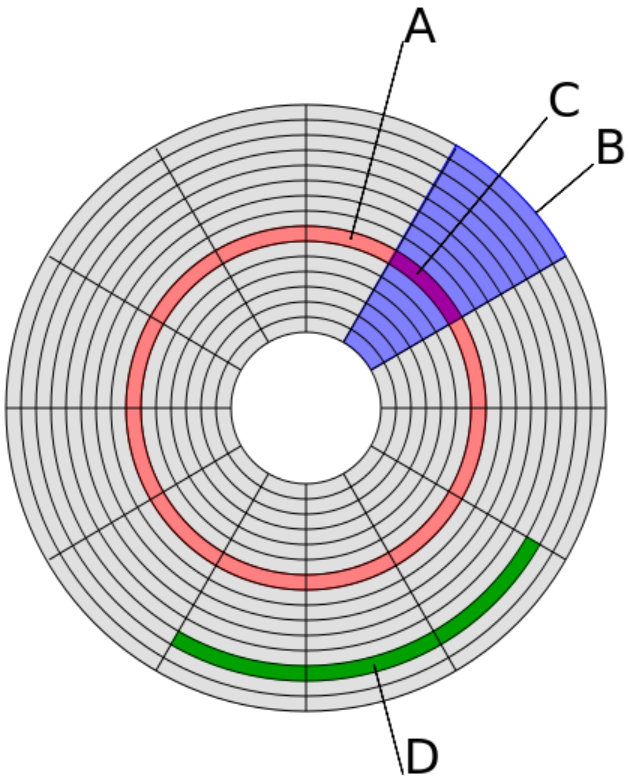
Hoy vamos a hablar del **sistema de archivos** de Windows NT, llamado **NTFS** (**new technology file system**). No está hecho desde cero, sino que está basado en otro sistema de archivos. Pero ¿Para qué hay que aprender ésto? Bueno, sobre todo para la informática **forense** donde podremos recuperar archivos y datos borrados o para ver huellas digitales.

Vamos desde un principio. Para almacenar información, el disco estará **dividido** en pequeñas proporciones, donde la **más pequeña** es denominada **clúster**. En este caso, el tamaño más pequeño de un clúster NTFS es de **512 Bytes**. Un archivo puede ocupar tantos como necesite, pero si es más pequeño que el **tamaño del clúster**, va a ocupar uno completo por lo menos. Es decir que aunque el archivo pese 4 Byte, en el disco va a ocupar un espacio de 512.

Otra cosa a tener en cuenta. La cantidad de éstos en cada disco, es limitado -no sé si dentro de cada partición o dentro de cada disco- a $\mathbf{2^{32}-1}$. Supuestamente soportarían hasta un espacio de 256 TeraByte, aunque dependen de la versión de NTFS.

Los nombres de los archivos dentro de un espacio NTFS están codificados en **Unicode**, y si queremos saber de éstos googleen pero lo básico sería que es **multilingüe** soportando muchos símbolos de distintas lenguas.



(imagen extraída de Wikipedia)
A – Pista
B – Sector geométrico
C – Sector de una pista
D – Grupo de Clústers

Algo interesante es que **Microsoft no liberó el código**, por lo que otros sistemas operativos tuvieron que realizar una técnica denominada **ingeniería inversa** para dar soporte de NTFS. Ingeniería inversa lo veremos más adelante pero es bastante entretenido y nos sirve para poder ver como funciona un mecanismo, desde su funcionamiento mismo.

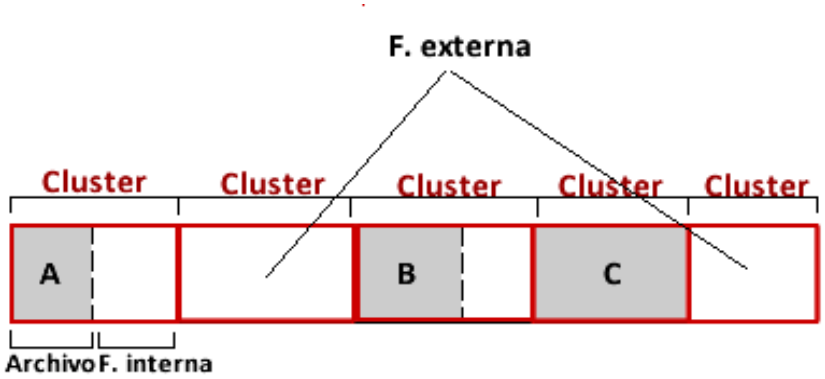
Bueno, volviendo al tema del clúster, hay 3 cosas que tenemos que saber:

