Chương 1: TỔNG QUAN

1.1_ THÔNG TIN VÀ DỮ LIỆU

*** * * Thông Tin - Information * * ***

- Thông tin là tất cả những gì mang lại hiểu biết cho con người.
- Thông tin được thể hiện dưới dạng vật lý là những tín hiệu.
- Thông tin được thể hiện dưới dạng toán, văn, ngôn ngữ là những dữ liệu (ký hiệu, chữ viết, chữ số, hình ảnh, âm thanh,…)

*** * * Dữ Liệu - Data * * ***

- Định nghĩa chung: Dữ liệu là thông tin dưới dạng ký hiệu, chữ viết, chữ số, hình ảnh, âm thanh hoặc dạng tương tự. [1]
- Dữ liệu máy tính: Là thông tin được lưu trữ trong máy tính. Trong thời đại số hiện tại thì mặc định Dữ liệu chính là Dữ liệu máy tính.
- Dữ liệu mêta (metadata) : là dữ liệu của dữ liệu, nó cung cấp thông tin về các dữ liệu khác. [2]

1.2_ TẦM QUAN TRỌNG CỦA THÔNG TIN – DỮ LIỆU

Tùy vào thông tin gì mà mức độ quan trọng tương ứng sẽ như thế nào. Có những thông tin mang tầm quan trọng đặc biệt, có thể nói là quan trọng hơn cả các tài sản vật chất. Thông tin giúp chúng ta có được những hiểu biết, những kiến thức, và từ đó giúp chúng ta có thể giải quyết các vấn đề trong cuộc sống. Cuộc sống tinh thần của mỗi người thể hiện qua việc liên tục tiếp thu - xử lý và truyền đạt các thông tin. Nếu không tồn tại thông tin thì cũng có thể xem như không tồn tại cuộc sống! Thông tin cũng là một trong những nguồn tài nguyên quan trọng nhất với tổ chức, doanh nghiệp. Trên thương trường cũng như các loại chiến trường khác, bên nào có thông tin nhanh hơn và chính xác hơn thì bên đó sẽ chiến thắng.

Trong thời đại thông tin hiện tại thì việc tiếp thu, xử lý và truyền đạt các thông tin của con người được hỗ trợ rất nhiều bởi các thiết bị công nghệ thông tin (chủ yếu là các loại máy tính), nhờ đó hiệu quả sẽ cao hơn cực kỳ đáng kể. Vì khả năng xử lý thông tin của con người rất hạn chế nên trong nhiều trường hợp chúng ta buộc phải dùng các hệ thống máy tính mới có thể làm việc được. Với mức độ bùng nổ thông tin như hiện tại thì không chỉ các tổ chức, doanh nghiệp mới cần dùng các thiết bị công nghệ thông tin để xử lý, mà các cá nhân cũng vậy – cũng có thể nói rằng cuộc sống của đại đa số người hiện tại phụ thuộc vào các thiết bị công nghệ, và sẽ rất không ổn nếu không có sự trợ giúp của chúng (chẳng hạn chỉ cần thiếu cái điện thoại là nhiều người sẽ khó mà sống & làm việc bình thường).

Thông tin khi lưu trữ trong máy tính được gọi là dữ liệu. Do đó các dữ liệu máy tính cũng đóng vai trò vô cùng quan trọng với các tổ chức, doanh nghiệp cũng như cá nhân. Trong các tổ chức thì lượng dữ liệu máy tính được lưu trữ có thể cực kỳ lớn (không thể lưu trữ dưới dạng giấy tờ, sổ sách) và phải xử lý tự động bằng máy tính mới có thể đạt được hiệu quả về tốc độ và độ chính xác, nhiều dữ liệu trong đó có tầm quan trọng ở mức quyết định khả năng hoạt động của tổ chức. Với cá nhân thì việc số hóa thông tin thành dữ liệu máy tính rồi sau đó xử lý cũng là việc thường xuyên, không trực tiếp thì gián tiếp. Ví dụ như ta vẫn hay duyệt web, soạn thảo các văn bản, gửi e-mail, thanh toán online, nghe gọi nhắn tin qua điện thoại ...; và không thể khác bởi sẽ bị lạc hậu, cô lập, chậm hơn, mất khả năng cạnh tranh và đời sống tinh thần sẽ rất hạn chế.

