Chương 1

THÔNG TIN VÀ XỬ LÝ THÔNG TIN

1.1. THÔNG TIN VÀ XỬ LÝ THÔNG TIN

1.1.1. Khái niệm về thông tin (Informations)

Đối với chúng ta, thông tin là một khái niệm rất quen thuộc và được sử dụng hàng ngày. Thông tin giúp con người hiểu tốt hơn về đối tượng mà mình quan tâm, nhằm đưa ra các giải pháp giải quyết vấn đề một cách tối ưu trong những điều kiện có thể. Trong tin học, thông tin được định nghĩa đơn giản như sau: *Thông tin là tất cả những gì đem lại hiểu biết, nhận thức cho con người về đối tượng mình quan tâm.*

Bản thân thông tin không phải là một thực thể vật chất, nghĩa là không mang năng lượng nội tại, nhưng khi được tham gia vào các quá trình hoạt động của con người, nó lại thể hiện khả năng vật chất của mình, tức là góp phần làm gia tăng năng lượng, mà với sự hỗ trợ của công nghệ thông tin và kỹ thuật truyền thông hiện đại, độ gia tăng đó ngày càng trở nên đáng kể, thậm chí còn vượt xa dự kiến của con người.

Thông tin mang lại cho chúng ta sự hiểu biết, làm tăng kiến thức (do vật mang tin đem lại) về một sự vật, một hiện tượng hay một quá trình nào đó. Vật mang tin có thể có nhiều dạng thức khác nhau song lượng kiến thức mà nó hàm chứa lại là không đổi. Giá trị của lượng thông tin nhận được còn phụ thuộc trực tiếp vào đối tượng nhận tin. Cụ thể là trước cùng một tin, không phải đối tượng nhận tin nào cũng thu được một lượng thông tin như nhau; kết quả còn phụ thuộc vào nhiều yếu tố liên quan đến tin đó.

Thông tin luôn tồn tại khách quan. Thông tin có thể tạo ra, phát sinh, truyền đi, lưu trữ, chọn lọc. Thông tin cũng có thể bị méo mó, sai lệch do nhiễu hay do người xuyên tạc.

Thông tin không đồng nhất với tín hiệu. Cùng một thông tin có thể được thể hiện bằng các tín hiệu khác nhau. Ví dụ: Tiếng đại bác có thể là tín hiệu chiến tranh, nhưng cũng có thể là tín hiệu chào mừng một nguyên thủ quốc gia.

1.1.2. Dữ liệu (data)

Dữ liệu sau khi tập hợp lại và xử lý sẽ cho ta thông tin. Nói cách khác dữ liệu là nguồn gốc, là vật mang thông tin, là vật liệu sản xuất ra tin. Thông tin chứa đựng ý nghĩa còn dữ liệu là các dữ kiện không có cấu trúc và không có ý nghĩa rõ ràng nếu không được tổ chức và xử lý. Cùng một thông tin có thể được biểu diễn bằng những dữ liệu khác nhau. Dữ liệu có nguồn gốc tự nhiên không có tính quy ước còn dữ liệu do con người tự đặt ra để mã hóa thông tin thì phải có quy ước về cách biểu diễn thông tin.

Ví dụ: cùng biểu diễn một đơn vị nhưng trong hệ thập phân số một được biểu diễn bởi 1, còn trong hệ đếm La-mã thì lại dùng ký hiệu I.

Mỗi dữ liệu có thể được thể hiện bằng những tín hiệu vật lý khác nhau. Cũng là gật đầu, đối với nhiều nước trên thế giới là tín hiệu thể hiện sự đồng tình, nhưng ngược lại đối với người Hy Lạp gật đầu để biểu lộ sự bất đồng.

Mỗi tín hiệu có thể dùng để thể hiện các thông tin khác nhau. Nâng ly rượu, chúc mừng ngày vui của lứa đôi nhưng cũng có thể để giã từ, gợi nhớ mối tình một thuở.

Dữ liệu có nhiều dạng khác nhau, có thể là:

- Tín hiệu vật lý (physical signal): tín hiệu điện, tín hiệu sóng điện-từ, tín hiệu ánh sáng, tín hiệu âm thanh, nhiệt độ, áp suất...
- Các số liệu (number) là dữ liệu bằng số mà ta đã quen với tên gọi số liệu. Đó là các số liệu trong các biểu bảng tính toán, thống kê, tài chính...
- Các ký
  hiệu khắ
- Văn bản,
  thông tư,
- Âm than
- Hình ảnh
- Đồ họa (
- ....

Với sự bùn
Người ta ước tí
tăng gấp đôi. Y
nay là việc biết
nghĩa là từ các
ẩn có giá trị mà
phát triển và nh
việc đó chính là
liệu (*Knowledge*
cho phép ta lấy
(*Datamining*).

1.1.3. Độ đo thô

Thông tin
Mỗi đối tượng c
Tính bất định n
tượng đó. Độ bấ
quan chặt chẽ đ
của sự kiện (biế
của nó có giá tr
suất của nó bằng

Người ta có
hành vi, trạng th
càng cao. Năm
bất định của sự

- Các ký hiệu (symbol) như các chữ viết (character) và các ký hiệu khắc trên tre, nứa, đá, trên bia, trên vách núi...
- Văn bản, chữ viết (text, character): sách báo, truyện, thông báo, thông tư, công văn...
- Âm thanh (sound): tiếng nói, âm nhạc, tiếng ồn
- Hình ảnh (image): phim ảnh, tivi, camera, tranh vẽ
- Đồ họa (graphic)
- ....