- **LCN** (Logical Cluster Number): Es un número que indica una **posición** en el disco. Empieza con el cero.
- **VCN** (Virtual Cluster Number): De este no estoy seguro (corríjanme si me equivoco), pero se supone que es el **número de clúster** de un programa, siendo el primero cero, sin importar su lugar en el disco.
- **Extent**: Es una **seguidilla** de clústers **contigüos**. Pero como cada archivo puede tener divisiones por todos lados, un archivo pequeño puede tener **varios** Extent. Para los programadores, podemos ver esto como una estructura dinámica.

Siento que estoy yendo demasiado rápido y profundo para ustedes, así que voy a intentar de no alejarme demasiado.
Fíjense bien en lo que puse sobre el extent. Un archivo puede tener varios y claramente son discontínuos entre sí. Es como si una parte del archivo estuviese en el LCN 5, luego salta del 30 al 32, y luego hubiese otro extent de 100 a 500 para finalizar el archivo. Así que la computadora debería **encargarse** de ir de un LCN a otro para leer un archivo. Y seguramente ya lo saben, pero es

un proceso en el cual **perdemos velocidad**. Si un archivo fuese **sólo un extent**, sería más rápido y fácil. Esto es lo que logramos con la famosa **desfragmentación**, porque unimos todos los bloques para ponerlos **contínuos** en un disco y no usar demasiados recursos por todo este embrollo.

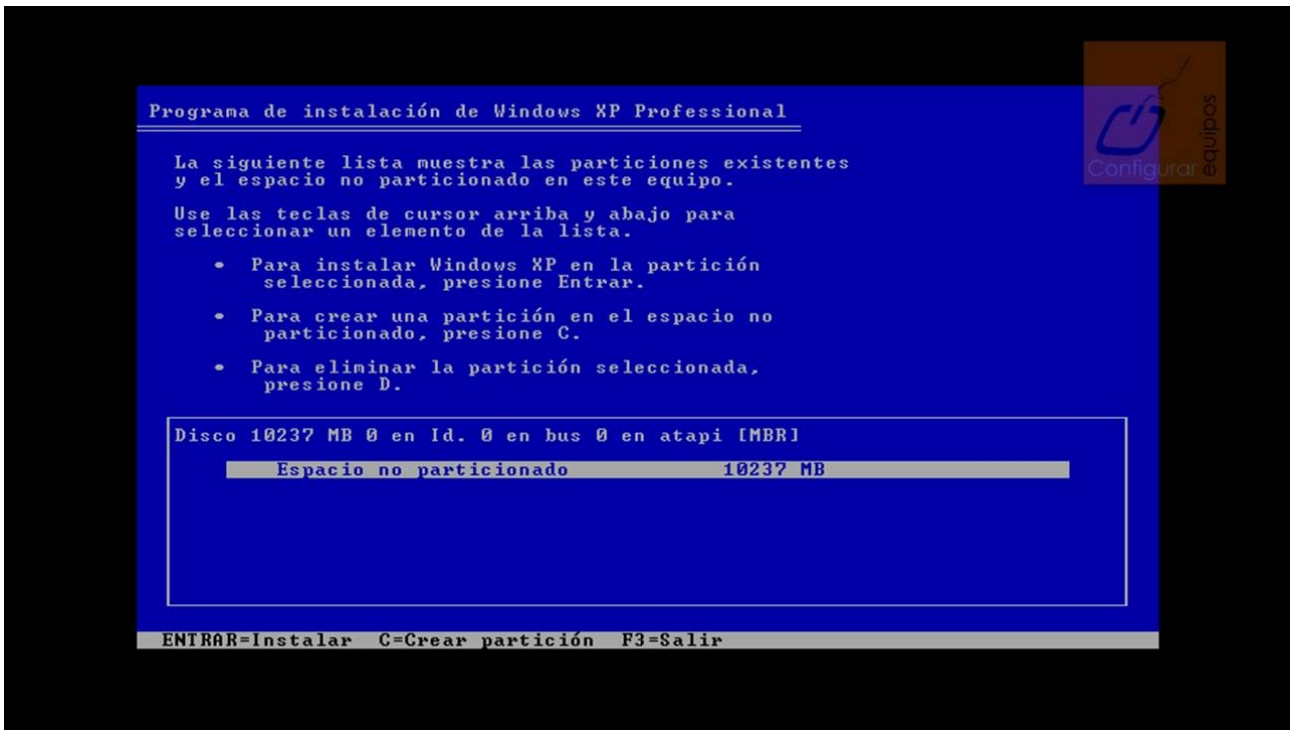
Y esa **fragmentación** que sucede, teniendo los extent desprendidos por todo el disco, se llama **fragmentacion externa**. La otra fragmentación se llama **fragmentación interna** y sucede cuando -como vimos anteriormente- un archivo pesa menos que el **tamaño menor** de un clúster. Esto último no se puede arreglar desfragmentando un disco, sino que debemos crear clústers de menor tamaño.



