

Márcio Fernandes Justino

***Fragmentação de Arquivos em File Carving em
Sistemas de Arquivos NTFS***

São Paulo – SP

Maio / 2013

Márcio Fernandes Justino

***Fragmentação de Arquivos em File Carving em
Sistemas de Arquivos NTFS***

Projeto de pesquisa apresentado como requisito parcial para a aprovação na disciplina Metodologia do Trabalho Científico do Curso de Computação Forense da Universidade Presbiteriana Mackenzie.

Orientadora:
Ivete Irene dos Santos

UNIVERSIDADE PRESBITERIANA MACKENZIE
INSTITUTO DE COMPUTAÇÃO
PÓS GRADUAÇÃO EM COMPUTAÇÃO FORENSE

São Paulo – SP

Maio / 2013

...

Prof. xxx
Departamento xxx
Orientador

Prof. yyy
Departamento yyy

Prof. zzz
Departamento zzz

*Dedico a meus pais, cujo exemplo
de honestidade e trabalho tem marcado
minha vida, à minha esposa que me apoiou
nesta caminhada e à minha filha que
acompanhou todo este trabalho
ainda no ventre da mãe.*

Resumo

A proposta desta pesquisa é explanar e determinar uma melhor forma de localização de fragmentos de arquivos não alocados durante o processo de file carving em uma perícia forense digital abordando o conceito da metodologia de identificação de fragmentos de arquivos e os benefícios que o mesmo proporciona para a análise em uma investigação de uma imagem digital quando possível localizar suas partes, permitindo assim sua identificação.

Palavras-chave: NTFS, fragmentação, carving, identificação, partes, arquivos.

Abstract

...

Sumário

Lista de Figuras	p. vii
Lista de Tabelas	p. viii
1 Introdução	p. 1
1.1 Justificativa	p. 2
1.2 File Carving	p. 3
1.3 Fragmentação de Arquivos	p. 3
1.4 Hipótese(s)	p. 4
1.5 Objetivo Geral	p. 4
1.6 Objetivo Específico	p. 4
1.7 Metodologia	p. 5
2 Sistema de Arquivos	p. 6
2.1 Discos Rígidos	p. 6
2.2 O Sistema NTFS	p. 7
2.3 Estrutura de Dados	p. 7
2.3.1 Cluster	p. 8
2.3.2 Master File Table	p. 9
2.4 Alocação de Arquivos	p. 10
2.4.1 Registros da MFT	p. 10
2.4.2 Registros de Arquivos	p. 12
2.4.3 Registros de Diretório	p. 12

2.4.4	Remoção de Arquivos	p. 13
2.5	Análise NTFS	p. 14
3	File Carving	p. 15
3.1	Assinatura de Arquivo	p. 15
3.2	Arquivos Fragmentados	p. 16
3.2.1	Fragmentação Linear	p. 17
3.2.2	Fragmentação Não Linear	p. 17
3.2.3	Arquivo Parcial	p. 17
3.2.4	Recuperação de Arquivos Fragmentados	p. 18
3.3	Aspectos de Carving	p. 18
3.3.1	Cabeçalho e Rodapé	p. 18
3.3.2	Carving em Estrutura de Arquivo	p. 20
3.4	Número Mágico	p. 20
4	File Carving Avançado	p. 21
4.1	Diferenciação	p. 21
4.2	Fragmentação	p. 21
4.3	Ponto de Fragmentação	p. 21
5	Ferramentas de File Carving	p. 22
5.1	Scalpel	p. 22
5.2	Foremost	p. 22
6	Considerações Finais	p. 23
	Referências Bibliográficas	p. 24
	Anexos	p. 25