Với sự bùng nổ của thông tin, núi dữ liệu ngày càng khổng lồ. Người ta ước tính: lượng thông tin trên thế giới cứ sau 20 tháng lại tăng gấp đôi. Yếu tố thành công trong mọi lĩnh vực hoạt động ngày nay là việc biết sử dụng thông tin một cách có hiệu quả. Điều đó có nghĩa là từ các dữ liệu sẵn có, phải tìm ra được những thông tin tiềm ẩn có giá trị mà trước đó chưa được phát hiện, tìm ra những xu hướng phát triển và những yếu tố tác động lên chúng. Thực hiện các công việc đó chính là thực hiện quá trình phát hiện tri thức trong cơ sở dữ liệu (*Knowledge Discovery in Database - KDD*) mà trong đó kỹ thuật cho phép ta lấy được các tri thức chính là kỹ thuật khai phá dữ liệu (*Datamining*).

1.1.3. Độ đo thông tin

Thông tin có liên quan chặt chẽ đến khái niệm về độ bất định. Mỗi đối tượng chưa xác định hoàn toàn đều có một độ bất định nào đó. Tính bất định này chưa cho biết một cách đích xác và đầy đủ về đối tượng đó. Độ bất định sẽ giảm đi khi nhận thêm thông tin và có liên quan chặt chẽ đến khái niệm xác suất - độ đo khả năng có thể xảy ra của sự kiện (biến cố). Nếu một biến cố không bao giờ xảy ra, xác suất của nó có giá trị bằng 0. Nếu có một biến cố chắc chắn xảy ra, xác suất của nó bằng 1. Đại lượng xác suất có giá trị đoạn [0,1].

Người ta có thể định lượng tin tức bằng cách đo độ bất định của hành vi, trạng thái. Xác suất xuất hiện của tin càng thấp thì lượng tin càng cao. Năm 1948, C. Shannon đưa ra công thức sau để xác định độ bất định của sự kiện (tính lượng tin) gọi là Entropi:

$$H = \sum_{i=1}^{n} p_i \log_2(p_i)$$

Trong đó P_i là xác suất xuất hiện sự kiện i của hệ, hệ có n khả năng khác nhau. Ví dụ: khi gieo xấp ngửa đồng xu, lượng tin xuất hiện mặt ngửa sẽ là:

$$H = - \log_2(1/2) = 1$$

Tin tức chỉ xuất hiện khi tối thiểu có hai trạng thái, hai khả năng. Do vậy trường hợp hai trạng thái (xấp, ngửa của đồng xu) đã cho ta đơn vị đo thông tin H = 1 bit. Lượng thông tin chứa trong một bit là vừa đủ để nhận biết một trong hai trạng thái có xác suất xuất hiện như nhau. Tại mỗi thời điểm trong một bit chỉ lưu giữ được hoặc là chữ số 0 hoặc là chữ số 1. Từ bit là từ viết tắt của Binary Digit (chữ số nhị phân). Trong tin học ta thường dùng một số đơn vị bội của bit như bảng 1.1.

Bảng 1.1

Tên gọi	Viết tắt	Giá trị
Byte	B	8 bit
Kilobyte	KB	1024 bytes = 2^{10}B
Megabyte	MB	1024 KB = 2^{10}KB
Gigabyte	GB	1024MB = 2^{10}MB

1.1.4. Mã hóa thông tin

Trong xử lý thông tin tự động, dạng mã quan trọng được dùng là dạng mã nhị phân. Thông tin được mã hóa trên bảng chữ cái gồm hai ký hiệu là chữ số 0 và chữ số 1. Ví dụ, với bảng chữ mã chuẩn của Mỹ ASCII (*American Standard Code for Information Interchange*), mỗi ký tự (chữ cái, chữ số, dấu, ký hiệu đặc biệt) tương ứng với một mã 8 bit (dãy liên tiếp 8 chữ số 0 và 1).

Toàn bộ bảng mã ASCII có 128 ký hiệu được mã hóa, trong đó có:

- 26 chữ cái La-tinh in thường (a, b, c,..., x, y, z)
- 26 chữ cái La-tinh in hoa (A, B, C,...,X, Y, Z)

- 10 chữ số thập phân
- 10 ký tự toán học thông dụng (+, -, *, /, =, >, <)
- Các dấu chính tả (?,..[,], {,}, %, #, &, $,;,...)
- Một số ký hiệu điều khiển, ví dụ CR dùng để điều khiển máy in chuyển đến cột đầu tiên của dòng in.

Bảng 1.2

Ký tự	Mã ASCII	Ký tự	Mã ASCII
0	0011 0000	A	0100 0001
2	0011 0010	B	0100 0010
3	0011 0011	Y	0101 1001
4	0011 0100	Z	0101 1010
5	0011 0101	...	...
6	0011 0110	a	0110 0001
7	0011 0111	b	0110 0010

Bảng mã ASCII mở rộng có 256 ký tự được mã hóa. Mỗi mã của ký tự là một từ 8 bit. Mỗi mã ký tự được lưu trữ trong một byte. Mỗi nước có một bảng mã riêng. Thường các bảng mã đó bao gồm ASCII và phần mở rộng dùng để cài đặt các ký tự riêng của mỗi nước (bảng 1.2).

