

ĐỀ CƯƠNG MÔN HỌC

CTT410 – An Toàn và Phục Hồi Dữ Liệu

1. THÔNG TIN CHUNG

Tên môn học (tiếng Việt): *An Toàn và Phục Hồi Dữ Liệu*

Tên môn học (tiếng Anh): *Data Recovery and Safety*

Mã số môn học: *CTT410*

Thuộc khối kiến thức: *Chuyên ngành*

Số tín chỉ: *4*

Số tiết lý thuyết: *45*

Số tiết thực hành: *30*

Số tiết tự học: *90*

Các môn học tiên quyết *Cấu trúc dữ liệu & thuật toán, Hệ điều hành.*

GV phụ trách *Thái Hùng Văn – Email: thvan@fit.hcmus.edu.vn*

2. MÔ TẢ MÔN HỌC (COURSE DESCRIPTION)

Thông tin /dữ liệu có tầm quan trọng đặc biệt, có thể nói là quan trọng hơn cả các tài sản vật chất; nếu không tồn tại thông tin thì cũng có thể xem như không tồn tại cuộc sống. Và một điều cực kỳ nguy hiểm luôn có khả năng xảy ra trong thời đại số hiện nay là *hư hỏng mất mát dữ liệu*, thiệt hại xảy ra có thể ảnh hưởng nghiêm trọng đến các cá nhân, doanh nghiệp, các tổ chức lớn. Môn học trình bày các kiến thức cơ bản về tổ chức dữ liệu trên các loại thiết bị số và giới thiệu một số hình thức bảo vệ sự an toàn của dữ liệu cũng như những cách thức cơ bản để có thể phục hồi lại dữ liệu khi đã bị hư hỏng mất mát.

3. MỤC TIÊU MÔN HỌC (COURSE GOALS)

Sinh viên học xong môn học này có khả năng :

Mục tiêu	Mô tả (mức tổng quát)	CDR ` của chương trình
G1	Hiểu, biết cấu trúc và cơ chế lưu trữ, vận hành của các thiết bị chứa dữ liệu cơ bản.	
G2	Hiểu các hình thức tổ chức dữ liệu máy tính, từ các dữ liệu cơ bản đến các định dạng tập tin và hệ thống tập tin.	
G3	Khảo sát, thực nghiệm được các hình thức truy xuất dữ liệu trực tiếp trên thiết bị không qua hệ điều hành, thông qua đó thiết lập được một số hàng rào bảo mật và phục hồi một số trường hợp hư hỏng dữ liệu.	
G4	Xây dựng được các hàm /chương trình tự động phục hồi dữ liệu trong một số tình huống phổ thông.	
G5	Thiết kế và xây dựng được các hình thức tổ chức hệ thống tập tin chuyên biệt phục vụ cho mục đích an toàn dữ liệu.	

4. KẾ HOẠCH GIẢNG DẠY LÝ THUYẾT

STT	Tên chủ đề	Các thực nghiệm	Hoạt động dạy /học	Hoạt động đánh giá
1	Tổng quan về An toàn và Phục hồi Dữ liệu	TN#1.1 ~> #1.5	Thuyết giảng Bài tập nhóm	BTVN#1.1 ~> BTVN#1.6
2	Các thiết bị lưu trữ dữ liệu thông dụng – Những sự cố và cách khắc phục	TN#2.1 ~> #2.3	Thuyết giảng Bài tập nhóm	BTVN#2.1 ~> BTVN#2.5 ĐAMH#1, ĐAMH#2
3	Hệ thống tập tin và các kiến trúc tổ chức phổ biến	TN#3.1 ~> #3.4	Thuyết giảng Bài tập nhóm	BTVN#3.1 ~> BTVN#3.3 ĐAMH#3
4	Các hư hỏng logic trên hệ thống tập tin và cách khắc phục	TN#4.1 TN#4.2	Thuyết giảng Bài tập nhóm	BTVN#4.1 ~> BTVN#4.4
5	Phục hồi tập tin theo định dạng	TN#5.1 TN#5.2	Thuyết giảng Thảo luận nhóm	ĐAMH#4
6	Xây dựng các cơ chế bảo vệ dữ liệu	TN#6.1 TN#6.2	Thuyết giảng Bài tập nhóm	BTVN#6.1, BTVN#6.2
7	Thiết lập hệ thống tập tin chuyên biệt	TN#7.1 TN#7.2	Thuyết giảng Bài tập nhóm	BTVN#7.1 ĐAMH#5