Lista de Figuras

1.1	Internautas ativos em residências e no trabalho e horas navegadas - 2012 [IBOPE//NetRatings 2012]	p. 1
1.2	Hipótese de pesquisa	p. 4
2.1	Disco rígido moderno [Wikipédia 2013]	p. 6
2.2	Entrada na MFT - Cabeçalho e espaço reservado aos diferentes tipos de atributos. Para o exemplo, a entrada possui 3 atributos.	p. 8
2.3	Estrutura do NTFS	p. 8
2.4	Registro MFT - Representação da estrutura da MFT [T. 2013].	p. 10
2.5	Registro MFT - Representação de um atributo e o tamanho de dados.	p. 10
2.6	Registro de arquivo da MFT [Svensson 2005].	p. 12
2.7	Registro de diretório	p. 13
2.8	Registro de diretório e a “Index Allocation Buffer”.	p. 13
3.1	Exemplo de Fragmentação Linear	p. 17
3.2	Exemplo de Fragmentação Não Linear	p. 17
3.3	Estrutura de cabeçalho de um arquivo do tipo PNG [FileFormat 2013]	p. 19
3.4	Área de dados de um arquivo do tipo PNG [Kloet 2007]	p. 19

Lista de Tabelas

2.1	Tamanho do Cluster Padrão Para Formatos NTFS.	p. 8
2.2	Registros reservados do NTFS na Master File Table.	p. 9
2.3	Registros reservados do NTFS na Master File Table.	p. 11

1 Introdução

Juntamente com o avanço da tecnologia computacional e da internet, veio o aumento do número de pessoas conectadas trocando informações, seja em nível pessoal ou organizacional. Segundo o Centro de Estudos sobre as Tecnologias da Informação e da Comunicação (cetic.br), o número de usuários domésticos e no trabalho tem aumentado juntamente com o tempo em que os mesmos permanecem conectados à internet. A figura 1.1 mostra a evolução desses números até o presente momento.

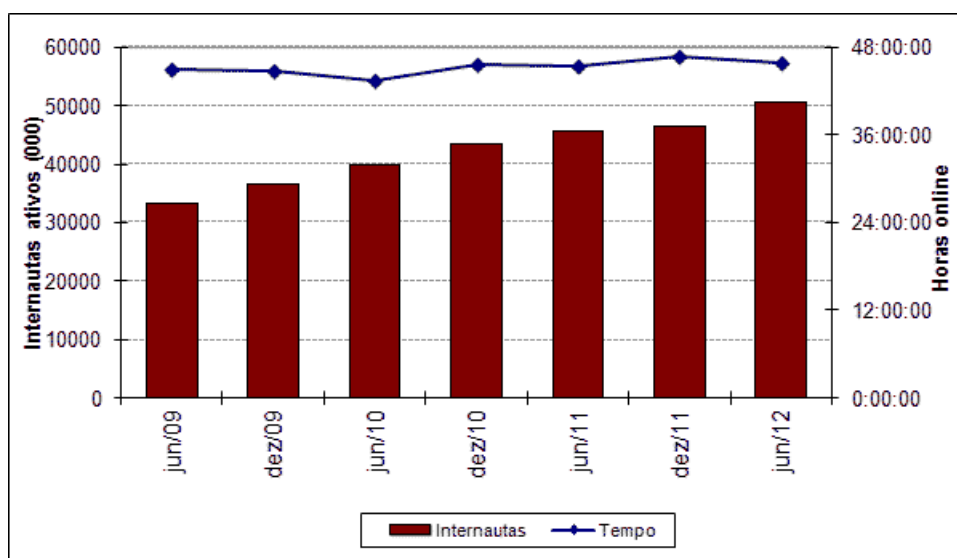


Figura 1.1: Internautas ativos em residências e no trabalho e horas navegadas - 2012 [IBOPE//NetRatings 2012]

Nesse meio, existem usuários que promovem o cibercrime¹ ou atividades ilegais na rede². As informações computacionais são armazenadas em discos rígidos³ usando apropriados sistemas de arquivos que são suportados pelo sistema operacional instalado no computador. Existem diversos sistemas de arquivos para armazenamento de arquivos no mercado, e um dos mais comuns atualmente é o NTFS [Mahant 2012].

¹ crimes cibernéticos tendo sistemas informatizados como meio de ação [CyberCitizen 2012].

² o termo rede será usado ao longo deste texto podendo representar a internet como um todo ou a ligação de mais de computadores entre si.

