Pontifícia Universidade Católica
do Rio de Janeiro

Claudio Gomes de Mello

Codificação livre de prefixo para cripto-compressão

Tese de Doutorado

Tese apresentada ao Programa de Pós–graduação em Informática do Departamento de Informática da PUC–Rio como requisito parcial para obtenção Do título de Doutor em Informática

Orientador: Prof. Ruy Luiz Milidiú

Rio de Janeiro
Setembro de 2006

Pontifícia Universidade Católica
do Rio de Janeiro

Claudio Gomes de Mello

Codificação livre de prefixo para cripto-compressão

Tese apresentada ao Programa de Pós–graduação em Informática do Departamento de Informática do Centro Técnico Científico da PUC–Rio como requisito parcial para obtenção Do título de Doutor em Informática. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Ruy Luiz Milidiú
Orientador
Departamento de Informática — PUC–Rio

Prof. Weiler Alves Finamore
CETUC (Centro de Estudos em Telecomunicações da PUC-Rio)
— PUC–Rio

Prof. Marco Antonio Casanova
Departamento de Informática — PUC–Rio

Prof. Ricardo Dahab
Instituto de Computação — Unicamp

Prof. José Antônio Moreira Xexéo
Seção de Engenharia de Computação e Telemática — IME

Prof. José Engenio Leal
Coordenador Setorial do Centro Técnico Científico — PUC–Rio

Rio de Janeiro, 18 de Setembro de 2006

Claudio Gomes de Mello

Ficha Catalográfica

Mello, Claudio Gomes de

Codificação livre de prefixo para cripto-compressão / Claudio Gomes de Mello; orientador: Ruy Luiz Milidiú. — Rio de Janeiro : PUC–Rio, Departamento de Informática, 2006.

v., 110 f: il. ; 29,7 cm

1. Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática.

Inclui referências bibliográficas.

1. Informática – Tese. 2. Segurança da Informação. 3. Compressão de Dados. 4. Códigos de Prefixo. 5. Substituição Homofônica. I. Milidiú, Ruy Luiz. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Agradecimentos

Ao meu orientador Prof. Ruy Luiz Milidiú por todo o apoio, motivação e sabedoria. O Prof. Ruy soube guiar esse trabalho de pesquisa desde o seu início e foi essencial naqueles momentos em que o trabalho apresentava obstáculos para sua continuidade. Agradeço ao Prof. Ruy por ter passado sua experiência como pesquisador e por sua dedicação.

À minha esposa Dane Schemidt de Mello por seu apoio, sem o qual um trabalho de tanto tempo não teria como ser concluído.

Aos meus companheiros Zé Rodrigues e Carlos Eduardo que desenvolveram duas dissertações de mestrado na mesma área, contribuindo para a troca de experiências e para a evolução do trabalho.

Aos meus colegas da PUC–Rio: Frederico, Artur, Flavio, Raul, Lorenza, Duarte entre outros, que criaram um ótimo ambiente de trabalho.

Ao Exército Brasileiro pela oportunidade de cursar o doutorado, sem a qual este trabalho não existiria.

À Seção de Engenharia de Computação e Telemática (SE/8) do Instituto Militar de Engenharia (IME), pela compreensão e apoio, cedendo inclusive algumas horas de trabalho para que eu pudesse desenvolver esse trabalho.

Aos meus companheiros de IME pelo apoio total na reta final, em especial Salles, Luiz Henrique, Cecilio, Carla e Freire.