1.3_ CÁC ẢNH HƯỞNG – THIỆT HẠI KHI CÓ SỰ CỐ TRÊN DỮ LIỆU

Để giải quyết tốt các vấn đề thì các dữ liệu cần thiết phải đầy đủ và chính xác. Nếu có sự cố hư hỏng sai sót trên dữ liệu thì các vấn đề sẽ không thể giải quyết được hoặc giải quyết không chính xác. Dĩ nhiên tùy thuộc vào vấn đề là gì mà sự ảnh hưởng sẽ nghiêm trọng tới mức nào. Với một doanh nghiệp mà hàng loạt dữ liệu kinh doanh đột nhiên bị mất thì có khả năng ảnh hưởng đến sự tồn vong chứ không đơn giản là mất rồi thì làm lại. Các thông tin tài khoản trong ngân hàng bị mất hoặc sai đi thì thậm chí cũng có thể ảnh hưởng tới cả nền kinh tế và kéo theo cả nền chính trị của một quốc gia. Các cá nhân thì

cũng có thể có những dữ liệu cực kỳ quí giá mà khi sự cố xảy ra thì thiệt hại cũng có thể nghiêm trọng, ảnh hưởng đến uy tín và công việc.

Một số tác hại cụ thể của việc hư hỏng mất mát dữ liệu:

* Với doanh nghiệp:

+ Năm 2014 ở khu vực Châu Á – TBD thiệt hại 229 tỉ USD. [1]

+ Nếu mất dữ liệu trong vòng 5 ngày, 80% doanh nghiệp bị phá sản. [2]

* Với cá nhân:

+ Năm 2014 ở Châu Á – TBD, chi phí khắc phục của người tiêu dùng lên đến 10.8 tỉ USD. [1]

+ Năm 2012, việc mất dữ liệu trong điện thoại gây ra số thiệt hại khoảng 30 tỉ USD. [2]

* Có khi thiệt hại không chỉ là tiền bạc mà còn gây ảnh hưởng nghiêm trọng đến uy tín, danh dự và an ninh tính mạng.

* Khi đến các trung tâm phục hồi dữ liệu để nhờ cứu thì ngoài việc mất tiền, còn có thể bị lộ thông tin và có khi còn hại hơn ☹

---

[1]: http://cand.com.vn/the-gioi-so/Ma-doc-va-that-thoat-du-lieu-gay-thiet-hai-229-ti-uSd-cho-cac-doanh-nghiep-khu-vuc-Chau-a-Thai-Binh-duong-371953/

[2]: http://voz.vn/2015/08/21/hang-chuc-ti-usd-thiet-hai-moi-nam-do-mat-du-lieu/

1.4_ KHẢ NĂNG XẢY RA HƯ HỎNG MẤT MÁT DỮ LIỆU

Cùng với sự phát triển của công nghệ thì các rủi ro có thể phát sinh với dữ liệu cũng gia tăng nhanh chóng. Dữ liệu có thể bị mất mát hư hỏng vì rất nhiều lý do: thiết bị lưu trữ hoạt động sai hoặc bị hư hỏng vật lý, phần tử phá hoại trực tiếp tạo ra sự cố (hoặc gián tiếp thông các chương trình mã độc), phần mềm điều hành sai, chính bản thân người sở hữu thao tác sai,… Nói chung khả năng có thể phát sinh ra sự cố là không hề nhỏ và cũng khó tránh khỏi (ví dụ như thiết bị lưu trữ sẽ phải bị trục trặc sau một khoảng thời gian nhất định), nhất là đối với các cá nhân ở Việt Nam còn sử dụng nhiều thiết bị chưa có độ an toàn cao và chưa có nhiều kiến thức về an toàn dữ liệu.