TIP: si van a desfragmentar, háganlo sin softwares dentro del sistema operativo porque así puede tocar los archivos del sistema. Sino, los archivos dentro de la carpeta de Windows y demases yerbas, no serían desfragmentados.

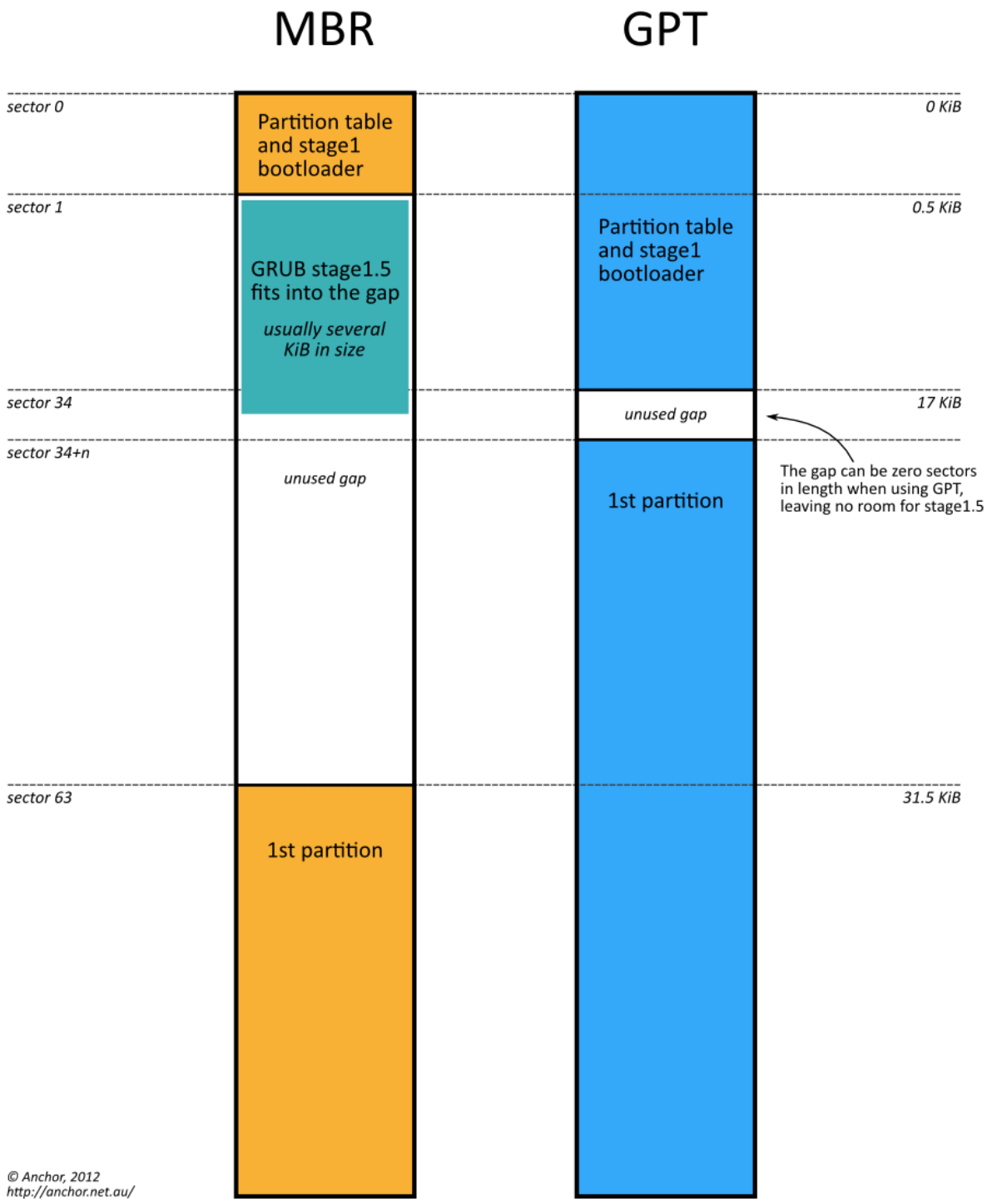
¿Qué pasa cuando formateamos? Me es difícil entender que es un punto a aclarar, pero la mentira es el oro de los tontos y por eso me pude encontrar con muchos técnicos que no saben la mitad de las cosas que hacen y hablan porque nadie les cobra por ello. Necesito desmentir para quienes no tienen en claro de qué se trata.

