

Universidade Federal de Minas Gerais

Departamento de Matemática - ICEX

**Modelo satelital para geração de chaves
criptográficas perfeitamente seguras**

Rodrigo dos Santos Marques Porto

Novembro de 2012

Rodrigo dos Santos Marques Porto

Modelo satelital para geração de chaves criptográficas
perfeitamente seguras

Dissertação apresentada ao corpo docente de Pós-Graduação em Matemática do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do título de Mestre em Matemática.

Orientador: Prof. Dr. Jeroen van de Graaf

Belo Horizonte

2012



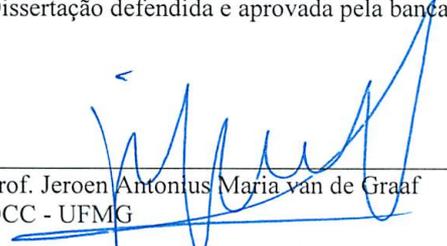
Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Secretaria de Pós-Graduação em Matemática
(31) 3409.5963 FAX 3409.5797
e-mail: pgmat@mat.ufmg.br // www.mat.ufmg.br/pgmat

FOLHA DE APROVAÇÃO

Modelo Satelital para Geração de Chaves Criptográficas Perfeitamente Seguras

RODRIGO DOS SANTOS MARQUES PORTO

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:



Prof. Jeroen Antonius Maria van de Graaf
DCC - UFMG



Prof. Marcelo de Oliveira Terra Cunha
UFMG



Prof. Anderson Clayton Alves Nascimento
UnB

Belo Horizonte, 30 de novembro de 2012.

Agradecimentos

Aos meus pais, Luiza e Waldemar, que tanto me ensinaram com suas palavras e seus exemplos.

Ao meu irmão que me mostrou que quando queremos mudar, não há o que nos impeça. O parabenizo pelo ser humano que se tornou.

Aos meus amores, Priscila e Alice, que abdicaram de seus horários de lazer para me confortar nos momentos de dificuldade, tanto pelos problemas encontrados na dissertação, quanto pelas batalhas profissionais travadas nesse período.

Ao meu orientador, Jeroen van de Graaf, pelo apoio, motivação, compreensão e momentos de boas conversas.

Ao professor Paulo Antônio Fonseca Machado, com o qual nunca tive aula formal, mas que me ajudou muito no início do mestrado, me indicando o caminho a ser tomado, tendo, inclusive, me orientando e me apresentando ao professor Jeroen.

Aos professores do Departamento de Matemática, que nunca se negaram a sanar minhas dúvidas, tendo contribuído imensamente para minha formação.

Aos colegas e amigos que formei no Departamento de Matemática, em especial à Luiza Vidigal, ao Luiz Gustavo, ao Danilo Sanção e ao José Jozelmo, os quais me presentearam com boas conversas, conselhos e discussões durante nossos estudos.

À Andrea e à Kelli, pela ajuda com toda a burocracia durante o curso.

Aos meus amigos da Polícia Federal, que lutam ao meu lado diariamente em busca de uma sociedade justa e que tanto me incentivaram a retomar os estudos. Juntos “*somos fortes na linha avançada, sem da luta os embates temer*” (trecho do Hino do Departamento de Polícia Federal).

“Quando você perceber que, para produzir, precisa obter a autorização de quem não produz nada; quando comprovar que o dinheiro flui para quem negocia não com bens, mas com favores; quando perceber que muitos ficam ricos pelo suborno e por influência, mais que pelo trabalho, e que as leis não nos protegem deles, mas, pelo contrário, são eles que estão protegidos de você; quando perceber que a corrupção é recompensada, e a honestidade se converte em auto-sacrifício; então poderá afirmar, sem temor de errar, que sua sociedade está condenada.”

Ayn Rand

Resumo

Esta dissertação revisita o artigo *Unconditional Security in Cryptography*, de Stefan Wolf [31], abordando um modelo de geração e distribuição de chaves que permite a utilização do *One-Time Pad* com a segurança que lhe é peculiar, o que se pensou inviável por muito tempo. Esse modelo, pautado em um cenário realista, utiliza as interferências inerentes da comunicação satelital para gerar sequências de *bits* aleatórios, as quais, após escolhidas de forma adequada para privilegiar os participantes legítimos da comunicação, são submetidas às fases de correção de erros e de aumento da privacidade. Dessa forma, os participantes legítimos passam a possuir, em comum, uma chave criptográfica altamente segura. Esse elevado nível de segurança pode ser alcançado mesmo na presença de um participante não autorizado que intercepte toda a comunicação, ainda que este possua vantagens tecnológicas sobre os demais, fato que será provado com a utilização de conceitos de Teoria da Informação.

Palavras-chave: Criptografia, segurança incondicional, troca de chaves secretas, One-Time Pad

Abstract

This dissertation revisits the article *Unconditional Security in Cryptography* by Stefan Wolf [31], which addresses a model of key generation and key distribution that allows using a *One-Time Pad* with its characteristic security properties, something considered impossible for a long time. This model, based on a realistic scenario, uses the noise inherent to satellite communication in order to generate correlated random strings which, after having been chosen appropriately to give advantage to the legitimate participants, are submitted to an error correction and privacy amplification phase. In this manner, the legitimate parties obtain a shared, highly secure, cryptographic key. This high level of security can be achieved even in the presence of an eavesdropper whose command of technology is superior compared to that of the legitimate participants, a property that is proven by using concepts from Information Theory.

Keywords: Criptography, unconditional security, secret-key exchange, One-Time Pad.

Lista de Figuras

1.1	Entropia Binária	21
1.2	Entropia e Informação Mútua de duas variáveis aleatórias	27
1.3	Entropia Conjunta de duas variáveis aleatórias	27
1.4	Entropia e Informação Mútua de três variáveis aleatórias	28
1.5	Função Convexa	31
1.6	Divisões da Criptologia	36
1.7	Sistema de Criptografia Simétrica	37
1.8	Segurança incondicional do <i>One-Time Pad</i>	42
1.9	A prova do Teorema de Shannon	44
2.1	Canal Simétrico Binário simples	47
2.2	Modelo de Wyner alterado	48
2.3	Canal Simétrico Binário duplo	49
2.4	Modelo com canal público sem ruído	50
2.5	Modelo Satelital com Canal Autêntico interceptado por Eva	52
2.6	Equivalente Matemático do Modelo Satelital	53
2.7	Repeat-Code Protocol	56

2.8	Cálculo das Probabilidade.	57
2.9	Parity-Check Protocol	63
2.10	Incerteza de Eva após a fase <i>Information Reconciliation</i>	66
2.11	Binary-Search Protocol	67
2.12	Privacy Amplification	74
3.1	Cenário do RFID Tag	81

Lista de Tabelas

1.1	Níveis de Segurança	38
1.2	Tempo Estimado de Segurança	38
2.1	Valores de $n!$ (aproximação de <i>Stirling</i>)	60
2.2	Valores Referência com Variação de N	62
2.3	Informação obtida por Eva de acordo com o parâmetro <i>seg</i>	76

Sumário

Agradecimentos	iii
Resumo	vi
Abstract	vii
Lista de Figuras	ix
Lista de Tabelas	x
Introdução	12
1 Base Teórica	15
1.1 Teoria da Informação	15
1.1.1 Entropia	16
1.1.2 Entropia Conjunta e Entropia Condicional	22
1.1.3 Informação Mútua	25
1.1.4 Representação Gráfica	26
1.1.5 Propriedades da Entropia	30

1.2	Criptografia	33
2	O Modelo Satelital	45
2.1	Modelos Iniciais	45
2.2	Modelando a Comunicação Satelital	51
2.2.1	Cenário	51
2.2.2	Modelagem Matemática	52
2.3	O Protocolo e as Fases de Estabelecimento da Chave Secreta	53
2.3.1	Advantage Distillation	54
2.3.2	Information Reconciliation	65
2.3.3	Privacy Amplification	68
2.4	Estabelecendo a Comunicação	76
3	Outros Modelos	79
3.1	RFID Tags	79
3.1.1	Funcionamento do Protocolo	80
3.2	Áudio Ambiente	82
3.2.1	Funcionamento do Protocolo	83
3.3	Sistematização dos Modelos	84
	Conclusão	85
	Referências Bibliográficas	88

Introdução

A utilização de sistemas de criptografia tem se tornado cada vez mais comum na sociedade contemporânea, situação motivada, entre outras, pela difusão da *internet*. Entretanto, a maior parte dos sistemas criptográficos têm sua segurança baseada na capacidade computacional atual, ou seja, é pautada pela dificuldade dos protocolos e *hardwares* em realizar determinadas operações matemáticas, principalmente envolvendo curvas elípticas, logaritmo discreto ou fatoração de números muito grandes.

Esse tipo de segurança não nos permite criptografar informações altamente sensíveis, como as que necessitam de sigilo superior a 50 anos, uma vez que o desenvolvimento de um protocolo eficiente para resolução das operações matemáticas acima citadas, ou mesmo a criação de um novo modelo de *hardware*, a exemplo do computador quântico, permitiria a quebra da criptografia utilizada em sistemas baseados em segurança computacional.

Dessa forma, dependendo da aplicação do sistema de criptografia, deve-se optar por um que possua segurança incondicional, ou seja, que independa do *hardware* ou protocolo utilizado, o que torna a sua segurança atemporal.

Como exemplo de uma chave criptográfica com segurança incondicional citamos o *One-time Pad*, proposto por Vernam em 1926, no qual uma mensagem $M = [m_1, m_2, \dots, m_N]$ e uma chave $K = [k_1, k_2, \dots, k_N]$, ambas de comprimento N , dão origem a um texto cifrado C , através de uma soma *XOR*, i.e.:

$C = [c_1, c_2, \dots, c_N] = [m_1 \oplus k_1, m_2 \oplus k_2, \dots, m_N \oplus k_N]$, onde \oplus denota a soma *XOR*, ou adição módulo 2.

A segurança incondicional do *One-time Pad* foi provada apenas em 1946, por Claude Shannon, no artigo “*A Mathematical Theory of Cryptography*”. Esse artigo foi classificado pelo governo estadunidense, tamanha a importância de seu conteúdo. Dois anos mais tarde, Shannon publicou o artigo [26], no qual desenvolveu uma estrutura matemática para o estudo da segurança de sistemas criptográficos, definindo o que vem a ser segurança perfeita.

Definição ([25] e [31]). Uma cifra é chamada de perfeitamente segura se o texto cifrado não revela informação sobre a mensagem, i.e., se

$$I(M; C) = 0,$$

onde M é a mensagem, C é o texto cifrado e I a informação mútua.

A partir dessa definição, Shannon provou a segurança do *One-time Pad (OTP)* e estabeleceu o seguinte teorema:

Teorema ([25] e [31]). *Para todo sistema criptográfico perfeitamente seguro, e unicamente decodificado (ou seja, $H(M|CK) = 0$), tem-se:*

$$H(K) \geq H(M),$$

onde M é a mensagem, K a chave criptográfica, C o texto cifrado e H denota a entropia de Shannon.

A partir deste teorema, observa-se que para estabelecermos um sistema criptográfico perfeitamente seguro temos que pagar um preço alto, pois a chave a ser utilizada não pode ser menor que a mensagem que se quer criptografar. Esse fato, em geral, acaba inviabilizando a utilização de cifras perfeitamente seguras.

Entretanto, devemos observar que uma das hipóteses do teorema de Shannon é que o adversário tenha um acesso perfeito ao canal utilizado para a comunicação cifrada, o que na prática nem sempre ocorre, pois toda comunicação está sujeita a ruídos.

O fato de limitarmos a informação a que o adversário tem acesso permite a criação de modelos para troca de chaves OTP e a consequente realização de comunicação criptografada com segurança incondicional.

Um desses modelos é o modelo satelital, no qual utiliza-se o ruído natural da comunicação satelital, conforme os artigos [16] e [31], para geração de uma chave OTP entre dois personagens, que conseguem estabelecer um canal de comunicação seguro, mesmo que o canal seja interceptado por um terceiro que possua vantagens técnicas em relação a esses.

Durante o texto estudaremos este modelo, tomando como referência principal o artigo de Stefan Wolf [31], no qual o autor trabalha com o modelo satelital para geração de chaves seguras, e aplica uma determinada sequência de *bits* às fases *Advantage Distillation*, *Information Reconciliation* e *Privacy Amplification*, obtendo, ao fim, uma chave criptográfica que possibilitará a comunicação com segurança incondicional. Para atingirmos este objetivo serão utilizados conceitos de Teoria da Informação e de Criptografia, os quais serão abordados antes da explicação do modelo em si.

Ao terminarmos de analisar o modelo principal deste trabalho serão apresentados outros modelos para geração e troca de chaves criptográficas, como pode ser encontrado no artigo [6], no qual os autores trabalham com um modelo de RFID e no artigo [21], no qual os autores trabalham com o aumento de segurança da comunicação celular com a utilização de som ambiente.

Capítulo 1

Base Teórica

Neste capítulo serão abordados alguns conceitos e resultados essenciais para o perfeito entendimento do texto. Inicialmente faremos uma introdução à Teoria da Informação, com base nas referências [8], [23] e [25]. Para tanto utilizaremos alguns resultados básicos de probabilidade, encontrados na referência [9]. Em seguida passaremos a uma breve exposição de conceitos relacionados à Criptografia e, para tanto, nos utilizaremos das referências [5], [22], [25] e [31].

1.1 Teoria da Informação

O estudo da Teoria da Informação visa responder prioritariamente a questões relacionadas à compressão e transmissão de dados, problemas enfrentados por Claude Shannon em meados do século XX e formalizados em um famoso artigo intitulado “*A Mathematical Theory of Communication*”, [25].

Em seu artigo, Shannon apresenta os conceitos de entropia (em analogia ao termo utilizado na termodinâmica) e de capacidade de canal, o primeiro como resposta aos

problemas de compressão e o segundo aos problemas relacionados à transmissão de dados.

Nesta introdução ao estudo da Teoria da Informação nos focaremos nos conceitos de entropia e de informação mútua, necessários para mensurarmos a quantidade de informação trafegada por um canal de comunicação.

1.1.1 Entropia

A entropia de uma variável aleatória corresponde ao seu grau de incerteza. Para facilitar o entendimento deste termo iniciaremos com alguns exemplos de sua utilização.

Exemplo 1.1. Suponha que existam 16 objetos distintos e queremos saber a melhor forma de identificá-los, a fim de facilitar a recuperação de um deles mediante a menor quantidade possível de perguntas, cujas respostas podem ser apenas “*sim*” ou “*não*”.

Uma das possíveis soluções para esse desafio seria numerar todos os objetos de 01 a 16 e, para recuperar um específico, bastaria perguntar: “É o objeto XX?”, onde XX representaria um número de 01 a 16.

Obviamente esta não seria uma maneira eficiente de realizarmos esta tarefa, pois necessitaríamos de até 16 perguntas para recuperarmos um dos objetos.

Uma outra possibilidade seria tomarmos uma palavra de 4 *bits* para cada um dos objetos, em que o primeiro seria o 0001, o segundo seria 0010 e assim sucessivamente, até o último que seria 0000. Dessa forma, para recuperar um objeto específico, bastariam quatro perguntas do tipo: “O bit X é 1?”, em que X tomaria valores de 1 a 4.

No exemplo acima, os 16 objetos ocorriam de forma equiprovável, o que facilita

sobremaneira o entendimento de como chegamos a quantidade mínima de perguntas a realizar. Vamos analisar agora um exemplo parecido, entretanto com ocorrência não equiprovável dos objetos.

Exemplo 1.2. Suponha, conforme o Exemplo 1.1, que existam 16 objetos (nem todos distintos), e queremos saber a melhor forma de identificá-los, a fim de facilitar a recuperação de um deles mediante a menor quantidade possível de perguntas, cujas respostas podem ser apenas “*sim*” ou “*não*”.

Esses objetos seguem a seguinte distribuição:

- Existem 8 objetos do tipo **A**.
- Existem 4 objetos do tipo **B**.
- Existem 2 objetos do tipo **C**.
- Existe 1 objeto do tipo **D**.
- Existe 1 objeto do tipo **E**.

Com isso temos as seguintes probabilidades para cada um dos objetos:

- **A**: $p(\mathbf{A}) = \frac{1}{2}$
- **B**: $p(\mathbf{B}) = \frac{1}{4}$
- **C**: $p(\mathbf{C}) = \frac{1}{8}$
- **D** e **E**: $p(\mathbf{D}) = p(\mathbf{E}) = \frac{1}{16}$

Se fossemos utilizar a mesma solução adotada no Exemplo 1.1, necessitaríamos de 4 perguntas binárias (respostas “*sim*” ou “*não*”). Entretanto, podemos otimizar

a solução utilizando uma quantidade diferente de *bits* para identificar cada um dos objetos. Assim, teríamos as seguintes representações:

$$\mathbf{A} \rightarrow 0, \mathbf{B} \rightarrow 10, \mathbf{C} \rightarrow 110, \mathbf{D} \rightarrow 1110, \mathbf{E} \rightarrow 1111$$

O detalhamento de como se obtém essa representação pode ser obtido na referência [8], Huffman Codes.

Com essa alteração, o número médio de perguntas feitas será:

$$\frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \frac{1}{8} \times 3 + \frac{1}{16} \times 4 + \frac{1}{16} \times 4 = \frac{15}{8} = 1,875$$

uma redução superior a 50% da quantidade média de perguntas.

Observe a semelhança entre este problema e a necessidade enfrentada por Shannon em achar uma forma otimizada de comprimir dados para posteriormente enviá-los por um canal de comunicação limitado, i.e., que não suporta uma quantidade infinita de dados em um determinado espaço de tempo.

Para calcularmos a quantidade média de perguntas necessárias nos Exemplos 1.1 e 1.2, nos utilizamos de forma natural de uma média ponderada. Lembremos agora do conceito de esperança de uma variável aleatória discreta.

Definição 1.1 (Esperança, [9]). Seja X uma variável aleatória discreta em um conjunto \mathcal{X} , com distribuição de probabilidade $p(x) = P[X = x]; x \in \mathcal{X}$.

$$EX = \sum_i x_i p(x_i) = \sum_i x_i P[X = x_i]$$

Este valor está bem definido quando a série converge absolutamente, ou seja, $\sum_i |x_i| p(x_i) < \infty$, e representa uma média ponderada, no qual os pesos são as probabilidades.

A partir da definição de esperança, vemos que o cálculo da quantidade média de perguntas necessárias nada mais é que o cálculo da esperança da variável aleatória

discreta que representa a quantidade de perguntas necessárias, ou melhor, da quantidade de *bits* necessários para identificarmos de forma única os objetos dos Exemplos 1.1 e 1.2.