* Thống kê các nguyên nhân gây ra sự cố (khảo sát của Western Digital năm 2012 [3]):

+ Lỗi phần cứng hoặc hệ thống: 44%

+ Người dùng bất cẩn: 32% (ở VN tỉ lệ này chắc hẳn phải cao hơn:)

+ Lỗi phần mềm: 14%

+ Virus: 7% (tỉ lệ này ngày càng gia tăng)

Hiện tại, xu hướng người khác phá hoại bằng cách ra tay trực tiếp hoặc thông qua các chương trình mã độc đang tăng nhanh (các Ransomware mã hóa dữ liệu tống tiền đã xuất hiện khắp nơi và chỉ riêng dòng Cyptowall đã giúp tội phạm mạng thu hơn 325 triệu USD trong năm 2014 [4])

1.5_ KHẢ NĂNG BẢO VỆ VÀ PHỤC HỒI DỮ LIỆU

Khi sự cố hư hỏng hoặc mất mát các dữ liệu quí xảy ra, thật ra không phải sẽ gây ra thiệt hại. Trên thực tế khá nhiều trường hợp có thể "cứu"! Việc phục hồi lại dữ liệu đôi khi cũng khá dễ dàng và người sử dụng bình thường có thể tự làm được trong một khoảng thời gian ngắn, không phải nhờ đến các trung tâm phục hồi dữ liệu vừa tốn kém thời gian và tiền bạc vừa mang nỗi lo lắng vì người ngoài đã đọc được các dữ liệu của mình. Tuy nhiên trên thực tế không nhiều người biết cách tự cứu, và thậm chí vì không đủ kiến thức nên lại còn có thể gây ra các thiệt hại nghiêm trọng hơn – khiến cho mức độ hư hỏng trở nên nặng nề hơn và khả năng phục hồi lại giảm hẳn đi!

Dĩ nhiên là cũng có những trường hợp việc phục hồi dữ liệu rất khó khăn, có thể phải dùng đến các thiết bị chuyên dụng hoặc các thuật toán thông minh nhân tạo; và cũng có những trường hợp là bất khả. Nhìn chung, có rất nhiều tình huống hư hỏng mất mát dữ liệu và đa số là có thể cứu thành công nếu như có đầy đủ trang thiết bị cần thiết và người sử dụng có kiến thức tương đối tốt về an toàn dữ liệu cá nhân.

Việc ngăn chặn các tác động xấu đối với dữ liệu thường cũng không quá khó khăn và khả năng bảo vệ được dữ liệu cũng không hề nhỏ - vấn đề nằm ở chỗ nhiều người không biết! Có nhiều hình thức để thực hiện việc này, trong đó có những cách cơ bản đơn giản nhưng cũng khá hiệu quả như cài đặt các tường lửa, sao lưu thường xuyên và không thực thi các mã đáng ngờ trên hệ thống có dữ liệu quan trọng.

1.6_ MỘT SỐ HÌNH THỨC BẢO VỆ THÔNG TIN /DỮ LIỆU

❖ Mã hóa

- Ngoài hình thức mã hóa bình thường cho từng dữ liệu cụ thể, có thể triển khai mã hóa trên toàn hệ thống file *(vd như tự tạo các định dạng tích hợp .ZIP, .RAR, .ISO theo kiến trúc khác biệt)*

❖ Che giấu, nhúng (ẩn) dữ liệu

- Dữ liệu thường được truy xuất cuối cùng ở cấp hệ điều hành, và nếu có những "bất thường đặc biệt" thì có thể hệ thống từ chối truy xuất hoặc thậm chí xem là không tồn tại. (xem demo)
- Dữ liệu cũng có thể được nhúng (và "mất tích") trong một nội dung số (âm thanh, phim ảnh,..)