³ unidade física de armazenamento de dados em um computador

“O sistema de arquivos NTFS, assim como outros sistemas de arquivos, não foi criado com a computação forense em mente mas sim com a quantidade de informações no computador que pode ser usada em uma investigação” [Svensson 2005, p. 32]

Muitos criminosos se utilizam do artifício de excluir ou remover rastros de seus atos criminosos apagando os arquivos criados, manipulados ou alterados, acreditando que com isso seu crime seria perfeito, sem rastros da comprovação de seus atos. A técnica de file carving possibilita a análise de tais arquivos, provendo um avanço investigativo com a possibilidade de extrair provas de arquivos não mais alocados, porém que ainda estejam presentes fisicamente nos discos.

Segundo Caloyannides (2004, p. 26), a operação de remoção de um arquivo não faz absolutamente nada. Esta meramente altera um simples caractere na tabela de alocação do arquivo em questão indicando ao computador que o espaço desse arquivo foi tomado permitindo que os dados do arquivo possam ser sobrescritos no futuro, se necessário. Seguindo raciocínio semelhante, a operação de formatação não remove os dados sensíveis dos arquivos. A formatação faz com que os ponteiros da tabela de alocação que indicam onde os arquivos estão sejam liberados, perdendo assim sua localização pelo sistema de arquivos, mas mantendo os dados intactos em seu sistema. De acordo com Caloyannides (2004, p. 33), os arquivos são alocados no disco em unidades mínimas chamadas clusters⁴. Se o sistema de arquivos não necessita de todo o cluster para armazenar as informações do arquivo, este irá marcar o final do arquivo na porção final de seus dados dentro do cluster, deixando uma porção do cluster com dados considerados lixo, não sobrescritos pelo sistema. Essa situação gera o que é chamado de slack space⁵. Por fim, o processo de formatação e remoção de um arquivo, juntamente com as áreas de slack space do disco, provoca o surgimento de áreas não alocadas, o que caracteriza o processo de file carving.

1.1 Justificativa

Segundo Memon (2011, p. S2), um dos primeiros desafios em file carving pode ser encontrado na tentativa de se recuperar arquivos fragmentados. O processo de file carving é de suma importância para a investigação forense computacional e envolve a identificação de arquivos perdidos, corrompidos ou removidos do equipamento investigado. A dispersão desses arquivos não mais indexados pela tabela de alocação de arquivos do sistema de arquivos NTFS torna o processo de identificação dos arquivos um desafio para a investigação e identificação de ilícitos.

⁴Unidade mínima que pode ser acessada pelo sistema de arquivos, constituído de setores e de tamanho variável dependendo do sistema de arquivos e da opção de formatação.

⁵Espaço não alocado em um cluster que pode conter informações de outros arquivos que não foram sobrescritas.

“Durante uma investigação forense digital, muitas peças diferentes de dados são preservadas para investigação, das quais imagens de discos rígidos (HD’s) são as mais comuns. Essas imagens contêm os dados alocados para arquivos, bem como os dados não alocados. Os dados não alocados ainda podem conter informações relevantes para uma investigação, sob a forma de (partes de) intencionalmente excluídos ou arquivos temporários removidos automaticamente. Infelizmente, esses dados nem sempre são facilmente acessíveis: uma sequência de caracteres da pesquisa sobre os dados brutos pode recuperar (partes de) documentos de texto interessantes, mas ele não vai ajudar para obter a informação presente em, por exemplo, imagens ou arquivos compactados. Além disso, as sequências de caracteres exatas para procurar não podem ser conhecidas antecipadamente. Para obter esta informação, os arquivos apagados precisam ser recuperados” [Kloet 2007, p. 1].

1.2 File Carving

A técnica de file carving é frequentemente utilizada durante investigações digitais. Conforme Memon (2011, p. S2), essa é uma técnica em que arquivos de dados são extraídos de um dispositivo digital sem o auxílio de tabelas de arquivo ou outros meta-dados do disco.

“Na pratica forense, o processo de file carving pode recuperar arquivos que foram removidos e que tiveram sua entrada de diretório realocada para outros arquivos, desde que seus setores de dados não tenham sido sobrescritos ainda” [Garfinkel 2007].