Para expressarmos essa variável aleatória discreta de uma forma mais técnica, observemos que no Exemplo 1.1 são necessários 4 *bits* para representarmos os 16 objetos, independente de qual seja, pois os mesmos são equiprováveis. Já no Exemplo 1.2 necessitamos de 1 *bit* para o objeto **A**, o qual ocorre com probabilidade $\frac{1}{2}$; de 2 *bits* para o objeto **B**, que ocorre com probabilidade $\frac{1}{4}$; de 3 *bits* para o objeto **C**, que ocorre com probabilidade $\frac{1}{8}$; e de 4 *bits* para representar os objetos **D** e **E**, que ocorrem com probabilidade $\frac{1}{16}$ cada.

Partindo disso, percebemos que no Exemplo 1.1 necessitamos de $\log_2 16 = 4$ *bits* e no Exemplo 1.2 de $\log_2 \left(\frac{1}{p(x)} \right)$ *bits* na representação de cada objeto, no qual x é um dos objetos de **A** a **E**.

Dessa forma, ao considerarmos a variável aleatória $W(x_i) = \log_2 \left(\frac{1}{P[X = x_i]} \right)$, podemos obter a quantidade de *bits* necessários para identificarmos de forma única os objetos dos Exemplos 1.1 e 1.2 a partir do cálculo da esperança $EW(X)$. Vamos relembrar como calcular a esperança de uma função real.

Definição 1.2 (Esperança de uma função, [9]). Seja X uma variável aleatória, com função de probabilidade $p(x_i)$ e $\varphi(x)$ uma função real, então:

$$E\varphi(X) = \sum_i \varphi(x_i)p(x_i)$$

Utilizando a Definição 1.2 podemos calcular $EW(X)$.

$$EW(X) = \sum_{x \in \mathcal{X}} p(x) \log_2 \left(\frac{1}{p(x)} \right)$$

Vejamos agora a definição de entropia de uma variável aleatória discreta.

Definição 1.3 (Entropia, [8]). A entropia $H(X)$ de uma variável aleatória discreta X em um conjunto \mathcal{X} , com probabilidade $p(x) = P[X = x]; x \in \mathcal{X}$ é:

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log p(x)$$

Por trabalharmos com o sistema binário na transmissão de informações, utilizaremos variáveis aleatórias binárias e, com isso, o logaritmo na base 2. Assim a entropia será expressa em *bits*.

A partir da definição de entropia podemos concluir que, para uma variável aleatória binária nas condições da Definição 1.3, temos:

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log_2 p(x) = E[-\log_2 p(X)] = E \left[\log_2 \left(\frac{1}{p(X)} \right) \right]$$

Finalizando os Exemplos 1.1 e 1.2, a quantidade média “ótima” de perguntas necessárias está limitada pela entropia da variável aleatória correspondente aos objetos e esta entropia mais um *bit*, podendo ser igual à entropia caso as probabilidades sejam potências de 2, ou seja:

$$H(X) \leq L(X) < H(X) + 1$$

onde $H(X)$ é a entropia da variável aleatória X e $L(X)$ a quantidade média de perguntas. Uma prova desta afirmação pode ser encontrada em [8].

Vejamos mais um exemplo da utilização do conceito de entropia.

Exemplo 1.3. Considere que uma variável aleatória Y assume os seguintes valores:

$$Y = \begin{cases} 1, & \text{com probabilidade } p, \\ 0, & \text{com probabilidade } 1-p. \end{cases}$$

A incerteza ou entropia dessa variável aleatória será:

$$H(Y) = -p \cdot \log_2 p - (1 - p) \cdot \log_2(1 - p).$$

Considere agora $p = \frac{1}{4}$, teremos:

$$\begin{aligned} H(Y) &= -\frac{1}{4} \log_2 \frac{1}{4} - \frac{3}{4} \log_2 \frac{3}{4} \\ &\simeq \frac{1}{2} + 0,31 \\ &\simeq 0,81 \text{ bit.} \end{aligned}$$

Repare que para $p = \frac{1}{2}$ temos $H(Y) = 1$, ou seja, a incerteza da variável aleatória é máxima, não se sabe nada a seu respeito. Em contrapartida, para $p = 1$ e $p = 0$ temos $H(Y) = 0$, ou seja, não há incerteza sobre Y .

A Figura 1.1 representa o gráfico da entropia de uma variável binária X em função de sua probabilidade em assumir um dos valores, no caso $X = 1$.

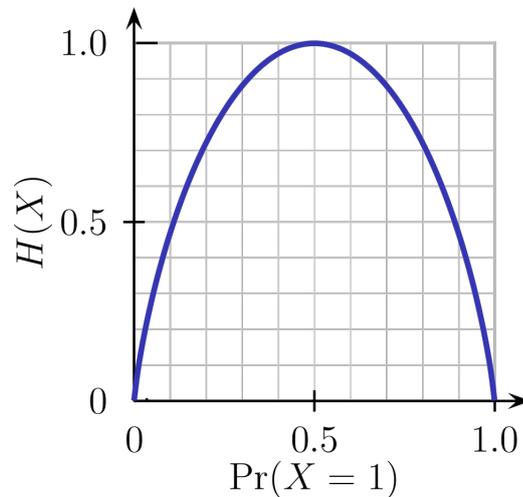


Figura 1.1: Entropia Binária

A partir de agora utilizaremos apenas “log” no lugar de “log₂”. Caso surja a necessidade de utilizarmos outra base logarítmica, esta será informada expressamente.

1.1.2 Entropia Conjunta e Entropia Condicional

Na seção anterior tratamos da entropia de variáveis aleatórias simples. Nesta seção trabalharemos com a entropia de mais de uma variável aleatória, e com o relacionamento entre essas quantidades.

Inicialmente vejamos como calcular a entropia de um par (X, Y) de variáveis aleatórias discretas, as quais podem ser tratadas como um vetor de variáveis aleatórias.

Definição 1.4 (Entropia Conjunta, [8]). Sejam X e Y variáveis aleatórias discretas que possuem distribuição conjunta igual a $p(x, y)$. A entropia conjunta, $H(X, Y)$ é dada por:

$$H(X, Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y) = -E[\log p(X, Y)].$$

Também podemos definir a entropia de uma variável aleatória discreta condicionada ao conhecimento do valor de outra. Neste caso, estamos medindo o grau de incerteza de uma variável aleatória, dado que conhecemos uma outra variável aleatória.

Definição 1.5 (Entropia Condicional, [8]). Sejam X e Y variáveis aleatórias discretas que possuem distribuição conjunta igual a $p(x, y)$. Definimos entropia condicional de Y em relação a X , $H(Y|X)$ como:

$$\begin{aligned} H(Y|X) &= \sum_{x \in \mathcal{X}} p(x) H(Y|X = x) \\ &= - \sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x) \\ &= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y|x) \\ &= -E[\log p(Y|X)]. \end{aligned}$$

A partir dos conceitos de Entropia Conjunta e Entropia Condicional podemos nos indagar sobre a relação entre estes conceitos, pois, intuitivamente, percebemos que a entropia de uma variável aleatória somada a entropia de uma variável aleatória condicionada a ocorrência da primeira, deve resultar na entropia conjunta dessas variáveis. Este fato está expresso abaixo sob a forma de um teorema.

Teorema 1.1 (Regra da Cadeia, [8]).

$$H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$$

Demonstração. Basta lembrarmos que pela definição de probabilidade condicional (ver [9]) $p(x, y) = p(y)p(x|y)$, e que ao somarmos as probabilidades conjuntas em relação a uma variável, obtemos a probabilidade da outra (ver [23] e [25]), ou seja:

$$\sum_j p(x_i, y_j) = p(x_i) \quad e \quad \sum_i p(x_i, y_j) = p(y_j)$$

Assim obtemos:

$$\begin{aligned} H(X, Y) &= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y) \\ &= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log [p(x)p(y|x)] \\ &= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x) - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y|x) \\ &= - \sum_{x \in \mathcal{X}} p(x) \log p(x) - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y|x) \\ &= H(X) + H(Y|X) \end{aligned}$$

□

Vejam agora a generalização da Regra da Cadeia para um número maior de variáveis aleatórias.

Corolário 1.1. *Sejam X_1, X_2, \dots, X_n variáveis aleatórias. Podemos definir:*

$$H(X_1, X_2, \dots, X_n) = H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) + \dots + H(X_n|X_1, X_2, \dots, X_{n-1})$$

Demonstração. Basta lembrarmos do Teorema da Probabilidade Composta (ver [9]), aplicarmos a função logarítmica (multiplicada por -1) e em seguida calcularmos a sua esperança.

$$\begin{aligned} P(X_1, X_2, \dots, X_n) &= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\dots P(X_n|X_1, X_2, \dots, X_{n-1}) \\ -\log P(X_1, X_2, \dots, X_n) &= \\ &= -[\log P(X_1) + \log P(X_2|X_1) + \log P(X_3|X_1, X_2) + \dots + \log P(X_n|X_1, X_2, \dots, X_{n-1})] \\ -E[\log P(X_1, X_2, \dots, X_n)] &= \\ &= -E[\log P(X_1) + \log P(X_2|X_1) + \log P(X_3|X_1, X_2) + \dots + \log P(X_n|X_1, X_2, \dots, X_{n-1})] \\ H(X_1, X_2, \dots, X_n) &= H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) + \dots + H(X_n|X_1, X_2, \dots, X_{n-1}) \end{aligned}$$

□

Ainda não analisamos como se comportam 3 variáveis aleatórias quando queremos a entropia conjunta de duas delas condicionada a ocorrência da terceira.

Corolário 1.2. *Sejam X, Y e Z variáveis aleatórias, temos:*

$$H(X, Y|Z) = H(X|Z) + H(Y|X, Z)$$

Demonstração.

$$\begin{aligned}
H(X, Y|Z) &= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} p(x, y, z) \log p(x, y|z) \\
&= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} p(x, y, z) \log \left(\frac{p(x, y, z)}{p(z)} \right) \\
&= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} p(x, y, z) \log \left(\frac{p(x, y, z)}{p(z)} \times \frac{p(x, z)}{p(x, z)} \right) \\
&= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} p(x, y, z) \log \left(\frac{p(x, z)}{p(z)} \times \frac{p(x, y, z)}{p(x, z)} \right) \\
&= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} p(x, y, z) \log (p(x|z)p(y|x, z)) \\
&= - \sum_{x \in \mathcal{X}} \sum_{z \in \mathcal{Z}} p(x, y, z) \log p(x|z) - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} p(x, y, z) \log p(y|x, z) \\
&= H(X|Z) + H(Y|X, Z)
\end{aligned}$$

□

1.1.3 Informação Mútua

Nesta seção trabalharemos com a quantidade de informação que uma variável aleatória pode conter a respeito de outra. Esta grandeza é chamada de Informação Mútua.

Definição 1.6 (Informação Mútua, [8]). Sejam X e Y variáveis aleatórias com distribuição conjunta de probabilidade $p(x, y)$ e probabilidades marginais $p(x)$ e $p(y)$, respectivamente. A Informação Mútua $I(X; Y)$ é dada por:

$$I(X; Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} = E_{p(x, y)} \log \frac{p(X, Y)}{p(X)p(Y)}$$

Assim, $I(X; Y)$ expressa a quantidade de informação que X contém de Y e vice-versa.

Agora vamos estabelecer a relação entre Informação Mútua e Entropia.

Teorema 1.2 ([8]). *A Informação Mútua de duas variáveis aleatórias X e Y pode ser obtida a partir da Entropia destas pela seguinte relação:*

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

Demonstração.

$$\begin{aligned} I(X; Y) &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \\ &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log \frac{p(x|y)}{p(x)} \\ &= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x) + \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x|y) \\ &= - \sum_{x \in \mathcal{X}} p(x) \log p(x) - \left(- \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x|y) \right) \\ &= H(X) - H(X|Y) \end{aligned}$$

□

Agora nos resta analisar a informação que uma variável aleatória contém sobre outra condicionada à ocorrência de uma terceira.

Definição 1.7 (Informação Mútua Condicional, [8]). *Sejam X , Y e Z variáveis aleatórias. Definimos a informação mútua condicional de X e Y em relação a Z como:*

$$I(X; Y|Z) = H(X|Z) - H(X|Y, Z) = E_{p(x,y,z)} \log \frac{p(X, Y|Z)}{p(X|Z)p(Y|Z)}$$

1.1.4 Representação Gráfica

Visando facilitar o entendimento e a manipulação dos conceitos de Entropia e Informação Mútua, apresentaremos estas grandezas informacionais em diagramas de

Venn. Além disso apresentaremos uma abordagem diferente para essas grandezas, facilitando suas “visualizações”.

Sejam X , Y e Z , variáveis aleatórias. Podemos representar suas quantidades de informação da seguinte maneira.

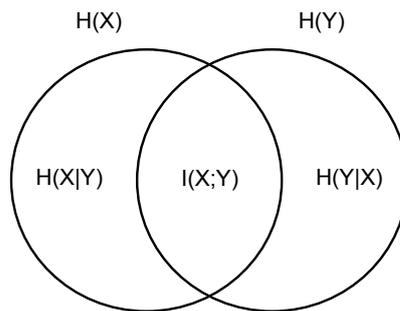


Figura 1.2: Entropia e Informação Mútua de duas variáveis aleatórias. (Fonte: [31])

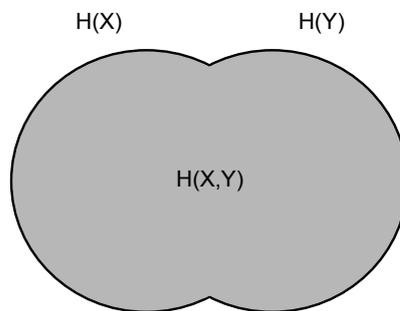


Figura 1.3: Entropia Conjunta de duas variáveis aleatórias. (Fonte: [31])

Repare que ao trabalharmos com três variáveis aleatórias, Figura 1.4, surge a grandeza $R(X;Y;Z)$, a qual também poderia ser representada por $I(X;Y;Z)$, mas não o foi por um motivo em especial, está grandeza assume valores negativos. Como veremos a seguir $I(X;Y) \geq 0$ e, por isso, optou-se por representar a referida grandeza por $R(X;Y;Z)$, evitando possíveis confusões. No momento oportuno será explicado o porquê dessa grandeza assumir valores negativos.

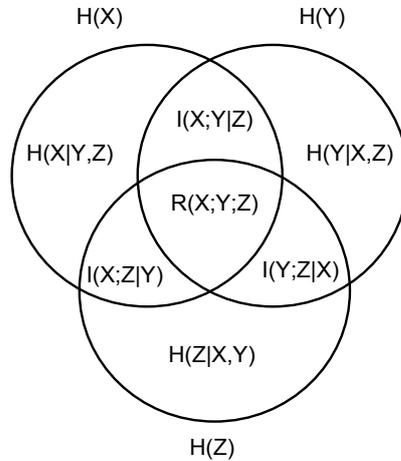


Figura 1.4: Entropia e Informação Mútua de três variáveis aleatórias. (Fonte: [31])

Esta representação gráfica nos leva a pensar uma forma de calcular a relação entre as grandezas informacionais por meio de operações de conjuntos. De fato é isso que será feito, mas antes vamos apresentar uma correlação entre a simbologia de Shannon e a de conjunto, devida a Raymond W. Yeung, e introduzida no artigo [33]. No artigo, Yeung constrói um espaço de medida (*I-medida*) e formaliza as substituições apresentadas.

Sejam X e Y variáveis aleatórias e considere as seguintes substituições:

$$\begin{aligned}
 H/I &\longrightarrow \mu \\
 , &\longrightarrow \cup \\
 ; &\longrightarrow \cap \\
 | &\longrightarrow -
 \end{aligned}$$

Adicionando a hipótese de $X \cup Y = \Omega$ (o espaço amostral tradicional) passaremos

a ter:

$$H(X) = \mu(X)$$

$$H(Y) = \mu(Y)$$

$$H(X, Y) = \mu(X \cup Y)$$

$$H(X|Y) = \mu(X - Y) \quad ; (X - Y) = (X \cap Y^C)$$

$$H(Y|X) = \mu(Y - X)$$

$$I(X; Y) = \mu(X \cap Y)$$

$$H(X|Y) + H(Y|X) = \mu((X \cap Y)^C)$$

e

$$\mu(\emptyset) = 0$$

Repare que não há diferença entre informação mútua e entropia, tanto que a entropia é referenciada na literatura como informação mútua própria.

Como exemplo da utilização desta notação e sua consequente vantagem iremos apresentar outra demonstração para o Teorema 1.2.

Exemplo 1.4. Sejam X e Y variáveis aleatórias, teremos:

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

Demonstração.

$$\begin{aligned} I(X; Y) &= \mu(X \cap Y) = \mu(X) + \mu(Y) - \mu(X \cup Y) \\ &= \mu(X) - \mu(X - Y) = H(X) - H(X|Y) \\ &= \mu(Y) - \mu(Y - X) = H(Y) - H(Y|X). \end{aligned}$$

□

1.1.5 Propriedades da Entropia

Salientaremos algumas propriedades da entropia, as quais facilitarão demonstrações futuras e servirão como exercício para a consolidação dos conceitos.

Começaremos com a abordagem de algumas propriedades da esperança matemática, incluindo a Desigualdade de Jensen, propriedade de funções convexas muito utilizada no cálculo de esperança e nas demonstrações de Teoria da Informação.

Propriedades da Esperança

Teorema 1.3 ([9]). *Sejam X e Y variáveis aleatórias definidas em Ω .*

E.1 *Se $X = c$ (i.e., $X(x) = c, \forall x \in \Omega$), então $EX = c$ ($c = cte$).*

E.2 *Se $X \leq Y$ então $EX \leq EY$, se as esperanças estão bem definidas. (Basta uma das esperanças ser finita, ou $EY = -\infty$, ou $EX = +\infty$).*

E.3 *Linearidade*

(i) *Se EX está bem definida, então $E(aX + b) = aEX + b$ para todo $a, b \in \mathbb{R}$.*

(ii) *$E(aX + bY) = aEX + bEY$, quando o termo à direita da igualdade tem sentido.*

Demonstração. As demonstrações destas três propriedades serão omitida, pois envolvem conceitos que não são foco deste trabalho. Caso haja interesse, podem ser encontradas em [9]. □

Antes de apresentarmos a próxima propriedade da esperança (Desigualdade de Jensen), iremos relembrar a definição de função convexa, fundamental para o entendimento da propriedade.

Definição 1.8 (Função Convexa, [8]). Uma função $f(x)$ é chamada de convexa num intervalo (a, b) se para cada $x_1, x_2 \in (a, b)$ e $0 \leq \lambda \leq 1$,

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$$

A função é chamada de estritamente convexa se a igualdade ocorre apenas em $\lambda = 0$ ou $\lambda = 1$.

Observação 1.1. Uma função f será côncava se $-f$ for convexa.

Observação 1.2. Como exemplo de função convexa temos $x^2, e^x, |x|$, e de função côncava temos $\log x$.

