

1 Introdução

1.1 Motivação

Um problema comum que existe em um sistema de comunicação é o envio de informação através de um canal com apagamento, em que certos símbolos recebidos são distintos de todos os possíveis símbolos transmitidos (tais símbolos irreconhecíveis são ditos apagados). Um exemplo de um sistema de comunicação em que ocorre tal problema é a comunicação realizada por meio da Internet, na qual a troca de informação é feita através de pacotes.

Inicialmente, a estratégia mais utilizada para solucionar o problema de perda de pacotes é empregar um canal de retorno (*feedback channel*) entre o receptor e o transmissor, para controlar a retransmissão dos pacotes corrompidos ou perdidos. Este método de controle de erros na transmissão de dados é conhecido como ARQ (*Automatic Repeat Request*) [29]. Existem situações em que tal estratégia não é apropriada, como por exemplo a transmissão do tipo *multicast*, já que, requer-se um grande número de retransmissões para que os receptores sejam capazes de receber a informação original por completo [2, 5, 26]. Recentemente foi proposta uma solução para o problema de transmissão de informação através de canais com apagamento, usando uma classe códigos conhecida por “códigos fontanais” [7].

A primeira implementação prática de códigos fontanais inventada por M. Luby, é conhecida como códigos LT (Luby Transform) [6]. Códigos LT são capazes de recuperar, com probabilidade maior do que $(1 - \delta)$, um conjunto de k símbolos de entrada a partir de quaisquer $k + O(\sqrt{k} \ln^2(k/\delta))$ símbolos codificados recebidos — para isso uma média de $O(k \ln(k/\delta))$ operações¹ por símbolo são realizadas.

¹considera-se que uma operação corresponde a uma operação ou-exclusivo (*XOR*) ou uma operação que copia um símbolo de um registrador para outro.

Posteriormente, foram propostos os códigos Raptor por Shokrollahi [21], que são uma extensão de códigos LT. A idéia principal do código Raptor é primeiro codificar a mensagem original (símbolos de entrada) usando um código de bloco de taxa fixa (por exemplo, um código LDPC), e em seguida, codificar esses novos símbolos (símbolos intermediários) usando um código LT.

Códigos Fontanais são uma classe de códigos criados para a transmissão de dados de maneira eficiente através de canais com apagamento. Estes códigos são capazes de gerar um número ilimitado de símbolos codificados a partir dos dados de entrada, por esta razão tais códigos são denominados, na literatura inglesa, por “*rateless codes*” — nesta dissertação denominaremos tais códigos por códigos com taxa versátil [2, 6, 26].

1.2 Sistema de Comunicação

Em um sistema de comunicação, o objetivo consiste em transmitir uma mensagem através de um canal de comunicação, de modo que o receptor seja capaz de determinar esta mensagem com confiabilidade, diante das adversidades impostas pelo canal. No entanto, as comunicações reais são sempre afetados por ruído, que é sem dúvida um dos fatores básicos que limitam a taxa das comunicações.

A estratégia utilizada para garantir uma transmissão fidedigna de informação através de canais não ideais é codificar a informação a ser transmitida, inserindo redundância, para reduzir os efeitos do ruído sobre a mensagem original.

O modelo básico de sistemas de comunicações utilizado no presente trabalho, seguindo tal abordagem, encontra-se ilustrado na Figura 1.1.

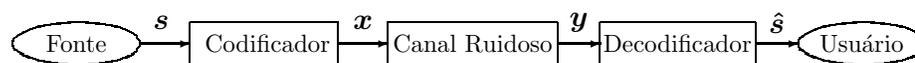


Figura 1.1: Diagrama em Blocos de um Sistema de comunicações.

Neste modelo, é suposto que a fonte utilizada é binária, e que portanto a mensagem \mathbf{s} que se deseja transmitir, consiste de um bloco de bits de tamanho k , representada pelo vetor:

$$\mathbf{s} = \left(s_1, s_2, \dots, s_k \right). \quad (1-1)$$

A mensagem \mathbf{s} é enviada a um codificador que adicionando redundância, a transforma em uma mensagem codificada \mathbf{x} que é representada pelo vetor:

$$\mathbf{x} = (x_1, x_2, \dots, x_n) \quad (1-2)$$

sendo esta enviada através do canal. Diz-se que a razão $R = k/n$ é a Taxa de Transmissão de Informação do codificador. A mensagem recebida \mathbf{y} à saída do canal, pode não ser uma réplica da mensagem codificada enviada \mathbf{x} , ou seja, $\mathbf{y} \neq \mathbf{x}$. O decodificador então usa a redundância introduzida pelo codificador para recuperar, desejavelmente sem erro, a mensagem original \mathbf{s} .

A teoria para o desenvolvimento de bons códigos corretores de erros foi introduzida por Claude Shannon, através do Teorema de Codificação de Canal, que garante ser possível realizar transmissões com probabilidades de erro arbitrariamente baixas, desde que a taxa de codificação seja menor que a capacidade do canal [20].

