

ABORDAGEM ENTRÓPICA DE CODIFICAÇÃO DE FONTE USANDO COMPRESSÃO DE DADOS

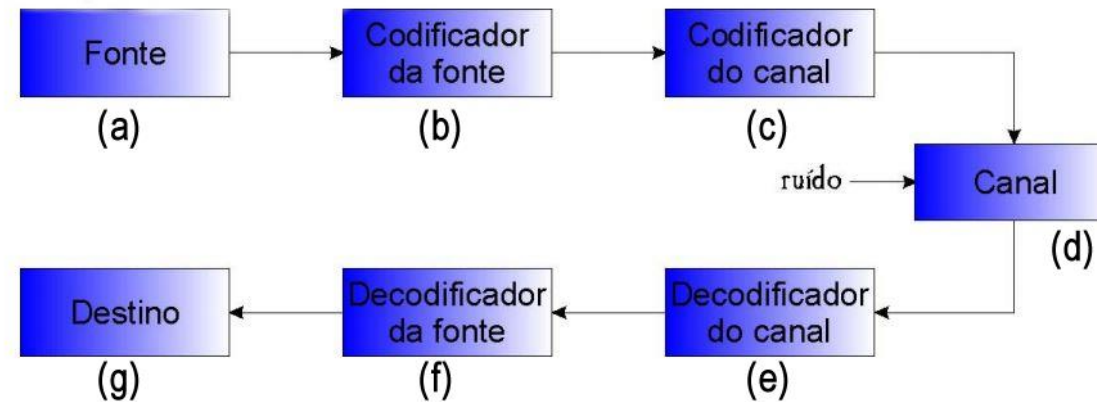
Mestrandos:

Kaio Gomes, Eduardo Sallum e Everton Skeika

INTRODUÇÃO

- Modelo de transmissão de dados.

Figura 1 - Processo de transmissão de dados



Fonte: Wolsky (2003)

INTRODUÇÃO

- A transmissão de dados em redes locais (*LAN*) seguindo o padrão *Ethernet*.
 - Codificação da fonte: *ASCII*.
 - Tamanho: 1 *byte* ou 8 *bits* para representando um determinado símbolo.

Figura 2 – Codificação ASCII

Binário	Carácter
0011 0000	0
0011 0001	1
0011 0010	2
0011 0011	3
0011 0100	4
0011 0101	5
0011 0110	6
0011 0111	7
0011 1000	8
0011 1001	9

Fonte: Autoria própria

INTRODUÇÃO

- A transmissão de dados em redes locais (*LAN*) seguindo o padrão *Ethernet*.
 - *PDU: frame ethernet* ou quadro *ethernet* (Pacote de dados).
 - Tamanho mínimo: 512 *bits* ou 64 *bytes*.
 - Tamanho máximo do frame: 12.144 *bits* ou 1518 *bytes*.
 - Tamanho mínimo destinado a dados: 46 *bytes*.
 - Tamanho máximo destinado a dados: 1500 *bytes*.
 - Modelo *TCP/IP* de comunicação: agregado na parte destinada a dados do *frame ethernet*, sendo 40 *bytes* de cabeçalho *TCP/IP* e no máximo 1460 *bytes* de dados.

INTRODUÇÃO

- O objetivo: propor uma abordagem para transmissão de dados em redes locais (Ethernet).
 - Aumentar a eficiência da transmissão de dados.
 - Em um mesmo pacote de dados transmitir mais informações.

ABORDAGEM PROPOSTA

- Características da abordagem entrópica de codificação de fonte proposta:
 - Algoritmo de compressão de dados;
 - Fonte e Destino precisam lidar com a compressão de dados.

ABORDAGEM PROPOSTA

- Codificação da fonte com compressão de dados, usando o algoritmo de Huffman:
 - Levantamento estatístico da frequência de ocorrência dos símbolos codificados;
 - A relação entre a frequência e o número de *bits* para representar um determinado símbolo;
 - Letras do alfabeto inglês e suas frequências (ZIM, 1962).