Teorema 1.4 (Desigualdade de Jensen, [8]). *Seja X uma variável aleatória e f uma função convexa. Então:*

$$Ef(X) \geq f(EX)$$

Demonstração. Dado x_0 e o ponto $f(x_0)$ do gráfico de f , pela convexidade existe uma reta L que passa por $f(x_0)$ e deixa a curva f acima dela (ver Figura 1.5):

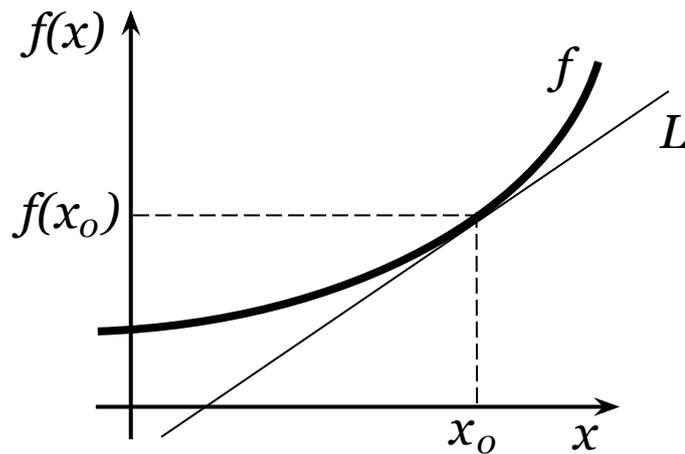


Figura 1.5: Função Convexa

Seja para algum $\lambda \in \mathbb{R}$, $y - f(x_0) = \lambda(x - x_0)$ a equação desta reta L . Então

$$f(x) \geq L(x) = f(x_0) + \lambda(x - x_0), \forall x.$$

Portanto

$$Ef(X) \geq EL(X) = f(x_0) + \lambda(EX - x_0).$$

Tomando-se $x_0 = EX$, vem $Ef(X) \geq f(EX)$. □

Vejam agora o que ocorre no caso de a função ser côncava.

Teorema 1.5 (Desigualdade de Jensen para funções côncavas). *Seja X uma variável aleatória e f uma função côncava,*

$$Ef(X) \leq f(EX)$$

Demonstração. A prova é análoga a do Teorema 1.4. □

De posse dessas propriedades, passaremos a expor as propriedades da entropia úteis ao entendimento do presente texto.

Propriedades da Entropia

Nas propriedades apresentadas a seguir considere X e Y variáveis aleatórias com valores em um conjunto \mathcal{X} e \mathcal{Y} respectivamente.

P.1 (Ver referência [8]) $H(X) \geq 0$.

Demonstração. $0 \leq p(x) \leq 1$, logo $\log \frac{1}{p(x)} \geq 0$. □

P.2 (Ver referência [31]) $H(X) \leq \log |\mathcal{X}|$, onde $|\mathcal{X}|$ representa a cardinalidade do conjunto \mathcal{X} .

Demonstração. Utilizando a Desigualdade de Jensen para funções côncavas vem:

$$H(X) = E \left[\log \left(\frac{1}{P(X)} \right) \right] \leq \log \left(E \left[\frac{1}{P(X)} \right] \right) = \log |\mathcal{X}|.$$

□

P.3 (Ver referência [31]) $I(X; Y) \geq 0$, com igualdade se, e somente se, X e Y são independentes.

Demonstração. Utilizando a substituição apresentada por Yeung em [33], obtemos:

$$I(X; Y) = \mu(X \cap Y)$$

Como $H(X) = \mu(X) \geq 0$ e $H(Y) = \mu(Y) \geq 0$, temos o resultado. □

P.4 (Ver referência [23]) $H(X|Y) \leq H(X)$, com igualdade se, e somente se, X e Y são independentes.

Demonstração. Basta utilizarmos a Propriedade **P.3** e o Teorema 1.2, assim,

$$I(X; Y) = H(X) - H(X|Y) \geq 0$$

e temos o resultado. □

Observação 1.3. Na Propriedade **P.2** a igualdade ocorre quando temos uma distribuição uniforme em \mathcal{X} , ou seja, $p(x) = \frac{1}{|\mathcal{X}|} \forall x \in \mathcal{X}$.

1.2 Criptografia

Nesta seção serão abordados alguns conceitos de criptografia necessários para o entendimento do texto. Para um estudo mais profundo do tema sugerimos as referências [11] e [27], nas quais é apresentada a história da evolução da criptografia, a

referência [22] que possui uma linguagem simples para quem não possui familiaridade com o tema e as referências [20] e [24] que apresentam conceitos mais técnicos sobre a criptografia, sendo que a referência [24] aborda, inclusive, protocolos criptográficos.

A criptografia surge como uma arte no Egito antigo, cerca de 2000 anos antes de Cristo, dependendo profundamente da evolução da matemática para se tornar uma ciência, o que ocorre com Shannon em 1948.

Apesar de utilizarmos comumente o termo criptografia, o mais correto seria usarmos criptologia, a ciência relacionada ao desenvolvimento de métodos, esquemas e modelos de comunicação secreta e segura. A criptologia, termo mais genérico, pode ser dividida em criptografia e criptoanálise. A primeira trata do estudo das escritas secretas, com foco em esconder o significado de uma mensagem. A segunda é a ciência, e às vezes arte, de decifrar mensagens criptografadas, sem o conhecimento de todos os parâmetros conhecidos pelos comunicantes.

Nesta exposição nos restringiremos à criptografia, a qual, segundo as referências [20] e [22], pode ser dividida em:

- **Simétrica** - inclui todo o tipo de criptografia anterior a 1976 (criação do conceito de chave-pública), sendo a criptografia tradicional, e consiste basicamente de métodos de encriptar e decriptar mensagens com a utilização de uma chave secreta trocada entre as partes que realizarão a comunicação.

Definição 1.9 (Sistema de Chave Simétrica, [20]). Considere um sistema criptográfico com as seguintes transformações de encriptação e decriptação: $\{E_e : e \in \mathcal{K}\}$ e $\{D_d : d \in \mathcal{K}\}$, respectivamente, onde \mathcal{K} é o espaço das chaves criptográficas. A criptografia será simétrica se para cada par de chaves (e, d) for computacionalmente simples determinar d sabendo apenas e , e determinar e a partir de d . Em geral, $e = d$.

- **Assimétrica** - em 1976 Whitfield Diffie, Martin Hellman e Ralph Merkle desenvolveram um sistema de troca de chave no qual uma das partes possui uma chave pública (todos a conhecem) e uma chave privada (só conhecida pelo proprietário). Esse novo sistema não possui o inconveniente de troca antecipada de chaves, apesar de não atingir, em geral, os níveis de segurança atingidos na criptografia simétrica. A criptografia assimétrica permitiu o desenvolvimento de novas aplicações para a criptografia, como é o caso da assinatura digital.

Definição 1.10 (Sistema de Chave Pública, [20]). Seja $\{E_e : e \in \mathcal{K}\}$ um conjunto de transformações de encriptação e seja $\{D_d : d \in \mathcal{K}\}$ o conjunto de transformações de decriptação correspondente, onde \mathcal{K} é o espaço de chaves criptográficas. Considere qualquer par de transformações de encriptação e decriptação correspondentes (E_e, D_d) e suponha que cada par tem a propriedade de ao se conhecer E_e ser computacionalmente improvável, dado um texto cifrado $c \in \mathcal{C}$, descobrir-se a mensagem $m \in \mathcal{M}$, tal que $E_e(m) = c$. Essa propriedade implica que dado e é improvável se determinar a correspondente chave de decriptação d .

Entretanto, segundo Singh, em [27], o responsável pelo desenvolvimento da criptografia assimétrica teria sido James Ellis, criptógrafo e militar inglês, em 1969. Seus estudos teriam sido classificados pelo governo inglês e, dessa forma, esse acontecimento só foi publicizado anos mais tarde.

Ainda podemos dividir a criptografia simétrica em criptografia por blocos e por “*stream*” de dados. Na primeira são formados blocos de dados para em seguida serem criptografados e enviados, enquanto na segunda os dados são criptografados individualmente, conforme são gerados. Antes de apresentarmos mais algumas definições, apresentaremos o modelo que será utilizado ao longo do texto para explicarmos o funcionamento dos protocolos criptográficos.

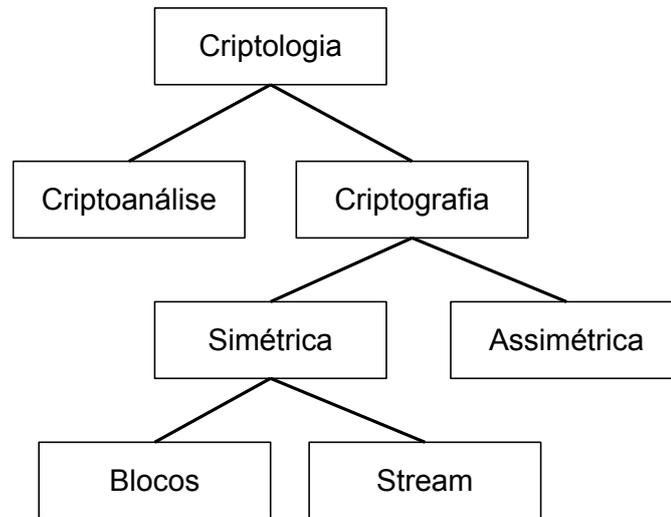


Figura 1.6: Divisões da Criptologia. (Fonte: [22])

Basicamente teremos o estabelecimento de comunicação entre duas pessoas, Alice (**A**) e Bob (**B**), com a possível interceptação dessa comunicação por uma terceira pessoa, Eva (**E**). Para estabelecerem essa comunicação, Alice e Bob trocarão uma chave criptográfica K , que será usada para criptografar uma mensagem (também chamada de mensagem em claro, ou texto em claro) M . A mensagem criptografada (ou texto criptografado) C poderá, eventualmente, ser conhecida por Eva.

Vejamos na Figura 1.7 um sistema de criptografia simétrica utilizando este modelo.

Repare que a mensagem é transmitida por um canal inseguro, como o telefone ou a *internet*, enquanto a chave criptográfica é transmitida por um canal seguro.

Mas o que poderia ser considerado um canal seguro? Um canal seguro é aquele que permite o estabelecimento de comunicação de maneira que todos que possuam acesso aos dados trafegados por ele sejam conhecidos pelos demais participantes da comunicação, além de não permitir a adulteração dos dados ali trafegados. Ou seja, obter um canal de comunicação seguro não é uma tarefa muito simples.

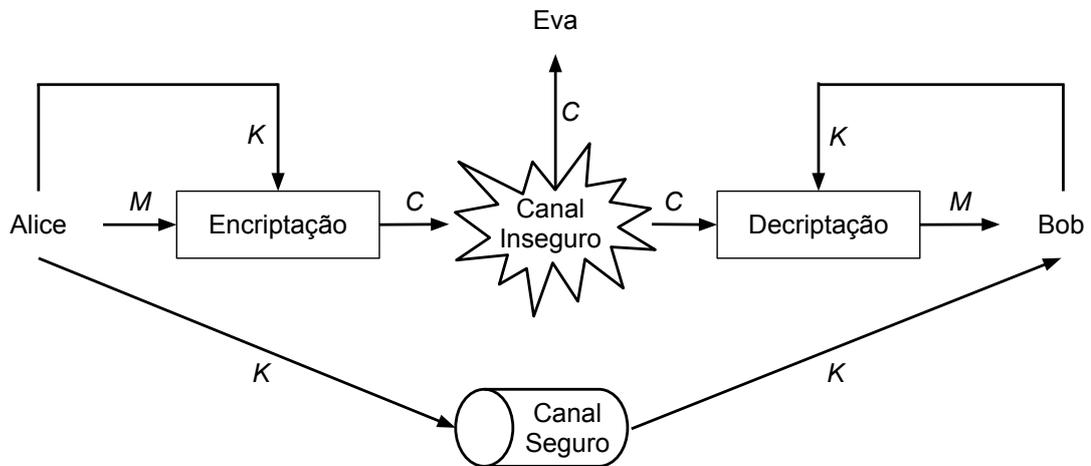


Figura 1.7: Sistema de Criptografia Simétrica

Como exemplo de um canal seguro podemos citar um encontro pessoal em um ambiente controlado, ambiente do qual se tem total controle, tanto de entrada e saída de pessoal, quanto de equipamentos e informações.

Dessa forma, uma solução para a distribuição de chaves criptográficas, por um canal seguro, é a utilização de livros com chaves a serem utilizadas em datas específicas pelos participantes, os quais devem ser entregues em um encontro pessoal e em um momento anterior à transmissão da mensagem.

Obviamente esta solução não é prática e acarreta um aumento nos custos de empresas que lidam diariamente com a transmissão de dados sigilosos, como é o caso das instituições financeiras.

Assim, uma solução eficiente para esta necessidade de troca de chaves é a criptografia assimétrica, que passou a ser utilizada para a troca de chaves da criptografia simétrica. O único inconveniente é que na criptografia assimétrica temos que trabalhar com chaves muito maiores para atingir a mesma segurança da criptografia simétrica. Vejamos uma tabela, adaptada da referência [22], de comparação dos níveis de segurança atingidos por estes tipos de criptografia.

Tipo de Criptografia	Tipo de Algoritmo	Sistema de Criptografia	Níveis de Segurança (bit)			
			80	128	192	256
Assimétrica	Fatoração de Inteiro	RSA	1024	3072	7680	15360
	Logaritmo Discreto	DH, DSA, Elgamal	1024	3072	7680	15360
	Curvas Elípticas	ECDH, ECDSA	160	256	384	512
Simétrica	Chave Simétrica	AES, 3DES	80	128	192	256

Tabela 1.1: Níveis de Segurança. (Fonte: [22], adaptada)

Para termos uma noção do tempo, estimado para um ataque por força bruta (ver referência[22]), que levaria para se obter êxito ao atacar um sistema de criptografia simétrica, vejamos a tabela a seguir, retirada da referência [22].

Tamanho da Chave	Segurança Estimada
56–64 bits	Poucas horas ou dias.
112–128 bits	Muitas décadas, considerando a não existência de computadores quânticos.
256 bits	Muitas décadas, mesmo considerando a existência de computadores quânticos, entretanto com os algoritmos conhecidos atualmente.

Tabela 1.2: Tempo Estimado de Segurança. (Fonte: [22])

Pela Tabela 1.1 percebemos o quão custoso se torna realizarmos a troca de mensagens com a utilização de criptografia assimétrica, o que acaba por justificar a utilização deste tipo de criptografia apenas para a troca de chaves criptográficas.

Assim, passamos à exposição de alguns conceitos relacionados à segurança de sistemas de criptografia, iniciando com a definição do Princípio de Kerckhoffs. Este princípio estabelece o que devemos supor ser de conhecimento de um possível atacante.

Definição 1.11 (Princípio de Kerckhoffs, [22]). Um sistema criptográfico deve ser seguro no caso de um atacante (Eva) saber todos os detalhes sobre o sistema, com exceção da chave criptográfica. Em particular, o sistema deve ser seguro quando o atacante conhece os algoritmos de encriptação e decriptação.

O Princípio de Kerckhoffs deixa bem claro que não podemos supor que um sistema é seguro simplesmente por não divulgarmos seus detalhes, o que chamamos de “segurança por obscuridade”. A história já nos mostrou o quão inseguros são os sistemas criptográficos protegidos pela obscuridade.

A segurança criptográfica pode ser dividida em dois tipos:

Definição 1.12 (Segurança Computacional, [20]). Um sistema criptográfico é computacionalmente seguro se o nível computacional necessário para burlá-lo (usando o melhor ataque conhecido) excede, por uma margem considerável, os recursos computacionais de um adversário hipotético. A segurança computacional está condicionada aos *hardwares* e algoritmos existentes em determinado momento, podendo ser superada subitamente.

Definição 1.13 (Segurança Incondicional, [22] e [20]). Um sistema criptográfico possui segurança incondicional se não pode ser quebrado, a não ser por uma quantidade infinita de operações, ou seja, mesmo assumindo que um adversário possui

recursos computacionais ilimitados, o sistema continua seguro, pois a questão é se existe, ou não, informação suficiente e disponível para burlar o sistema. A segurança incondicional é provada matematicamente com o auxílio da Teoria da Informação.

As Definições 1.12 e 1.13 possuem uma diferença fundamental: o tempo de segurança que conferem ao sistema. Um sistema criptográfico que utiliza segurança computacional deve prever em quanto tempo essa segurança será superada, pois, por depender da tecnologia existente em determinado período, esse sistema se tornará inevitavelmente obsoleto. Enquanto isso, ao se projetar um sistema criptográfico com segurança incondicional, a robustez do sistema se torna atemporal, mantendo o sigilo da comunicação eternamente.

Mas trabalhar simplesmente com estes conceitos se torna um tanto quanto abstrato. Por isso vamos recorrer à definição de *sigilo perfeito*, uma definição mais específica, introduzida por Shannon em [26]. Para facilitar o entendimento da definição, utilizaremos a interpretação dada por Stefan Wolf em [31].

Definição 1.14 (Sigilo Perfeito, [31]). Uma cifra é chamada de perfeitamente segura se o texto cifrado não revela informação sobre a mensagem, i.e., se

$$I(M; C) = 0$$

onde M é a mensagem e C é o texto cifrado.

Um exemplo de uma cifra perfeitamente segura, i.e., com segurança incondicional, é o *One-time Pad*, proposto em 1917 por Gilbert Vernam (AT&T) e melhorado em seguida por Joseph Mauborgne (oficial do Exército dos Estados Unidos) (ver [11]).

Definição 1.15 (*One-time Pad - OTP*, [5]). Considere n um inteiro e $\mathcal{M} = \mathcal{C} = \mathcal{K} = \{0, 1\}^n$. Seja $k \in \mathcal{K}$ uma chave escolhida conforme o algoritmo de amostragem definido em \mathcal{K} . Nessas condições temos:

- Ciframento: Calcule $C = e_K(M)$ como o resultado da operação XOR (ou exclusivo *bit-a-bit*) de K e M .
- Deciframento: Calcule $M = d_K(C)$ como o resultado da operação XOR (ou exclusivo *bit-a-bit*) de K e C .

Antes de provarmos que o OTP possui segurança incondicional, vamos observar seu funcionamento.

Considere a mensagem formada pela sequência $M = [m_1, m_2, \dots, m_n]$ de n bits e a chave $K = [k_1, k_2, \dots, k_n]$ também de tamanho n , a qual possui distribuição uniforme no conjunto \mathcal{K} e é independente de M . Realizando a soma XOR, *bit-a-bit*, de M e K , obtemos a mensagem cifrada C .

$$C = [c_1, c_2, \dots, c_n] = [m_1 \oplus k_1, m_2 \oplus k_2, \dots, m_n \oplus k_n] = M \oplus K$$

Perceba que a mensagem cifrada C é obtida de forma única a partir de M e K , ou seja, $H(C|M, K) = 0$, o mesmo ocorrendo para a mensagem M , $H(M|C, K) = 0$ e para a chave K , $H(K|M, C) = 0$. Além disso, a informação mútua entre as variáveis aleatórias C e M , dado que K é conhecido, é igual a n , ou seja, $I(C; M|K) = n$.

