

# Formatul Fisierelor PE (Portable Executable)

**Autor:** Ionut Popescu (Nytro)

**Website:** <https://rstforums.com/forum/> (Romanian Security Team)

## 1. Introducere

**Portable Executable** este formatul de fisiere folosit de sistemul de operare Microsoft Windows pentru fisierele executabile si alte tipuri de fisiere care contin cod executabil, cod masina - care este executat direct de catre procesor.

Acest format este folosit in special de catre fisierele executabile cu extensia “.exe” si de catre bibliotecile de functii ce sunt incarcate de catre aceste executabile avand extensia “.dll” dar si de alte tipuri de fisiere cum ar fi controalele ActiveX (extensia “.ocx”) sau driverele/modulele sistemului de operare (extensia “.sys”).

Numele de “Portable” provine de la faptul ca formatul este acelasi pentru mai multe arhitecturi: x86, x86-64, IA32 si IA64, este “portabil”. Practic acest format este o versiune modificata a formatului COFF (Common Object File Format) folosit de sistemele UNIX in timp ce sistemele Linux folosesc formatul ELF (Executable and Linkable Format).

Acest format a aparut pentru prima oara odata cu sistemul de operare Windows NT 3.1 si a fost creat pentru a pastra compatibilitatea cu vechiul format, MS-DOS, sistemul anterior sistemului Windows.

## 2. Structura generala a fisierelor PE

Din motive de compatibilitate cu formatul folosit de MS-DOS (sistemul de operare al celor de la Microsoft de dinainte de Windows), fisierele PE contin practic un mic program pentru MS-DOS care afiseaza mesajul: *“This program cannot be run in DOS mode.”*, acesta fiind singurul lucru pe care il face, astfel, daca se incearca executarea unui fisier PE pe sistemul MS-DOS, se va afisa acel mesaj.

Formatul foloseste mai multe anteturi/headere pentru a defini datele folosite. Asadar fiecare fisier PE va contine atat un header MS-DOS necesar programului MS-DOS, cat si un header PE, necesar programului nostru.

Headerele sunt un grup de octeti, numar fix sau variabil, care definesc datele ce vor urma, niste octeti care ofera informatii despre structurile ce urmeaza, structuri pe care le definesc.

Pentru o detaliere mai simpla a acestui format voi afisa definitiile acestor headere ca structuri in C/C++. Astfel se vor putea observa mai usor dimensiunile fiecarui camp din structuri. Structurile pot fi regasite in SDK-uri in fisierul “WinNT.h”.

Ca structura generala un fisier PE este destul de simplu, este alcatuit din programul MS-DOS (denumit adesea "stub") care e alcatuit dintr-un header MS-DOS si programul MS-DOS, din headerele PE (vom vedea ca nu este unul singur) care include si tabelul de sectiuni, urmate de sectiunile propriu-zise, adica datele si instructiunile programului nostru.

Pe intelesul tuturor, primii octeti dintr-un fisier PE il reprezinta un mic program MS-DOS, urmat de niste headere care definesc datele ce urmeaza, urmate de datele propriu-zise ale fisierului PE: sectiuni de date si de cod.

### Structura fisierelor PE

Header MS-DOS	Program MS-DOS
Program MS-DOS	
Header PE	Headere PE
Header optional PE	
Tabel sectiuni	
Sectiunea 1	Date si cod
Sectiunea 2	
...	
Sectiunea n	

Asadar prima parte a oricarui fisier executabil o reprezinta un mic program MS-DOS, adesea in jur de 240 de bytes. Deschis intr-un Hex Editor, acest program MS-DOS compus din header si cod arata cam asa (intreg programul MS-DOS, header si cod):

	00	01	02	03	04	05	06	07	08	09	0a	0b	0c	0d	0e	0f	
00000000	4d	5a	90	00	03	00	00	00	04	00	00	00	ff	ff	00	00	MZ .....
00000010	b8	00	00	00	00	00	00	00	40	00	00	00	00	00	00	00	.....0.....
00000020	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....8...
00000030	00	00	00	00	00	00	00	00	00	00	00	00	f0	00	00	00	.....!
00000040	0e	1f	ba	0e	00	b4	09	cd	21	b8	01	4c	cd	21	54	68	..°..'í!..Lí!Th
00000050	69	73	20	70	72	6f	67	72	61	6d	20	63	61	6e	6e	6f	is program cannc
00000060	74	20	62	65	20	72	75	6e	20	69	6e	20	44	4f	53	20	t be run in DOS
00000070	6d	6f	64	65	2e	0d	0d	0a	24	00	00	00	00	00	00	00	mode....\$.....
00000080	63	8a	9f	9f	27	eb	f1	cc	27	eb	f1	cc	27	eb	f1	cc	çŠŸŸ'ëñì'ëñì'ëñì
00000090	2e	93	62	cc	16	eb	f1	cc	27	eb	f0	cc	55	e8	f1	cc	."bì.ëñì'ësiUëñì
000000a0	2e	93	63	cc	26	eb	f1	cc	2e	93	64	cc	20	eb	f1	cc	."cì&ëñì."dì ëñì
000000b0	2e	93	72	cc	d1	eb	f1	cc	2e	93	75	cc	c4	eb	f1	cc	."rìÑëñì."uìÄëñì
000000c0	2e	93	65	cc	26	eb	f1	cc	2e	93	60	cc	26	eb	f1	cc	."eì&ëñì."ì&ëñì
000000d0	52	69	63	68	27	eb	f1	cc	00	00	00	00	00	00	00	00	Rich'ëñì.....
000000ef	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....

Un fisier PE deschis intr-un Hex Editor

Daca veti deschide un executabil/DLL cu Notepad sau orice alt editor, veti obtine un rezultat asemanator. Scopul acestui tutorial este de a intelege ce reprezinta acele date.

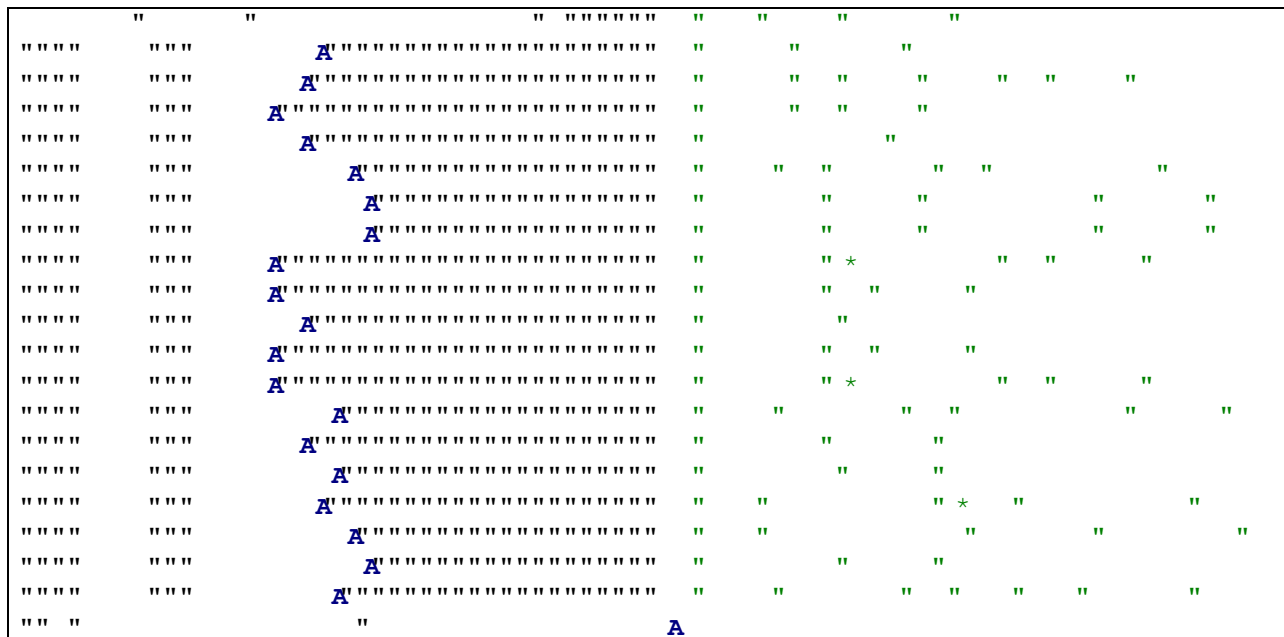
### 2.1. Header-ul MS-DOS

Headerul care contine informatii despre acest mic program este o structura (C/C++) numita IMAGE\_DOS\_HEADER. Structura are dimensiunea de 64 de octeti si ca orice header contine informatii despre programul MS-DOS, informatii necesare pentru ca programul sa poata fi executat cu succes.