Formatear es una tarea simple. Digamos que un disco lleva cuenta de dónde están todos los archivos. Si tenemos 1TB y la mitad es un espacio lleno, el disco pondrá un **marcador** a los 500 GB, donde dice que para un lado tenemos espacio **ocupado** y para el otro tenemos espacio **libre** para que el sistema operativo pueda usarlo. Si de repente queremos **formatear** para poder tener más espacio libre, lo único que se hará es **cambiar la referencia** para decir que todo el disco es **utilizable**. ¿Entonces? Entonces que **la información no se borra**, sigue allí atormentándonos contra nuestra voluntad. Teóricamente, si sacamos el disco, lo formateamos desde afuera y no lo tocamos más, la información puede ser extraída con la misma facilidad que antes -aunque tengamos que entender esto, antes-.



Ahora, supongamos que el disco **se utiliza** luego de eso, como que instalamos un sistema operativo y algunos programas más. Le damos al análisis forense y nos encontramos que muchos de los archivos siguen íntegros, pero otros los vemos por la **mitad**. Sucede que el disco escribe encima de esos lugares porque lo ve como **disponible** y la información se irá **perdiendo** de a poco (ya veremos cómo recuperarla, más adelante ;)).

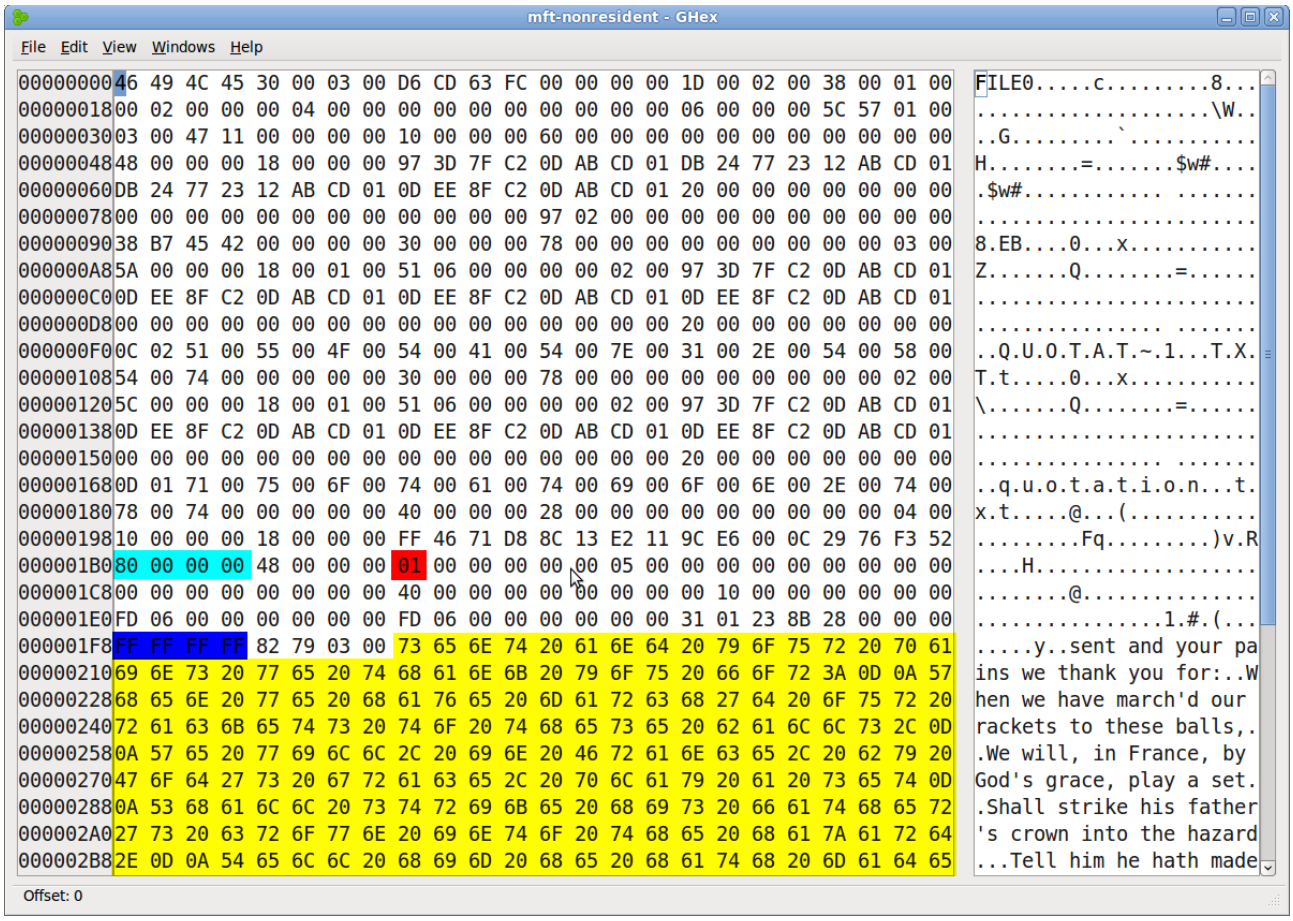
Pero el formateo no hace eso únicamente, y el disco no está **repleto de datos**. En los **primeros sectores del disco, partiendo desde el sector número cero, tenemos el MBR (Master Boot Record).** Se llama así porque es el punto donde una PC va a intentar arrancar o bootear todo el sistema y obviamente tenemos el **código de arranque**.