Como $H(K) = \log_2 |\mathcal{K}| = \log_2 2^n = n$ (lembre-se que K é variável aleatória com distribuição uniforme) e $I(C; K|M) = I(M; K|C) = n$, obtemos $I(M; K) = I(M; C) = 0$. Vejamos a representação gráfica dessas quantidades na Figura 1.8.

Repare que, conforme informado na seção 1.1.4, a quantidade $R(M; C; K)$ assume valor negativo, o que pode ser facilmente explicado pelo fato de termos a

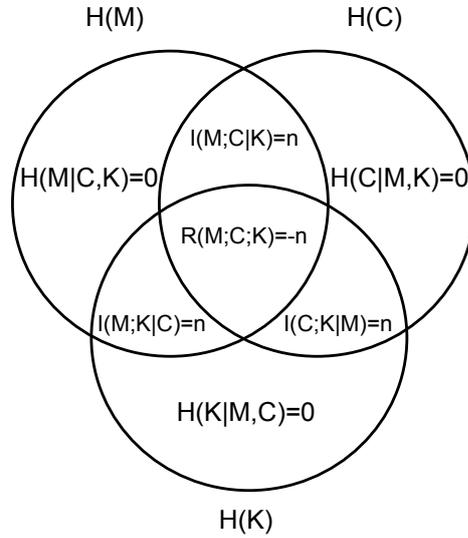


Figura 1.8: Segurança incondicional do *One-Time Pad*. (Fonte: [31])

necessidade de $I(M; K) = I(C; K) = 0$. Assim vem:

$$\begin{aligned}
 I(M; K) &= I(M; K|C) + R(M; C; K) = 0 \\
 I(M; K|C) &= -R(M; C; K) \\
 I(M; K|C) \geq 0 &\Rightarrow R(M; C; K) \leq 0.
 \end{aligned}$$

É importante ressaltar que a segurança do OTP depende da utilização de cada chave K uma única vez, pois caso contrário, basta fazermos uma soma XOR de um resultado conhecido com um que queremos conhecer para que descubramos a chave e a mensagem alvo do ataque. Observe:

Exemplo 1.5. Suponha que queiramos descobrir uma mensagem M_A criptografada com uma chave K utilizando o OTP, ou seja, $M_A \oplus K = C_A$. Sabendo que o sistema utiliza a mesma chave K , basta pegarmos uma outra utilização desse sistema, como $M_B \oplus K = C_B$ e realizarmos uma soma XOR de ambas, obtendo:

$$C_A \oplus C_B = M_A \oplus K \oplus M_B \oplus K = M_A \oplus M_B$$

Assim, conhecendo alguns *bits* de M_B , encontramos os *bits* correlatos de M_A .

Após a apresentação do *One-Time Pad*, podemos enunciar o famoso Teorema de Shannon, o qual estabelece um critério simples para determinarmos se um sistema criptográfico pode ter segurança incondicional. Vejamos:

Teorema 1.6 (Teorema de Shannon, [25] e [31]). *Para todo sistema criptográfico perfeitamente seguro, e unicamente decodificado (ou seja, $H(M|C, K) = 0$), tem-se:*

$$H(K) \geq H(M)$$

onde M é a mensagem, K a chave criptográfica e C o texto cifrado.

Demonstração. Observe a Figura 1.9. Como $b \geq a$, pois, como o sistema criptográfico é perfeito, pela Propriedade P3 da entropia, $I(C; K) \geq 0$ e $I(M; C) = 0$.

Assim, obtemos:

$$H(K) \geq b - a + c \geq a - a + c = H(M).$$

□

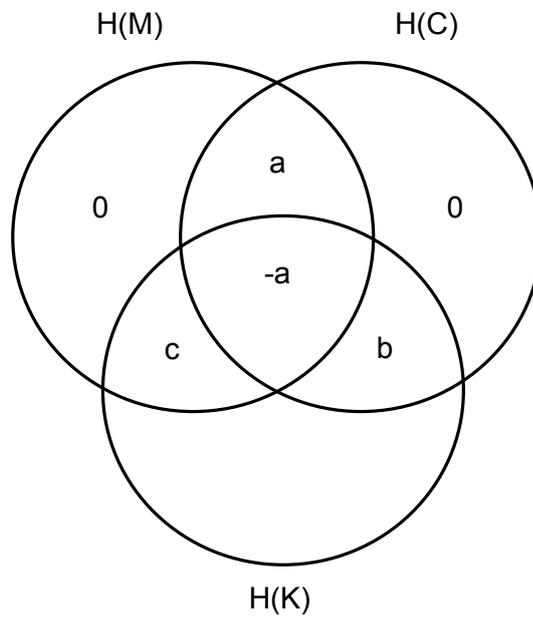


Figura 1.9: A prova do Teorema de Shannon. (Fonte: [31])

Após apresentarmos os conceitos básicos de Teoria da Informação e de Criptografia necessários ao entendimento do modelo de geração de chaves criptográficas, foco deste texto, passaremos, a partir da próxima seção, a estudar o modelo em si.

Capítulo 2

O Modelo Satelital

No capítulo anterior mostramos conceitos básicos de Teoria da Informação e de Criptografia, os quais, a partir de agora, serão utilizados na exposição de alguns modelos de comunicação criptografada, quando serão apresentados problemas e possíveis soluções para, ao final, chegarmos a um modelo de comunicação com segurança criptográfica incondicional; o modelo satelital.

Construiremos o modelo satelital de forma natural, partindo de um modelo simples e ideal, ao qual serão adicionadas, aos poucos, hipóteses reais, tratadas uma a uma.

Todos esses modelos são trabalhados por Stefan Wolf em [31], no qual nos basearemos.

2.1 Modelos Iniciais

Como vimos no capítulo anterior, pelo Teorema de Shannon só é possível atingirmos a segurança incondicional com a utilização de uma chave criptográfica de

tamanho maior ou igual ao tamanho da mensagem a ser criptografada, o que sem dúvida torna inviável a sua utilização de forma eficiente, já que a troca de chaves previamente seria muito custosa.

Entretanto, Shannon supôs que o adversário possui acesso perfeito ao texto cifrado, o que na prática dificilmente ocorre. Dessa forma, podemos trabalhar com modelos em que o acesso do adversário ao texto cifrado seja limitado. Mas como seria esse limite?

O limite pode se dar por diferentes motivos, por exemplo: ruído no canal de comunicação, armazenamento finito de informações por parte do adversário e o princípio da incerteza da mecânica quântica.

Iremos nos ater ao estudo da limitação originada pelo ruído existente no canal de comunicação, o qual em geral deve ser tratado pelo receptor da mensagem, visando possibilitar seu entendimento. Assim o ruído passará de vilão, por dificultar a transmissão de mensagens, a aliado, já que permitirá o estabelecimento de um canal de comunicação com criptografia incondicional.

Vamos iniciar com a exposição de um modelo simples de comunicação estabelecida entre Alice e Bob, os quais estão ligados por um canal binário autêntico e inseguro. O canal autêntico permite que os interlocutores se reconheçam e por ser inseguro, permite a influência de um adversário. Consideraremos adversários passivos, os quais apenas coletam informações, não sendo possível a manipulação ou inserção de mensagens no canal interceptado. Como exemplo de canal autêntico e inseguro citamos a comunicação por rádio frequência.

Será utilizada ao longo do texto a expressão “canal interceptado” como aquele que é escutado ou copiado por alguém sem a anuência dos demais participantes da comunicação desenvolvida com o auxílio desse canal. Tecnicamente, a mensagem é interceptada, e não o canal, entretanto é comum a utilização do termo canal inter-

ceptado para se referir à interceptação de mensagem transportada pelo canal. Salvo menção expressa, nos modelos utilizados, Eva será a responsável por interceptar as mensagens do canal de comunicação.

Suponhamos que não existe ruído entre Alice e Bob, e que Eva intercepta essa comunicação por um Canal Simétrico Binário (BSC), com probabilidade de erro $\varepsilon > 0$ (veja a Figura 2.1).



Figura 2.1: Canal Simétrico Binário simples. (Fonte: [31])

Nesse modelo podemos estabelecer a seguinte regra para a troca de mensagens: Alice manda uma sequência de N bits independentes x_i ($1 \leq i \leq N; i \in \mathbb{N}$), os quais, concatenados, formarão M . Se faz necessário que a distância de Hamming do conjunto ao qual M pertence seja 1, pois caso contrário Eva poderia corrigir alguns erros. No momento adequado levaremos em consideração a possibilidade de correção de erros por parte de Eva.

A probabilidade de erro de M' , concatenação dos bits interceptados por Eva, será:

$$p = 1 - (1 - \varepsilon)^N$$

que converge para 1 em N , tornando a comunicação entre Alice e Bob segura.

No artigo [32], Wyner parte de um modelo mais realista em que Alice e Bob estão conectados por um canal com capacidade limitada de memória e Eva consegue se

conectar a Bob pelo mesmo tipo de canal. Com isso, Eva obtém um sinal atenuado tanto no trecho entre Alice e Bob quanto entre Bob e Eva. Como a limitação da capacidade de memória não é objeto deste texto, substituiremos essa limitação por um Canal Simétrico Binário, sem prejuízo do entendimento do modelo (ver Figura 2.2).

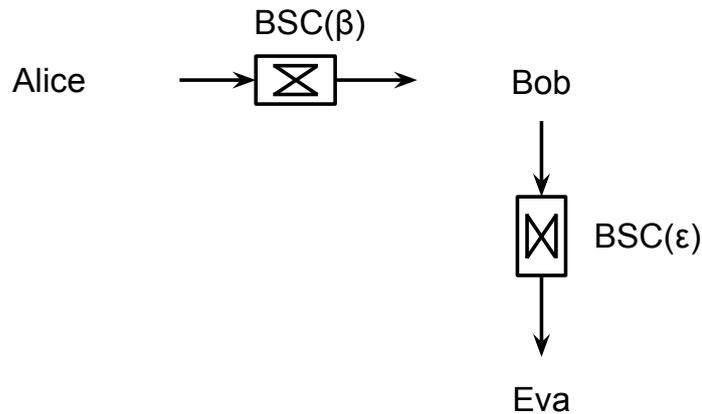


Figura 2.2: Modelo de Wyner alterado. (Fonte: [31])

O modelo de Wyner foi generalizado por Csiszár e Körner no artigo [7], chegando-se a um modelo mais real, em que tanto Bob quanto Eva estão conectados a Alice por meio de um Canal Simétrico Binário, com probabilidade de erro β e ε respectivamente (ver Figura 2.3).

Neste modelo, caso Bob possua vantagem sobre Eva ($\beta < \varepsilon$), será possível estabelecer uma comunicação segura entre Alice e Bob, mas em caso contrário ($\beta > \varepsilon$), a comunicação segura se torna impossível.

Considere que Alice envia a sequência M , de N bits, recebida por Bob como M_B e por Eva como M_E . Considere ainda que $0 < \beta, \varepsilon \leq \frac{1}{2}$, pois caso tenhamos uma taxa de erro maior que $1/2$ bastaria invertermos os bits recebidos, ficando com taxas

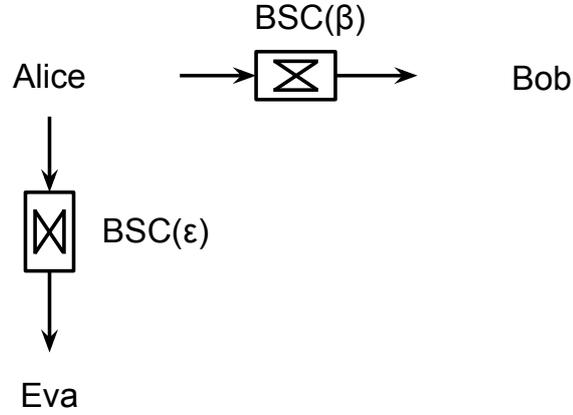


Figura 2.3: Canal Simétrico Binário duplo. (Fonte: [31])

de erro de $\frac{1}{2} - \beta$ ou $\frac{1}{2} - \epsilon$, respectivamente. Assim temos:

$$I(M; M_B) = \sum_{i=1}^N I(M_i; M_{B_i}) = N \cdot (1 - h(\beta))$$

$$I(M; M_E) = \sum_{i=1}^N I(M_i; M_{E_i}) = N \cdot (1 - h(\epsilon))$$

$$\frac{I(M; M_B)}{I(M; M_E)} = \frac{1 - h(\beta)}{1 - h(\epsilon)}$$

- **Caso 1** ($\beta < \epsilon$): Teremos $h(\beta) < h(\epsilon)$ e assim:

$$\frac{I(M; M_B)}{I(M; M_E)} = \frac{1 - h(\beta)}{1 - h(\epsilon)} > 0 \Rightarrow I(M; M_B) > I(M; M_E)$$

Como Bob possui mais informação sobre M que Eva, podemos estabelecer uma comunicação segura.

- **Caso 2** ($\beta > \epsilon$): Teremos $h(\beta) > h(\epsilon)$ e assim:

$$\frac{I(M; M_B)}{I(M; M_E)} = \frac{1 - h(\beta)}{1 - h(\epsilon)} < 0 \Rightarrow I(M; M_E) > I(M; M_B)$$

Como Bob possui menos informação sobre M que Eva, não podemos estabelecer uma comunicação segura.

Para resolver este problema da necessidade de vantagem por parte de Bob, Maurer apresentou em [16] um modelo alternativo no qual é adicionado um canal público, autêntico, inseguro, porém sem ruídos, por exemplo o telefone. Esse novo canal permite uma interação entre Alice e Bob, por meio da confirmação dos *bits* recebidos por Bob, o que lhe garante vantagem, mesmo com um Canal Simétrico Binário (por onde os *bits* da mensagem são trafegados), com probabilidade de erro superior ao de Eva (ver Figura 2.4).

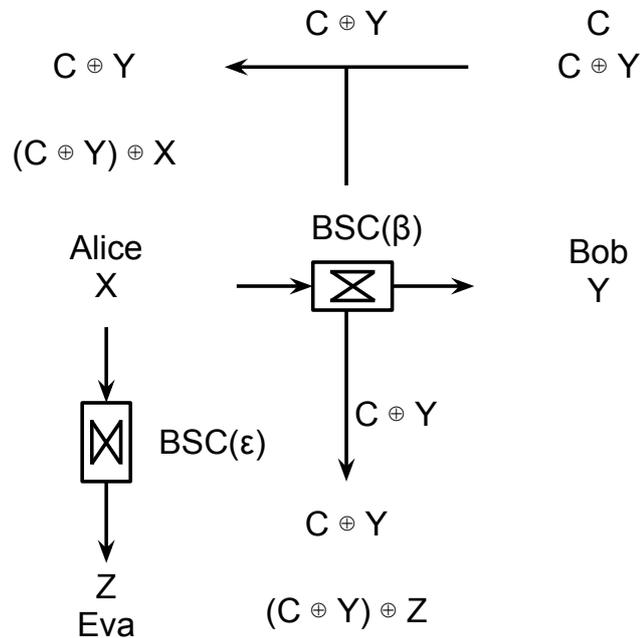


Figura 2.4: Modelo com canal público sem ruído. (Fonte: [31])

Para garantirmos essa vantagem de Bob sobre Eva basta estabelecermos a seguinte regra de troca de mensagens: Alice envia um *bit* aleatório X para Bob pelo canal ruidoso, sendo recebido por Bob como Y ($Y = X$ ou $Y = \bar{X}$) e por Eva como Z ($Z = X$ ou $Z = \bar{X}$). De posse de Y e de um *bit* C da mensagem a ser enviada, Bob calcula $Y \oplus C$ e envia pelo canal sem ruído para Alice. Assim, Alice calcula $X \oplus (Y \oplus C)$, enquanto Eva calcula $Z \oplus (Y \oplus C)$.

Assumindo β e ε independentes e $\beta, \varepsilon \ll 1$, podemos calcular a taxa de erro de Eva ao receber C :

$$Err = \beta(1 - \varepsilon) + \varepsilon(1 - \beta) = \beta + \varepsilon - 2\beta\varepsilon \approx \beta + \varepsilon.$$

Como a probabilidade de erro de Alice ao receber C é β , temos que a taxa de erro de Alice será menor ou igual ao de Eva, conferindo-lhe a necessária vantagem.

Esse modelo de comunicação será utilizado para atingirmos o objetivo da comunicação com segurança incondicional.

2.2 Modelando a Comunicação Satelital

Com base no modelo interativo desenvolvido por Ueli Maurer e apresentado na seção anterior, passaremos a expor o Modelo Satelital de comunicação criptográfica, estudado por Maurer e Stefan Wolf em [18].

2.2.1 Cenário

Considere o seguinte cenário: um satélite emite sinais aleatórios, os quais são recebidos por Alice, Bob e Eva. Alice e Bob querem estabelecer uma chave criptográfica para trocarem mensagens de forma segura. Para negociarem a chave a ser utilizada se comunicam por um canal autêntico, inseguro e sem ruídos. Este canal é interceptado por Eva, permitindo que ela receba todas as informações trocadas por Alice e Bob, mas não permitindo que Eva insira ou modifique informações sem que seja detectada por Alice e Bob (ver Figura 2.5).

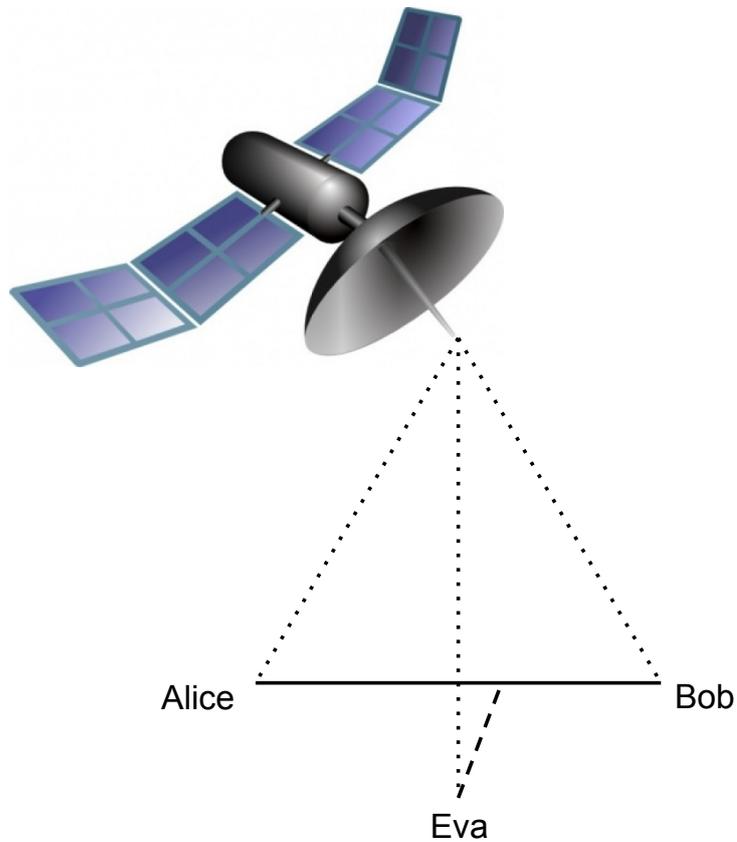


Figura 2.5: Modelo Satelital com Canal Autêntico interceptado por Eva

2.2.2 Modelagem Matemática

Devido às interferências naturais da comunicação satelital e à qualidade das antenas utilizadas pelos receptores, podemos modelar o canal satelital como um canal simétrico binário (ver Figura 2.6).