Pentru non-programatori, primii 64 de octeti din fisier ii reprezinta niste date care reprezinta headerul MS-DOS, niste informatii necesare pentru a putea executa acel program pe sistemul MS-DOS. Pentru a putea intelege aceste notiuni, e necesara o intelegere la nivel de baza a structurilor din C/C++.

Practic datele din fisier se mapeaza peste aceasta structura, astfel, primii 2 octeti dintr-un fisier PE vor fi reprezentati de campul "e\_magic" din structura (daca veti deschide un fisier executabil sau o biblioteca de functii .dll intr-un editor de text, veti observa ca primele 2 caractere din fisier il reprezinta sirul de caractere "MZ", 2 caractere insumand 2 octeti, ceea ce in structura reprezinta campul "e\_magic"), urmatorii 2 octeti reprezinta campul "e\_cblp" si tot asa. Asadar, primii 64 de octeti ii reprezinta un header care contine informatiile definite mai jos.

In cazul in care nu sunteti familiari cu limbajele C/C++, nu va sperati, veti intelege despre ce e vorba.



Structura care defineste header-ul MS-DOS

Unul dintre lucrurile de baza pe care trebuie sa le cunoasteti atunci cand programati pentru Windows il reprezinta tipurile de date<sup>1</sup>. Astfel avem:

- *BYTE* – Un numar pe 8 biti (**1 byte**), definit ca "unsigned char"
- *CHAR* – Un numar pe 8 biti (**1 byte**), definit ca "char"
- *DWORD* – Un numar pe 32 de biti (**4 bytes**) fara semn, definit ca "unsigned long"
- *LONG* – Un numar pe 32 de biti (**4 bytes**) cu semn, definit ca "long"
- *ULONGLONG* – Un numar pe 64 de biti (**8 bytes**) fara semn, definit ca "unsigned long long"
- *WORD* – Un numar pe 16 biti (**2 bytes**) fara semn, definit ca "unsigned short"

Chiar daca sunteti cunoscatori ai limbajelor C/C++ va trebui sa va familiarizati cu tipurile de date specifice sistemului de operare Windows. Nu este nimic complicat, sunt practic doar "alte nume" pentru tipurile de date pe care le cunoasteti deja.

Practic avem nevoie sa stim aceste informatii deoarece fisierele PE se mapeaza perfect peste aceste structuri. Astfel, daca vom citii primii 64 de bytes dintr-un executabil, acestia vor fi exact acest header, iar daca vom citi 64 de bytes intr-o astfel de structura vom putea avea acces simplu si rapid la fiecare dintre campurile acestui header.

Astfel, primii 2 octeti (tipul WORD ocupa 2 bytes) ai fiecarui fisier PE vor reprezenta campul **e\_magic** (un WORD are 2 octeti), urmatorii 2 octeti vor fi **e\_cblp** si tot asa.

Sa citim o astfel de structura si sa vedem cum arata. Am scris un mic program care citeste un fisier PE, "kernel32.dll" si afiseaza informatiile din headerul sau PE.

Vom vedea ca majoritatea campurilor au valoarea 0 deoarece nu sunt folosite. Pe noi ne intereseaza doar doua dintre aceste campuri: **e\_magic** si **e\_lfanew**.

```

"
  "
" /
" " " " " " " " " "
" " " * " " " " " "
" /
" "
" " " "
" "
" " ./"
" " " " 'A'
" " " " " 'A'
" " " " " 'A'
" "
  
```

<sup>1</sup> Windows Data Types (Windows) - <http://msdn.microsoft.com/en-us/library/windows/desktop/aa383751%28v-vs.85%29.aspx>

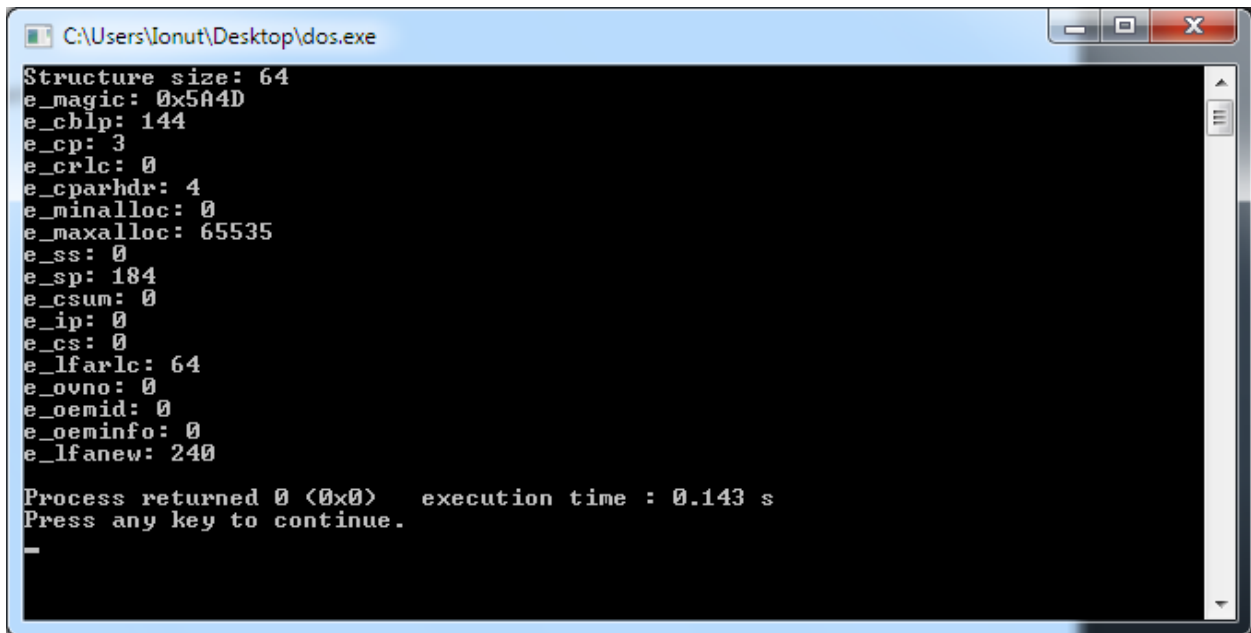
```

" "
" " " " " "
" " " /A'
" " " /"
" " " " " " " " # /A'
" " " A'
" "
" " " "
" " " " " " " " " "
" " " " " " " * " " " " " "
" " " " " " " " " " " "
" " " " " " " "
" " " " " " " " /" /A'
" " " " //
" " " " " " # /A'
" " " " A'
" "
" " /A'
" " " " " " " " " "
" " " " " " " " " "
" " " " " " " " " /A' //A'
" " " " " /A'
" " " " /A'
" " " " /A'
" " " " /A'
" " " " /A'
" " " " /A'
" " " " /A'
" " " " /A'
" " " " /A'
" " " " /A'
" " " " /A'
" " " " /A'
" " " " /A'
" " " " /A'
" " " " /A'
" " " " /A'
" " " " /A'
" " " " A'

```

Programul care afiseaza headerul DOS al unui fisier PE

O executie a acetui program ar produce rezultate asemanatoare:



```
C:\Users\Ionut\Desktop\dos.exe
Structure size: 64
e_magic: 0x5A4D
e_cblp: 144
e_cp: 3
e_crlc: 0
e_cparhdr: 4
e_minalloc: 0
e_maxalloc: 65535
e_ss: 0
e_sp: 184
e_csum: 0
e_ip: 0
e_cs: 0
e_lfarlc: 64
e_ovno: 0
e_oemid: 0
e_oeminfo: 0
e_lfanew: 240

Process returned 0 (0x0)   execution time : 0.143 s
Press any key to continue.
_
```

Campul **e\_magic**, primii doi octeti ai fiecarui fisier PE, reprezinta o “semnatura”, un identificator care ne spune ca avem un fisier MS-DOS. Daca ati deschis vreodata un executabil in Notepad ati observat ca incepe cu literele “MZ”. Aceste litere provin de la numele inginerului Microsoft - Mark Zbykowski.<sup>2</sup>

Pentru programatori, aceste litere apar in fisier ca fiind octetii 0x4D (M) si 0x5A (Z). De observat ca in screenshot-ul de mai sus, numarul este **0x5A4D**, deoarece procesoarele Intel sunt little-endian si in memorie numerele sunt memorate in ordinea inversa a octetilor (de exemplu, numarul **3**, intr-o variabila **int**, care ocupa 4 bytes, va fi memorat in memorie ca “**03 00 00 00**”, mai exact primul octet de la adresa la care e memorat numarul in memorie va fi “cel mai putin semnificativ”). In WinNT.h:

```
”
”””” %           ”           .....
```

Campul **e\_lfanew** este cel care ne intereseaza in mod special, deoarece acesta ne indica unde se termina programul MS-DOS si unde incepe headerul PE. Astfel vom putea sari peste acest program si vom ajunge la ceea ce ne intereseaza.