1.3 Fragmentação de Arquivos

Conforme Menom (2011, p. S2), um dos primeiros desafios do processo de investigação utilizando file carving é justamente a tentativa de recuperar os fragmentos de arquivos não alocados. A fragmentação de arquivos é um desafio para o processo de file carving e é de suma importância para a recuperação de arquivos perdidos em processos de investigação digital. Tendo em vista o presente desafio, tem-se a necessidade de se determinar qual a melhor técnica para identificação de fragmentos de arquivos no processo de file carving.

1.4 Hipótese(s)

Pretende-se com a análise dos metadados⁶ de arquivos (cabeçalho e rodapé), determinar um padrão de identificação de pontos de fragmentação de arquivos, visando a identificação, com maior consistência e confiabilidade, de suas partes, reduzindo consequentes falsos positivos apresentados, com certa frequência, durante o processo de file carving, prejudicando a identificação de possíveis evidências comprobatórias. A figura 1.2 demonstra as hipóteses de pesquisa:

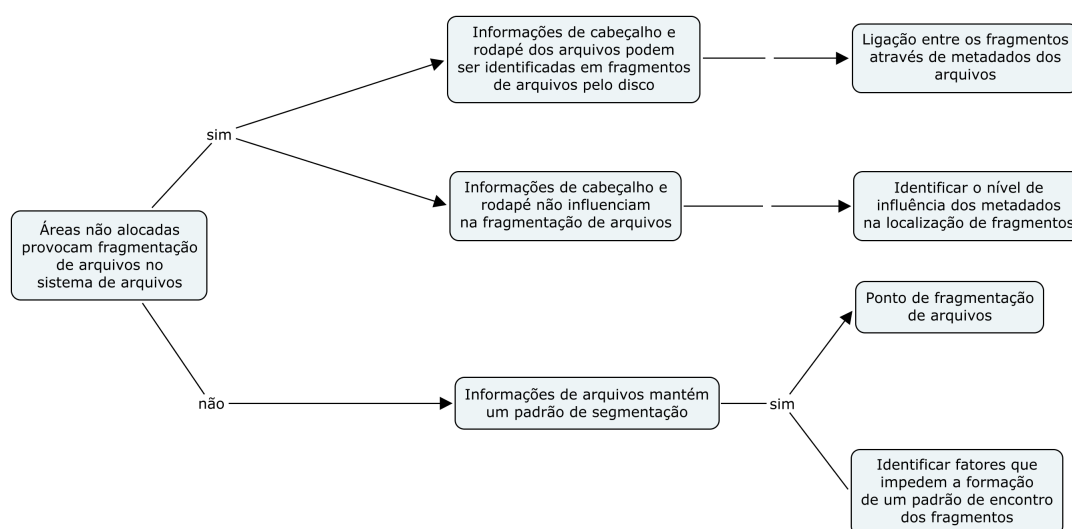


Figura 1.2: Hipótese de pesquisa

1.5 Objetivo Geral

Visto a necessidade de mitigar o impacto da fragmentação no processo de file carving em uma investigação digital, busca-se determinar uma melhor metodologia de localização dos fragmentos de arquivos, esclarecer uma maneira de localização dos fragmentos de um arquivo dentro da área não alocada em sistemas de arquivos NTFS.

1.6 Objetivo Específico

Para chegar ao objetivo principal e determinar uma melhor metodologia de localização de fragmentos de arquivos é necessário entender primeiramente e de forma mais detalhada alguns itens específicos:

- Verificar como são identificados os arquivos no sistema de arquivo NTFS.
- Verificar como um arquivo fragmentado é armazenado em um sistema NTFS.
- Levantar uma padronização entre os fragmentos de arquivos para melhor localização.
- Identificar formas de localização de fragmentos dos arquivos não alocados.
- Verificar o uso de entropia para localização de fragmentos de arquivos.

Verificar assim a forma como os arquivos são registrados nos sistemas de arquivos NTFS, o processo de diferenciação de tipos de arquivos para determinar o início e o fim de um arquivo (área de cabeçalho, área de dados, de metadados e ponto de fim de arquivo), podendo então encontrar certos padrões que possam permitir a identificação de partes de um arquivo fragmentado no sistema de arquivos.

