SELECTIVE ENCRYPTION TEXT FILES WITH HUFFMAN CODING

VAN T. PHU - New Jersey Institute of Technology
Advisor: Prof. TOM LOOKABAUGH - University of Colorado - Boulder

Albert Dorman Honors College
NJIT NEW JERSEY INSTITUTE OF TECHNOLOGY

Abstract

Selective encryption is the technique of encrypting some parts of a compressed data file while leaving others unencrypted. Selective encryption is not a new idea. It has been proposed in several applications, especially in the commercial multimedia industry. However, selective encryption of losslessly compressed text files has not been explored, and that is the focus of our project. Through the project, we carefully studied how selective encryption can achieve a high level of effectiveness. By this, we mean a strategy in which even a small fraction of encrypted bits can cause a high ratio of damage to a file if an attacker attempts to decode it without decrypting the secured portions. In this project, we combined the encrypting and compressing processes to consider the choices of which types of bits are most effective in the selective encryption sense when they are changed. And so, instead of encrypting the whole file bit by bit, we changed only these highly sensitive bits. Moreover, by combining the compression and encryption tasks and reducing the total encryption work required, we can achieve a savings in system complexity.

Methods

Nest the encrypting process into the encoding process while compressing a data file.

1. Huffman coding algorithm

- Fix-to-variable data compression scheme that encodes data based on the frequency of occurrence of each character.
- Used to applied both compression and encryption.

2. Levenshtein distance algorithm

- A measurement of the difference between two strings by calculating the minimum number of Substitution, Deletion and Insertion operations to convert the source string to the target string.
- Used to measure the damage that the encryption process made to the file.

char	binary
'A'	0
'C'	100
'B'	101
'F'	1100
'E'	1101
'D'	111

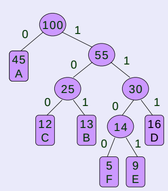


Figure 3: *Huffman binary tree* `

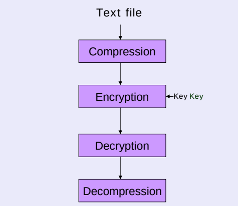


Figure 1*: Regular encryption scheme*

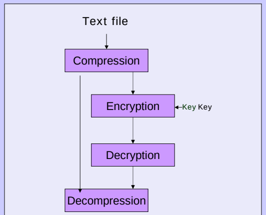


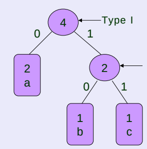
Figure 2*: Selective encryption scheme*

Hypothesis

Hypothesis:
Encrypting bits for some internal node choices are more effective (higher DSID per encrypted bit) than others.
Definition:
Efficiency = %damage / % encryption

Experiments

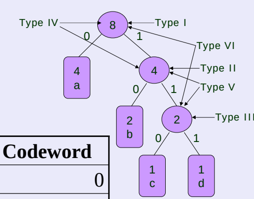
Simple case 1: 50% 'a', 25% 'b' and 25% 'c'.



Char	Codeword
a	0
b	10
c	11

Figure 4: *Binary tree for simplecase 1*

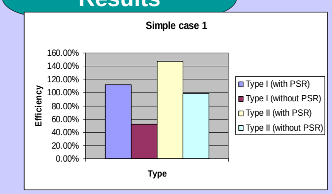
Simple case 2: 50% 'a', 25% 'b', 12.5% 'c' and 12.5% 'd'.

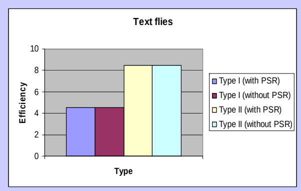


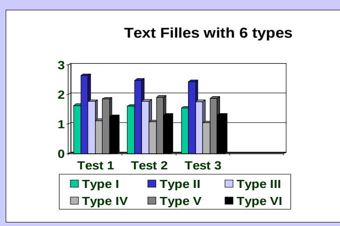
Char	Codeword
a	0
b	10
c	110
d	111

Figure 5: *Binary tree for simple case 2*

Results







	Test 1	Test 2	Test 3
Type I	1.63	1.61	1.54
Type II	2.64	2.49	2.44
Type III	1.77	1.78	1.76
Type IV	1.12	1.08	1.05
Type V	1.85	1.91	1.88
Type VI	1.25	1.30	1.31

Discussion

•!00% encryption does not guarantee 100% damage.

•Type II seems more efficient than others especially in real text cases.

•An error that is followed by another error in some cases would not result in the edit distance of two.

Ex: ab→ ba , aba→ba , abac→ bab
010→ 100 0100→1001 010011→ 100101
DSI 2 1 2
But bc→ cb , bcc→ cbb , bccb→ cbbc
DSID 2 3 3

•Spaces and other special character when being flipped would give "efficient" errors!

References

Ammeraal, Leendert. (1996). Algorithms and data structure in C++. New York, NY: John Wiley & Sons.

Berghel, Hal and Roach, David. (1996). An extension of Ukkonen's enhanced dynamic programming ASM algorithm. Retrieved on June 15, 2003 at: http://www.acm.org/~hlb/publications/asm/asm.html

Lelewer, Debra A. and Daniel S. Hirschberg. "Data Compression". AMC Computing Survey. Vol. 19, No. 3. September 1987.

Nelson, Mark. The Data Compression Book. M & T Publishing Inc.: NY. 1992.

Nelson, Mark. (1996). Priority Queues and the STL. Dr. *Dobb's journal*. Retrieved on June 16, 2003 at: http://www.dogma.net/markn/articles/pq_stl/priority.htm

Acknowledgments

Special thank to:
Professor Tom Lookabaugh.
REU Summer Program in Cyber Security at University of Colorado at Boulder.
Albert Dorman Honors College at NJIT
McNair program at NJIT.