Insa ce face acest “stub” (cum este numit acest mic program) MS-DOS? Cum afiseaza acel mesaj? Contine cod?

Dupa cum spuneam mai sus, headerul MS-DOS are 64 de bytes, asadar programul MS-DOS va incepe imediat dupa acest header, de la octetul 0x40 (64 in zecimal).

<sup>2</sup> MZ - <http://wiki.osdev.org/MZ>



Daca ne uitam in hex editor putem observa cu usurinta faptul ca sirul de caractere se afla la pozitia **0xE** (fata de inceputul codului executabil) si ca se termina cu caracterul "\$".

	00	01	02	03	04	05	06	07	08	09	0a	0b	0c	0d	0e	0f	
00000000	4d	5a	90	00	03	00	00	00	04	00	00	00	ff	ff	00	00	MZ .....ÿÿ..
00000010	b8	00	00	00	00	00	00	00	40	00	00	00	00	00	00	00	, .....@.....
00000020	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
00000030	00	00	00	00	00	00	00	00	00	00	00	00	f0	00	00	00	.....8...
0000004e	0e	1f	ba	0e	00	b4	09	cd	21	b8	01	4c	cd	21	54	68	..°..'í!..Lí!T:
00000050	69	73	20	70	72	6f	67	72	61	6d	20	63	61	6e	6e	6f	is program canno
00000060	74	20	62	65	20	72	75	6e	20	69	6e	20	44	4f	53	20	t be run in DOS
00000070	6d	6f	64	65	2e	0d	0d	0a	24	00	00	00	00	00	00	00	mode....\$.....

Sirul de caractere afisat de programul MS-DOS

## 2.2. Header-ul PE

Dupa cum spuneam la prezentarea header-ului MS-DOS, ultimul cand din aceasta structura, ultimii 4 bytes din cei 64, campul , reprezinta un numar pe 4 octeti care indica locatia "noului" header, indica exact pozitia in fisier de unde incepe header-ul PE.

Campul este un numar LONG, formatul fiind little-endian, astfel octetii sunt in ordine inversa. Desigur, citirea acestor 4 octeti intr-o variabila "long" se va face corect, iar variabila va contine valoarea corecta deoarece formatul din fisier este formatul in care e memorata acea valoare in memorie, pe procesoarele little-endian.

	00	01	02	03	04	05	06	07	08	09	0a	0b	0c	0d	0e	0f	
00000000	4d	5a	90	00	03	00	00	00	04	00	00	00	ff	ff	00	00	MZ .....ÿÿ..
00000010	b8	00	00	00	00	00	00	00	40	00	00	00	00	00	00	00	, .....@.....
00000020	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
0000003c	00	00	00	00	00	00	00	00	00	00	00	00	f0	00	00	00	.....\$...
00000040	0e	1f	ba	0e	00	b4	09	cd	21	b8	01	4c	cd	21	54	68	..°..'í!..Lí!Th
00000050	69	73	20	70	72	6f	67	72	61	6d	20	63	61	6e	6e	6f	is program canno
00000060	74	20	62	65	20	72	75	6e	20	69	6e	20	44	4f	53	20	t be run in DOS
00000070	6d	6f	64	65	2e	0d	0d	0a	24	00	00	00	00	00	00	00	mode....\$.....
00000080	63	8a	9f	9f	27	eb	f1	cc	27	eb	f1	cc	27	eb	f1	cc	cŠÿÿ'ëñì'ëñì'ëñì
00000090	2e	93	62	cc	16	eb	f1	cc	27	eb	f0	cc	55	e8	f1	cc	..`bì.ëñì'ëñìUëñì
000000a0	2e	93	63	cc	26	eb	f1	cc	2e	93	64	cc	20	eb	f1	cc	..`cìëñì. `dì ëñì
000000b0	2e	93	72	cc	d1	eb	f1	cc	2e	93	75	cc	c4	eb	f1	cc	..`rìñëñì. `uìäëñì
000000c0	2e	93	65	cc	26	eb	f1	cc	2e	93	60	cc	26	eb	f1	cc	..`eìëñì. `ìëñì
000000d0	52	69	63	68	27	eb	f1	cc	00	00	00	00	00	00	00	00	Rich'ëñì.....
000000e0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
000000f0	50	45	00	00	4c	01	04	00	15	3b	b8	50	00	00	00	00	PE...;P...
00000100	00	00	00	00	e0	00	02	21	0b	01	09	00	00	50	0c	00	....ä...!.....P..

Locatia din headerul MS-DOS (rosu) ce indica headerul PE (verde)



Dupa cum se poate vedea in exemplul de mai sus, header-ul PE se afla la locatia **0xf0** (240 zecimal) in executabil.

Imediat la aceasta locatie putem gasi noul header PE, care poate fi gasit in “WinNT.h” cu numele **IMAGE\_NT\_HEADERS**.

```
struct IMAGE_NT_HEADERS {  
  unsigned short Signature;  
  unsigned short SizeOfImage;  
  unsigned short SectionOffset;  
  unsigned short NumberOfSections;  
  unsigned short Reserved;  
  unsigned short ImageBase;  
  unsigned short SectionTableOffset;  
  unsigned short SectionTableSize;  
  unsigned short SectionTableEntrySize;  
  unsigned short SectionTableEntries;  
  unsigned short SectionTableEntryTableOffset;  
  unsigned short SectionTableEntryTableSize;  
  unsigned short SectionTableEntryTableEntrySize;  
  unsigned short SectionTableEntryTableEntries;  
  unsigned short SectionTableEntryTableEntryTableOffset;  
  unsigned short SectionTableEntryTableEntryTableSize;  
  unsigned short SectionTableEntryTableEntryTableEntrySize;  
  unsigned short SectionTableEntryTableEntryTableEntries;  
  unsigned short SectionTableEntryTableEntryTableEntryTableOffset;  
  unsigned short SectionTableEntryTableEntryTableEntryTableSize;  
  unsigned short SectionTableEntryTableEntryTableEntryTableEntrySize;  
  unsigned short SectionTableEntryTableEntryTableEntryTableEntries;  
  unsigned short SectionTableEntryTableEntryTableEntryTableEntryTableOffset;  
  unsigned short SectionTableEntryTableEntryTableEntryTableEntryTableSize;  
  unsigned short SectionTableEntryTableEntryTableEntryTableEntryTableEntrySize;  
  unsigned short SectionTableEntryTableEntryTableEntryTableEntryTableEntries;  
};
```

*Structura IMAGE\_NT\_HEADERS*

Un lucru pe care ar trebui sa il avem in considerare il reprezinta faptul ca exista aceste structuri sunt diferite pe versiunile de 32 de biti si de 64 de biti:

```
struct IMAGE_NT_HEADERS_32 {  
  unsigned short Signature;  
  unsigned short SizeOfImage;  
  unsigned short SectionOffset;  
  unsigned short NumberOfSections;  
  unsigned short Reserved;  
  unsigned short ImageBase;  
  unsigned short SectionTableOffset;  
  unsigned short SectionTableSize;  
  unsigned short SectionTableEntrySize;  
  unsigned short SectionTableEntries;  
};  
  
struct IMAGE_NT_HEADERS_64 {  
  unsigned short Signature;  
  unsigned short SizeOfImage;  
  unsigned short SectionOffset;  
  unsigned short NumberOfSections;  
  unsigned short Reserved;  
  unsigned short ImageBase;  
  unsigned short SectionTableOffset;  
  unsigned short SectionTableSize;  
  unsigned short SectionTableEntrySize;  
  unsigned short SectionTableEntries;  
};
```

*Diferenta dintre structura pe 32 de biti si cea pe 64 de biti*

Structura este aparent simpla, contine doar 3 campuri:

- **Signature**: un numar de 4 octeti care identifica aceasta structura ca fiind o structura PE, definitia sa poate fi gasita ca

```
#define IMAGE_NT_SIGNATURE 0x00004550 // PE00
```

*Numarul pe 4 octeti (little endian) care indica semnatura unui header PE*

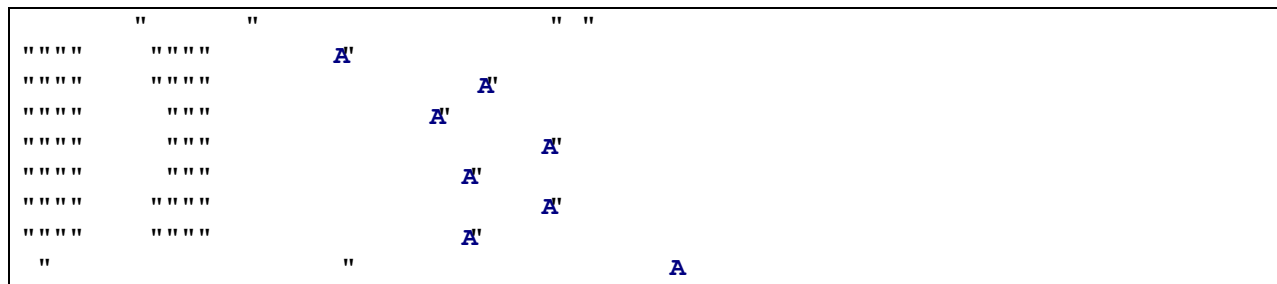
Se poate identifica usor in executabil, caracterele text “PE” identifica inceputul acestui header in fisier. De acolo incepe structura IMAGE\_NT\_HEADERS.

Urmatoarele doua campuri din aceasta structura sunt de asemenea structuri:

- **FileHeader**, o structura de tipul **IMAGE\_FILE\_HEADER**, simpla, care contine informatii referitoare la proprietatile fisierului

- **OptionalHeader**, o structura complexa, de tipul **IMAGE\_OPTIONAL\_HEADER** care contine mai multe informatii. Desi se gaseste in orice fisier executabil sau in orice biblioteca de functii, aceasta structura poate fi optionala pentru fisierele obiect, cum ar fi rezultatul compilarii unui singur fisier

Campul "**FileHeader**", structura **IMAGE\_FILE\_HEADER** este definita astfel:



Structura **IMAGE\_FILE\_HEADER**

Informatiile continute de aceasta structura (20 de octeti) sunt urmatoarele:

- **Machine**: Un numar care identifica arhitectura procesorului, de exemplu i386, IA64...

Exemple:



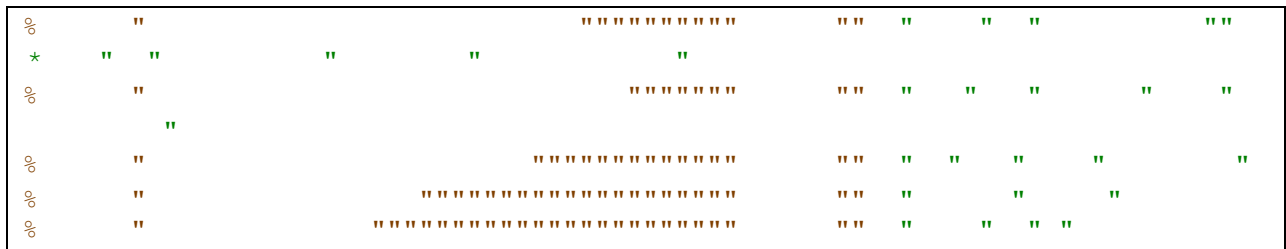
Exemple de arhitecturi ale procesorului pentru un fisier PE

In cazul exemplului nostru, imediat dupa primii 4 octeti care reprezinta semnatura PE (50 45 00 00), urmeaza 2 octeti care reprezinta valoarea acestui camp: 4c 01, adica numarul **0x014c**, ceea ce inseamna ca executabilul nostru este destinat arhitecturii procesoarelor **Intel 386** sau mai noi.

- **NumberOfSections**: Reprezinta numarul de sectiuni din fisier. Vom observa ca fisierele contin date impartite in sectiuni. In fisierul analizat de noi, imediat dupa cei 2 octeti care reprezinta campul "Machine", observam ca fisierul nostru contine **4** sectiuni.
- **TimeStamp**: Un numar pe 4 octeti (time\_t din C) care reprezinta data la care fisierul a fost creat, masurat in numarul de secunde trecute de la 1 Ian. 1970. Putem observa urmatoarea valoare pentru acest camp: **15 3b b8 50** adica numarul **0x50b83b15**, in zecimal 1354251029, ceea ce inseamna ca fisierul nostru a fost creat la data de "**Fri, 30 Nov 2012 04:50:29 GMT**".

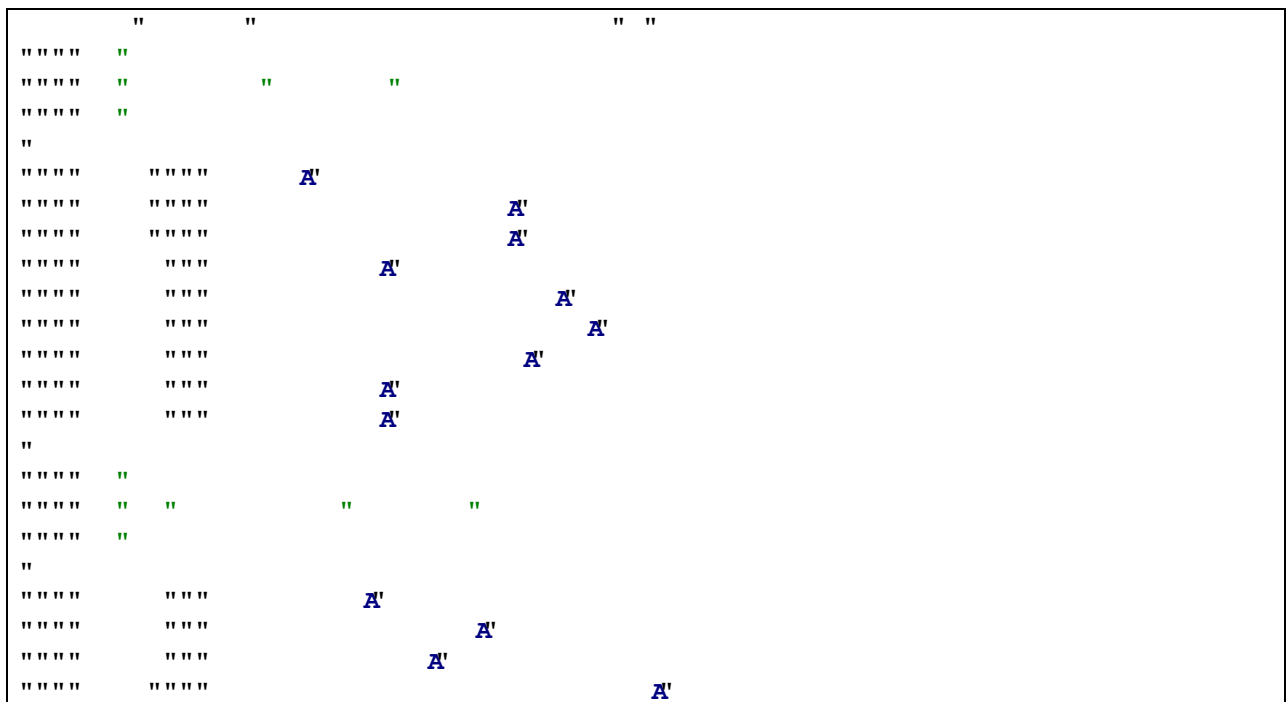
- **PointerToSymbolTable:** Acest camp ar trebui sa fie un pointer catre un tabel de simboluri COFF insa nu mai este folosit si ar trebui sa aiba valoarea 0.
- **NumberOfSymbols:** Numarul de simboluri din tabelul de mai sus care nu este folosit, asadar ar trebui sa aiba valoarea 0.
- **SizeOfOptionalHeader:** Contine marimea header-ului optional, cel care urmeaza imediat dupa aceasta structura
- **Characteristics:** Un set de flag-uri care indica informatii despre fisier: daca este executabil sau biblioteca de functii, sau altele.