1.7 Metodologia

Pesquisar e descrever o funcionamento do sistema de arquivos NTFS e identificar em sua estrutura de armazenamento de arquivos o ponto de fragmentação dos arquivos podendo dessa forma atingir o objetivo de determinar uma melhor forma de recuperação dos fragmentos de arquivos em áreas não alocadas, através das bibliografias levantadas juntamente com pesquisas já realizadas sobre assuntos correlatos à fragmentação de arquivos e file carving.

O método de pesquisa utilizado é o embasamento nas definições de documentação do sistema de arquivos NTFS junto ao criador (Microsoft), pesquisas correlatas referentes à fragmentação de arquivos e sua forma de identificação, pesquisar a ligação entre os fragmentos de dados para entender o ponto de fragmentação e identificar se há um padrão nessa relação de fragmentos de arquivos.

Para fins comprobatórios, serão utilizados dois softwares amplamente utilizados com a finalidade de identificação de arquivos não alocados que serão testados sobre os mesmos parâmetros para que se possa concretizar, através de dados significativos, a eficácia do processo de identificação de fragmentos de arquivos não alocados. As ferramentas em questão utilizadas são o Foremost e Scalpel.

2 *Sistema de Arquivos*

2.1 Discos Rígidos

Os discos rígidos (*HD's*) são usados para armazenar informações *não voláteis*¹ como programas, documentos, planilhas, anexos de e-mail, arquivos de sistema e outros criados pelo usuário.

Um disco rígido é composto por uma ou mais partições lógicas, representadas por um volume que pode ser formatado utilizando um sistema de arquivo, seja esse um sistema FAT, NTFS (*Computadores que utilizam Windows*), ext3, ext4 (*Computadores que utilizam Linux*), entre outros sistemas existentes atualmente [Svensson 2005].



Figura 2.1: Disco rígido moderno [Wikipédia 2013]

¹Refere-se a informações que permanecem em memória mesmo depois que a energia é cortada do dispositivo ou o mesmo é desligado. Já a informação volátil é armazenada em memória volátil, como a memória RAM, e que são perdidas quando a energia do dispositivo é cortada ou o mesmo é desligado.

2.2 O Sistema NTFS

Projetado pela Microsoft, o NTFS é o sistema de arquivos padrão de vários sistemas operacionais Microsoft desde a versão do Microsoft Windows NT (1993) até as versões mais recentes como o Windows8 (2012), com algumas evoluções ao longo do tempo. Sendo este o predecessor do antigo sistema FAT, é o sistema mais presente em investigações de sistemas devido a superioridade de utilização de sistemas operacionais Windows em todo o mundo.

O sistema de arquivos NTFS foi projetado para melhor confiabilidade, segurança e suporte de grandes dispositivos de armazenamento de dados. Dispõe do uso de estruturas genéricas que servem de envoltório para estruturas de dados com conteúdo específico. Desta forma, o NTFS se torna um projeto escalável pois a estrutura interna de dados pode mudar inúmeras vezes enquanto que a sua casca (a estrutura genérica) permanece constante. Um bom exemplo desse modelo é que todos os bytes² são alocados em arquivos no sistema [Carrier 2005].

2.3 Estrutura de Dados

Segundo Carrier (2005, p. 199), a única estrutura consistente no NTFS está presente nos primeiros setores do disco, contendo os setores de boot e código.

O coração do sistema de arquivos NTFS está na Master File Table (MFT³), pois esta contém as informações de todos os diretórios. Todo arquivo e diretório existente possui uma entrada na MFT, sendo que esta é uma estrutura bem simples de 1 KB de tamanho. A Entrada na MFT possui:

- Cabeçalho;
- Atributos; e
- Espaço não utilizado.

A figura 2.2 demonstra um registro de entrada na MFT.

²Termo binário que representa 8 unidades da menor unidade de informação (bit) que pode ser armazenada ou transmitida [Wikipédia 2012].

³Tabela de alocação de arquivos do sistema NTFS.