Nhận xét:

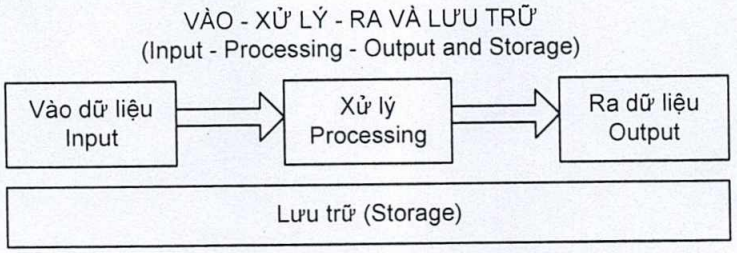
- Chữ cái hoa và chữ thường chênh nhau một khoảng là 32. Mã chữ cái 'A' là 65 thì mã chữ cái 'a' là 97 = 65 + 32.
- Cụ thể là mã nhị phân cho chữ cái:
- 'A' sẽ là 40H = 0100 0000
- 'a' sẽ là 61H = 0110 0001
- Còn chữ số '0' sẽ có mã là 30H = 0011 0000
- Các mã từ 0 đến 31 được dùng để mã hóa các ký tự điều khiển. Mã số 127 cũng là mã điều khiển, không biểu diễn ký tự đọc được.
- *Ví dụ:* Mã 7, BELL: tạo ra tiếng chuông

- Mã 27, ESC (thoát khỏi): có nhiều tác dụng được quy ước cụ thể trong từng trường hợp...

1.1.5. Chu trình xử lý thông tin

Quá trình xử lý thông tin bằng máy tính hay bằng con người có thể được minh họa theo qui trình:

Trước tiên ta phải đưa dữ liệu đầu vào nào đó vào máy tính thông qua các thiết bị vào như bàn phím, máy quét ảnh, sau đó máy tính hoặc bản thân chúng ta sẽ thực hiện quá trình xử lý để lấy ra thông tin. Thông tin ra sẽ được đưa ra dưới dạng dữ liệu ra thông qua màn hình, máy in, máy vẽ... Ngoài ra dữ liệu vào và ra cũng như quá trình xử lý đều cần phải lưu trữ lại để dùng tiếp cho các lần sau. Quá trình đó gồm các công đoạn chính như hình 1.1:



Hình 1.1: *Chu trình xử lý thông tin*

Xử lý thông tin bao gồm nhiều quá trình xử lý dữ liệu để lấy ra thông tin phục vụ theo ý muốn của con người. Có thể kể ra một số quá trình đó như sau:

- *Truyền tin:* yêu cầu nhanh nhất, chất lượng, tiết kiệm chi phí, giá thành và chính xác nhất.

- *Lọc nhiễu, khử nhiễu:* Trong quá trình truyền tin, thông tin có thể bị sai lệch do nhiễu đường truyền, cự ly truyền, kỹ thuật truyền nên cần phải lọc nhiễu để khôi phục lại tin tức trung thực ban đầu. Mặt khác qua quá trình lọc, ta có thể lấy những thông tin cần thiết trong hàng loạt các thông tin dư thừa.

- *Lưu trữ thông tin:* phục vụ cho việc tra cứu, tìm kiếm, sao chép.

- Mã hóa thông tin, bảo mật thông tin, giải mã thông tin:

Máy tính điện tử (MTĐT) là công cụ xử lý thông tin tự động, không cần sự tham gia trực tiếp của con người. Tuy nhiên, MTĐT không thể quyết định được khi nào thì phải làm gì, mà phải tuân theo một tập hợp lệnh (chương trình) định sẵn do con người đề ra và được lưu trữ trong bộ nhớ của máy. Chương trình đó sẽ thay thế con người điều khiển MTĐT làm việc.

Quá trình xử lý dữ liệu bằng MTĐT có thể tóm tắt như sau:

- Trước hết đưa chương trình cần thực hiện (do người lập sẵn) vào bộ nhớ của máy

- Máy bắt đầu xử lý dữ liệu nhập từ môi trường ngoài vào bộ nhớ (thông qua một thiết bị nhập dữ liệu)

- Máy thao tác dữ liệu, ghi kết quả trong bộ nhớ.

- Đưa kết quả từ bộ nhớ ra môi trường ngoài (thông qua một thiết bị xuất dữ liệu)

Trước đây khi thông tin còn ít và chưa phức tạp thì con người có thể tự xử lý được. Song ngày nay với sự bùng nổ của thông tin, nếu không có MTĐT chắc chắn chúng ta khó có thể xử lý nổi những bài toán quản lý phức tạp. MTĐT ra đời đã giúp con người xử lý thông tin một cách tự động và hợp lý, đặc biệt nhiều hệ trợ giúp quyết định với các công cụ khai tuyển dữ liệu và phát hiện cơ sở tri thức đã giúp các nhà lãnh đạo và quản lý có thể chọn phương án tối ưu, giúp cho việc giải quyết các bài toán đặt ra một cách tốt hơn và hiệu quả hơn.

1.1.6. Khái niệm tin học và công nghệ thông tin

Tin học (Computer Science): là ngành khoa học nghiên cứu các phương pháp, các công nghệ và các kỹ thuật xử lý thông tin một cách tự động.