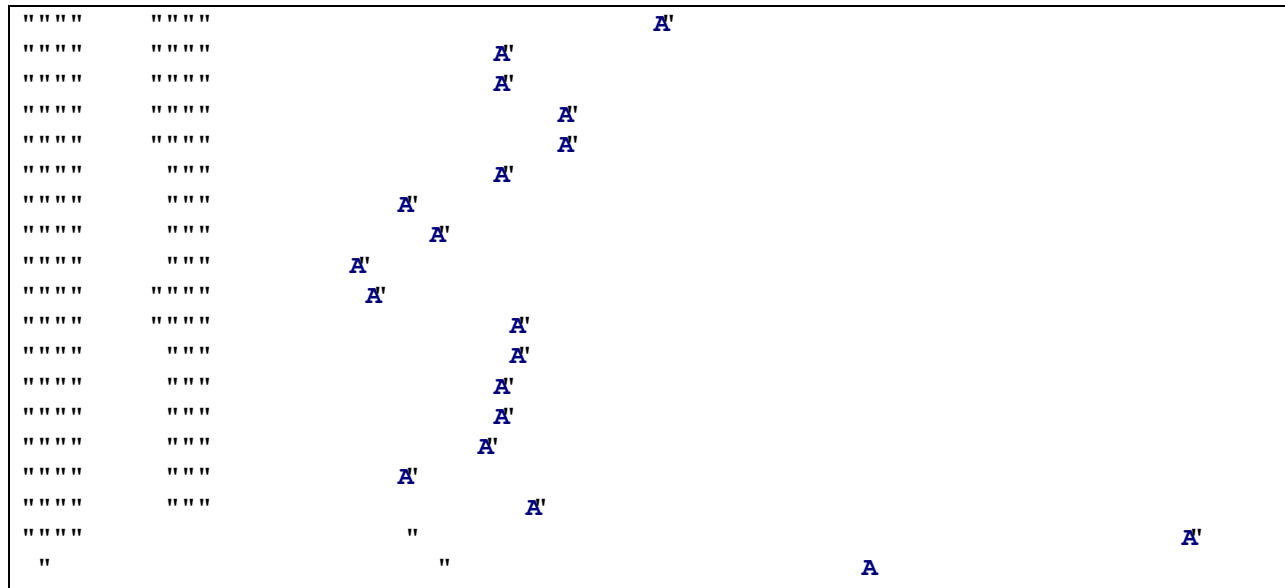
Exemple:



*Caracteristici posibile pentru un fisier PE*

Campul **OptionalHeader** (224 de octeti), structura **IMAGE\_FILE\_HEADER**, este definita astfel:





Structura IMAGE\_OPTIONAL\_HEADER

Intr-un fisier PE, un exemplu de astfel de structura, cei 224 de octeti din care este alcatuita:

	00	01	02	03	04	05	06	07	08	09	0a	0b	0c	0d	0e	0f	
00000080	63	8a	9f	9f	27	eb	f1	cc	27	eb	f1	cc	27	eb	f1	cc	cšÿÿ'ëñì'ëñì'ëñì
00000090	2e	93	62	cc	16	eb	f1	cc	27	eb	f0	cc	55	e8	f1	cc	."bì.ëñì'ësiUëñì
000000a0	2e	93	63	cc	26	eb	f1	cc	2e	93	64	cc	20	eb	f1	cc	."cìëñì."dì ëñì
000000b0	2e	93	72	cc	d1	eb	f1	cc	2e	93	75	cc	c4	eb	f1	cc	."rìÑëñì."uìÄëñì
000000c0	2e	93	65	cc	26	eb	f1	cc	2e	93	60	cc	26	eb	f1	cc	."eìëñì."ìëñì
000000d0	52	69	63	68	27	eb	f1	cc	00	00	00	00	00	00	00	00	Rich'ëñì.....
000000e0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
000000f0	50	45	00	00	4c	01	04	00	15	3b	b8	50	00	00	00	00	PE..L....;P....
00000100	00	00	00	00	e0	00	02	21	0b	01	09	00	00	50	0c	00	...à..!.....P..
00000110	00	e0	00	00	00	00	00	00	6f	cd	04	00	00	10	00	00	..à.....cí.....
00000120	00	00	0c	00	00	00	de	77	00	10	00	00	00	10	00	00	.....Pw.....
00000130	06	00	01	00	06	00	01	00	06	00	01	00	00	00	00	00	.....
00000140	00	40	0d	00	00	10	00	00	c7	ff	0d	00	03	00	40	01	..@.....Çÿ.....@.
00000150	00	00	04	00	00	10	00	00	00	00	10	00	00	10	00	00	.....
00000160	00	00	00	00	10	00	00	00	c0	51	0b	00	b1	a9	00	00	.....ÀQ..±@..
00000170	74	fb	0b	00	f4	01	00	00	00	70	0c	00	28	05	00	00	tú..ó...p..(...
00000180	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
00000190	00	80	0c	00	b0	b0	00	00	b4	59	0c	00	38	00	00	00	..e.."".'Y..@...
000001a0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
000001b0	00	00	00	00	00	00	00	00	90	28	08	00	40	00	00	00	..... (.@...
000001c0	00	00	00	00	00	00	00	00	00	10	00	00	fc	0d	00	00	.....ú...
000001d0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
000001e0	00	00	00	00	00	00	00	00	2e	74	65	78	74	00	00	00	.....text...
000001f0	15	4a	0c	00	00	10	00	00	00	50	0c	00	00	10	00	00	..J.....P.....
00000200	00	00	00	00	00	00	00	00	00	00	00	00	20	00	00	60	.....

Partea selectata reprezinta structura IMAGE\_OPTIONAL\_HEADER

Structura **IMAGE\_FILE\_HEADER** este alcatuita din urmatoarele campuri:

- **Magic:** Un numar pe 2 octeti care identifica practic tipul de executabil, de cele mai multe ori PE32 (valoarea 0x10b) dar poate fi si PE32+ (valoarea 0x20b), executabile ce permit adresarea pe 64 de biti in timp ce limiteaza dimensiunea executabilului la 2 GB
- **MajorLinkerVersion, MinorLinkerVersion:** Dupa cum spune numele, aceste campuri de cate un octet identifica versiunea link-erului folosita. Cum aceasta structura, IMAGE\_FILE\_HEADER nu se regaseste in fisierele obiect, inseamna ca structura e creata la linkare, de catre linker

Vom observa ca datele dintr-un fisier PE sunt impartite in sectiuni. Astfel avem sectiuni pentru cod (cel mai adesea o sectiune cu numele “.text”), sectiuni pentru date initializate sau date neinitializate (cel mai adesea numele “.bss”) care contin variabilele globale dar si variabilele statice, sectiuni pentru date “read-only” ca “.rdata” ce contin siruri de caractere sau constante.

- **SizeOfCode:** Indica dimensiunea sectiunii de cod (“.text”) sau dimensiunea tuturor sectiunilor de cod in cazul in care sunt mai multe
- **SizeOfInitializedData:** Indica dimensiunea sectiunii de date initializate sau dimensiunea tuturor sectiunilor de date initializate in cazul in care sunt mai multe
- **SizeOfUninitializedData:** Indica dimensiunea sectiunii de date neinitializate sau dimensiunea tuturor sectiunilor de date neinitializate in cazul in care sunt mai multe
- **AddressOfEntryPoint:** Un camp foarte important, o valoare RVA (Relative Value Address) adica un pointer catre o locatie relativa la adresa la care este incarcat modulul (.exe, .dll) in memorie (vezi mai jos campul ImageBase). E adresa de unde se incepe executia codului pentru un

memorie, trebuie aliniate la dimensiunea specificata de acest camp, de cele mai multe ori 0x1000 (4096 de octeti – dimensiunea unei pagini de memorie). Astfel, chiar daca o sectiune ar trebui sa inceapa de la adresa 0x2001, acesta nu este un multiplu de 0x1000 si sectiunea va fi incarcata in memorie la adresa 0x3000.