❖ Đặt mật khẩu

- Mọi người vẫn được khuyến cáo: *phải đặt sao cho phức tạp khó đoán, không dùng chung mật khẩu cho các hệ thống khác nhau, phải thay đổi thường xuyên, không nên ghi lại,...* và đành chịu thua -> Thật ra cách giải quyết cũng khá đơn giản: ghi lại dưới hình thức mã hóa có tung "hỏa mù"!

❖ Sao lưu

Ngoài việc sao lưu thủ công những gì quan trọng một cách thường xuyên (ra thiết bị lưu trữ khác), cần biết tới các dịch vụ "đưa lên mây" (qua các đĩa Dropbox, GoogleDrive, OneDrive,..)

❖ Giữ an toàn cho các thiết bị phần cứng

- Trước mắt là cần tránh các loại sốc, ẩm
- Tránh trộm, cướp,...!

1.7_ MỘT SỐ HÌNH THỨC PHỤC HỒI DỮ LIỆU CƠ BẢN

❖ Xử lý trên phần cứng

- Khi thiết bị chứa dữ liệu bị lỗi vật lý.
- Thông thường cần am tường về cấu trúc thiết bị và cơ chế vận hành
- Vẫn có nhiều trường hợp không cần nhiều kiến thức & kinh nghiệm nhưng ít người biết!

❖ Xử lý trên phần mềm

- Khi thiết bị chứa dữ liệu không bị lỗi vật lý.
- Có thể chỉ đơn giản sử dụng các phần mềm cứu dữ liệu có sẵn (nên làm trên bản image của volume)
- Có nhiều trường hợp không thể dùng phần mềm có sẵn để cứu, khi này cần nhiều kiến thức về tổ chức dữ liệu trong máy tính.

1.8_ CÁC HÌNH THỨC TRUY XUẤT TRỰC TIẾP DỮ LIỆU

❖ Qua công cụ

Có khá nhiều công cụ cho ta khả năng truy xuất đến sector trên đĩa cũng như các can thiệp ở cấp không cần nhờ đến hệ điều hành, điển hình là phần mềm WinHex. Khi thao tác cần chú ý các vấn đề:

- Phân biệt Logical Volumes và Physical Media.
- Cẩn trọng khi thay đổi nội dung sector, vì có thể khiến hệ điều hành không còn "hiểu" được dữ liệu.
- Trong trường hợp muốn cứu dữ liệu, để an toàn, có thể thao tác trên Image của Volume

❖ Qua hình thức lập trình

- ***Mức BIOS (Basic Input /Ouput System):***

Bộ điều khiển đĩa đưa ra các khả năng cho phép truy xuất ở mức vật lý. Các chức năng này được thực hiện thông qua ngắt 13h:

Chức năng	Input	Output
Reset disk (dùng để reset đĩa sau một tác vụ gặp lỗi)	AH = 0 , DL = số hiệu đĩa (0 = đĩa A , 1 = đĩa B , 80H = đĩa cứng ...)	Không
Lấy mã lỗi của tác vụ đĩa gần nhất	AH = 01 , DL = đĩa vật lý (80H lấy lỗi của đĩa mềm, 7FH lấy lỗi của đĩa cứng)	AL chứa mã lỗi, cũng là giá trị tại 0:0441.
Đọc sector	AH = 2, DL = số hiệu đĩa , DH = số đầu đọc ghi , CH = số track (cylinder) , CL = số sector, AL = số sector cần đọc/ghi (không vượt quá số sector trên 1 track) , ES:BX = địa chỉ của buufer chứa thông tin.	Cờ Carry =1 nếu có lỗi chứa trong thanh ghi AH.
Ghi sector	AH = 3, còn lại tương tự như chức năng đọc	Như chức năng đọc
Verify sector	AH = 4, các thanh ghi khác tương tự	Như chức năng đọc