1.3 Modelo de canais com apagamentos

Canais com apagamento foram introduzidos por P. Elias [4], na qual um pacote é recebido sem erro ou é convertido em um apagamento.

Em um canal com apagamento, o conjunto $\mathcal{B} = \{a_0, \dots, a_{q-1}, a_q = ?\}$, de possíveis símbolos à saída do canal, possui um símbolo a mais do que o alfabeto de entrada $\mathcal{A} = \{a_0, \dots, a_{q-1}\}$. O alfabeto de saída \mathcal{B} é composto pelos “ q ” símbolos de \mathcal{A} e por um símbolo adicional que corresponde ao “símbolo apagado” denotado como “?”. É importante mencionar que existe uma probabilidade de transição que relaciona um símbolo de entrada e um símbolo de saída do canal, a qual é definida como

$$P(Y_j = a_m | X_j = a_i) = \begin{cases} 1 - P_a, & m = i \in \{0, q-1\} \\ P_a, & m = q \end{cases}$$

Um canal ruidoso (que introduz erro) usado com um bom código detetor de erro, pode ser aproximado por um canal com apagamento, ou seja, caso um pacote de informação recebido contenha erros detectáveis, pelo código detetor de erro em uso, o pacote é descartado e interpretado como apagado.

Um modelo de canal com apagamento formado por um alfabeto de

entrada não-binário de tamanho $q = 2^\ell$, onde ℓ é o número de bits de cada pacote que vai ser transmitido, gerado pela fonte de informação, é mostrado na Figura 1.2. Nela encontra-se ilustrado um canal quaternário ($q = 2^2$) com apagamento, o qual possui uma probabilidade $(1 - P_a)$ de transmitir o pacote de entrada sem erro, e uma probabilidade P_a de entregar ao receptor o símbolo “?”, o qual significa um apagamento.

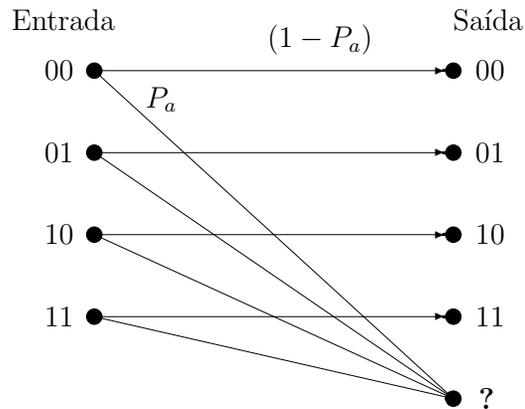


Figura 1.2: Canal quaternário com apagamento.

A capacidade de um canal com apagamento q-ário é:

$$C = (1 - P_a)\ell \quad \text{bits/uso-do-canal} \quad (1-3)$$

Os códigos de bloco que são tradicionalmente utilizados como códigos detetores de erros, para detectar símbolos não reconhecíveis, produzindo assim um canal com apagamento, são os códigos *RS* (Reed-Solomon) [16]. Mais recentemente surgiram os códigos Tornado [27] — estes, códigos de taxa fixa ($R = k/n$), apresentam severas limitações, por exemplo, k e n são limitados a valores razoavelmente pequenos. Além disso, ao contrário do que acontece com códigos fontanais, para o uso de códigos *RS* ou códigos Tornado deve-se estimar a probabilidade P_a de apagamento do canal e determinar a taxa R do código, antes que o processo de codificação tenha início [8].

1.4 Fontanas Digitais

O conceito de Fontana Digital (*Digital Fountain*) introduzido em [26] surgiu da intenção de se desenvolver um esquema de codificação no qual, a partir de um subconjunto qualquer de pacotes codificados com cardinalidade igual à do conjunto de pacotes que constituem os dados originais, o receptor

seja capaz de recuperar a mensagem originalmente transmitida. O nome fontana digital surgiu a partir da observação de que o codificador pode ser visto como um chafariz que jorra um número arbitrariamente grande de gotas de água (pacotes codificados²).

Uma definição de Fontana Digital Ideal é apresentada a seguir.

Definição 1.1 (Fontana Digital Ideal) (*Ideal Digital Fountain*)

Uma Fontana Digital Ideal é um esquema capaz de produzir, a partir de k símbolos de entrada \mathbf{s} , uma seqüência de símbolos codificados \mathbf{x} que obedece as seguintes propriedades [10],

1. O número de símbolos codificados denotado por $|\mathbf{x}|$ é um número arbitrariamente grande (possivelmente infinito).
2. Existe um receptor capaz de reconstruir a mensagem original \mathbf{s} de tamanho k , a partir de qualquer conjunto $\{x_{i_1}, x_{i_2}, \dots, x_{i_j}, \dots, x_{i_k}\}$ de k símbolos codificados. (Essa reconstrução deve ser rápida, de preferência linear em k). ■

Os códigos LT e códigos Raptor são chamados códigos fontanais, já que eles são uma boa aproximação para o conceito de fontanas digitais, no sentido que o número de pacotes necessários para uma decodificação bem sucedida aproxima-se do valor de k .

