# Các bước đọc ổ đĩa dạng NTFS

## Bước 1: Lấy file cần đọc

* Đọc volume hệ thống tương ứng trong máy tính theo định dạng tên file truyền vào là [\\\\.\\X:](file:///\\\\.\\X:) với X là tên của ổ đĩa cần đọc.
* Chương trình cần phải chạy với phân quyền người quản trị để đọc được các thông tin ổ đĩa.
* Nếu không tồn tại thì thoát chương trình.

## Bước 2: Đọc Boot Sector

* Đọc 512-byte đầu tương ứng với VBR (Volume Boot Record)
* Các thông tin chính cần quan tâm:

|  |  |  |
| --- | --- | --- |
| Địa chỉ | Kích thước | Thông tin |
| 0x3 | 8 | OEM ID (Sẽ là ‘NTFS ’) |
| 0xB | 2 | Bytes/sector |
| 0xD | 1 | Sectors/cluster |
| 0xE | 2 | Số sector dự trữ |
| 0x28 | 8 | Tổng số sector |
| 0x30 | 8 | Cluster bắt đầu của $MFT |
| 0x38 | 8 | Cluster bắt đầu của $MFTMirr |
| 0x40 | 1 | Kích thước của một $MFT record |

* Kích thước của mỗi một MFT record ở trên sẽ là một con số dạng bù 2, và sẽ cần phải lấy 2 lũy thừa với trị tuyệt đối của giá trị trên để có kích thước thực tế.

VD: *(Theo ví dụ tham khảo phần giải đáp của thầy Long)*

Tại 0x40 có là 0xF6, là một con số bù 2 nên quy đổi ra hệ thập phân là -10.

Do đó, kích thước là (byte)

* Để cho thuận tiện, cluster bắt đầu của MFT đầu tiên nên được quy đổi ra byte: (Cluster bắt đầu) \* (bytes/sector) \* (sectors/cluster)

## Bước 3: Đọc MFT

* Đưa con trỏ đọc đến vị trí bắt đầu đã tính ở trên.
* Đọc MFT đầu tiên để lưu giá trị tổng số sector để lưu các MFT record.  
  (Tại offset 0x118, độ dài là 8)
* Tính tổng số MFT record mà có trong ổ đĩa.  
  (Lấy tổng số sector đó chia cho sectors/record, thường mỗi record sẽ chiếm 1024 bytes)
* Lần lượt đọc qua các MFT record tiếp theo.

## Bước 4: Đọc các MFT record header

* Quá trình đọc MFT record cần chú ý đọc qua 4 phần chính, phần header, attribute Standard Information, attribute File Name, attribute Data.
* Ở phần đầu tiên là header

|  |  |  |
| --- | --- | --- |
| Offset | Size | Thông tin |
| 0x0 | 4 | Dấu hiệu nhận biết MFT entry.  - “FILE”: MFT record bình thường.  - “BAAD”: MFT record lỗi. |
| 0x16 | 2 | Giá trị cờ báo  - 0x01: MFT record của một file.  - 0x02: MFT record của một thư mục. |
| 0x2C | 4 | Số hiệu của MFT record này. |
| 0x14 | 2 | Vị trí bắt đầu của attribute đầu tiên (Standard Information) |

## Bước 5: Standard Information

* Đọc và tính vị trí bắt đầu của Standard Information ở byte thứ 0x14, 0x15 của MFT.
* Kiểm tra xem mã thuộc tính (4 bytes đầu tiên) có phải là 0x10 hay không.
* Lấy phân quyền của file tại 0x20 🡪 0x23 (tính từ đầu standard information).

## Bước 6: File name

* Nhảy qua hết phần Standard Information
* Kiểm tra mã thuộc có phải là 0x30 hay không, nếu không thì nhảy tiếp 1 lần nữa phần Attribute List.
* Đọc số hiệu của record cha tại 0x0 🡪 0x5 (tính từ vị trí bắt đầu của file name)
* Đọc phân quyền của file tại 0x38 🡪 0x3B.
* Đọc kiểu định dạng tên (0x41) và độ dài tên (0x40).
* Phần tên được lưu N bytes đã đọc từ 0x42.

## Bước 7: Data

* Tương tự như File Name, đọc và nhảy qua các thuộc tính tương ứng đến thuộc tính có mã là 0x80.
* Tại đây, nếu dữ lớn hơn 700 bytes sẽ được lưu dưới dạng non-resident, ngược lại là resident.
* Nếu thuộc tính này lưu dạng resident
  + Vị trí bắt đầu nội dung (0x14, 0x15)
  + Độ dài nội dung (0x10 🡪 0x13)
  + Phần nội dung đọc từ vị trí của vị trí bắt đầu nội dung đã tính trên.
* Nếu thuộc tính lưu dưới dạng non-resident, phần dữ liệu này sẽ được lưu dưới dạng các data-runs.

## Bước 8: Nối các record con đến record cha của nó

* Mỗi record được đọc đều được đánh số hiệu tương ứng.
* Lần lượt duyệt qua các record, các record con cha và con sẽ có lần lượt có số hiệu và số hiệu record cha giống nhau.
* Nếu record có số hiệu cha và chính nó giống nhau tức là vị trí gốc của thư mục.

# Cách đọc các data-runs

Lần lượt các data-run bao gồm phần header, length, offset.

Các data-run sẽ được xếp liền kề nhau.

Header có dạng 0xXY với:

* Y: số byte của phần length (bắt đầu từ sau header)
* X: số byte của phần offset (bắt đầu sau phần length)

Ví dụ, giả sử chúng ta có các dãy byte sau:

11 03 07 32 AD E1 4C 1A 01 00 00 ….

Ở đây chúng ta có 2 data-runs:

* Data-run 1:
  + 1 byte length, 03 clusters long.
  + 1 byte offset, cluster 07.
* Data-run 2:
  + 2 bytes length, AD E1 clusters long.
  + 3 bytes offset, cluster 4C 1A 01.

Phía sau data-run thứ 2 là 00, tức là không còn data-run nào khác nữa.