Além disso, como queremos garantir que o modelo é seguro e que independe da qualidade do adversário ao qual é submetido, iremos assumir que a antena de Eva é de qualidade superior às antenas de Alice e Bob. Mesmo assim, mostraremos que é possível obter um protocolo criptográfico que garanta segurança incondicional para a comunicação entre Alice e Bob.

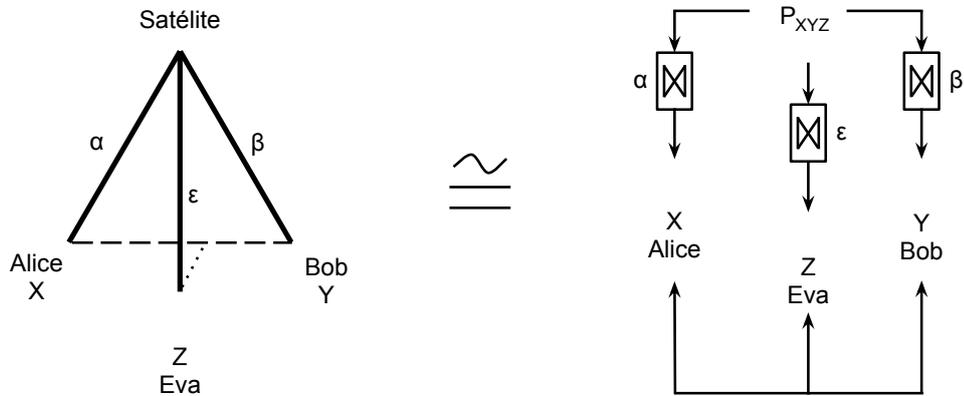


Figura 2.6: Equivalente Matemático do Modelo Satalital

Nesse modelo representaremos a sequência de N bits aleatórios recebidos por Alice, $[X_1, X_2, \dots, X_N]$, como X^N e, analogamente, Y^N para Bob e Z^N para Eva.

Além disso, iremos supor que a taxa de erro dos participantes seja:

$$0 < \alpha, \beta, \varepsilon < \frac{1}{2}, \text{ onde } \varepsilon < \min(\alpha, \beta).$$

2.3 O Protocolo e as Fases de Estabelecimento da Chave Secreta

Conforme apresentamos na seção anterior, Alice e Bob necessitam estabelecer uma chave segura para realizarem uma comunicação, mas toda a comunicação é interceptada por Eva, a qual possui vantagem sobre os participantes legítimos, uma vez que possui equipamentos melhores.

Para resolver esse problema Maurer propôs um protocolo que consiste de três fases: *Advantage Distillation*, *Information Reconciliation* e *Privacy Amplification*¹.

¹Destilação de Vantagem, Reconciliação de Informação e Aumento de Privacidade (tradução nossa).

Na primeira fase busca-se minimizar a vantagem de Eva por possuir melhor antena. Isso será possível com a exploração da troca de informações entre Alice e Bob pelo canal autêntico e sem ruídos. Assim, os participantes legítimos podem combinar quais sequências de *bits* recebidos do satélite devem utilizar.

Em seguida, Alice e Bob precisam corrigir os erros da sequência de *bits* obtida na fase anterior.

Finalizando o protocolo, os participantes legítimos transformam a sequência de *bits* que possui, eventualmente, algumas partes conhecidas por Eva, em uma sequência menor, mas com alto nível de privacidade.

2.3.1 Advantage Distillation

Consideremos N realizações da variável aleatória, ou seja, os três participantes recebem uma sequência com N *bits* enviados pelo satélite. A informação mútua da sequência de *bits* dos participantes, tomados dois a dois, será:

$$I(X^N; Y^N) = \sum_{i=1}^N I(X_i; Y_i) = N.(1 - h(\alpha(1 - \beta) + (1 - \alpha)\beta)),$$

$$I(X^N; Z^N) = \sum_{i=1}^N I(X_i; Z_i) = N.(1 - h(\alpha(1 - \varepsilon) + (1 - \alpha)\varepsilon)),$$

$$I(Y^N; Z^N) = \sum_{i=1}^N I(Y_i; Z_i) = N.(1 - h(\beta(1 - \varepsilon) + (1 - \beta)\varepsilon)).$$

Assim, considerando $\varepsilon < \alpha, \beta < \frac{1}{2}$, obtemos:

$$I(X^N; Y^N) < \min\{I(X^N; Z^N), I(Y^N; Z^N)\}.$$

O que comprova que Eva possui vantagem em relação a Alice e Bob, pois possui mais informação sobre a sequência de *bits* X^N do que Bob e mais informação sobre a sequência de *bits* Y^N que Alice.

Para alterarmos essa situação, apresentaremos dois protocolos capazes de proporcionar vantagem a Alice e Bob, mesmo com essa superioridade inicial de Eva.

Através do canal autêntico e sem ruídos, Alice e Bob irão trocar informações para identificar *bits* das sequências X^N e Y^N que sejam iguais, com uma probabilidade muito alta.

Repeat-Code Protocol

Considere a sequência de N *bits* recebidos pelos três participantes. Alice escolhe aleatoriamente um *bit* C e computa:

$$C^N \oplus X^N := [C \oplus X_1, C \oplus X_2, \dots, C \oplus X_N],$$

tal que C^N corresponde à sequência, de tamanho N , formada pela repetição do *bit* aleatório C , ou seja, $C^N = [C, C, \dots, C]$. Em seguida, Alice envia esta nova sequência para Bob, o qual computará:

$$(C^N \oplus X^N) \oplus Y^N.$$

Bob aceitará o *bit* C' se obtiver $(C^N \oplus X^N) \oplus Y^N = C'^N$, situação que corresponde a $X^N = Y^N$ ou $X^N = Y^N \oplus 1^N$ (ver Figura 2.7).

O problema deste método é que conforme se aumenta N , a probabilidade de $C = C'$ aumenta, mas também aumenta a chance de Eva “corrigir” possíveis erros de sua sequência ao computar: $(C^N \oplus X^N) \oplus Z^N$.

Mesmo com esse empecilho, Alice e Bob ainda conseguem obter vantagem sobre Eva. Para vermos isso vamos utilizar a seguinte notação: α_{be} = probabilidade de Alice enviar o *bit* x e Bob receber como b , enquanto Eva recebe como e . Ou seja, Alice envia o *bit* x resultante de $C \oplus X$ pelo canal público, o qual é recebido por Bob como b e por Eva como e , podendo ser x ou \bar{x} . Assim teremos (ver Figura 2.8):

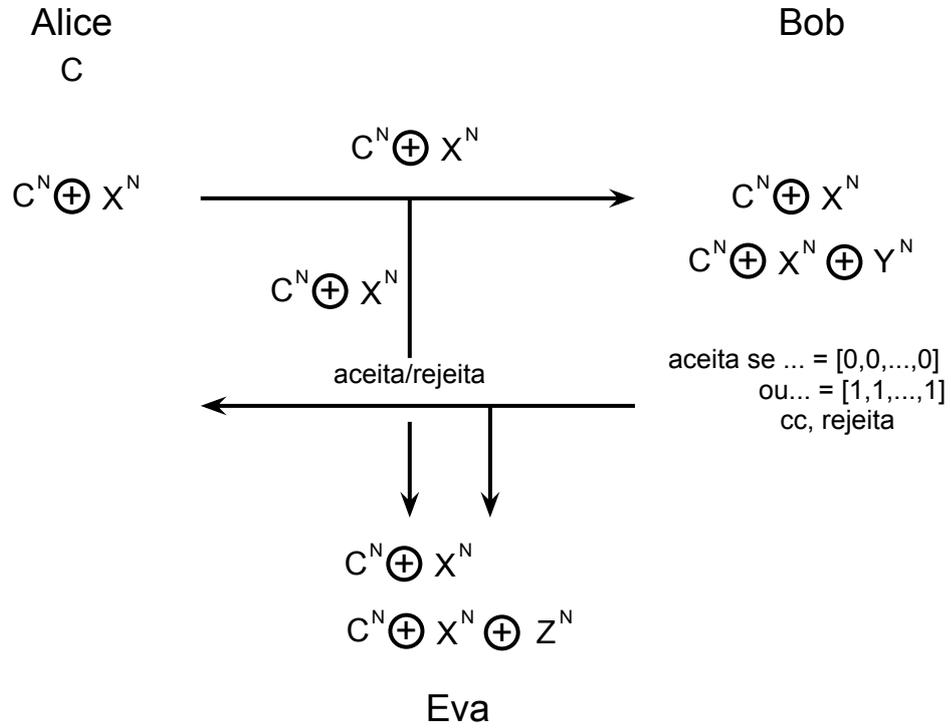


Figura 2.7: Repeat-Code Protocol. (Fonte: [31])

$$\begin{aligned} \alpha_{xx} &= (1 - \alpha)(1 - \beta)(1 - \varepsilon) + \alpha\beta\varepsilon, \\ \alpha_{x\bar{x}} &= (1 - \alpha)(1 - \beta)\varepsilon + \alpha\beta(1 - \varepsilon), \\ \alpha_{\bar{x}x} &= (1 - \alpha)(1 - \varepsilon)\beta + \alpha\varepsilon(1 - \beta), \\ \alpha_{\bar{x}\bar{x}} &= (1 - \alpha)\beta\varepsilon + \alpha(1 - \beta)(1 - \varepsilon). \end{aligned}$$

Como sempre podemos adicionar erro ao sistema, podemos supor que $\alpha = \beta$ e com isso teremos:

$$\begin{aligned} \alpha_{xx} &= (1 - \alpha)^2(1 - \varepsilon) + \alpha^2\varepsilon, \\ \alpha_{x\bar{x}} &= (1 - \alpha)^2\varepsilon + \alpha^2(1 - \varepsilon), \\ \alpha_{\bar{x}x} &= \alpha_{\bar{x}\bar{x}} = \alpha(1 - \alpha). \end{aligned}$$

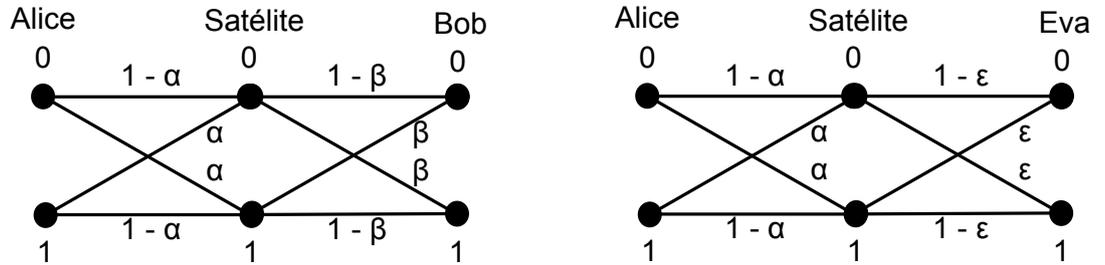


Figura 2.8: Cálculo das Probabilidade.

A probabilidade de Bob aceitar a sequência de *bits* enviada por Alice será:

$$\begin{aligned}
 P_{aceita} &= (\alpha_{xx} + \alpha_{x\bar{x}})^N + (\alpha_{\bar{x}x} + \alpha_{\bar{x}\bar{x}})^N \\
 &= [(1 - \alpha)^2 + \alpha^2]^N + [2\alpha(1 - \alpha)]^N
 \end{aligned}$$

Para calcularmos a probabilidade γ de Bob aceitar o bloco de tamanho N enviado por Alice e obter o valor de C errado, isto é, $C = C' \oplus 1$, fazemos:

$$\begin{aligned}
 \gamma &= \frac{1}{P_{aceita}} \cdot (\alpha_{\bar{x}x} + \alpha_{\bar{x}\bar{x}})^N \\
 &= \frac{1}{P_{aceita}} \cdot (2\alpha - 2\alpha^2)^N
 \end{aligned}$$

que corresponde à situação em que todos os *bits* da sequência de Bob são diferentes dos *bits* da sequência de Alice (uma soma XOR das sequências resultaria em 1^N).

Agora falta calcularmos a probabilidade de Eva obter insucesso ao “adivinhar” o *bit* C . Esse insucesso de Eva ocorre quando, ao receber a sequência de *bits* enviada por Alice e calcular o *bit* C (utilizando a relação $C^N = C^N \oplus X^N \oplus Z^N$), Eva encontra um valor diverso do C de Alice, isto é, a sequência de Eva diverge da sequência da Alice em mais da metade dos *bits*. Além disso temos o caso em que Eva obtém metade dos *bits* como 1 e a outra metade como 0, tendo que escolher aleatoriamente

qual adotar. Como queremos mostrar que este método deixa Eva em desvantagem, vamos calcular apenas este último caso, o qual nos fornecerá um limite inferior para a probabilidade de Eva obter um insucesso.

Sem perda de generalidade, consideremos N par. A probabilidade δ de Bob aceitar o bloco de tamanho N enviado por Alice e Eva adivinhar o *bit* C incorretamente será:

$$\delta \geq \frac{1}{2} \cdot \frac{1}{P_{aceita}} \cdot \binom{N}{N/2} \left(\alpha_{xx}^{N/2} \alpha_{x\bar{x}}^{N/2} + \alpha_{\bar{x}x}^{N/2} \alpha_{\bar{x}\bar{x}}^{N/2} \right)$$

A partir deste ponto, passamos a uma nova abordagem deste protocolo, pois divergimos dos autores quando utilizam a hipótese de N suficientemente grande, o que no nosso entendimento cria um viés na análise, pois como mencionado anteriormente, uma sequência muito grande poderia ser facilmente corrigida por Eva. Na prática temos que utilizar sequências não superiores a 30 *bits*, o que não pode ser considerada uma sequência “suficientemente grande”, conforme abordado pelos autores.

Pela abordagem original, obtém-se o valor de $N!$ pela aproximação de *Stirling*, a qual utiliza um limite em que N tende ao infinito. Tendo em vista o fato de N , não poder ser um número grande, passaremos a uma nova abordagem do modelo, utilizando uma melhoria da aproximação de *Stirling* para números pequenos. Observe o lema a seguir.

Lema 2.1 ([19], [13] e [10]). *Para qualquer valor par de N , temos:*

$$\binom{N}{N/2} \approx \frac{1}{\sqrt{2\pi N}} \cdot 2^{N+1} \cdot e^{-\frac{1}{4N}}.$$

Demonstração. Pela aproximação de *Stirling* temos:

$$\lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n} = 1 \Rightarrow n! \approx \sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n$$

Mas no volume 1 do livro *The art of computer programming*, referência [13], Donald Knuth melhora essa aproximação chegando a:

$$n! \approx \sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n \cdot e^{\frac{1}{12n} - \frac{1}{360n^3}}$$

Entretanto, no artigo [10] o autor mostra que, mesmo para valores pequenos, basta utilizarmos a seguinte aproximação:

$$n! \approx \sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n \cdot e^{\frac{1}{12n}}$$

Para compararmos a aproximação de *Stirling* original, com as dadas pelas referências [10] e [13], montamos a Tabela 2.1. Repare que a aproximação dada por [10] nos fornece valores bem próximos do real, com taxa de erro de 0,2% para $n = 1$, sendo esse a taxa de erro máxima obtida. Para $n = 3$, a taxa de erro já é inferior a 0,01%.

Assim, utilizando a aproximação de *Stirling* obtida por Felício em [10], vem:

$$\binom{N}{N/2} = \frac{N!}{((N/2)!)^2} \approx \frac{N^N \cdot \sqrt{2\pi N} \cdot e^N}{e^N \cdot (N/2)^N \cdot \pi N} \cdot \frac{e^{\frac{1}{12N}}}{e^{\frac{1}{3N}}} = \frac{1}{\sqrt{2\pi N}} \cdot 2^{N+1} \cdot e^{-\frac{1}{4N}}$$

□

Utilizando o Lema 2.1 no cálculo da probabilidade δ , obtemos:

$$\delta \geq \frac{1}{P_{aceita}} \cdot \frac{2^N}{\sqrt{2\pi N \cdot \sqrt[2N]{e}}} \cdot \left(\sqrt{\alpha_{xx} \cdot \alpha_{x\bar{x}}}^N + \sqrt{\alpha_{\bar{x}x} \cdot \alpha_{\bar{x}\bar{x}}}^N \right)$$

com:

$$\begin{aligned} \alpha_{xx} \cdot \alpha_{x\bar{x}} &= [(1 - \alpha)^2(1 - \varepsilon) + \alpha^2\varepsilon] \cdot [(1 - \alpha)^2\varepsilon + \alpha^2(1 - \varepsilon)] \\ &= \alpha^2(1 - \alpha)^2 + (1 - 2\alpha)^2\varepsilon(1 - \varepsilon) \\ \alpha_{\bar{x}x} \cdot \alpha_{\bar{x}\bar{x}} &= [\alpha(1 - \alpha)]^2 = \alpha^2(1 - \alpha)^2 \end{aligned}$$

n	n!	Aproximação de <i>Stirling</i>	Aproximação de <i>Stirling</i> por Felício [10]	Aproximação de <i>Stirling</i> por Knuth [13]
1	1	0,922137009	1,002274449	0,999494217
2	2	1,919004351	2,000652048	1,999957497
3	6	5,836209591	6,000599142	5,999981829
4	24	23,50617513	24,00102389	23,9999822
5	120	118,019168	120,0026371	119,9999704
6	720	710,0781846	720,0091873	719,999928
7	5040	4980,395832	5040,040582	5039,999766
8	40320	39902,39545	40320,21779	40319,99903
9	362880	359536,8728	362881,3779	362879,9952
10	3628800	3598695,619	3628810,051	3628799,971

Tabela 2.1: Valores de n! (aproximação de *Stirling*)

Para mostrarmos em que situações podemos garantir que $\delta > \gamma$, calcularemos a razão $\frac{\delta}{\gamma}$.

$$\begin{aligned} \frac{\delta}{\gamma} &\geq \frac{1}{P_{aceita}} \cdot \frac{2^N}{\sqrt{2\pi N} \cdot \sqrt[2N]{e}} \cdot \left(\sqrt{\alpha_{xx} \cdot \alpha_{x\bar{x}}^N} + \sqrt{\alpha_{\bar{x}x} \cdot \alpha_{\bar{x}\bar{x}}^N} \right) \cdot \left(\frac{1}{P_{aceita}} \cdot 2^N \cdot (\alpha(1-\alpha))^N \right)^{-1} \\ &\geq \frac{1}{\sqrt{2\pi N} \cdot \sqrt[2N]{e}} \cdot \frac{\left(\sqrt{\alpha_{xx} \cdot \alpha_{x\bar{x}}^N} + \sqrt{\alpha_{\bar{x}x} \cdot \alpha_{\bar{x}\bar{x}}^N} \right)}{(\alpha(1-\alpha))^N} \end{aligned}$$

Substituindo as probabilidades já calculadas e fazendo $\alpha(1-\alpha) = \zeta$, obtemos:

$$\begin{aligned} \frac{\delta}{\gamma} &\geq \frac{1}{\sqrt{2\pi N} \cdot \sqrt[2N]{e}} \cdot \frac{\left(\sqrt{\zeta^2 + (1-2\alpha)^2 \varepsilon(1-\varepsilon)^N} + \zeta^N \right)}{\zeta^N} \\ &\geq \frac{1}{\sqrt{2\pi N} \cdot \sqrt[2N]{e}} \cdot \left(\sqrt{1 + \frac{(1-2\alpha)^2 \varepsilon(1-\varepsilon)^N}{\zeta^2}} + 1 \right) \end{aligned}$$

Definindo $D(N) := \sqrt{2\pi N \cdot \sqrt[2N]{e}}$ e substituindo o valor de ζ , vem:

$$\frac{\delta}{\gamma} \geq \frac{1}{D(N)} \cdot \left(\sqrt{1 + \frac{(1-2\alpha)^2 \varepsilon (1-\varepsilon)^N}{(\alpha(1-\alpha))^2}} + 1 \right)$$

A igualdade ocorre quando:

$$D(N) = \left(\sqrt{1 + \frac{(1-2\alpha)^2 \varepsilon (1-\varepsilon)^N}{(\alpha(1-\alpha))^2}} + 1 \right)$$

ou seja, $\delta \geq \gamma$ quando:

$$\left(\sqrt[2N]{D(N)} - 1 \right)^2 - 1 \leq \frac{(1-2\alpha)^2 \varepsilon (1-\varepsilon)}{(\alpha(1-\alpha))^2}$$

Defina o valor de segurança $Seg(N) := \left(\sqrt[2N]{D(N)} - 1 \right)^2 - 1$. Assim, montamos a Tabela 2.2.