5. NỘI DUNG MÔN HỌC

*** TÓM TẮT:**

(1) Các kiến thức về cấu trúc và cơ chế lưu trữ - vận hành của các thiết bị chứa dữ liệu cơ bản, các hình thức tổ chức dữ liệu máy tính và các định dạng cơ bản; (2) Ứng dụng phục hồi các dữ liệu máy tính bị hư hỏng mất mát; (3) Vận dụng /thiết lập được các hình thức bảo vệ & bảo mật các dữ liệu máy tính có sẵn, tổ chức được một hệ thống dữ liệu phù hợp với nhu cầu an toàn & bảo mật

*** CHI TIẾT:**

Chương I. Tổng quan

I.1 Thông tin và Dữ liệu

- Định nghĩa chung
- Dữ liệu máy tính
- Dữ liệu meta

I.2 Tác hại của việc hư hỏng dữ liệu

- Với cá nhân
- Với doanh nghiệp
- Với các tổ chức lớn

I.3 Khả năng xảy ra hư hỏng mất mát dữ liệu

- Các trường hợp có thể xảy ra
- Các nguyên nhân

I.4 Khả năng bảo vệ và phục hồi dữ liệu

- Khả năng bảo vệ dữ liệu
- Khả năng phục hồi dữ liệu

I.5 Các vấn đề chính trong bảo vệ và phục hồi dữ liệu

I.6 Một số hình thức bảo vệ thông tin /dữ liệu

- Mã hóa
- Che giấu, nhúng (ẩn) dữ liệu
- Đặt mật khẩu
- Sao lưu
- Giữ an toàn cho thiết bị

I.7 Một số hình thức phục hồi dữ liệu cơ bản

- Trên phần cứng

- Trên phần mềm

I.8 Các hình thức truy xuất trực tiếp dữ liệu

- Qua công cụ (Active Disk Editor /WinHex)
- Qua hình thức lập trình

Các Thực nghiệm :

TN#1.1	Làm « Hư » & Phục hồi lại đĩa (Active Disk Editor /Winhex)
TN#1.2	Bảo vệ dữ liệu bằng cách mã hóa (MS Office, WinRar)
TN#1.3	Thiết lập mật khẩu an toàn
TN#1.4	Kết nối thiết bị lưu trữ
TN#1.5	Phục hồi dữ liệu bằng các <i>Data Recovery Tool</i>

Các Bài tập :

BTVN#1.1	Khảo sát (qua Internet) và liệt kê ít nhất 5 tình huống hư hỏng mất mát thông tin có gây thiệt hại, cách ngăn ngừa /khắc phục tương ứng.
BTVN#1.2	Thực tập cứu các file đã xóa bằng các <i>Data Recovery Tool</i>
BTVN#1.3	Nghiên cứu sử dụng Active Disk Editor /WinHex và làm lại TN#1
BTVN#1.4	Sưu tầm, khảo sát và lập các thống kê so sánh sơ phác trên ít nhất 3 <i>Data Recovery Tool</i> (Getdataback Recovery, OnTrack EasyRecovery, Wondershare Data Recovery,...)
BTVN#1.5	Viết chương trình mã hóa nội dung file và chương trình giải mã tương ứng
BTVN#1.6	Viết chương trình khóa & mở Volume bằng cách mã hóa /giải mã một phần vùng hệ thống.

Các chủ đề mở rộng:

MR#1.1	Nghiên cứu và trình bày về Active Disk Editor /Winhex (tập trung cho phần hướng dẫn sử dụng)
MR#1.2	Nghiên cứu và trình bày về lập trình truy xuất sector (trên môi trường khác với môi trường đã được giới thiệu)
MR#1.3	Xác định các trục trặc có thể phát sinh khi lập trình truy xuất sector và cách giải quyết tương ứng
MR#1.4	Nghiên cứu và trình bày một phương pháp cứu dữ liệu cụ thể.
MR#1.5	Nghiên cứu và trình bày một phương pháp mã hóa dữ liệu khác với phương pháp XOR

Chương II. Các thiết bị lưu trữ phổ dụng – Những sự cố và cách khắc phục

II.1 Các thiết bị lưu trữ dữ liệu thông dụng

- Internal HDD
- External HDD

- SSD & SSHD
- USB flash
- Memory Card

II.2 Các hư hỏng trực trực phổ biến

- Trên Internal HDD
- Trên External HDD
- Trên SSD
- Trên USB flash và Memory Card