Khía cạnh khoa học của tin học chính là phương pháp, còn khía cạnh kỹ thuật của tin học chính là công nghệ chế tạo các thiết bị MTĐT cũng như sản xuất ra các chương trình phần mềm (software) hệ thống và ứng dụng.

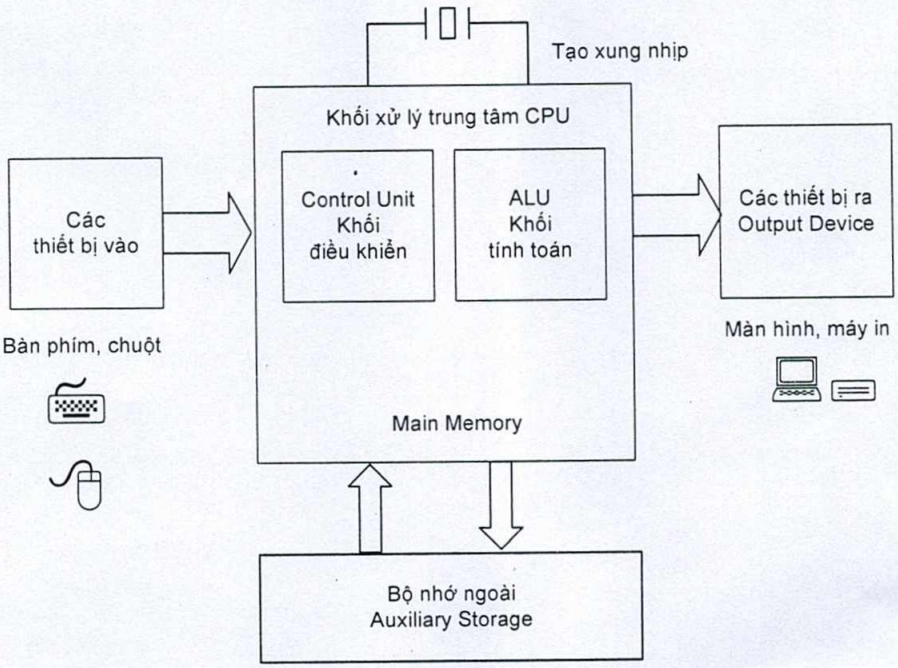
Một thuật ngữ khác bao quát hơn và chính xác hơn đó là Công nghệ thông tin (Information Technology).

Công nghệ thông tin là tập hợp các phương pháp khoa học, các phương tiện và công cụ kỹ thuật hiện đại - chủ yếu là kỹ thuật máy tính và viễn thông nhằm tổ chức khai thác và sử dụng có hiệu quả các nguồn tài nguyên thông tin rất phong phú và tiềm năng trong mọi lĩnh vực hoạt động của con người và xã hội.

1.2. CẤU TRÚC MÁY TÍNH ĐIỆN TỬ

1.2.1. Cấu trúc chung của máy tính điện tử

Hệ thống máy tính nói chung (xem hình 1.2), bao gồm:



Hình 1.2: *Sơ đồ khối cấu trúc của máy tính điện tử*

- Khối xử lý trung tâm: CPU
- Bộ nhớ trong: RAM, ROM
- Bộ nhớ ngoài: đĩa mềm, đĩa cứng, đĩa CD.

- Các thiết bị
- Các thiết bị

1.2.2. Khối xử lý

CPU (Central
diễn ra việc xử l
MTĐT. CPU có n
đồng thời điều khi
trưng sau:

- Tốc độ xử l
vi tính ngày
phép tính trc
- Lượng thông
vi xử lý thườ
- CPU hiểu và

Lệnh bao gồm

Mã lệnh là một
những việc hoặc nhí

Phần địa chỉ củ
xác định địa chỉ củ
hạng là đối tượng tá

CPU có ba bộ
điều khiển và một số

a. Khối tính toán số

Khối ALU (Ari
tác, các phép tính qu
(cộng, trừ, nhân, chi
các phép tính quan hệ

Chương 6

PHỤC HỒI DỮ LIỆU ĐIỆN TỬ

Trong các vụ án nói chung, các đối tượng phạm tội luôn có hành vi "tẩy xóa" dữ liệu điện tử để tránh bị cơ quan điều tra phát hiện và đưa ra xét xử. Các vụ án thực tế cho thấy các đối tượng sử dụng các phương pháp xóa sạch dữ liệu, FORMAT ổ đĩa, FDISK và dùng các chương trình đặc biệt khác nhằm làm mất dữ liệu không cho điều tra viên thu thập để làm chứng cứ pháp lý trước tòa. Các đối tượng phạm tội đều hiểu được việc dùng email, truy nhập những trang Web cấm, những trang Web khủng bố trong hoạt động của chúng đều có thể để lại các dấu vết điện tử là đầu mối để cơ quan điều tra "lần ra" tung tích của chúng do đó chúng đều "cố gắng" loại bỏ các thông tin này một cách vĩnh viễn và triệt để.

Trong các trang Web của các đối tượng hải ngoại như ykien.net, thongluan.org đều có các mục dạy đối tượng vượt tường lửa (dùng Proxy), xóa dấu vết khi truy nhập mạng (cookies), ẩn danh khi vào Internet. Điều này gây khó khăn cho công tác điều tra.