Aquí, en **Windows**, aparece la **tabla de particiones** donde nos da información de los arranques de los sistemas operativos alojados en el sistema. A parte, tenemos una firma del disco, para que sea identificado.
En Windows 8, tenemos algo nuevo denominado **GPT** (**Tabla de partición GUID**) y tiene que ver con un nuevo **BIOS** de **Intel**. No nos vamos a meter mucho ahora, pero es un supuesto arranque más seguro, que obviamente no tardo en ser **vulnerado**.



Y me falta explicar una sola cosa: Todos los archivos, directorios y metadatos de un disco están almacenados en el **MFT** (**Master file Table**). Windows siempre intentará llegar hasta la MFT para saber en donde leer un archivo específico. Algo así como una gran **base de datos**.
Por cada uno de los ficheros, existe un registro en la **MFT y sólo uno**.

Si tenemos una foto, irá a un registro de la MFT a ver por qué parte del disco se encuentran los datos y metadatos de dicho archivo, si tenemos acceso de lectura y otras cosas. Si eliminamos esta foto, el espacio de la MFT queda liberado y disponible para su uso. Mientras vamos creando y

guardando más información, el MFT queda desfragmentado y agrandado por todo el disco.



Muy bien. Quizás todavía no estamos en tema demasiado porque la práctica la vemos un poco lejos, pero ya aplicaremos todo el conocimiento.
Muchas gracias por todos los que leen y siguen en tema. Estoy muy contento de haber llegado a esta instancia a casi 8 meses de haber empezado este curso.

Pueden seguirme en Twitter: @RoaddHDC

Cualquier cosa pueden mandarme mail a: r0add@hotmail.com

Para donaciones, pueden hacerlo en bitcoin en la dirección siguiente:
1HqpPJbbWJ9H2hAZTmpXnVuoLKkP7RFSvw

Roadd.