De posse das taxas de erro iniciais, α e ε , podemos obter o valor de N necessário para garantir a vantagem de Alice e Bob em relação a Eva, pois:

$$Seg(N) \leq \frac{(1-2\alpha)^2 \varepsilon (1-\varepsilon)}{(\alpha(1-\alpha))^2}.$$

Sejam, por exemplo, $\alpha = 0,3$ e $\varepsilon = 0,2$. Nesse caso teremos $Seg(N) \leq 0,5805$ e assim, $N \geq 10$. Já para $\alpha = 0,2$ e $\varepsilon = 0,15$, teremos $Seg(N) \leq 1,793$ e assim, $N \geq 4$. Entretanto, quando a probabilidade de erro ε é muito menor que α , temos que utilizar uma sequência maior, o que acaba por inviabilizar a utilização do modelo. Sejam, por exemplo, $\alpha = 0,3$ e $\varepsilon = 0,05$. Nessa situação teremos $Seg(N) \leq 0,1056$, que ocorre apenas com $N \geq 58$, dificultando a aplicação do modelo.

Assim, fica demonstrado que, quando a diferença entre as probabilidades de erro não é muito alta, o “*Repeat-Code Protocol*” confere vantagem a Alice e Bob em relação a Eva.

$N(par)$	$D(N)$	$Seg(N)$
2	4,016906679	2,016906679
4	5,336583818	1,082446594
6	6,401196554	0,754540217
8	7,314870303	0,585225715
10	8,127318812	0,481102916
12	8,866012901	0,410238096
14	9,547929654	0,358698173
16	10,1844077	0,319414931
18	10,78345808	0,288411848
20	11,35098665	0,263275795
22	11,8914945	0,242455033
24	12,40850488	0,224904963
26	12,9048368	0,209895767
28	13,38278741	0,19690163
30	13,84425794	0,185533784

Tabela 2.2: Valores Referência com Variação de N

Parity-Check Protocol

Alice e Bob estipulam o tamanho N da sequência de *bits* a ser considerada. Alice calcula a paridade dos dois primeiros *bits* e envia para Bob, ou seja, Alice calcula $X_1 \oplus X_2$ e envia o resultado para Bob. Caso a paridade dos dois primeiros *bits* de Bob seja igual a recebida de Alice ($X_1 \oplus X_2 = Y_1 \oplus Y_2$), Bob envia um sinal de aceite da paridade e, nesse caso, ambos separam o primeiro *bit*, X_1 e Y_1 , respectivamente, para a próxima rodada do protocolo, caso contrário, cada um descarta os dois *bits*, passando para o próximo par de *bits*. Ao terminar toda a sequência de N *bits*,

inicia-se a segunda rodada com os *bits* separados na primeira rodada. A quantidade de rodadas necessárias vai depender das taxas de erro iniciais (α , β e ε). Observe a Figura 2.9.

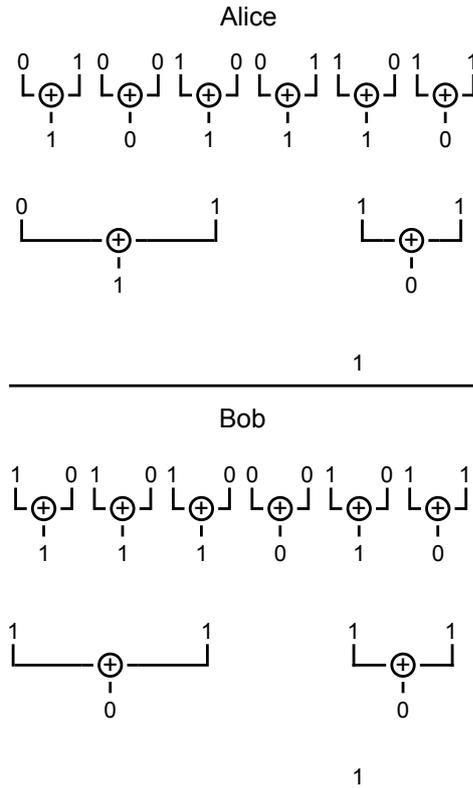


Figura 2.9: Parity-Check Protocol. (Fonte: [31])

Assim como no protocolo anterior, consideraremos a probabilidade de erro da comunicação satelital de Alice e Bob iguais, ou seja, $\alpha = \beta$. Além disso, conforme explicamos anteriormente, temos: $0 < \varepsilon < \alpha = \beta < \frac{1}{2}$. Para facilitar o entendimento adotaremos a seguinte notação:

$X_{a,b} = (x_a, x_b)$ significa que os *bits* a e b da sequência da Alice são x_a e x_b , dessa forma, $Y_{1,2} = (0, 1)$ significa que os *bits* 1 e 2 da sequência de Bob são 0 e 1, respectivamente. Conforme adotado anteriormente, representaremos a sequência de

Alice por X , a de Bob por Y e a de Eva por Z .

Vamos calcular a probabilidade de Bob aceitar o *bit* de paridade enviado por Alice, o que ocorre quando a dupla de *bits* de Alice e Bob são iguais ou diferentes, um a um, i.e.: $X_{1,2} = Y_{1,2} = (k, k)$ ou $X_{1,2} = (k, k)$ e $Y_{1,2} = (\bar{k}, \bar{k})$, $k \in \{0, 1\}$.

$$P_{aceita} = [(1 - \alpha)^2 + \alpha^2]^2 + [2\alpha(1 - \alpha)]^2.$$

A probabilidade γ de Bob aceitar a paridade e separar um *bit* diferente do *bit* de Alice para a próxima rodada do protocolo é:

$$\gamma = \frac{1}{P_{aceita}} \cdot [2\alpha(1 - \alpha)]^2.$$

Passemos ao cálculo da probabilidade δ de Bob aceitar o *bit* de paridade e Eva obter um insucesso, ou seja, Eva separar um *bit* diferente do *bit* da Alice. Para facilitar, dividiremos em duas situações. Na primeira, Eva obtém a mesma paridade de Alice e Bob, mas escolhe o *bit* diferente do *bit* de Alice, i.e., $X_{1,2} = (k, k)$ e $Z_{1,2} = (\bar{k}, \bar{k})$, $k \in \{0, 1\}$. Na segunda situação, Eva obtém a paridade diferente de Alice e, por isso, tem que adivinhar o *bit* a ser considerado.

$$\begin{aligned} \delta_1 &= \frac{1}{P_{aceita}} \cdot [\alpha(1 - \varepsilon) + \varepsilon(1 - \alpha)]^2 \\ \delta_2 &= \frac{1}{2} \cdot \frac{1}{P_{aceita}} \cdot 2 \cdot [(1 - \alpha)(1 - \varepsilon) + \alpha\varepsilon] \cdot [(1 - \alpha)\varepsilon + \alpha(1 - \varepsilon)] \\ \delta_2 &= \frac{1}{P_{aceita}} \cdot [(1 - \alpha)(1 - \varepsilon) + \alpha\varepsilon] \cdot [(1 - \alpha)\varepsilon + \alpha(1 - \varepsilon)] \\ \delta &= \delta_1 + \delta_2 \\ \delta &= \frac{1}{P_{aceita}} \cdot [\alpha(1 - \varepsilon) + \varepsilon(1 - \alpha)] \cdot [\alpha(1 - \varepsilon) + \varepsilon(1 - \alpha) + (1 - \alpha)(1 - \varepsilon) + \alpha\varepsilon] \\ \delta &= \frac{1}{P_{aceita}} \cdot (\alpha + \varepsilon - 2\alpha\varepsilon) = \frac{1}{P_{aceita}} \cdot [(\alpha(1 - \varepsilon) + \varepsilon(1 - \alpha))] \end{aligned}$$

Para avaliarmos se δ é maior que γ façamos $\frac{\delta}{\gamma}$:

$$\frac{\delta}{\gamma} = \frac{\alpha(1 - \varepsilon) + \varepsilon(1 - \alpha)}{[2\alpha(1 - \alpha)]^2} = \frac{\alpha + \varepsilon(1 - 2\alpha)}{4\alpha^2(1 - \alpha)^2}$$

$$\lim_{\varepsilon \rightarrow 0} \frac{\delta}{\gamma} = \frac{1}{4\alpha(1 - \alpha)^2},$$

e como $0 < \alpha < \frac{1}{2}$, $\delta > \gamma$.

A principal vantagem deste método reside na sua eficiência, independente das probabilidades de erro iniciais consideradas, enquanto o método anterior funciona em uma faixa específica de taxas de erro iniciais, na qual a diferença entre as taxas não é muito alta.

2.3.2 Information Reconciliation

A fase de *Advantage Distillation* nos permitiu aumentar a vantagem de Bob em relação à Eva e ao final dela, Alice e Bob possuem sequências de *bits*, S_A e S_B , respectivamente, que são muito parecidas, mas que ainda podem conter divergências em alguns *bits*. Durante a fase chamada de *Information Reconciliation*, realizaremos a correção dos erros ainda existentes, trocando informações sobre as sequências S_A e S_B , tornando-as iguais com alta probabilidade.

Assim, ao final da fase, Bob deve saber exatamente qual a sequência de *bits* de Alice, mas para isso terá que pagar um preço, o qual pode inviabilizar todo o esforço para estabelecer uma chave secreta. O preço é a troca de informações suficientes para se realizar a correção dos erros, informações que também são recebidas por Eva. Com isso, temos que minimizar a troca dessas informações, evitando que Eva também consiga corrigir os erros de sua sequência em relação a de Alice.

Para garantir a segurança necessária nesta fase, Bob deverá ser capaz de obter S_A a partir de S_B e E , sendo E as informações trocadas entre Alice e Bob para a

correção dos erros. Entretanto, para Eva não deve ser possível obter S_A a partir de E e S_E , sendo S_E a sequência de *bits* obtida por Eva após a fase *Advantage Distillation*. Assim, devemos garantir que:

$$0 \approx H(S_A|S_B, E) \geq H(S_A|S_B) - H(E)$$

$$H(E) \geq H(S_A|S_B),$$

além disso devemos ter:

$$H(S_A|S_E, E) \geq H(S_A|S_E) - H(E) > 0$$

$$H(E) < H(S_A|S_E).$$

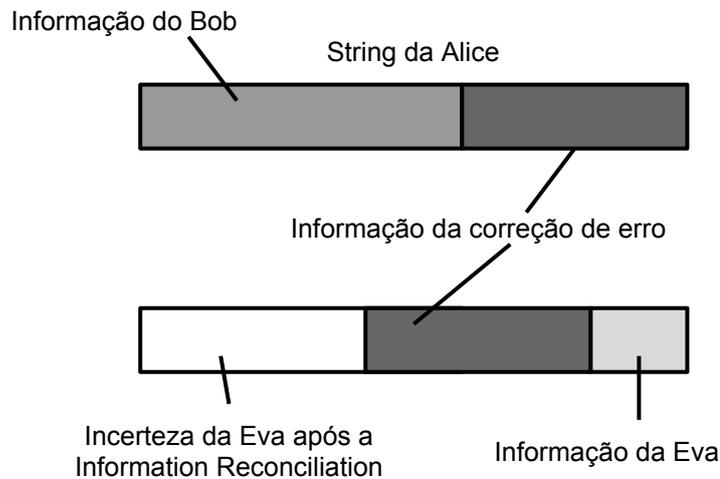


Figura 2.10: Incerteza de Eva após a fase *Information Reconciliation*. (Fonte: [31])

Apresentaremos um protocolo capaz de corrigir os erros ainda existentes entre S_A e S_B , sem revelar muita informação para Eva. Este protocolo é trabalhado em [31] e em [3].

Binary-Search Protocol

Neste protocolo Alice e Bob calculam a paridade de subsequências aleatórias de tamanho combinado previamente e as comparam. Caso a paridade seja diferente temos uma quantidade ímpar de erros nas subsequências e iniciaremos o processo para corrigir esse erro, dividindo a subsequência ao meio e calculando a paridade das novas subsequências obtidas. Esse processo continua até se encontrar o *bit* divergente. Veja a Figura 2.11.

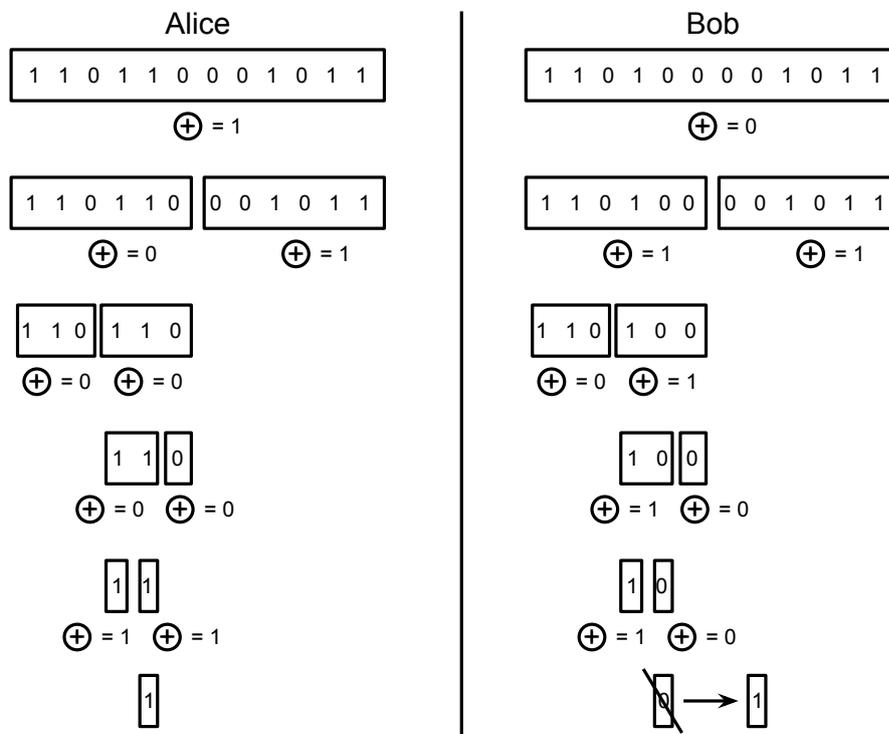


Figura 2.11: Binary-Search Protocol. (Fonte: [31])

No exemplo da Figura 2.11, fica evidente que Eva terá certeza dos três *bits* que compõem a quarta subsequência (os *bits* 110) e por isso devemos excluí-los da sequência original antes de continuarmos a conferência. Assim, visando diminuir a quantidade de informação que Eva possuirá a respeito da sequência S_A , devemos

excluir os *bits* que compõem a subsequência de número $\lceil s \rceil$, tal que $s = \log n$, sendo n a quantidade de *bits* da subsequência escolhida aleatoriamente. Com isso, no exemplo da Figura 2.11, o correto seria excluir todos os *bits* da terceira sequência (os *bits* 110110).

Deve-se observar que o fato de encontrarmos a mesma paridade para uma subsequência específica não significa que essa subsequência esteja sem erro, mas sim que a quantidade de erros é par, podendo, inclusive, ser zero. Por isso, para encontrarmos os erros, devemos realizar buscas por subsequências aleatórias.

Esse protocolo é eficiente para uma quantidade pequena de erros entre S_A e S_B , pois para cada erro encontrado devem ser trocados $\lceil \log k \rceil$ *bits* para corrigi-los, sendo k a quantidade de bits de S_A .

Ao final desta fase, Alice e Bob terão, com alta probabilidade, uma sequência S da qual Eva tem alguma informação, que corresponderá a *bits* oriundos da fase *Advantage Distillation*, e paridades da fase *Information Reconciliation*. Agora restará à dupla aumentar o sigilo desta sequência.

2.3.3 Privacy Amplification

O objetivo desta fase é aumentar a privacidade da sequência S , a qual é conhecida tanto por Alice quanto por Bob. Para isso precisamos aumentar os efeitos dos erros existentes na sequência S_E de Eva. Por não conhecermos a sequência S_E , podemos fazê-lo aplicando a sequência S a uma função f que a “embaralhe”.

As funções *hash* realizarão essa tarefa, levando S em S' , onde S tem dimensão n e S' tem dimensão r , i.e., $f : \{0,1\}^n \rightarrow \{0,1\}^r$. Além disso, será utilizada uma classe específica de funções *hash*, a classe universal de funções *hash*, conferindo aleatoriedade à escolha da função a ser utilizada.