Ngay cả dữ liệu của cơ quan Công an cũng có thể gặp các rủi ro như: Đĩa cứng bị hư hỏng ở một sector nào đó, tập tin, phần mềm bị lỗi, hoặc do vô tình xóa mất tập tin hay thư mục nào đó, hoặc do đối tượng xâm nhập một cách cố ý để phá hoại thông tin.

Ngoài ra thông tin còn bị mất do các yếu tố kỹ thuật như sự phá hoại: Virus, Trojan hay một hình thức phá hoại nào đó do một loại đối tượng nào đó gây ra.

Chương này trình bày một vài chương trình khôi phục dữ liệu một cách nhanh chóng, chính xác và toàn diện nhằm giúp cán bộ kỹ thuật

có thể sử dụng trong quá trình điều tra các vụ án liên quan đến công nghệ thông tin.

Nội dung bao gồm:

1. Trình bày khái niệm, qui trình thu thập, phục hồi dữ liệu điện tử

2. Trình bày cơ chế lưu trữ thông tin trên đĩa cứng

3. Một số chương trình để phục hồi dữ liệu khi bị xóa hoặc hủy vô tình hay cố ý.

Việc sử dụng tốt các chương trình này cho phép cán bộ điều tra có thể phân tích dữ liệu trên các máy tính đã tịch thu của đối tượng.

6.1. QUI TRÌNH THU THẬP, PHỤC HỒI VÀ PHÂN TÍCH DỮ LIỆU ĐIỆN TỬ

6.1.1. Nguyên tắc chung

Để thu được dữ liệu điện tử, trước tiên phải thu được các thiết bị có bộ nhớ kỹ thuật số đã được sử dụng làm phương tiện gây án như: máy tính, máy điện thoại di động, máy scan, máy in... và xác định những thiết bị này có phải là phương tiện để chuẩn bị gây án, lưu chứng cứ phạm tội, truy cập vào website bị tấn công, viết và chuyển email lừa đảo hoặc các tài liệu hay không. Trên cơ sở đó quyết định có cần thu giữ máy tính hoặc các thiết bị có bộ nhớ kỹ thuật số nào và có thể tìm ở đâu. Do máy tính có thể lưu trữ rất nhiều thông tin nên cần nghiên cứu việc tìm kiếm dữ liệu sao cho hiệu quả và kịp thời. Trong công tác trinh sát cần thu được các user và password thông qua việc xét hỏi để có thể mở được máy tính, thiết bị di động và các dữ liệu đã được mã hóa.

Khi tìm và thu giữ các thiết bị lưu trữ chứng cứ điện tử, chúng ta cần tập trung vào hai nguồn sau đây:

1. Tìm và thu giữ phần cứng, phần mềm, tài liệu, các ghi chú của người sử dụng, email, nickname, username và password các loại, các thiết bị truyền thông lưu trữ thông tin như điện thoại di động, PDA.

ị án liên quan đến công

phục hồi dữ liệu điện tử
đĩa cứng
u khi bị xóa hoặc hủy vô

ɔ phép cán bộ điều tra có
thu của đối tượng.

ɔhải thu được các thiết bị
phương tiện gây án như:
n, máy in... và xác định
để chuẩn bị gây án, lưu
tấn công, viết và chuyển
ên cơ sở đó quyết định có
nhớ kỹ thuật số nào và có
t nhiều thông tin nên cần
ệu quả và kịp thời. Trong
password thông qua việc
di động và các dữ liệu đã

hứng cứ điện tử, chúng ta

ı, tài liệu, các ghi chú của
và password các loại, các
ện thoại di động, PDA.

2. Tìm nguồn cung cấp dịch vụ: các ghi chép về dịch vụ, hóa đơn và thông tin về:

- Nhà cung cấp dịch vụ vệ tinh GPS, điện thoại di động
- Nhà cung cấp dịch vụ lưu trữ dữ liệu
- Nhà cung cấp dịch vụ không dây, mạng
- Ngân hàng phát hành, ngân hàng thanh toán thẻ tín dụng, hiệp hội thẻ, đơn vị chấp nhận thẻ, nhà cung cấp dịch vụ thanh toán qua mạng, ngân hàng tiền ảo như e-gold, webmoney, ngân lượng...
- Nhà cung cấp dịch vụ cáp quang.
- Nhà cung cấp dịch vụ Internet, ISP, IXP
- Nhà cung cấp dịch vụ điện thoại cố định, di động
- Nhà cung cấp dịch vụ nội dung số.

Trong quá trình thu giữ phải chú ý bảo vệ khu vực hiện trường để tìm vân tay, các dấu vết truyền thống khác có liên quan khi cần thiết. Để bảo vệ dữ liệu, tránh việc truy cập máy tính từ xa phải chú ý đến các thiết bị ngoại vi của máy tính. Cần sử dụng thiết bị dò tìm để phát hiện trong khu vực có máy tính hoặc thiết bị nào đang hoạt động hay không. Đối với máy ATM cần thu cả dữ liệu của máy videocamera bảo vệ. Trong đó cần chú ý thời gian được ghi trong máy, so sánh với thời gian có hình của đối tượng nghi vấn và thời gian thực để kiểm tra logfile của những thanh toán có nghi vấn.

6.1.2. Các chú ý khi tiến hành thu giữ máy tính và các thiết bị kỹ thuật số tại hiện trường

1. Chụp ảnh và ghi hình hiện trường

2. Cho phép tất cả các máy in hoàn tất quá trình in

3. Kiểm tra để đảm bảo máy tính đã tắt, không được bật máy tính trong bất cứ trường hợp nào. Chú ý một số máy tính xách tay có thể được bật lên khi mở nắp màn hình.

