1
Introdução

A cifragem e a compressão são requisitos essenciais para a transmissão e armazenamento de dados em meios digitais. A transferência digital de dados é um serviço amplamente utilizado atualmente, e o custo da utilização da rede para a transferência é proporcional à quantidade dos dados enviados. Logo, para a diminuição do volume de dados transferidos, são utilizadas técnicas de compressão. E, para garantir a segurança de um canal de dados, ou de uma mídia digital de armazenamento, são aplicadas técnicas criptográficas.

A proposta deste trabalho é apresentar algoritmos onde as funções de cifragem e compressão de dados são realizadas simultâneamente. Os algoritmos propostos são baseados nos códigos de Huffman Canônicos (22), na técnica de substituição homofônica (5, 11) e no uso de distribuições diádicas (11).

São apresentadas análises teóricas e experimentos práticos para avaliar o sigilo dos métodos propostos e o desempenho em relação às taxas de compressão e velocidade de codificação/decodificação.

O objetivo é alcançar eficiência computacional em comparação com soluções serializadas, tendo em vista a relação custo x benefício associada ao ganho de se manter as funcionalidades de indexação e busca, típicas de um sistema de recuperação de informação - Information Retrieval Systems (IRS), e que são perdidas não soluções comprime ⇒ cifra ⇒ decifra ⇒ descomprime. Ou seja, executar tão ou mais rápido, provendo mais funcionalidades.

O foco das aplicações que se beneficiam destes algoritmos é o armazenamento e distribuição eletrônica segura de grandes coleções textuais. As coleções são armazenadas em servidores e distribuídas em meio digital (ex: CD ou DVD) ou remotamente. O usuário legítimo da informação recebe uma senha de decodificação, através de um meio seguro, para poder ter acesso aos dados locais ou remotos codificados. Os testes realizados mostraram que a utilização de técnicas criptográficas não prejudicaram significativamente as taxas de compressão.

Nos capítulos 2, 3 e 4 alguns conceitos são definidos para serem utilizados durante esse trabalho. No capítulo 2 são definidos alguns conceitos básicos sobre teoria da informação como entropia, redundância e códigos de prefixo. No

capítulo 3 são apresentados alguns conceitos sobre segurança de sistemas que serão utilizados durante a análise da segurança dos algoritmos. E no capítulo 4 são introduzidos os códigos de Huffman e códigos de Huffman Canônicos.

Nos capítulos subsequentes, são apresentados os algoritmos desenvolvidos nessa tese. O capítulo 5 apresenta a idéia de algoritmos cripto-compressores e faz uma revisão de trabalhos correlatos.

A primeira contribuição desta tese é o algoritmo ADDNULLS — Inserção Seletiva de Nulos (13, 14), apresentado no capítulo 6. 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 (15), apresentado no capítulo 7. Este algoritmo cria uma nova árvore homofônica baseada na árvore de Huffman canônica original para o texto de entrada. Os resultados de alguns experimentos são mostrados.

A terceira contribuição desta tese é o algoritmo RHUFF — Huffman Randomizado (18), apresentado no capítulo 8. 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 (17, 19, 20, 21), apresentado no capítulo 9. 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.

Finalmente, no capítulo 10 são apresentadas as conclusões sobre esse trabalho e como esta pesquisa pode ser continuada.