II.3 Các giải pháp khắc phục

- Trên Internal HDD
- Trên External HDD
- Trên SSD
- Trên USB flash và Memory Card

Các Thực nghiệm

TN#2.1	<i>Phân biệt chính xác các thiết bị trên, khảo sát các hình thức kết nối để truy xuất DL</i>
TN#2.2	<i>Vệ sinh các điểm tiếp xúc, trợ nguồn, kết nối trực tiếp</i>
TN#2.3	<i>Mở và khảo sát đầu đọc, mâm đĩa và các phụ kiện bên trong ổ cứng – cách thay thế</i>

Các Bài tập :

BTVN#2.1	<i>Lập bảng so sánh các đặc điểm chính của các thiết bị trên</i>
BTVN#2.2	<i>Lập bảng so sánh chi tiết HDD, SSD và SSHD</i>
BTVN#2.3	<i>Thí nghiệm mở /gỡ board HDD (đưa SV đĩa cứng để mang về làm)</i>
BTVN#2.4	<i>Nghiên cứu cách chuyển đổi qua lại giữa sector logic và sector vật lý</i>

Các Đồ án:

ĐAMH#1a	<i>Tìm hiểu và trình bày về tổ chức Partition trên đĩa. So sánh các ưu khuyết điểm giữa MBR và GPT</i>
ĐAMH#1b	<i>Viết chương trình truy xuất trực tiếp sector logic /vật lý bất kỳ, trình bày các thông tin lấy được từ bảng Parttion /BootSector và tạo Image cho đĩa cần cứu chữa (nếu có thể thì thực hiện giao diện tương tự WinHex và thiết kế tính năng tự chạy tiếp phần việc còn dở dang khi có sự cố gây gián đoạn – ví dụ như mất điện)</i>

Chương III. Hệ thống tập tin và các kiến trúc tổ chức phổ biến

III.1 Kiến trúc hệ thống tập tin tổng quát

- Khái niệm
- Thiết kế mô hình
- Tổ chức triển khai lên không gian lưu trữ
- Cài đặt các chức năng cơ bản

III.2 Kiến trúc hệ thống tập tin FAT

- Tổ chức Cluster
- Bảng quản lý cluster
- Bảng thư mục
- Boot Sector

III.3 Các kiến trúc hệ thống tập tin cao cấp

- Kiến trúc exFAT
- Kiến trúc NTFS
- Kiến trúc Ext

Các Thực nghiệm

TN#3.1	Sử dụng Active Disk Editor /Winhex xác định các thông số cơ bản của Volume ở định dạng FAT
TN#3.2	Xác định định dạng của Vol và các thông số của volume FAT32
TN#3.3	Đọc nội dung tập tin bằng cách truy xuất trực tiếp
TN#3.4	Xóa và Phục hồi tập tin bằng cách truy xuất trực tiếp

Các Bài tập :

BTVN#3.1	Viết chương trình xác định định dạng của Vol và các thông số tương ứng
BTVN#3.2	Viết chương trình lấy nội dung tập tin ở mức trực tiếp trên kiến trúc FAT
BTVN#3.3	Sử dụng Active Disk Editor /WinHex khảo sát & truy xuất tập tin trên kiến trúc NTFS /exFAT /ext

Các Đồ án:

ĐAMH#2	Viết chương trình phục hồi nội dung tập tin đã bị xóa trên kiến trúc FAT32
ĐAMH#3	Nghiên cứu & trình bày về 1 hệ thống tập tin đang /sẽ dùng nhiều. Phân tích / nhận xét các đặc điểm liên quan đến sự an toàn dữ liệu trên hệ thống đó.

Các chủ đề mở rộng:

MR#3.1	Khảo sát một công cụ cứu dữ liệu trên Linux, trình bày chi tiết một số trường hợp cụ thể
--------	--

MR#3.2	Nghiên cứu, trình bày và so sánh một số hệ thống tập tin tiên tiến
MR#3.3	Nghiên cứu và trình bày về tổ chức dữ liệu trên SmartPhone và các giải pháp bảo vệ /phục hồi.