- **FileAlignment:** De asemenea, ca si campul anterior, sectiunile trebuie sa fie aliniate si in fisierul de pe disc, la un multiplu de valoarea indicata de acest camp. Implicit are valoarea 512 bytes, dar poate fi orice putere a lui 2 intre 512 si 64K.
- **MajorOperatingSystemVersion, MinorOperatingSystemVersion:** Versiunea sistemului de operare destinat executie fisierului. De exemplu 5.1 e Windows XP, 6.0 e Windows Vista si 6.1 e Windows 7.
- **MajorImageVersion, MinorImageVersion:** Versiunea fisierului
- **MajorSubsystemVersion, MinorSubsystemVersion:** Versiunea subsystemului (Windows Console, Windows...) – Vezi campul “Subsystem”
- **Win32VersionValue:** Nu e folosit, ar trebui sa fie 0
- **SizeOfImage:** Dimensiunea imaginii incarcate in memorie, inclusiv headerele. Trebuie sa fie un multiplu de “SectionAlignment”
- **SizeOfHeaders:** Dimensiunea tuturor headerelor: header DOS, header PE, headerle sectiunilor, rotunjita la un multiplu al campului “FileAlignment”
- **Checksum:** O suma pe 4 octeti, verificata la incarcare pentru DLL-urile care sunt incarcate in timpul procesului de boot, DLL-urile incarcate intr-un proces critic si pentru toate driverele. Algoritmul este incorporat in ImagHelp.dll
- **Subsystem:** Subsistemul fisierului. De exemplu, poate fi:

☺	"		"	.....	"	"	)	"	"	"
☺	"		"	.....	"	"	"	"	"	"
"	"	"	"	.....	"	"	"	"	"	"
☺	"		"	.....	"	"	"	"	"	"

*Subsistemul sau tipul fisierului PE*

Mai exact, are valoarea 1 daca este un driver, valoarea 2 daca este un executabil grafic (foloseste ferestre) sau 3 daca este o aplicatie in linie de comanda.

- **DllCharacteristics:** Diverse trasaturi cum ar fi: poate fi mutat la o adresa diferita, nu foloseste SEH (Structured Exception Handler) sau e necesara fortarea verificarii integritatii
- **SizeOfStackReserve:** Ce dimensiune sa fie rezervata, pastrata pentru stiva folosita de executabil. Nu este alocata, ci este doar specificata ca dimensiune maxima
- **SizeOfStackCommit:** Cat spatiu sa fie alocat pe stiva initial, si cat spatiu sa fie alocat cand mai e necesara o alocare, pana cand se atinge dimensiunea specificata de campul anterior
- **SizeOfHeapReserve:** Cat spatiu sa fie rezervat pentru heap
- **SizeOfHeapCommit:** Cat spatiu sa fie alocat initial si cat sa fie alocat atunci cand este necesar, pana se ating dimensiunea campului anterior
- **LoaderFlags:** Camp rezervat, nefolosit, ar trebui sa fie 0
- **NumberOfRvaAndSizes:** Numarul de intrari in tabelul urmator, DataDirectory, cel mai adesea 16 (0x10). Inainte de a parcurge acest tabel trebuie verificata aceasta valoare
- **DataDirectory:** Un vector [NumberOfRvaAndSizes] de structuri de tipul "IMAGE\_DATA\_DIRECTORY".

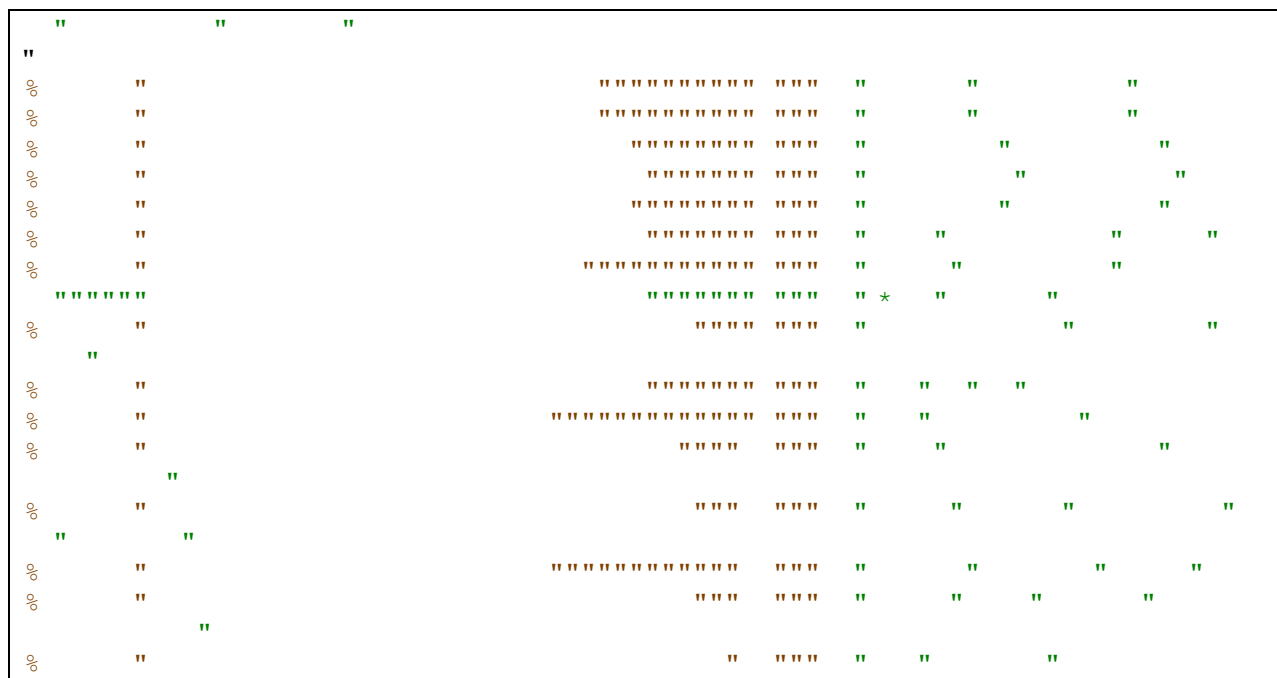
Campul **DataDirectory** contine de obicei **16** structuri de tipul **IMAGE\_DATA\_DIRECTORY**, structura definita astfel:



*Structura IMAGE\_DATA\_DIRECTORY*

Campul "VirtualAddress" reprezinta adresa RVA (relativa la ImageBase, adica adresa la care este incarcat in memorie fisierul nostru) unde se afla datele respective iar campul "Size" reprezinta dimensiunea datelor.

Datele continute de acest **vector**, de DataDirectory, sunt predefinite. Astfel, indecsii elementelor de tipul IMAGE\_DATA\_DIRECTORY din vector sunt urmatorii:



*Indicii elementelor din structura DataDirectory*

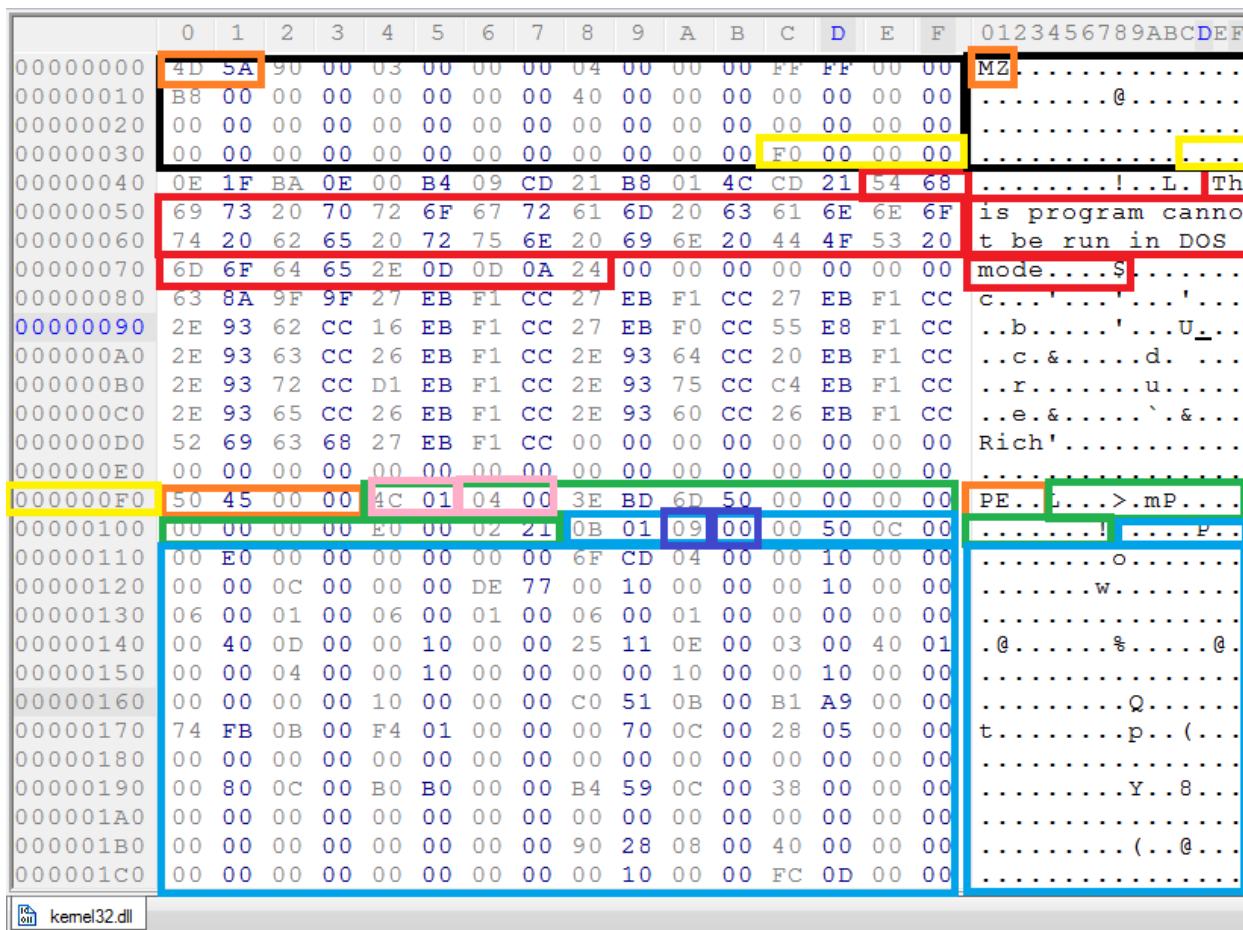
Mai exact, primul element din vector (indicele 0) va contine tabelul de functii exportate, urmatorul va contine functiile importate, urmatorul resursele continute de fisier etc. Elementele din vector nu sunt obligatorii, daca un element nu este prezent, atunci va avea valoarea 0.

