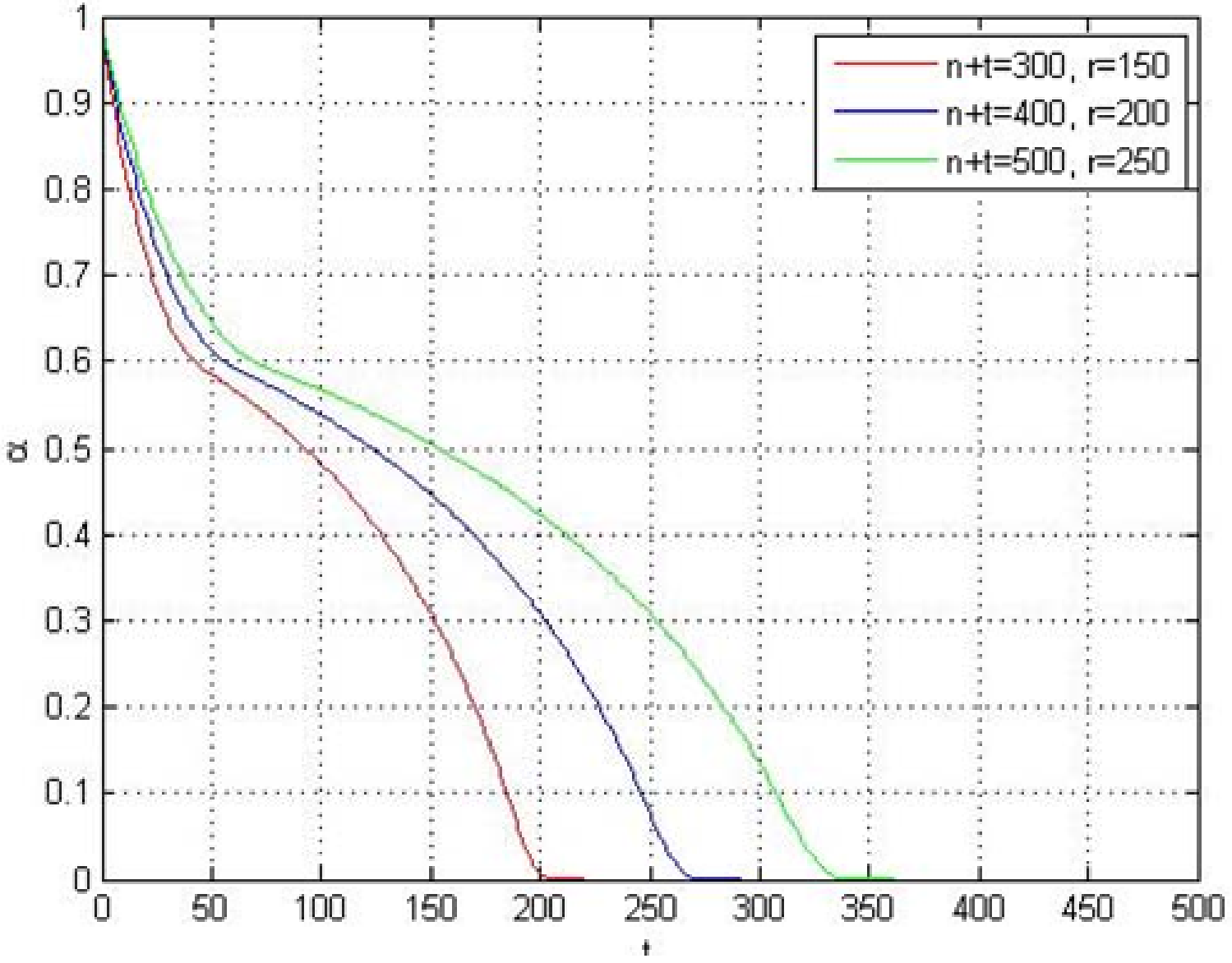
$$I(\mathbf{M};\mathbf{X}) \leq \epsilon \ . \tag{2}$$

the simulator of binary channel with random insertions provides

$$\frac{1}{n} I(\mathbf{M};\mathbf{Y}) \leq \frac{\alpha \cdot \epsilon}{n} \ , \tag{3}$$

$$\alpha = 1 - \frac{1}{n} \log_2(u(n+t,r,t)) \ , \tag{4}$$

where $u(n+t,r,t)$ is number of certain equally likely subsequences.


n+t=300, r=150
n+t=400, r=200
n+t=500, r=250
α
t
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
50
100
150
200
250
300
350
400
450
500

Concluding Notes

Thank You Very Much for the Attention,

and

QUESTIONS Please!