

**ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**

NGUYỄN HỮU HIẾU

**THUẬT TOÁN NÉN TẬP THỰC THI BDC
VÀ ỨNG DỤNG**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Thái Nguyên - 2016

**ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**

NGUYỄN HỮU HIẾU

**THUẬT TOÁN NÉN TỆP THỰC THI BDC
VÀ ỨNG DỤNG**

**Chuyên ngành: Khoa học máy tính
Mã số: 60480101**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC

PGS.TS NGUYỄN VĂN TAM

Thái Nguyên – 2016

LỜI CAM ĐOAN

Tôi xin cam đoan rằng:

Số lượng và kết quả nghiên cứu trong luận văn này là hoàn toàn trung thực và chưa từng được sử dụng hoặc công bố trong bất cứ công trình nào khác.

Mọi sự giúp đỡ cho việc thực hiện luận văn này đã được cảm ơn và các thông tin trích dẫn trong luận văn đều được ghi rõ nguồn gốc.

Tác giả luận văn

Nguyễn Hữu Hiếu

LỜI CẢM ƠN

Lời đầu tiên học viên xin gửi lời cảm ơn trân thành tới các Thầy, Cô giáo trường Đại học Công nghệ thông tin và Truyền thông, Đại học Thái Nguyên, các thầy cô đã tận tình dạy bảo, truyền đạt các kiến thức quý báu cho học viên trong suốt quá trình học tập.

Xin trân trọng cảm ơn Khoa, Phòng, Ban và các cán bộ đã tạo điều kiện tốt nhất cho học viên học tập và hoàn thành đề tài tốt nghiệp của mình.

Đặc biệt, học viên xin được gửi lời cảm ơn sâu sắc đến Thầy giáo hướng dẫn học viên PGS. TS Nguyễn Văn Tam, Thầy đã tận tình chỉ bảo giúp đỡ học viên trong suốt quá trình nghiên cứu để hoàn thành luận văn.

Cuối cùng học viên xin gửi lời cảm ơn gia đình, bạn bè, đồng nghiệp đã giúp đỡ, động viên ủng hộ học viên rất nhiều trong toàn bộ quá trình học tập cũng như nghiên cứu để hoàn thành luận văn này.

Trân trọng cảm ơn!

MỤC LỤC

MỞ ĐẦU.....	1
1. Đặt vấn đề.....	1
2. Đối tượng và phạm vi nghiên cứu:.....	1
3. Ý nghĩa khoa học, thực tiễn của đề tài:.....	2
4. Phương pháp nghiên cứu:.....	2
5. Bố cục của luận văn:	2
CHƯƠNG I	3
GIỚI THIỆU CHUNG VỀ MỘT SỐ CÔNG NGHỆ NÉN.....	3
1.1. Tầm quan trọng của nén dữ liệu trong truyền tin.....	3
1.2. Nguyên tắc nén dữ liệu	4
1.3. Một số phương pháp nén dữ liệu	4
1.3.1. Phương pháp mã hóa độ dài loạt (Run - Length Encoding)	4
1.3.2. Phương pháp mã hóa Huffman	5
1.3.3. Phương pháp nén LZW	9
1.3.4. Chọn phương pháp nén	15
Kết luận:	16
CHƯƠNG 2.....	17
CÔNG NGHỆ NÉN DELTA.....	17
2.1. Giới thiệu về công nghệ nén Delta	17
2.1.1. Khái niệm về nén	17
2.1.2. Bộ nén Delta	18
2.1.3. Tính hiệu quả.....	18
2.2. Nền tảng	19
2.2.1. Nền tảng chung	19
2.2.2. Bộ nén LZ77- Nền tảng của bộ nén Delta	21