4. Rút điện và các thiết bị khác ra khỏi máy tính. Chú ý máy tính ở chế độ standby vẫn có thể truy cập từ xa và thay đổi dữ liệu.

5. Đánh dấu các cổng và dây cáp để có thể lắp ráp lại máy tính sau đó.

6. Lập biên bản đầy đủ, đúng qui định pháp lý để chứng cứ không bị bác bỏ trước tòa.

7. Tìm nhật ký, sổ tay hoặc những mẩu giấy ghi chép username, password thường đi kèm theo hoặc ở gần máy tính.

8. Lấy lời khai những người có liên quan để thu thông tin về username, password và không để những người không phận sự đến gần máy tính và nguồn điện.

9. Chụp ảnh và ghi hiện trường đã đánh số vị trí các loại dấu vết trong đó có các thiết bị kỹ thuật số.

10. Để lưu lại dữ liệu đang hoạt động trên DDRAM, tắt nguồn điện ở đằng sau máy tính mà không tắt bất kỳ một chương trình nào. Khi rút dây nguồn điện phải rút phần nối với máy tính chứ không rút ổ cắm điện để tránh mất dữ liệu đang được đọc ở DDRAM.

6.2. PHỤC HỒI DỮ LIỆU TRÊN MÁY TÍNH VÀ CÁC THIẾT BỊ SỐ

Để sử dụng tốt các công cụ phục hồi chúng ta cần nghiên cứu cấu trúc đĩa từ và tổ chức lưu trữ thông tin trên đó.

6.2.1. Cơ chế lưu trữ thông tin trên đĩa cứng và các dạng khôi phục dữ liệu

6.2.1.1. Cơ chế lưu trữ thông tin trên đĩa cứng

Thông tin được lưu trữ trên đĩa cứng gồm nhiều đĩa từ mỏng. Mỗi đĩa từ được chia thành hai mặt (Side): side 0 và side 1. Mỗi một Side được chia làm nhiều vòng tròn đồng tâm, các vòng tròn đồng tâm này được gọi là Track (rãnh) và mỗi Track chia thành nhiều cung, mỗi cung được gọi là Sector

Trong hệ điều hành MS DOS, mỗi Sector có 512 bytes. Tập hợp nhiều Sector được gọi là Cluster. Số lượng sector trên một cluster ở các hệ điều hành là khác nhau, (Hình 6.1).

thể lắp ráp lại máy tính

áp lý để chứng cứ không

giấy ghi chép username,
tính.
uan để thu thông tin về
ời không phận sự đến gần

ı số vị trí các loại dấu vế

trên DDRAM, tắt nguồi
kỳ một chương trình nào
i máy tính chứ không rút
c ở DDRAM.

VÀ CÁC THIẾT BỊ SỐ

húng ta cần nghiên cứu cả
đó.

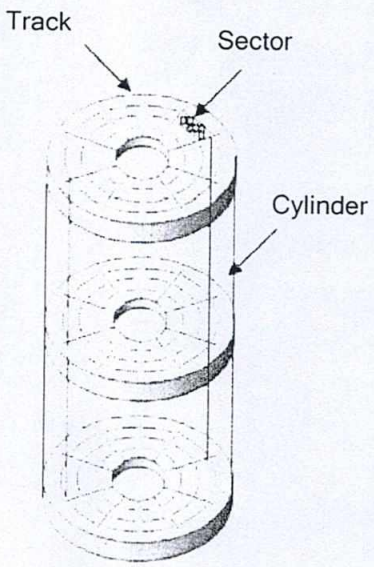
ứng

cứng

gồm nhiều đĩa từ mỏng. M
le 0 và side 1. Mỗi một Si
các vòng tròn đồng tâm n
chia thành nhiều cung, m

Sector có 512 bytes. Tập h
g sector trên một cluster ở c

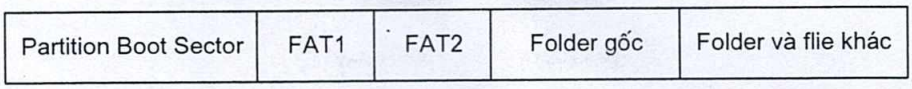
Thông tin được tổ chức và lưu trữ trên từng sector của đĩa cứng vật lý gọi là địa chỉ dữ liệu, mỗi một tập tin được lưu trữ trên đĩa đều có một địa chỉ nhất định và các địa chỉ này được quản lý bởi một danh mục, danh mục này được quy định bởi bảng tham số đĩa.



Hình 6.1: *Cấu trúc ổ đĩa cứng*

Các hệ điều hành đưa ra hệ thống tệp tin FAT và NTFS để quản lý thông tin trên đĩa.