In cazul in care v-ati pierdut printre atatea structuri, aveti mai jos o imagine de ansamblu a acestor headere: structurile si cateva campuri luate ca exemplu.

Datele colorate in aceasta imagine sunt:

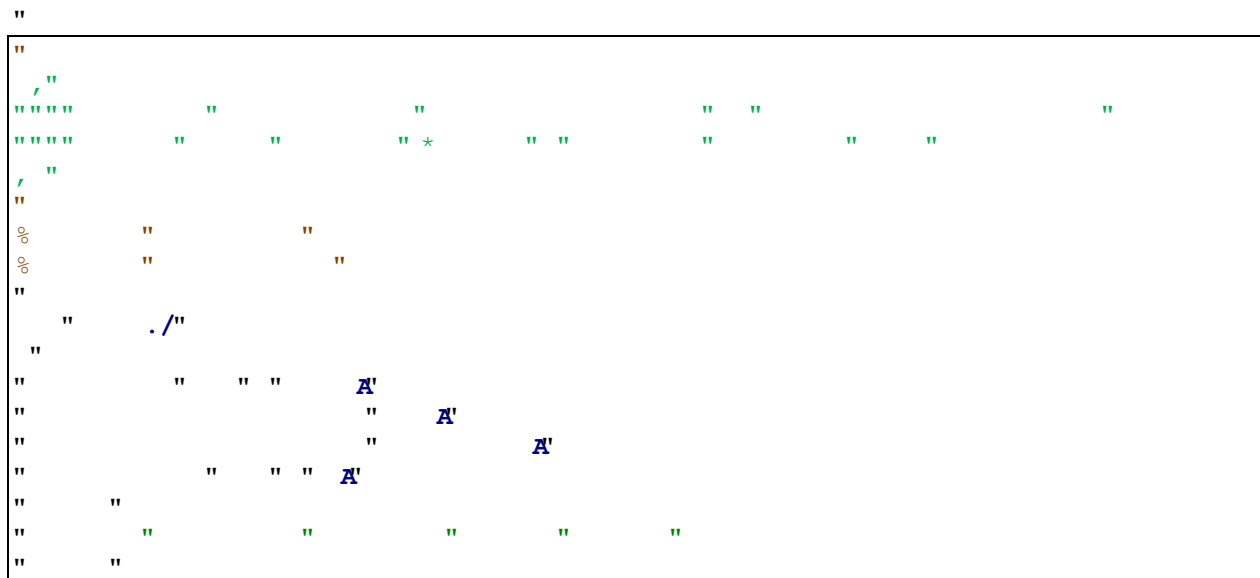
1. Portocaliu (cel de sus): primii 2 octeti din fisier, campul "e\_magic" din structura IMAGE\_DOS\_HEADER
2. Negru: Headerul complet MS-DOS, intreaga structura IMAGE\_DOS\_HEADER
3. Galbel: Ultimii 4 octeti din headerul MS\_DOS, campul "e\_lfanew", practice locatia la care se afla headerul PE, dupa cum se poate vedea mai jos, tot cu galben, acolo incepe headerul PE
4. Rosu: Sirul de caractere afisat de programul MS-DOS
5. Portocaliu (cel de jos): primii 4 octeti din header-ul PE, campul Signature din structura IMAGE\_NT\_HEADERS, indica inceputul headerelor PE
6. Verde: Structura IMAGE\_FILE\_HEADER, imediat dupa campul prezentat mai sus
7. Roz: Doua campuri din structura IMAGE\_FILE\_HEADER, primele doua elemente ale structurii
8. Albastru: Structura IMAGE\_OPTIONAL\_HEADER (incompleta, nu a incaput tot)
9. Mov: Ca exemplu, elementele "Major/minor linker version", ceea ce indica versiunea 09.00





*Imaginea de ansamblu a unui fisier PE*

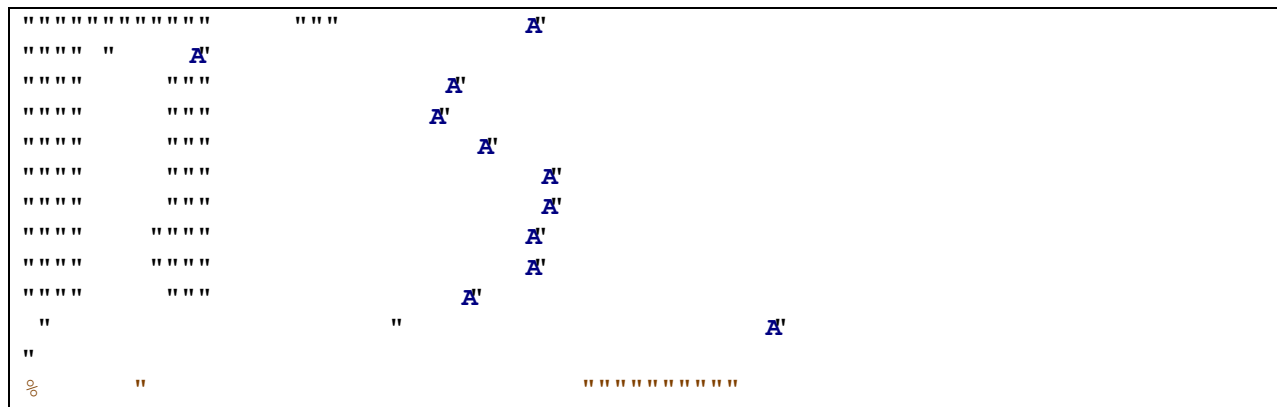
Putem verifica ce contin fiecare dintre aceste campuri cu un program simplu, care afiseaza datele din FileHeader, din OptionalHeader dar si vectorul DataDirectory.