Chương IV. Các hư hỏng logic trên hệ thống tập tin và cách khắc phục

IV.1 Các nguyên nhân

IV.2 Lỗi thông số mô tả

- Sai thông tin mô tả Partition
- Sai thông số Volume

IV.3 Lỗi trên các bảng quản lý

- Lỗi trên bảng quản lý cluster
- Lỗi trên bảng thư mục

IV.4 Lỗi do tác động bên ngoài

- Volume bị format
- Dữ liệu bị xóa
- Dữ liệu bị mã hóa

IV.5 Phục hồi tập tin theo định dạng

- Giới thiệu
- Các định dạng thông dụng
- Thiết lập các module phân loại khối dữ liệu
- Kết nối các khối dữ liệu
- Lọc các dữ liệu dư thừa

Các Thực nghiệm

TN#4.1	Sử dụng Winhex khảo sát bảng Partition /BootSector & phân tích các lỗi logic
TN#4.2	Khảo sát tình trạng Volume sau khi định dạng
TN#4.3	Khảo sát định dạng (header) của file .WAV
TN#4.4	Khảo sát dấu hiệu nhận biết các vị trí cluster bắt đầu dữ liệu và định dạng dữ liệu tương ứng

Các Bài tập :

BTVN#4.1	Viết chương trình tự xác định lỗi và tái tạo lại khi có 1 thông số định dạng bị sai trên BootSector
----------	---

BTVN#4.2	Viết chương trình tự xác định các lỗi và phục hồi lại khi có 1 loạt phần tử bị sai trên 2 bảng quản lý cluster theo dạng danh sách liên kết kết hợp chỉ mục (các phần tử sai không ở cùng vị trí)
BTVN#4.3	Tương tự BTVN#1.4, nhưng các thống kê so sánh ở mức chi tiết (tự tạo ra các tình huống lỗi)
BTVN#4.4	Viết chương trình tự xác định và phục hồi các lỗi về tổ chức dữ liệu trên Vol

Các Đồ án:

DAMH#4	Viết chương trình tự xác định và tái tạo lại nội dung của ít nhất 2 loại file đã không còn thông tin mô tả
--------	--

Chương V. Xây dựng các cơ chế bảo vệ dữ liệu

V.1 Sử dụng các hỗ trợ có sẵn

- Các hỗ trợ đăng nhập
- Các hỗ trợ mã hóa dữ liệu
- Các tường lửa

V.2 Xây dựng công cụ mã hóa

- Mã hóa với khóa công khai
- Mã hóa với mật khẩu “động”
- Khóa mềm và khóa cứng

V.3 Thiết lập các hệ thống chuyên biệt

- Giới thiệu
- Tổ chức khối dữ liệu và Bảng quản lý khối
- Tổ chức Bảng quản lý thư mục
- Tổ chức các thông số và thông tin quản lý
- Cài đặt các chức năng thiết yếu

V.4 Thiết lập không gian lưu trữ đặc biệt

- Tổ chức một hệ thống tập tin trong một tập tin
- Tổ chức một hệ thống tập tin trên nhiều thiết bị
- Tổ chức hệ thống tập tin nhúng trong các tập tin tin hiệu

Các Thực nghiệm

TN#5.1	Khảo sát hình thức mật khẩu « động »
TN#5.2	Khảo sát hình thức đưa một phần dữ liệu ra thiết bị Removeable

TN#5.3	Khảo sát sơ lược các file .ZIP, .ISO
TN#5.4	Khảo sát hình thức đưa một phần dữ liệu ra thiết bị Removeable

Các Bài tập :

BTVN#5.1	Viết chương trình đưa một phần dữ liệu ra thiết bị Removeable và lấy vào trở lại
BTVN#5.2	Khảo sát và trình bày về các secure dongle (khóa cứng)
BTVN#5.3	Thiết lập một mô hình hệ thống tập tin đơn giản theo mô tả về nhu cầu và cấu hình cho trước

Các Đồ án:

ĐAMH#5	Viết chương trình cho phép tạo một hệ thống tập tin trong một tập tin (như các file .ZIP, .ISO) và các chức năng cho phép lấy ra /đưa vào thêm các tập tin bên trong.
--------	---

Chương VI. An toàn và Phục hồi dữ liệu với công nghệ CDR

VI.1 Tổng quan về CDR

- Khái niệm
- Các ưu khuyết điểm

VI.2 An toàn dữ liệu

- Áp dụng trên các tập tin MS Office
- Áp dụng CDR trên các tập tin hình ảnh

VI.3 Phục hồi dữ liệu

- Áp dụng trên các tập tin MS Office
- Áp dụng CDR trên các tập tin hình ảnh

Các Thực nghiệm

TN#6.1	Khảo sát trên file MS Word
TN#6.2	Khảo sát trên file PNG

Các Bài tập :