Definição 2.1 (Classe Universal de Funções *Hash*, [14], [28] e [31]). Seja \mathcal{H} uma classe universal de funções h , mapeando um conjunto \mathcal{A} em um conjunto \mathcal{B} . Se $\forall x, y \in \mathcal{A}, x \neq y$, temos:

$$P_{h \in_U \mathcal{H}}[h(x) = h(y)] = \frac{1}{|\mathcal{B}|},$$

onde $h \in_U \mathcal{H}$ indica que h é escolhida aleatoriamente em \mathcal{H} segundo uma distribuição uniforme.

Observação 2.1. O fato de a classe de funções ser universal, garante a escolha uniforme da função $h \in \mathcal{H}$.

De posse do conceito de classe universal de funções *hash*, podemos descrever o protocolo a ser adotado por Alice e Bob, visando aumentar a privacidade da sequência de *bits* S obtida ao final da fase de correção de erros.

Alice escolhe uma função h de uma classe universal \mathcal{H} , predefinida, de funções *hash*, a qual mapeia uma sequência de n *bits* em uma sequência de r *bits*. Após a escolha, Alice envia a descrição de h para Bob pelo canal público e ambos computam $S' := h(S)$. Como a descrição de h foi enviada pelo canal público, Eva, possivelmente, conhecerá a função h .

Agora abordaremos uma forma de Alice obter r com vistas a tornar S' seguro com relação às interceptações de Eva. Para isso devemos analisar quanta informação Eva pode ter a respeito de S . Observe inicialmente que se considerarmos $U = u$ como sendo o que Eva sabe sobre S e considerando que Eva não conhece toda a sequência S , vem:

$$H(S|U = u) < n,$$

onde n é o tamanho da sequência S .

Assim, dizemos que Eva possui $n - H(S|U = u)$ *bits* de S . Como queremos

minimizar o conhecimento de Eva a respeito de S' , devemos ter:

$$H(S'|Pub, U = u) \approx r,$$

onde r é a quantidade de *bits* de S' e Pub a comunicação trocada entre Alice e Bob pelo canal público. Rapidamente esse fato leva à conclusão de que se Eva conhece t *bits* de S , S' precisa ter $n - t$ *bits*. Entretanto isso não pode ser generalizado, sendo verdade quando podemos determinar exatamente a quantidade de informação de S que é de conhecimento de Eva. O exemplo 2.1 ilustra uma situação em que esse fato não é verdade.

Exemplo 2.1. Considere:

$$P_{S|U=u}(s_0) = \frac{1}{2}, \quad s_0 \in \{0, 1\}^n,$$

$$P_{S|U=u}(s) = \frac{1}{2 \cdot (2^n - 1)}, \quad \forall s \neq s_0.$$

Nessas condições, a probabilidade de $S = s_0$ é muito maior que a de qualquer outra sequência. Assim teremos $H(S|U = u) \approx \frac{n}{2}$, mas não podemos retirar nenhuma quantidade de *bits* de S seguramente, pois Eva conhece S , e conseqüentemente $S' = h(S)$, com probabilidade $\frac{1}{2}$, ou seja, não podemos obter S' com alta privacidade.

Para resolvermos esse problema, recorreremos às definições de *Probabilidade de Colisão* e de *Entropia de Colisão*.

Definição 2.2 ([31] e [1]). Seja X uma variável aleatória com distribuição P_X . A *probabilidade de colisão* $P_C(X)$ é definida como:

$$P_C(X) := \sum_{x \in \mathcal{X}} (P_X(x))^2$$

$$= E[P_X(X)].$$

A *entropia de colisão*, ou *entropia de Rényi* (de ordem 2) de X é:

$$\begin{aligned} H_2(X) &:= -\log(P_C(X)) = -\log \sum_{x \in \mathcal{X}} (P_X(x))^2 \\ &= -\log E[P_X(X)] \end{aligned}$$

A *probabilidade de colisão* é a probabilidade de duas realizações independentes de uma variável aleatória apresentarem o mesmo resultado. Adaptando à nossa situação, corresponde a se descobrir corretamente a realização de uma determinada variável aleatória sem se conhecer sua distribuição de probabilidade.

Tanto a *probabilidade de colisão condicional*, quanto a *entropia de Rényi condicional*, podem ser definidas de forma natural para uma distribuição condicional. Ou seja, considerando o evento \mathcal{E} , a *probabilidade de colisão* e a *entropia de Rényi* de X condicionadas ao evento \mathcal{E} são calculadas para uma distribuição condicional $P_{X|\mathcal{E}}$ (ver a referência [1]). Assim, a *entropia de Rényi* de uma variável aleatória X , condicionada a uma variável aleatória Y é:

$$H_2(X|Y) = \sum_{y \in \mathcal{Y}} P_Y(y) H_2(X|Y = y)$$

Agora iremos obter um limite superior para a *entropia de Rényi*.

Lema 2.2 ([1]). *Dada a distribuição de probabilidade P_X , obtemos:*

$$H_2(X) \leq H(X),$$

com igualdade se, e somente se, P_X é uniformemente distribuída em um conjunto \mathcal{X} ou um subconjunto de \mathcal{X} . Além disso, para uma distribuição condicional $P_{X|Y}$, temos:

$$H_2(X|Y) \leq H(X|Y).$$

Demonstração. Utilizando a desigualdade de Jensen (ver 1.4), vem:

$$H_2(X) = -\log_2(E[P_X]) \leq E[-\log_2 P_X] = H(X).$$

Analogamente, substituindo X por $X|Y$, obtemos:

$$H_2(X|Y) = -\log_2(E[P_{X|Y}]) \leq E[-\log_2 P_{X|Y}] = H(X|Y).$$

□

Voltemos agora ao problema de se determinar um valor seguro de r para aumentar a segurança da sequência de *bits* S quando mapeada por h . O Teorema 2.1 mostra como a *entropia de Rényi* pode representar a quantidade de informação a ser excluída de S , dando origem a S' com alta privacidade.

Teorema 2.1 ([1]). *Sejam X uma variável aleatória de um conjunto \mathcal{X} com distribuição de probabilidade P_X e entropia de Rényi $H_2(X)$, e G uma variável aleatória correspondendo à escolha aleatória e uniforme de um elemento da classe universal de funções hash $\mathcal{X} \rightarrow \{0, 1\}^r$, com $Q = G(X)$. Nessas condições, temos:*

$$H(Q|G) \geq H_2(Q|G) \geq r - \log_2(1 + 2^{r-H_2(X)}) \geq r - \frac{2^{r-H_2(X)}}{\ln 2}.$$

Note que G é uma variável aleatória e a quantidade $H(Q|G) = H(G(X)|G)$ é uma média de todas as escolhas de funções g , com g um valor particular de G .

Demonstração. Pelo Lema 2.2, obtemos:

$$H(Q|G) \geq H_2(Q|G).$$

A partir daí vem:

$$H_2(Q|G) = H_2(G(X)|G) = \sum_g P_G(g) H_2(G(X)|G = g) \quad (2.1)$$

$$= \sum_g P_G(g) (-\log_2 P_C(G(X)|G = g)) \quad (2.2)$$

$$\geq -\log_2 \left(\sum_g P_G(g) P_C(G(X)|G = g) \right) \quad (2.3)$$

onde a última passagem vem da desigualdade de Jensen. A soma em (2.3) equivale à probabilidade de $g(x_1) = g(x_2)$, em que g é escolhida aleatoriamente de acordo com P_G e x_1 e x_2 são escolhidas aleatória e independentemente uma da outra e de g , de acordo com P_X . Assim, temos:

$$\sum_g P_G(g)P_C(G(X)|G = g) = P[G(X_1) = G(X_2)] = \quad (2.4)$$

$$= P[X_1 = X_2] + P[X_1 \neq X_2] \cdot P[G(X_1) = G(X_2)|X_1 \neq X_2] \leq \quad (2.5)$$

$$\leq P_C(X) + (1 - P_C(X)) \cdot 2^{-r} \leq 2^{-H_2(X)} + 2^{-r} = \quad (2.6)$$

$$= 2^{-r} \cdot (1 + 2^{r-H_2(X)}). \quad (2.7)$$

Na passagem (2.5) utilizamos o fato de a classe de funções consideradas ser universal, com $\frac{1}{|\mathcal{B}|} = 2^{-r}$. Em (2.6) utilizamos que $P_C(X) = 2^{-H_2(X)}$ e que $0 \leq 1 - P_C(X) \leq 1$. Voltando à desigualdade em (2.3) ficamos com:

$$H_2(Q|G) \geq -\log_2(2^{-r} \cdot (1 + 2^{r-H_2(X)})) \quad (2.8)$$

$$\geq r - \log_2(1 + 2^{r-H_2(X)}) \quad (2.9)$$

$$\geq r - \frac{2^{r-H_2(X)}}{\ln 2}. \quad (2.10)$$

Em (2.9) utilizamos a desigualdade $\log_2(1 + y) \leq \frac{y}{\ln 2}$. □

Agora vamos iniciar a adequação do Teorema 2.1 ao modelo de comunicação envolvendo Alice, Bob e Eva.

Corolário 2.1 ([1]). *Considere a distribuição de probabilidade $P_{S|U}$ e seja u um valor particular de U observado por Eva. Se a entropia de Rényi relativa a Eva $H_2(S|U = u)$ sobre S é pelo menos c , e Alice e Bob escolhem $S' = G(S)$ como chave secreta, onde G é escolhido aleatoriamente de uma classe universal \mathcal{H} de funções hash que mapeiam S em $\{0, 1\}^r$, então:*

$$H(S'|G, U = u) \geq r - \log_2(1 + 2^{r-c}) \geq r - \frac{2^{r-c}}{\ln 2}.$$

Vemos que quando $r < c$, a entropia de Eva em relação a S' é alta e sua distribuição se aproxima da distribuição uniforme. Em particular, sua informação a respeito de S' , especificamente $H(S') - H(S'|G, U = u)$, é muito pequena. Mais precisamente, sua informação a respeito de S' decresce exponencialmente com a diminuição de $r - c$. Infelizmente o resultado do Corolário 2.1 não pode ser generalizado para $H_2(S|U)$, sendo válido para um valor particular de $U = u$.

Vamos agora obter um resultado que nos garantirá a segurança desejada, sem que para isso tenhamos que saber com exatidão a quantidade de informação que Eva possui de S .

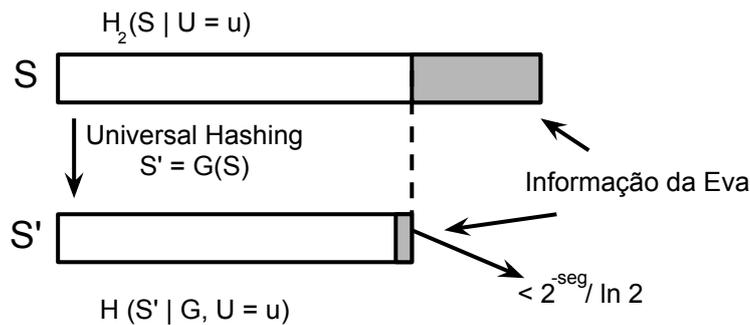


Figura 2.12: Privacy Amplification. (Fonte: [31])

Corolário 2.2 ([1]). *Considere S uma sequência de n bits com distribuição uniforme sobre $\{0, 1\}^n$. Seja $U = \varepsilon(S)$, onde ε é uma função arbitrária de Eva, tal que $\varepsilon : \{0, 1\}^n \rightarrow \{0, 1\}^t$, para um $t < n$. Considere $seg < n - t$ um valor de segurança positivo e $r = n - t - seg$. Se Alice e Bob escolhem $S' = G(S)$ como chave secreta, onde G é escolhido aleatoriamente de uma classe universal de funções hash que mapeiam S em $\{0, 1\}^r$, a informação de Eva sobre a chave secreta S' , dado G e U , satisfaz:*

$$I(S'; G, U) \leq \frac{2^{-seg}}{\ln 2}.$$

Note que, ao contrário do Corolário 2.1, este resultado considera a média de valores

de U e, além disso, a estratégia de Alice e Bob não depende da função ε utilizada por Eva, dependendo apenas de um limite superior de t .

Demonstração. Seja $u \in \{0, 1\}^t$ e c_u a quantidade de elementos $s \in \{0, 1\}^n$ que satisfazem $\varepsilon(s) = u$ (s é um possível valor tomado pela variável aleatória S). Dado $U = u$, todos os valores s tomados por S são equiprováveis, ou seja, $P_{S|U=u} = \frac{1}{c_u}, \forall s; \varepsilon(s) = u$. Assim, temos:

$$P_C(S|U = u) = c_u \cdot \left(\frac{1}{c_u}\right)^2 = \frac{1}{c_u}$$

e

$$H_2(S|U = u) = \log_2 c_u.$$

Utilizando o Corolário 2.1, obtemos:

$$H(S'|G, U = u) \geq r - \frac{2^{r - \log_2 c_u}}{\ln 2} \geq r - \frac{2^r}{c_u \cdot \ln 2}.$$

Utilizando $P_U(u) = c_u \cdot 2^{-n}$, vem:

$$I(S'; G, U) = H(S') - H(S'|G, U) \tag{2.11}$$

$$= r - \sum_{u \in \{0, 1\}^t} P_U(u) H(S'|G, U = u) \tag{2.12}$$

$$\leq r - \sum_{u \in \{0, 1\}^t} c_u \cdot 2^{-n} \cdot \left(r - \frac{2^r}{c_u \cdot \ln 2} \right) \tag{2.13}$$

$$= r(1 - c_u \cdot 2^{t-n}) + \frac{2^{-n+t+r}}{\ln 2} \tag{2.14}$$

$$= r(1 - c_u \cdot 2^{t-n}) + \frac{2^{-seg}}{\ln 2} \tag{2.15}$$

$$\leq \frac{2^{-seg}}{\ln 2}, \tag{2.16}$$

onde, em (2.15), assumimos que $c_u \geq 2^{n-t}$, pois c_u é a quantidade de elementos de um subconjunto de $\{0, 1\}^n$ que satisfaz $\varepsilon(s) = u$, $u \in \{0, 1\}^t$. \square

Aplicando o Corolário 2.2, podemos estabelecer um valor para r , baseados no tamanho n da sequência de *bits* estabelecida por Alice e Bob, da quantidade t de *bits* trocados publicamente nas fases de *Advantage Distillation* e *Information Reconciliation*, os quais são de conhecimento de Eva, e de um fator de segurança seg que eleva a segurança da chave estabelecida conforme aumenta.

seg	$I(S'; G, U)$
1	0,72134752
2	0,36067376
3	0,18033688
4	0,09016844
5	0,04508422
6	0,02254211
7	0,011271055
8	0,005635528
9	0,002817764
10	0,001408882

Tabela 2.3: Informação obtida por Eva de acordo com o parâmetro seg

Observe na Tabela 2.3 que para $seg = 10$, a informação obtida por Eva a respeito da chave criptográfica já é menor que 0,0015.

2.4 Estabelecendo a Comunicação

Até aqui trabalhamos com um sistema de troca de chaves criptográficas, faltando abordar como essas chaves serão utilizadas na transmissão das mensagens entre Alice e Bob.

No início do protocolo de estabelecimento de chave criptográfica, Alice e Bob tinham acesso a quantas realizações fossem necessárias de uma variável aleatória, X e Y , respectivamente, ou, interpretando de outra maneira, uma mesma variável aleatória que devido aos erros naturais da comunicação satelital pode apresentar divergências em uma mesma realização para diferentes receptores.

Alice e Bob iniciam a fase *Advantage Distillation* com sequências de *bits* semelhantes (cada *bit* correspondendo a uma realização das variáveis aleatórias) e terminam essa fase com duas sequências S_A e S_B parecidas, mas que ainda possuem algumas divergências em seus *bits*. Essas divergências são tratadas na fase *Information Reconciliation*, e ao final, tanto Alice quanto Bob possuem a mesma sequência S .

Em seguida, Alice e Bob, a fim de impedir que Eva entenda as mensagens que serão trocadas, aplicam a sequência em comum S à fase *Privacy Amplification*, obtendo uma sequência S' , menor que S , porém com segurança muito maior.

Conforme abordamos da seção anterior, S' é uma chave criptográfica com alta privacidade, pois Eva quase não possui informação sobre S' (lembre que a quantidade de informação de Eva sobre S' será limitada por um parâmetro de segurança s de acordo com a Tabela 2.3). Neste ponto, Alice e Bob podem utilizar o *One-Time Pad* com a chave criptográfica S' para realizarem a troca de mensagens com segurança incondicional.

Uma pergunta que surge naturalmente é o porquê desse modelo tornar viável a utilização do *One-Time Pad*. Esse modelo criou um protocolo para estabelecer chaves criptográficas de forma segura, as quais podem ter tamanho variado. Assim, Alice e Bob podem estabelecer quantas chaves forem necessárias, de acordo com as mensagens a serem transmitidas. Em geral, ao utilizarmos o *One-Time Pad*, precisamos definir anteriormente todas as chaves que serão utilizadas pelo sistema,

chaves que, conforme explicado anteriormente, só poderão ser utilizadas uma única vez.

Ou seja, esse modelo satelital torna a utilização do *One-Time Pad* viável pela dinamicidade que sua troca de chaves seguras confere ao sistema.

Capítulo 3

Outros Modelos

No capítulo anterior foi apresentado o modelo satelital para geração de chaves criptográficas altamente seguras, o qual viabiliza a utilização do *One-Time Pad* de maneira eficiente.

Neste capítulo apresentaremos dois modelos de comunicação segura que utilizam características do ambiente ao qual os dispositivos estão imersos para gerar uma chave criptográfica. O primeiro se baseia no ruído inerente da comunicação por radiofrequência, enquanto o segundo tem por base o som ambiente capturado por dispositivos próximos.

3.1 RFID Tags

A aplicação de *RFID Tags*¹ tem se tornado cada vez mais comum em nossa sociedade e sua utilização pode ser muito perigosa caso a comunicação envolvida nessa tecnologia não incorpore a devida segurança.

¹*Radio-Frequency IDentification Tags*: Etiquetas de Identificação por Radiofrequência (tradução nossa).

Exemplos comuns de aplicação do *RFID* são o comércio, com a possibilidade de reconhecimento dos produtos adquiridos pelos clientes, os pedágios de rodovias, em que a cobrança ocorre sem que o veículo tenha que ser parado, em pessoas, animais e equipamentos, de modo a viabilizar a sua localização, independente de recepção de sinais *GPS*², e em cartões para liberação de acesso a áreas restritas.

Em qualquer dessas aplicações, caso não seja implementada uma forma de comunicação segura, o risco de danos por má utilização se torna alto. Na liberação de acesso a áreas restritas, por exemplo, caso o utilizador tenha acesso à comunicação entre o transmissor e o receptor dos sinais do *RFID*, e consiga manipulá-la, este pode se passar por um usuário com acesso a todos os setores sensíveis de um banco ou de um órgão de segurança pública.