Resumo

Mello, Claudio Gomes de; Milidiú, Ruy Luiz. **Codificação livre de prefixo para cripto-compressão**. Rio de Janeiro, 2006. 110p. Tese de Doutorado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Cifragem e compressão de dados são funcionalidades essenciais quando dados digitais são armazenados ou transmitidos através de canais inseguros. Geralmente, duas operações sequenciais são aplicadas: primeiro, compressão de dados para economizar espaço de armazenamento e reduzir custos de transmissão, segundo, cifragem de dados para prover confidencialidade. Essa solução funciona bem para a maioria das aplicações, mas é necessário executar duas operações caras, e para acessar os dados, é necessário primeiro decifrar e depois descomprimir todo o texto cifrado para recuperar a informação. Neste trabalho são propostos algoritmos que realizam tanto compressão como cifragem de dados. A primeira contribuição desta tese é o algoritmo ADDNULLS — Inserção Seletiva de Nulos. Este algoritmo usa a técnica da esteganografia para esconder os símbolos codificados em símbolos falsos. É baseado na inserção seletiva de um número variável de símbolos nulos após os símbolos codificados. É mostrado que as perdas nas taxas de compressão são relativamente pequenas. A segunda contribuição desta tese é o algoritmo HHC — Huffman Homofônico-Canônico. Este algoritmo cria uma nova árvore homofônica baseada na árvore de Huffman canônica original para o texto de entrada. Os resultados dos experimentos são mostrados. A terceira contribuição desta tese é o algoritmo RHUFF — Huffman Randomizado. Este algoritmo é uma variante do algoritmo de Huffman que define um procedimento de cripto-compressão que aleatoriza a saída. O objetivo é gerar textos cifrados aleatórios como saída para obscurecer as redundâncias do texto original (confusão). O algoritmo possui uma função de permutação inicial, que dissipa a redundância do texto original pelo texto cifrado (difusão). A quarta contribuição desta tese é o algoritmo HSPC2 — Códigos de Prefixo baseados em Substituição Homofônica com 2 homofônicos. No processo de codificação, o algoritmo adiciona um bit de sufixo em alguns códigos. Uma chave secreta e uma taxa de homofônicos são parâmetros que controlam essa inserção. É mostrado que a quebra do HSPC2 é um problema NP-Completo.

Palavras–chave

Segurança da Informação. Compressão de Dados. Códigos de Prefixo. Substituição Homofônica.

Abstract

Mello, Claudio Gomes de; Milidiú, Ruy Luiz. **Crypto-compression prefix coding**. Rio de Janeiro, 2006. 110p. PhD Thesis — Department of Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Data compression and encryption are essential features when digital data is stored or transmitted over insecure channels. Usually, we apply two sequential operations: first, we apply data compression to save disk space and to reduce transmission costs, and second, data encryption to provide confidentiality. This solution works fine for most applications, but we have to execute two expensive operations, and if we want to access data, we must first decipher and then decompress the ciphertext to restore information. In this work we propose algorithms that achieve both compressed and encrypted data. The first contribution of this thesis is the algorithm ADDNULLS — Selective Addition of Nulls. This algorithm uses steganographic technique to hide the real symbols of the encoded text within fake ones. It is based on selective insertion of a variable number of null symbols after the real ones. It is shown that coding and decoding rates loss are small. The disadvantage is ciphertext expansion. The second contribution of this thesis is the algorithm HHC — Homophonic-Canonic Huffman. This algorithm creates a new homophonic tree based upon the original canonical Huffman tree for the input text. It is shown the results of the experiments. Adding security has not significantly decreased performance. The third contribution of this thesis is the algorithm RHUFF — Randomized Huffman. This algorithm is a variant of Huffman codes that defines a crypto-compression algorithm that randomizes output. The goal is to generate random ciphertexts as output to obscure the redundancies in the plaintext (confusion). The algorithm uses homophonic substitution, canonical Huffman codes and a secret key for ciphering. The secret key is based on an initial permutation function, which dissipates the redundancy of the plaintext over the ciphertext (diffusion). The fourth contribution of this thesis is the algorithm HSPC2 — Homophonic Substitution Prefix Codes with 2 homophones. It is proposed a provably secure algorithm by using a homophonic substitution algorithm and a key. In the encoding process, the HSPC2 function appends a one bit suffix to some codes. A secret key and a homophonic rate parameters control this appending. It is shown that breaking HSPC2 is an NP-Complete problem.

Keywords

Information Security. Data Compression. Prefix Codes. Homophonic Substitution.

Sumário

Lista de figuras

Lista de tabelas