ABORDAGEM PROPOSTA

Tabela 1 – Frequência de letras parte 1

Símbolo	Frequência	Codificação
e	12,702%	100
t	9,056%	000
a	8,167%	1110
o	7,507%	1101
i	6,966%	1011
n	6,749%	1010
s	6,327%	0111
h	6,094%	0110
r	5,987%	0101
d	4,253%	11111
l	4,025%	11110
c	2,782%	01001
u	2,758%	01000

Fonte: Autoria própria

Tabela 2 – Frequência de letras parte 2

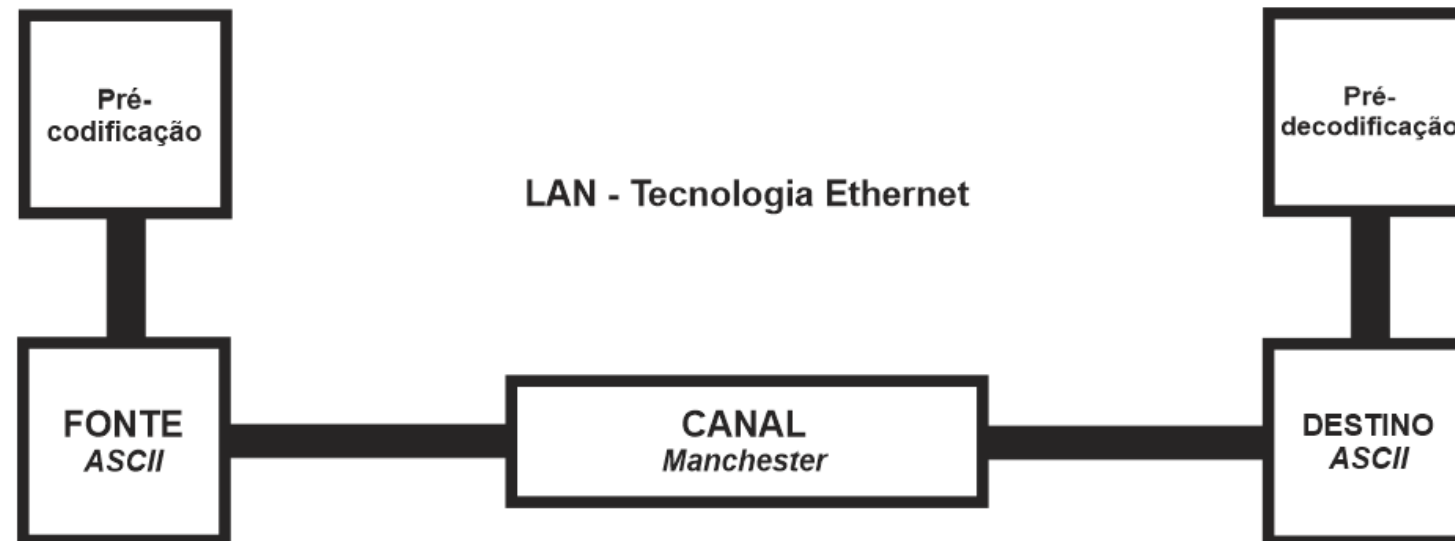
Símbolo	Frequência	Codificação
m	2,406%	00111
w	2,360%	00110
f	2,228%	00101
g	2,015%	110011
y	1,974%	110010
p	1,929%	110001
b	1,492%	110000
v	0,978%	001000
k	0,772%	0010011
j	0,153%	001001011
x	0,150%	001001010
q	0,095%	001001001
z	0,074%	001001000

Fonte: Autoria própria

ABORDAGEM PROPOSTA

- Funcionamento da abordagem proposta:

Figura 3 – Aplicação de pré-codificação e pré-decodificação



Fonte: Autoria própria

RESULTADOS

- Avaliação da abordagem entrópica com o uso do algoritmo de Huffman:
 - Comprimento médio de *bits* por símbolo: 4,21854 (4 *bits*);
 - Em cada pacote de dados, é possível enviar 1,89 vezes (2 vezes) mais símbolos do que com a codificação ASCII;
 - Eficiência da codificação: 99,19% (99%).

CONCLUSÃO

- Conhecendo a probabilidade de ocorrência de cada símbolo é possível aumentar a eficiência da codificação da fonte;
- Aplicação do algoritmo de Huffman para as letras do alfabeto inglês permitiu transmitir aproximadamente 2 vezes mais símbolos em um único pacote de dados.
- É possível enviar mais informações em um mesmo pacote de dados, sem alterar as tecnologias de redes LAN.

REFERÊNCIAS

- Kodituwakku, S. and Amarasinghe, U. (2010). Comparison of lossless data compression algorithms for text data. Indian journal of computer science and engineering, 1(4):416–425.
- Zim, H. S. (1962). Codes and secret writing (abridged edition). Scholastic Book Services.
- WOLSKY, E. Teoria da Informação. 2003. Disponível em: <https://www.ft.unicamp.br/~leobravo/TT%20081/apostila1.pdf>>. Acesso em: 18 set. 2019.