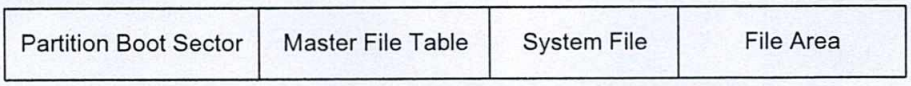
Hệ thống tệp FAT (File Allocation Table) gồm bảng FAT 12, 16 và 32 được sao lưu làm hai bản chứa thư mục gốc FAT cho biết tên tệp, thuộc tính, thời gian tạo lập, ngày tháng truy cập, số hiệu liên cung đầu tiên trong bảng cấp phát và kích thước tệp. Với FAT cluster đầu tiên là địa chỉ của cluster thứ hai được sử dụng để lưu trữ tập đó. Trong mục FAT đối với cluster thứ hai là địa chỉ của cluster thứ ba, cứ như thế tiếp tục cho đến khoản mục của cluster cuối cùng chứa mã kết thúc của tập. Vì chỉ có duy nhất bảng này cho biết cách tìm dữ liệu trên đĩa, cho nên DOS sẽ thành lập và duy trì hai bản sao của FAT để đề phòng một bản bị hỏng, (Hình 6.2).



Hình 6.2: *Cấu trúc bảng FAT*

Hệ thống FAT không có tính phân vùng nên người sử dụng đăng nhập đều có thể đọc sao chép hoặc thay đổi thông tin trên máy tính nên không có tính an toàn cao. Nhược điểm này được khắc phục bằng hệ thống file công nghệ mới NTFS (New Technology File System) có biện pháp mã hóa thư mục, tệp tin và phải dùng chứng chỉ số để giải mã. Định dạng phân vùng theo hệ thống NTFS là tạo một số tệp hệ thống và bảng tệp chính (MFT - Master File Table) chứa thông tin về mọi tệp và thư mục trên phân vùng NTFS.

NTFS được thiết kế cho các thao tác tập tin nhanh trên các ổ đĩa cứng có dung lượng lớn. NTFS bao gồm một hệ thống phục hồi tập tin và các thuộc tính được xây dựng sẵn để giải quyết an toàn và điều khiển truy xuất. Khi chúng ta định dạng một phân vùng (partition) trên một ổ đĩa sử dụng hệ thống tập tin NTFS, thì phân vùng này được khởi tạo như là một volume của NTFS. Volume này chứa MFT, chứa thông tin về từng tập tin có trong volume này. Thông tin này được lưu trữ trong các mẩu tin có kích thước 2,048 byte và hoạt động giống như một cơ sở dữ liệu quan hệ. Các tập tin được xác định bởi một con số, số này phụ thuộc vào vị trí của tập tin trong MFT và một số tuần tự đặc biệt, (Hình 6.3).



Hình 6.3: *Cấu trúc NTFS*

Như vậy, khi xóa một tập tin nào đó kể cả khi xóa sạch nó trong thùng rác của hệ điều hành thì hệ điều hành chỉ làm một nhiệm vụ đơn giản là xóa tên nó trong danh mục quản lý của bảng tham số đĩa (bảng FAT hoặc NTFS) hay chỉ xóa liên kết từ danh mục quản lý tập tin tới địa chỉ lưu trữ nó. Vì thế dữ liệu trên đĩa cứng tại những địa chỉ đó vẫn còn nguyên cho đến khi nó được ghi đè lên bởi một thông tin khác.

Folder và flie khác

người sử dụng đăng
tin trên máy tính
ợc khắc phục bằng
gy File System) có
ứng chỉ số để giải
tạo một số tệp hệ
chứa thông tin về

anh trên các ổ đĩa
ng phục hồi tập tin
ết an toàn và điều
ùng (partition) trên
ùng này được khởi
a MFT, chứa thông
n này được lưu trữ
ạt động giống như
nh bởi một con số,
một số tuần tự đặc

File Area

i xóa sạch nó trong
một nhiệm vụ đơn
g tham số đĩa (bảng
quản lý tập tin tới
hững địa chỉ đó vẫn
thông tin khác.

Từ đặc điểm này người ta đã tạo ra các phần mềm để tiến hành phục hồi lại những tệp tin bị xóa. Bản chất của việc phục hồi dữ liệu là tạo liên kết từ danh mục quản lý đến địa chỉ lưu trữ thông tin.

6.2.1.2. Các dạng khôi phục dữ liệu

Phục hồi dữ liệu trong máy tính thực chất là phục hồi dữ liệu trong ổ cứng, chủ yếu là các file đã bị xóa, bị ghi đè và giải mã những file bị mã hóa (encrypted file).

Các dạng khôi phục dữ liệu bao gồm:

1. Khôi phục file bị xóa: việc xóa file sẽ đánh dấu xóa trong Directory Entry và những thông tin liên quan trong bảng FAT hoặc MFT Entry. Về lý thuyết, khả năng khôi phục đầy đủ file này khá cao nhưng trên thực tế do đối tượng sau khi xóa đã cố gắng thực hiện các thao tác ghi đè lên dữ liệu hoặc cài đặt lại hệ điều hành để ghi đè dữ liệu vào các cluster được đánh dấu xóa nên kết quả không như mong muốn.

2. Phân vùng bị xóa (hoặc tạo lại) nhưng chưa định dạng format: hầu hết dữ liệu đều có thể khôi phục được trong trường hợp này vì FAT và MFT không bị ảnh hưởng khi người dùng xóa và tạo mới phân vùng.