BTVN#6.1	Viết chương trình loại bỏ các nội dung bất thường ra khỏi một file Office
BTVN#6.2	Viết chương trình loại bỏ các nội dung bất thường ra khỏi một file ảnh
BTVN#6.3	Viết chương trình phục hồi một số thông tin trong file Office (hình ảnh /văn bản)

- **Các Đề án:**

--	--

6. ĐÁNH GIÁ

Mã	Tên	Mô tả	Các chuẩn đầu ra	Tỉ lệ (%)
BTVN	Bài tập về nhà			20%
BTVN#1.5_6	VCT tích hợp 2 chức năng ứng với BTVN #1.5 và #1.6 (mã hóa / giải mã nội dung file, khóa & mở Volume bằng cách mã hóa /giải mã một phần vùng hệ thống)			6%
BTVN#3.1_2	VCT xác định định dạng của Vol và các thông số định dạng tương ứng, lấy nội dung tập tin ở mức trực tiếp trên kiến trúc FAT (16/32)			6%
BTVN#4.3	Khảo sát các chức năng và lập các thống kê so sánh 1 số Data Recovery Tool.			4%
BTVN#6.1	VCT dời 1 phần dữ liệu ra thiết bị khác và lấy vào trở lại			4%
DAMH	Đề án môn học			40%
DAMH#1a	Tìm hiểu và trình bày về tổ chức Partition trên đĩa. So sánh các ưu khuyết điểm giữa MBR và GPT. Viết chương trình truy xuất trực tiếp sector logic /vật lý bất kỳ, trình bày các thông tin lấy được từ bảng Parttion /BootSector và tạo Image cho đĩa cần cứu chữa (chương trình cần có khả năng tự chạy tiếp phần việc còn dở dang khi có sự cố gây gián đoạn – ví dụ như mất điện)	Nhóm tối đa 2SV		8%
DAMH#2	Viết chương trình phục hồi nội dung tập tin đã bị xóa	Nhóm tối đa 2SV		8%
DAMH#3	Nghiên cứu & trình bày về 1 hệ thống tập tin đang /sẽ dùng nhiều. Phân tích / nhận xét các đặc điểm liên quan đến sự an toàn dữ liệu trên hệ thống đó.	Nhóm tối đa 2SV		8%

DAMH#4	Viết chương trình tự xác định và tái tạo lại nội dung của ít nhất 2 loại file đã không còn thông tin mô tả	Nhóm tối đa 3SV		8%
DAMH#5	Viết chương trình cho phép tạo một hệ thống tập tin trong một tập tin (như các file .ZIP, .ISO) và các chức năng cho phép lấy ra /đưa vào thêm các tập tin bên trong.	Nhóm tối đa 3SV		8%
LTCK	Thi lý thuyết cuối kỳ	Tự luận		40%

7. TÀI NGUYÊN MÔN HỌC

Bài giảng **An toàn và Phục hồi Dữ liệu**, THÁI HÙNG VĂN, Khoa CNTT, ĐH KHTN

Diễn đàn môn học, các tài liệu cần thiết: <https://www.fb.com/groups/CuuDL> (SV cần đăng ký)

Tài liệu tham khảo

1. MSDN, <http://msdn.microsoft.com>
2. NTFS, FAT, <http://ntfs.com/>
3. Undelete a file in NTFS, <http://www.codeproject.com/file/NTFSUndelete.asp>
4. File Format, <http://www.wotsit.org>
5. Advanced System Format, <http://penta.zaniah.net/index.php?asf>
6. Linux-NTFS, <http://www.linux-ntfs.org/content/view/104/43/>
7. JPEG Header Format, http://www.fastgraph.com/help/jpeg_header_format.html
8. WAVE PCM soundfile format, <http://ccrma.stanford.edu/courses/422/projects/WaveFormat/>
9. ID3 tag, <http://id3.org>
10. Zip format, www.pkware.com/index.php?option=com_content&task=view&id=64&Itemid=107

Tài nguyên khác:

Phần mềm Active Disk Editor /WinHex /Hex Edit, Total Commander, Visual Studio,

Các phần mềm cứu dữ liệu (GetDataBack, Active@ File Recovery, Ontrack EasyRecovery, Recovery My Files, ..), tạo Image, backup

HDD /Flash disk phục vụ cho việc thí nghiệm



CÁC QUY ĐỊNH CHUNG

- Sinh viên cần tuân thủ nghiêm túc các nội quy và quy định của Khoa và Trường.
- Sinh viên không được vắng quá 50% số tiết học lý thuyết.

GV biên soạn

Thái Hùng Văn