Para proteger essa comunicação, o modelo apresentado por Chabanne e Fumaroli no artigo [6], utiliza o ruído característico da comunicação por rádio frequência para gerar uma chave criptográfica entre emissor e receptor. Esse modelo, assim como no modelo satelital apresentado no capítulo anterior, utiliza as fases de *Advantage Distillation*, *Information Reconciliation* e *Privacy Amplification* para impedir que um participante não autorizado tenha sucesso em sua interceptação.

3.1.1 Funcionamento do Protocolo

O *RFID Tag* consiste de um circuito integrado acoplado a uma antena capaz de transmitir dados a alguns metros de distância em resposta a uma solicitação. Assim, uma central de controle envia um sinal questionando os dados de determinada etiqueta de identificação, a qual responde, enviando os dados solicitados.

O protocolo para estabelecer a chave criptográfica entre transmissor e receptor é

²Global Position System: Sistema de Posicionamento Global (tradução nossa).

muito parecido com o do modelo satelital, diferindo, basicamente, por dois fatores. O primeiro está relacionado a obtenção da sequência aleatória que dará início ao protocolo, pois são utilizados os ruídos característicos da comunicação por rádio frequência. O segundo se refere a forma como se dá a comunicação, que ocorre sempre pelo canal com ruídos, o qual é público. Fazendo uma analogia com o modelo satelital, é como se Alice (ou Bob) estivesse conectada fisicamente ao satélite, e com isso não haveria erro gerado por essa conexão (ver Figura 3.1).

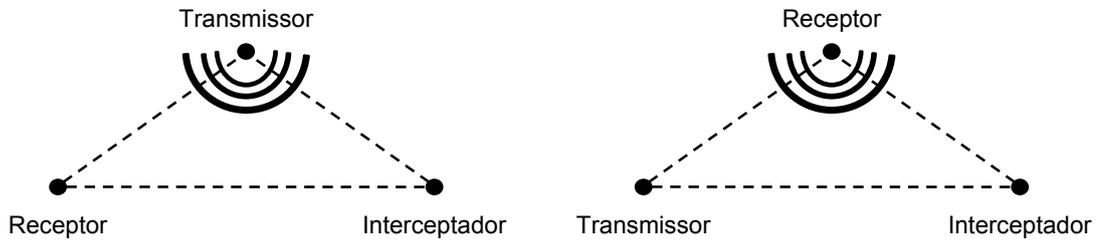


Figura 3.1: Cenário do RFID Tag

Considere \mathcal{R} o receptor, ou a etiqueta de identificação, \mathcal{T} o transmissor, ou a central de controle das etiquetas de identificação e \mathcal{I} um interceptador passivo. Suponha ainda que os canais de comunicação $\mathcal{R} \rightarrow \mathcal{T}$ e $\mathcal{R} \rightarrow \mathcal{I}$ sejam independentes.

A sequência S inicial terá que ser enviada por \mathcal{R} ou \mathcal{T} . Seja \mathcal{R} o primeiro a enviar a sequência, assim, $p_{\mathcal{R}} = 0$, $p_{\mathcal{T}} > 0$ e $p_{\mathcal{I}} > 0$. Assim como no modelo satelital, vamos considerar que o interceptador possui vantagens tecnológicas sobre quem está recebendo a sequência no momento, com isso, teremos $p_{\mathcal{T}} > p_{\mathcal{I}}$.

Para conferir vantagem à dupla transmissor/receptor, basta adotarmos a sequência recebida, pelo transmissor, a qual possui erros, como referência. As novas probabi-

lidades de erro serão:

$$\begin{aligned}
 p'_{\mathcal{T}} &= p_{\mathcal{R}} = 0 \\
 p'_{\mathcal{R}} &= p_{\mathcal{T}} \\
 p'_{\mathcal{I}} &= (1 - p_{\mathcal{T}})p_{\mathcal{I}} + p_{\mathcal{T}}(1 - p_{\mathcal{I}}) \\
 &= p_{\mathcal{T}} + p_{\mathcal{I}} - 2 \cdot p_{\mathcal{T}}p_{\mathcal{I}}
 \end{aligned}$$

Assim conseguimos deixar a dupla transmissor/receptor em situação de vantagem em relação ao interceptador, pois $p'_{\mathcal{I}} > p'_{\mathcal{R}}$, lembre que $0 < p_{\mathcal{I}} < p_{\mathcal{T}} < \frac{1}{2}$.

Em seguida a sequência S é aplicada às fases de *Information Reconciliation* e de *Privacy Amplification*, conforme o modelo satelital, obtendo-se ao final uma sequência extremamente segura de *bits*, que serão utilizados como chave criptográfica em um sistema criptográfico previamente escolhido pela dupla.

Chabanne e Fumaroli apresentam em [6], um exemplo da segurança obtida por esse modelo. Iniciam com $p_{\mathcal{T}} = 0,1$, $p_{\mathcal{I}} = 0,05$ e uma sequência de 600 *bits* enviada por \mathcal{R} . Com esses valores e o parâmetro de segurança $s = 7$, obtém-se uma sequência de 64 *bits* com elevada segurança, pois a informação obtida por \mathcal{I} está limitada superiormente por $\frac{2^{-7}}{\ln 2} \approx 0,0113$ *bit*.

3.2 Áudio Ambiente

A comercialização de *smartphones*, que têm a possibilidade de executar *softwares* desenvolvidos pelo próprio usuário, facilitou a realização de chamadas, troca de arquivos e troca de mensagens criptografadas a baixo custo.

Entretanto, um grande problema aparece ao se tentar sincronizar dois dispositivos: a possibilidade do pareamento ser realizado com dispositivo diverso daquele

que se pretendia. Esse problema pode ser resolvido com a incorporação de informações contextuais e sensoriais dos dispositivos.

No artigo *Using ambient audio in secure mobile phone communication* de Nguyen, Sigg, Huynh e Ji (referência [21]), os autores trabalham com um modelo capaz de gerar uma chave criptográfica a partir do som ambiente, sendo exigido apenas que os *smartphones* que participarão da comunicação estejam próximos, e assim sejam expostos a sons/sinais similares.

Os autores apresentam algumas dificuldades de se implementar este modelo na prática, o que não chega a prejudicar o modelo e muito menos diminuir as vantagens alcançadas com a proposta, a qual passamos a apresentar.

3.2.1 Funcionamento do Protocolo

Os dispositivos que irão partilhar uma chave criptográfica devem estar próximos e iniciar a captura de amostras de áudio através de seus microfones locais. Em seguida, cada dispositivo gera uma sequência de *bits* a partir das amostras capturadas, dando origem a uma “*impressão digital*”. Para gerar essa “*impressão digital*” são utilizadas várias características do sinal gravado, como os pontos em que o sinal atinge os maiores valores, as frequências desse sinal e as diferenças de energia entre amostras.

Essa sequência passa por um processo de correção de erros, dando origem a uma nova sequência S , pertencente a um conjunto \mathcal{X} , que poderá ser utilizada como chave criptográfica.

Durante a aplicação do modelo, os autores concluíram que os *smartphones* que vão participar da comunicação segura devem estar em uma esfera de 2 metros de diâmetro para conseguirem amostras de áudio similares o suficiente para a geração da chave criptográfica.

3.3 Sistematização dos Modelos

Os modelos apresentados durante o trabalho, apesar das diferentes abordagens, trazem muitas semelhanças, permitindo uma sistematização das etapas utilizadas para a geração de chaves criptográficas.

Inicialmente deve-se escolher uma fonte de *bits* aleatórios, os quais serão concatenados em uma sequência que servirá de base para a geração da chave criptográfica.

Em seguida, faz-se necessária uma fase que garanta vantagem aos participantes autênticos da comunicação, como a fase de *Advantage Distillation*. Dependendo do protocolo, essa fase pode ser dispensada, pois a vantagem já pode ser característica do sistema adotado.

Para garantir que a sequência dos comunicantes autênticos seja a mesma, utiliza-se uma fase de correção de erros, como a fase *Information Reconciliation*.

Finalizando a sistematização, é necessária uma fase que garanta que a chave criptográfica não seja conhecida por ninguém além dos participantes autênticos, como ocorre na fase *Privacy Amplification*.

Observamos que o modelo que utiliza áudio ambiente para geração de *bits* aleatórios não faz uso nem da fase de ampliação da vantagem e nem da fase de aumento da privacidade. Isso ocorre pelo fato de o modelo possuir um limite operacional que estabelece uma esfera de 2 metros de diâmetro, na qual devem estar todos os participantes autênticos. No artigo [21], os autores mostram que um participante fora desta esfera não consegue êxito na obtenção de informações da chave criptográfica gerada, pois terá acesso a amostras de áudios muito distintas das amostras dos participantes dentro da esfera.

Conclusão

Ao buscarmos sistemas criptográficos com segurança incondicional nos deparamos com o mais famoso, o *One-Time Pad*. Entretanto, conforme abordamos na Introdução e no Capítulo 1, a dificuldade de utilizarmos este sistema de criptografia vem do fato de o mesmo necessitar de chaves criptográficas de tamanho elevado e que sejam utilizadas uma única vez.

Esse empecilho poderia ser resolvido com a distribuição de “livros”, atualmente substituídos por arquivos com alguns gigabytes, com chaves de diversos tamanhos, mas isso tornaria o sistema muito oneroso e criaria uma falha crucial, pois caso esse “livro” fosse acessado por pessoa não autorizada, o sistema estaria vulnerável durante toda a utilização do “livro”.

Outra possível solução seria a utilização de um sistema criptográfico para a troca de chaves, como o RSA, mas nesse caso estaríamos diante de outro problema, a segurança da comunicação seria limitada pela segurança da troca de chaves, e o sistema estaria restrito à segurança computacional.

Assim, apresentamos alternativas para a geração de chaves criptográficas, impedindo que um personagem mal intencionado obtenha informações sobre as chaves criptográficas que o permitam decifrar a comunicação.

No Capítulo 2 analisamos o modelo satelital provando, com auxílio de conceitos

de Teoria da Informação, que é possível gerar chaves altamente seguras, viabilizando a utilização do *One-Time Pad*.

Em seguida, a fim de sistematizarmos as fases necessárias de um sistema de geração de chaves criptográficas com alta privacidade, foram abordados dois modelos, um baseado na utilização de ruídos inerentes à comunicação por rádio frequência e outro nos sons ambientes capturados por *smartphones* próximos.

Em geral não necessitamos de sistemas criptográficos com segurança incondicional, pois a segurança computacional, conferindo algumas décadas de segurança, pode ser suficiente. Entretanto existem situações nas quais não podemos abrir mão desse nível de segurança.

Uma situação que aborda a obrigatoriedade da segurança incondicional é apresentada por Jeroen van de Graaf no artigo *Voting with Unconditional Privacy by Merging Prêt-à-Voter and PunchScan*, referência [30]. O autor cria a seguinte hipótese: imagine que todos os cartões de voto da eleição de 1900 tivessem sido guardados e que, mesmo se utilizassem algum tipo de criptografia, fosse possível descobrir em quem Churchill votou, por exemplo, alguém ligado a um partido socialista ou até facista. Essa informação seria muito importante durante a 2ª Guerra Mundial, podendo, inclusive, mudar os rumos da guerra. O autor ainda aborda outra situação, na qual um ditador chega ao poder e passa a perseguir todos que votaram em candidatos de partidos da oposição.

Assim, considerando a importância do desenvolvimento de sistemas de criptografia que possibilitem atingir a segurança incondicional, buscaremos, em trabalhos futuros, a implementação do modelo satelital para geração de chaves criptográficas perfeitamente seguras. Com a implementação focaremos em quantificar o nível de segurança que se atinge partindo de distintos valores de parâmetros iniciais, sendo estes o tamanho inicial da sequência de *bits* que dará origem à chave

criptográfica e as taxas de erro dos três participantes da comunicação.

A exemplificação deste modelo tornará mais simples o seu entendimento. Além disso, a produção de tabelas com os resultados encontrados nas simulações servirá de base para outros pesquisadores que se interessem por desenvolver trabalhos nesta área.

Por fim, tentaremos manter um comparativo deste modelo com outros encontrados na literatura, de modo a facilitar a escolha do modelo mais adequado para implementação comercial de ferramenta que possibilite, de forma eficiente, a troca de informações criptografadas que atinjam a segurança incondicional.

Referências Bibliográficas

- [1] Bennett, C. H.; Brassard, G.; Crépeau, C.; Maurer, U. M. Generalized privacy amplification. *IEEE Transactions on Information Theory*, v. 41, n. 6, pp. 1915-1923, 1995.
- [2] _____; _____; Robert, J. M. Privacy Amplification by public discussion. *SIAM Journal on Computing*, v. 17, pp. 210-229, 1988.
- [3] Brassard, G.; Salvail, L. Secret-key reconciliation by public discussion. *Advances in Cryptology - EUROCRYPT '93: lecture notes in computer science*, v. 765, pp. 410-423, 1994.
- [4] Cachin, C. *Entropy measures and unconditional security in cryptography*. 1997. 144 f. Tese (Doutorado em Ciência da Computação) - ETH Zürich, Konstanz. 1997.
- [5] Castro, R. D.; Dahab, R.; Devegili, A. J. Introdução à Segurança Demonstrável. *VII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais*, pp. 103-152, 2007.
- [6] Chabanne, H.; Fumaroli, G. Noisy Cryptographic Protocols for Low-Cost RFID Tags. *IEEE Transactions on Information Theory*, v. 52, n. 8, pp. 3562-3566, 2006.

- [7] Csiszár, I.; Körner, J. Broadcast channels with confidential messages. *IEEE Transactions on Information Theory*, v. IT-24, pp. 339-348, 1978.
- [8] Cover, T. M.; Thomas, J. A. *Elements of Information Theory*. 2. ed. Massachusetts: Wiley-Interscience, ISBN 978-0-471-24195-9, 2006. 748 p.
- [9] James, Barry R. *Probabilidade: um curso em nível intermediário*. 3. ed. Rio de Janeiro: IMPA, ISBN 978-85-244-0101-5, 2008. 304 p. Coleção projeto Euclides, IMPA.
- [10] Felício, J. R. W. Fórmula de Stirling em Tempos de Maple. *Revista Matemática Universitária*, v. 17, pp. 40-51, 1994.
- [11] Kahn, D. *The Codebreakers: the story of secret writing*. New York: Scribner, ISBN 0-684-83130-9, 1996. 1181 p.
- [12] Lima, B. N. B.; Cioletti, L. M.; Terra Cunha, M. O.; Braga, G. A. *Entropia: introdução à Teoria Matemática da (des)Informação*. Notas para minicurso da II Bienal de Matemática da SBM. Salvador, 2004. Disponível em <http://www.mat.ufmg.br/~tcunha/entropia.pdf>. Acessado em 22 junho 2012.
- [13] Knuth, Donald E. *The art of computer programming: fundamental algorithms*, v. 1, 3. ed. Redwood City, USA: Addison Wesley Longman Publishing Co., ISBN 0-201-89683-4, 1997. 650 p.
- [14] Matos, A. B. *Tópicos Avançados em Algoritmos*. Porto, Portugal: Faculdade de Ciências da Universidade do Porto, 2008. Disponível em: <http://www.dcc.fc.up.pt/~acm/aulas/aa/t1.pdf>. Acessado em 20 de junho de 2012.

- [15] Maurer, U. M. Protocols for secret key agreement based on common information. *Advances in Cryptology - CRYPTO '92: lectures notes in computer science*, v. 740, pp. 461-470, 1993.
- [16] _____. Secret key agreement by public discussion from common information. *IEEE Transactions on Information Theory*, v. 39, n. 3, pp. 733-742, 1993.
- [17] _____; Wolf, S. The intrinsic conditional mutual information and perfect secrecy. *Information Theory. 1997. Proceedings, 1997 IEEE International Symposium on*, p. 88, 1997.
- [18] _____; Wolf, S. Towards characterizing when information-theoretic secret key agreement is possible. *Advances in Cryptology - ASIACRYPT '96: lectures notes in computer science*, v. 1163, pp. 196-209, 1996.
- [19] _____; Wolf, S. Unconditionally Secure Key Agreement and the Intrinsic Conditional Information. *Information Theory, IEEE Transactions on*, v. 45, n. 2, pp. 499-514, 1999.
- [20] Menezes, A. J.; Vanstone, S. A.; Oorschot, P. C. V. *Handbook of Applied Cryptography*. 1. ed. Boca Raton, USA: CRC Press, Inc. ISBN 0-8493-8523-7, 1996. 816 p.
- [21] Ngu Nguyen; Sigg, S.; Huynh, A.; Yusheng Ji. Using ambient audio in secure mobile phone communication. *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*, pp. 437-440, 2012.
- [22] Paar, C.; Pelzl, J. *Understanding Cryptography: a textbook for students and practitioners*. Bochum: Springer, ISBN 978-3-642-04100-6, 2010. 371 p.

- [23] Roman, S. *Coding and Information Theory*. California: Springer-Verlag, ISBN 0-387-97812-7, 1992. 486 p.
- [24] Schneier, B. *Applied Cryptography: protocols, algorithms and source code in C*. 2. ed. New York: John Wiley & Sons, ISBN 0-471-11709-9, 1996. 758 p.
- [25] Shannon, C. E. A mathematical theory of communication. *Bell System Technical Journal*, v. 27, pp. 379-423, 1948.
- [26] _____. Communication theory of secrecy system. *Bell System Technical Journal*, v. 28, pp. 656-715, 1949.
- [27] Singh, S. *The Code Book: the science of secrecy from ancient Egypt to quantum cryptography*. New York: Anchor Books, ISBN 978-0-385-49532-5, 2000. 410 p.
- [28] Stinson, D. R. Universal Hash Families and the Leftover Hash Lemma, and Applications to Cryptography and Computing. *Journal of Combinatorial Mathematics and Combinatorial Computing*, v. 42, pp. 3-31, 2002.
- [29] _____. Universal Hashing and authentication codes. *Advances in Cryptology - CRYPTO '91: Lecture Notes in Computer Science*, v. 576, pp. 74-85, 1992.
- [30] van de Graaf, J. Voting with Unconditional Privacy by Merging Prêt-à-Voter and PunchScan. *IEEE Transactions on Information Forensics and Security: especial issue on electronic voting*, v. 4, n. 4, pp. 674-684, 2009.
- [31] Wolf, S. Unconditional Security in Cryptography. *Lectures on Data Security*. Berlin: Springer Berlin Heidelberg, v. 1561, pp. 217-250, ISBN 978-3-540-65757-6, 1999.
- [32] Wyner, A. D. The wire-tap channel. *Bell System Technical Journal*, v. 54, n. 8, pp. 1355-1387, 1975.

- [33] Yeung, Raymond W. A new outlook on Shannon's information measures. *IEEE Transactions on Information Theory*, v. 37, n. 3, p. 466-474, 1991.