" " " /A'  
" /"  
" " " # /A'  
" " A'  
" " " /"  
" " " " /" /A'  
" " " //"  
" " " " # /A'  
" " A'  
" " " " "  
" " " " /" " /"  
" " " " " " " /A'  
" " A'  
" " " " " "  
" " " " " " "  
" " " " " " " "  
" " " " " " " /" /A'  
" " " " " " //"  
" " " " " " # /A'  
" " A'  
" " /A'  
" " " " " "  
" " " " " " //A'  
" " " " " " //A'  
" " " " " " "  
" " " " " " "  
" " " " " " " /A'  
" " " " " " " /A'  
" " " " " " " /A'  
" " " " " " " /A'







*Structura IMAGE\_SECTION\_HEADER*

In aceasta structura, gasim urmatoarele campuri:

- **Name:** Numele sectiunii, un sir de maxim 8 caractere. Atentie, sirul nu se termina obligatoriu cu NULL!
- **VirtualSize:** Dimensiunea sectiunii atunci cand este incarcata in memorie
- **VirtualAddress:** Adresa RVA (relativa la ImageBase) la care incepe sectiunea
- **SizeOfRawData:** Dimensiunea sectiunii in fisier, trebuie sa fie un multiplu al campului "fileAlignment" din OptionalHeader
- **PointerToRawData:** Adresa RVA la care incepe sectiunea in fisier
- **PointerToRelocations:** Ar trebui sa fie 0 pentru fisiere executabile, sau o alta valoare in cazul in care la incarcarea in memorie trebuie sa se faca "relocari". Relocarile pot fi necesare pentru DLL-uri. In cazul in care un DLL nu se poate incarca la adresa preferata, exista posibilitatea ca anumite adrese sa fie modificate pentru ca corespunde acestor modificari, acest lucru fiind realizat printr-un tabel de adrese care trebuie relocate in astfel de situatii
- **PointerToLinenumbers:** Informatii de debug care nu se mai folosesc, ar trebui sa aiba valoarea 0
- **NumberOfRelocations:** Daca sunt necesare relocari, aici se va putea gasi numarul de adrese care trebuiesc "relocate"
- **NumberOfLinenumbers:** Informatii de debug care nu se mai folosesc, ar trebui sa aiba valoarea 0

- **Characteristics:** Campul este o masca de biti care specifica mai multe detalii despre sectiune: daca aceasta contine cod, date initializate sau date neinitializate, permisiunile pentru sectiune cand aceasta este incarcata in memorie si multe altele

Cateva exemple de biti care pot fi setati pentru o sectiune:

0%	"	.....	" "	" "	" "
0%	"		.....	" "	" "
0%	" "	" "		" "	" "
0%	"		.....	" "	" "
0%	"	" "		" "	" "
0%	"		.....	" "	" "
0%	" "	" "		" "	" "
0%	"		.....	" "	" "
0%	"	" "		" "	" "
0%	"		.....	" "	" "
0%	"	" "		" "	" "
0%	"		.....	" "	" "
0%	"	" "		" "	" "

*Caracteristici posibile pentru sectiuni*

Pentru a intelege mai usor acest tabel de sectiuni, tabel care incepe imediat dupa headerele PE (IMAGE\_NT\_HEADERS), am detaliat fiecare sectiune. Se poate observa, ca si in structura, ca primii 8 octeti ai fiecari sectiuni (in tabel), il reprezinta numele.

Culorile reprezinta:

1. Rosu: sectiunea ".text" este sectiunea de cod
2. Negru: sectiunea ".data" reprezinta sectiunea de date initializate
3. Verde: sectiunea ".rsrc" va contine resurse (adesea icon-ul fisierului si versiunea/copyright-ul fisierului, dar poate contine multe alte informatii)
4. Albastru: sectiunea ".reloc" va contine acele "relocari" in cazul in care fisierul nostrul (in exemplu de fata o biblioteca .dll) nu poate fi incarcata la adresa specifica

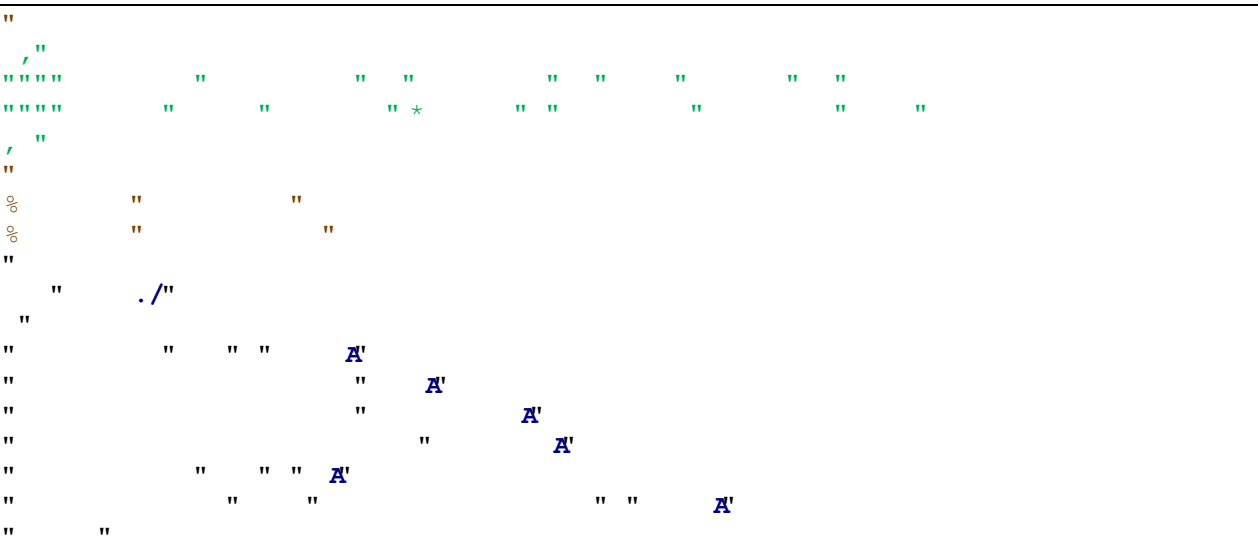
Pe langa aceste sectiuni, pot sa apara si altele:

4. Sectiunea ".bss" – date neinitializate globale
5. Sectiunea ".rdata" – date "read-only" sau alte sectiuni

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	
00000130	06	00	01	00	06	00	01	00	06	00	01	00	00	00	00	00	.....																
00000140	00	40	0D	00	00	10	00	00	25	11	0E	00	03	00	40	01	.@.....																
00000150	00	00	04	00	00	10	00	00	00	00	10	00	00	10	00	00	.....																
00000160	00	00	00	00	10	00	00	00	C0	51	0B	00	B1	A9	00	00	.....																
00000170	74	FB	0B	00	F4	01	00	00	00	70	0C	00	28	05	00	00	t.....																
00000180	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....																
00000190	00	80	0C	00	B0	B0	00	00	B4	59	0C	00	38	00	00	00	.....																
000001A0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....																
000001B0	00	00	00	00	00	00	00	00	90	28	08	00	40	00	00	00	.....																
000001C0	00	00	00	00	00	00	00	00	00	10	00	00	FC	0D	00	00	.....																
000001D0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....																
000001E0	00	00	00	00	00	00	00	00	2E	74	65	78	74	00	00	00	.....																
000001F0	15	4A	0C	00	00	10	00	00	00	50	0C	00	00	10	00	00	.J.....																
00000200	00	00	00	00	00	00	00	00	00	00	00	00	20	00	00	60	.....																
00000210	2E	64	61	74	61	00	00	00	F0	0F	00	00	00	60	0C	00	.data.....																
00000220	00	10	00	00	00	60	0C	00	00	00	00	00	00	00	00	00	.....																
00000230	00	00	00	00	40	00	00	C0	2E	72	73	72	63	00	00	00	.....@...																
00000240	28	05	00	00	00	70	0C	00	00	10	00	00	00	70	0C	00	(...p.....																
00000250	00	00	00	00	00	00	00	00	00	00	00	00	40	00	00	40	.....@..@																
00000260	2E	72	65	6C	6F	63	00	00	B0	B0	00	00	00	80	0C	00	.reloc.....																
00000270	00	C0	00	00	00	80	0C	00	00	00	00	00	00	00	00	00	.....@..B																
00000280	00	00	00	00	40	00	00	42	00	00	00	00	00	00	00	00	.....@..B																
00000290	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....																
000002A0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....																
000002B0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....																
000002C0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....																
000002D0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....																
000002E0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....																
000002F0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....																

Tabelul de sectiuni al fisierului "kernel32.dll"

Pentru a citi tabelul de sectiuni trebuie doar sa sarim peste headerele NT si citim "NumberOfSections" \* sizeof(IMAGE\_SECTION\_HEADER) octeti.







```
" " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
" " " " " " " " " " " " " " " " " "
```

Program care afiseaza tabelul de sectiuni al unui fisier PE

### 4. Concluzii

Articolul nu este nici pe departe complet, sunt prezentate doar notiunile de baza, intr-o maniera simpla si pe intelesul tuturor.

Pentru a putea intelege complet aceste notiuni sunt necesare notiuni de baza de C/C++, in special lucrul cu structuri, dar si o cunoastere de ansamblu a modului in care functioneaza un compilator.

Dupa cum se poate observa, nu este nici simplu dar nici foarte complicat sa intelegem cum functioneaza un fisier PE.