- ***Mức OS (Hệ điều hành – Operating System):***

Các chức năng của ngắt 13h của BIOS cho phép đọc bất kỳ 1 sector nào trên đĩa, tuy nhiên do các quy định thanh ghi phức tạp và phải làm việc với đĩa vật lý khó khăn nên hệ điều hành sẽ hỗ trợ một cách truy xuất khác rất thuận lợi hơn, đầu tiên vẫn là cấp ngắt (ngắt 25h và 26h) nhưng đĩa được chuyển thành volume logic, tham số đưa vào đây chỉ còn là sector logic. Các sector logic trong volume được đánh chỉ số bắt đầu từ 0, các volume cũng được biểu diễn theo chỉ số tương ứng với chữ cái biểu thị (vol A là 0, vol C là 2, vol D là 3,…)

Các hàm ngắt thật ra vẫn còn khá phức tạp nên hệ điều hành Windows tổ chức thêm các hàm API có tên là các chuỗi gợi nhớ gợi nghĩa và có thể sử dụng tham số giống hàm bình thường chứ không phải phải qua các thanh ghi như hàm ngắt, và volume được giả lập như một file. Do đó việc đọc ghi trên đĩa tương tự như việc đọc ghi file. Cách thức để thực hiện việc đọc ghi đĩa trên Windows như sau:

- Dùng hàm CreateFile để lấy handle của một đĩa
- Dùng hàm SetFilePointer để set vị trí của đầu đọc.

- Dùng hàm ReadFile/WriteFile để đọc và ghi đĩa.

Ngoài ra cũng có một số hàm khác với công năng tương tự.

CÁC THỰC NGHIỆM & BÀI TẬP

TN#1.1 Sử dụng Winhex làm hư /phục hồi lại đĩa (Memory card /Flash disk)

TN#1.2 Sử dụng công cụ mã hóa của MS Office và các công cụ Nén cho DL bất kỳ

TN#1.3 Đặt /lưu trữ mật khẩu an toàn

TN#1.4 Mở /gắn board HDD + kết nối với PC qua HDD box /switch

TN#1.5 Cứu các file đã xóa bằng Data Recovery Tool

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 Data Recovery Tool

BTVN#1.3 Nghiên cứu sử dụng 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 Data Recovery Tool.

Chủ đề mở rộng:

MR#1.1 Nghiên cứu và trình bày về 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 Xây dựng chương trình mã hóa & giải mã một tập tin thỏa:

i. Key mã hóa (password) do người dùng nhập vào được hash thành 1 chuỗi tối thiểu 100bit.

ii. Việc mã hóa /giải mã phải nhanh (thời gian giải mã trong RAM phải ở mức real time).

iii. *Tuy cùng 1 key mã hóa nhưng mỗi lần mã sẽ /phải được một kết quả khác nhau (về cả chiều dài lẫn nội dung nhị phân).*

iv. *Chương trình mã hóa có obfurcasted code để tăng độ an toàn, tránh không cho người khác dò ra cách làm và giải mã.*

v. *Mỗi khi chương trình chạy phải hỏi và kiểm tra 1 mật khẩu "động" và nếu đúng thì mới tiếp tục chạy.*

vi. *Chương trình có tính năng phát hiện các dấu hiệu bất thường và xử lý phù hợp (ví dụ: thấy nội dung không như ban đầu (bị crack hoặc có virus bám vào) thì tự khôi phục lại, thấy môi trường chạy không phải là môi trường "lành mạnh" (bị chạy trong sandbox hoặc trong chế độ debug, chương trình không nằm trong đĩa /máy được quy ước, ...) thì không chạy /tự hủy /gây lỗi hệ thống, ...)*

~.~

[1]: Điều 4 Luật Giao dịch điện tử

[2]: https://en.wikipedia.org/wiki/Metadata/

[3]: http://voz.vn/2015/08/21/hang-chuc-ti-usd-thiet-hai-moi-nam-do-mat-du-lieu/

[4]: http://nhipsongso.tuoitre.vn/tin/20151207/ma-doc-danh-cap-mat-khau-truoc-khi-ma-hoa-du-lieu-tong-tien/1016272.html