A seguir apresenta-se uma lista com vários métodos de transmissão de informação onde o uso dos códigos fontanais é recomendado. Estes exemplos [10] indicam a importância dos códigos fontanais que são o objeto de estudo desta dissertação.

Multicast. En uma transmissão *multicast*, uma fonte envia arquivos para um conjunto determinado de receptores. A principal vantagem de utilizar fontanas digitais neste tipo de transmissão é eliminar a necessidade de retransmissão. Por exemplo, se um usuário deixa de receber um determinado pacote, ele não precisa pedir retransmissão do mesmo, já que, qualquer conjunto de símbolos de saída, de cardinalidade igual à cardinalidade do arquivo original (conjunto de símbolos de entrada) pode ser

²Neste trabalho usaremos indistintamente os termos símbolos de saída, símbolos codificados e pacotes de saída. De forma análoga, os termos símbolos de entrada ou pacote de entrada

utilizado para recuperar o arquivo enviado.

Outra vantagem das fontanas digitais é que se torna trivial trabalhar com receptores operando a diferentes taxas de recepção. Por exemplo, se dois receptores que têm diferentes taxas de *download*, e compartilham um mesmo enlace — seqüência de *links*, exceto o último *link* do percurso de roteamento, os pacotes podem ser enviados à taxa mais alta ao longo do caminho, e são diminuídos no último *link*, ajustando-se assim à taxa do receptor mais lento. De maneira similar, é possível que dois receptores comecem a efetuar *downloads* em momentos diferentes, já que a ordem na qual são recebidos os pacotes não importa.

Transmissão de vídeo em tempo real. Para realizar uma transmissão em tempo real, a estratégia mais utilizada consiste em segmentar o arquivo original em várias partes. Enquanto a primeira parte é executada a segunda é baixada (*downloading*) e assim por diante. Para fazer uso desta estratégia deve-se ter em conta algumas considerações como por exemplo: o tamanho dos segmentos, a razão entre a taxa de *download* e a taxa de execução, etc.

Recepção paralela de arquivos. A recepção em paralelo de arquivos provenientes de varias fontes distintas é outra aplicação em que o uso de fontanas digitais é adequado. Nesta aplicação, cada fonte independentemente, pode produzir uma seqüência infinita de pacotes codificados. Como a disponibilidade de pacotes codificados é essencialmente infinita, a possibilidade de colisão (quando o mesmo pacote é enviado por múltiplas fontes) é nula. Além disso, o receptor pode encerrar a seção de recepção após receber o número de pacotes suficientes, sem preocupação alguma de onde os pacotes são provenientes.

1.5

Objetivos e organização do trabalho

Neste trabalho, apresentamos um estudo sobre o esquema de codificação fontanal, para canais com apagamento, conhecido como código LT. Outro código da família de códigos fontanais também abordado neste trabalho forma os códigos Raptor. O estudo inclui ainda a análise de diferentes distribuições de graus utilizadas por estes esquemas: a distribuição Sóliton Robusta, a distribuição Sóliton Robusta Melhorada, a distribuição Sóliton Robusta Truncada para códigos LT sistemáticos e a distribuição de Shokrollahi usada pelos códigos Raptor. É proposto também um esquema de códigos LT Bidimensionais com o intuito de encontrar uma implementação mais eficiente do que a

implementação dos códigos LT clássicos.

Esta dissertação encontra-se organizada em cinco capítulos e dois apêndices. As referências bibliográficas localizam-se imediatamente após o Capítulo 5. Em resumo, o conteúdo apresentado é o seguinte.

No Capítulo 2, apresentam-se os principais conceitos relacionados aos códigos LT, seu processo de codificação e decodificação. A forma de escolha dos parâmetros envolvidos na distribuição de graus Sóliton Robusta e análise de desempenho dos códigos LT em termos destes parâmetros são examinados.

No Capítulo 3 é analisado o processo de construção de uma boa distribuição de graus para os códigos LT clássicos, além de uma comparação de desempenho destes códigos quando a distribuição Sóliton Ideal e Sóliton Robusta são usadas.

No Capítulo 4 são apresentadas duas distribuições de graus recentemente propostas: as distribuições Sóliton Robusta Melhorada e Sóliton Robusta Truncada (esta última desenvolvida para uso com códigos LT sistemáticos), com intuito de melhorar ainda mais o desempenho dos códigos LT. Além disso, um modelo de códigos LT Bidimensionais é aqui introduzido.

No Capítulo 5, apresentamos as conclusões deste trabalho e algumas sugestões para trabalhos futuros.

Nos Apêndices A e B, apresenta-se a prova do Teorema 3.1 e um estudo detalhado sobre os códigos Raptor.