2.3. Thuật toán nén Delta	24
2.3.1. Nguyên lý của thuật toán nén Delta	24
2.3.2. Giới thiệu.....	25
2.3.3. Đặt vấn đề.....	27
2.3.4. Những nghiên cứu đầu tiên	28
2.3.5. Thuật toán cơ bản	30
2.3.6. Sự cải tiến và độ phức tạp của thuật toán	35
2.3.7. Xây dựng lại xâu đích	37
2.4. Giải pháp nén Delta dựa trên chữ ký	38
2.4.1. Thuật toán nén Delta dựa trên chữ ký.....	38
2.4.2. Đánh giá thuật toán qua thực thi rsync	40
2.5. Một số ứng dụng của thuật toán nén Delta	41
Kết luận:	44
CHƯƠNG 3.....	45
CHƯƠNG TRÌNH THỬ NGHIỆM.....	45
3.1. Bộ công cụ mã nguồn mở cho giải thuật nén Delta.....	45
3.1.1. Thư viện nén Delta: librsync.....	45
3.1.2. Hàm thư viện: rdiff.....	46
3.2. Xây dựng chương trình thử nghiệm.....	50
3.2.1. Mô hình thử nghiệm.....	50
3.2.2. Xây dựng chương trình	51
3.2.2.1. Module chương trình trên máy chủ	51
3.2.2.2. Module chương trình trên máy khách	53
3.2.3. Kịch bản thử nghiệm và đánh giá	54
Kết luận:	55
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	56
A. Kết luận:	56

B. Hướng phát triển.....	56
TÀI LIỆU THAM KHẢO.....	58
PHỤ LỤC.....	59

DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

Từ viết tắt	Tiếng anh
API	Application Programming Interface
BDC	Binary Delta Compression
COM	Component Object Model
DOS	Disk Operating System
HTTP	Hyper Text Transfer Protocol
LCS	Longgest common Subsequence
LZW	Lempel-Zip- Welch
MD5	Message Digest Algorithm 5
QAM	Quadrature Amplitude Modulation

DANH MỤC CÁC BẢNG

Bảng 1.1. Mô tả các trường hợp xuất hiện các ký hiệu khi nén.....	5
Bảng 1.2. Mô tả các thông tin với mã Huffman	7
Bảng 2.1. Các kết quả nén cho bộ dữ liệu gcc và emacs	24
Bảng 2.2. kết quả nén cho gcc và emacs bộ dữ liệu trong KB	41
Bảng 2.3. Kết quả nén cho emacs với kích thước block khác nhau trong KB ...	41
Bảng 3.1. Tên tệp và kích thước thử nghiệm trong chương trình.....	69

DANH MỤC CÁC HÌNH VẼ

Hình 1.1. Phương pháp nén RLE(1)	5
Hình 1.2. Phương pháp nén RLE(2)	5
Hình 1.3. Mã hóa Huffman	7
Hình 1.4. Mã hóa Huffman ngược	8
Hình 2.1. Bộ nén dữ liệu thông thường.....	17
Hình 2.2. Bộ nén Delta.....	18
Hình 2.3. Mô hình cập nhật Windows	43
Hình 3.1. Lưu đồ thuật toán cho thư viện hàm rdiff_sig	47
Hình 3.2. Lưu đồ thuật toán cho thư viện hàm rdiff_Delta	48
Hình 3.3. Lưu đồ thuật toán cho thư viện hàm rdiff_patch	49
Hình 3.4. Mô hình thử nghiệm.....	50
Hình 3.5. Chương trình trên server	51
Hình 3.6. Tạo tệp chữ ký bên máy chủ	52
Hình 3.7. Tạo tệp mới bên máy chủ.....	52
Hình 3.8. Tạo Delta giữa tệp cũ và mới bên chương trình máy chủ.....	53
Hình 3.9. Chương trình dành cho máy khách	53
Hình 3.10. Tạo tệp cập nhật bên máy khách.....	54
Hình 3.11. Sự đối lập của kích thước nén tập tin và sự giống nhau giữa tập tin.....	68
Hình 3.12. Sự đối lập giữa thời gian thực hiện và sự giống nhau của tập tin	69