3. Phân vùng bị format: Với phân vùng FAT, việc định dạng sẽ xóa bảng FAT, Boot Record và thư mục gốc (Root Directory) nhưng Partition Table và dữ liệu trong Allocation vẫn còn. Những file có dung lượng nhỏ hơn kích thước một cluster (32KB, mặc định của FAT32 hoặc theo tùy chọn khi định dạng), file được khôi phục hoàn toàn vì chúng không cần đến thông tin trong bảng FAT. Với những file có dung lượng lớn nằm trên nhiều cluster liên tiếp nhau, chúng sẽ bị phân mảnh khi có sự thay đổi nội dung theo thời gian. Việc tìm và nối các cluster có liên quan nhau là công việc khó khăn, nhất là với những file có dung lượng lớn và hay thay đổi. Một số phần mềm phục hồi dữ liệu chuyên dùng cho cảnh sát có khả năng khôi phục mà không cần phục hồi dữ liệu từ bảng FAT. Tuy nhiên nội dung những tệp tin sau khi tìm lại sẽ không đầy đủ, không thể đọc được, vì vậy chúng ta cần

một phần mềm khác có khả năng trích xuất những nội dung còn đọc được từ những tập tin này.

Với phân vùng NTFS, việc định dạng sẽ tạo MFT mới, tuy nhiên kết quả khôi phục sẽ tốt hơn phân vùng FAT vì NTFS không sử dụng bảng FAT để xác định cluster chứa dữ liệu của cùng tập tin.

4. Phân vùng bị format và cài đè hệ điều hành mới hoặc sử dụng Ghosh

Trường hợp này thực sự là khó khăn vì Directory Entry (FAT), MFT (NTFS) đã bị xóa. Giả sử có 10GB dữ liệu lưu trữ trên phân vùng 20GB, phân vùng này bị format và ghi đè 5GB dữ liệu mới. Như vậy không thể khôi phục những dữ liệu đã bị ghi đè mà chỉ có thể khôi phục dữ liệu từ 5GB trở về sau.

Đối với cảnh sát nhiều nước trên thế giới hiện nay, hai phần mềm: Encase và FTK thường được sử dụng để phục hồi dữ liệu và phân tích dữ liệu điện tử.

6.2.2. Cơ chế lưu trữ thông tin trên thẻ thông minh, thẻ từ và một số thiết bị khác

Các thẻ thông minh và thẻ từ thực hiện nhiều chức năng nhưng lại có những đặc điểm tương đồng. Cả hai loại thẻ này đều giao diện với một thiết bị đọc có khả năng hiểu được các thông tin được lưu trữ trong các vạch từ hay các chip máy tính gắn vào thẻ nhựa. Ứng dụng phổ biến nhất của các công nghệ này là thẻ tín dụng. Các công nghệ này tự bổ sung thêm nhiều loại ứng dụng khác do chúng có khả năng lưu trữ bất kỳ thông tin nào. Ở Mỹ, các ứng dụng này còn lưu thông tin của giấy phép lái xe, chìa khóa khách sạn, hộ chiếu, số an ninh xã hội...

Hiện nay có hai loại thẻ thông minh cơ bản:

Loại thẻ thứ nhất: là một thiết bị lưu trữ số có khả năng chứa đựng một lượng lớn thông tin.

Loại thứ hai: là một thẻ nhớ vi mạch, cơ bản giống như một máy tính nhỏ có khả năng thực hiện nhiều phép tính nhớ.

Chức năng của các thẻ này cho phép bảo vệ các thông tin được ghi nhớ. Các thiết bị đọc những thẻ này dựa vào khả năng liên lạc và phạm vi gần (tiện ích sử dụng: trao đổi trực tiếp về giá trị giữa các chủ thẻ, trao đổi giá trị thông qua mạng Internet, lưu trữ dữ liệu hoặc các tệp tin tương tự như một máy tính, các điện thoại không dây, các thiết bị sử dụng vệ tinh).

Các thẻ từ: Vạch từ có thể được nhận dạng bởi một vệt màu đen hoặc nâu chạy ngang tấm thẻ. Các thông tin tài khoản và người sử dụng có thể được lưu trữ trên các rãnh lớn trên vạch từ. Để đọc chính xác các thông tin này, các thiết bị đọc thẻ từ phải có khả năng đọc nhiều loại rãnh. Công nghệ này cũng có thể được sử dụng đối với một loại giấy tờ hay các loại thẻ có thể dùng được như thẻ vào siêu thị hay khóa cửa phòng.

6.3. CÁC CHƯƠNG TRÌNH PHỤC HỒI DỮ LIỆU TRÊN MÁY TÍNH

6.3.1. Chương trình Get Data Back for NTFS

Chương trình cho phép tìm lại các dữ liệu bị mất do fdisk, format hay bị các chương trình phá hoại gây ra như: virus, worm, lỗi phần mềm hoặc những tập tin đã xóa trên những ổ đĩa có phân vùng được định dạng theo định dạng NTFS.

Chương trình tương thích với Windows 2K/XP. Sau khi giải nén và cài đặt chương trình, chúng ta thực hiện các thao tác như sau:

6.3.1.1. Sử dụng những thiết lập mặc định

Để không phải thiết lập phức tạp khôi phục dữ liệu, ta có thể thực hiện như sau:

1. Trên giao diện chính của chương trình, nhấp chọn vào mục I don't know, use default settings, nhấp Next để tiếp tục.

2. Chương trình liệt kê các đĩa cứng và hiển thị số phân vùng của từng đĩa cứng này. Để phục hồi dữ liệu của phân vùng nào, chúng ta chỉ cần nhấp chọn phân vùng của ổ đĩa đó.