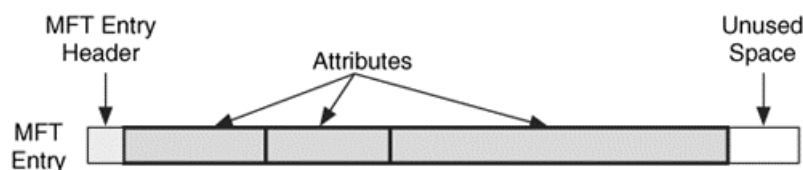


Figura 2.2: Entrada na MFT - Cabeçalho e espaço reservado aos diferentes tipos de atributos. Para o exemplo, a entrada possui 3 atributos.

2.3.1 Cluster

De acordo com Svensson (2005, p. 25), um disco rígido é dividido em setores sendo que o sistema de arquivos se utiliza de setores para formar um cluster. O tamanho do cluster depende da formatação do disco. A tabela 2.1 demonstra tamanhos padrões definidos de cluster para diferentes tamanhos e formas de formatação do disco NTFS:

Tamanho do Disco	Tamanho do Cluster Padrão
até 512 MB	512 bytes
513 MB - 1024 MB	1 KB
1025 MB - 2048 MB	2 KB
acima de 2048 MB	4 KB

Tabela 2.1: Tamanho do Cluster Padrão Para Formatos NTFS.

Como pode ser visto na tabela demonstrada por Svensson (2005, p. 25), quanto maior o tamanho do disco maior o tamanho do cluster.

Nos sistemas de arquivos NTFS, todos os clusters possuem um identificador chamado de *Logical Cluster Number* (LCN). O LCN é um número sequencial que representa todos os clusters do disco, do começo do disco até o seu final, iniciando-se em 0 (zero), referindo ao setor de boot do sistema de arquivos. O sistema de arquivo converte o identificador LCN em um endereço físico de disco (posição dos bytes onde o cluster está localizado no disco), multiplicando o identificador LCN com o tamanho do cluster.



Figura 2.3: Estrutura do NTFS

2.3.2 Master File Table

Segundo Svensson (2005, p. 26), todos os arquivos no sistema NTFS apresentam ao menos uma entrada no registro da *Master File Table* (MFT)⁴. Os primeiros 16 registros da MFT são reservados pelo sistema de arquivos para manter metadados⁵ específicos do NTFS. A tabela 2.2 detalha os 16 primeiros registros reservados ao NTFS na MFT:

Nome MTF	Registro	Descrição
\$MFT	0	NTFS's Master File Table. Contém um arquivo base para cada arquivo ou diretório do disco.
\$MFTMIRR	1	Uma cópia parcial da MFT que serve como backup para casos de falhas de um setor.
\$LOGFILE	2	Log de transação de arquivos.
\$VOLUME	3	Contém o número serial do volume e data de criação.
\$ATTRDEF	4	Definição de atributos.
.	5	Diretório raiz do disco.
\$BITMAP	6	Contém um mapeamento binário de todos os clustes do volume (usados e não usados).
\$BOOT	7	Registro de boot do drive ⁶ .
\$BADCLUS	8	Lista de setores com problemas no drive.
\$SECURE	9	Contém uma única descrição de segurança para todos arquivos do volume.
\$UPCASE	10	Mapeia caracteres em texto minúsculo para texto maiúsculo.
\$EXTEND	11	Extensões opcionais como as quotas ⁷ , pontos de reanálise de dados e identificador de objetos.
	12-15	Reservado para uso futuro.

Tabela 2.2: Registros reservados do NTFS na Master File Table.

O primeiro registro da tabela MFT é um registro referente à própria MFT em si. Esse registro é seguido por outro arquivo que representa a cópia parcial da MFT (um espelho da própria master file table, **\$MFTMIRR**). Esse registro de espelho da MFT funciona como um backup para importantes metadados de arquivos, usado para recuperação da MFT em caso de problemas.

O registro **\$BITMAP** que contém todo o mapeamento de clusters que estão em uso (alocados por arquivos). Todo cluster é identificado por um bit no registro **\$BITMAP**, sendo que quando o cluster está sendo usado o bit tem seu valor igual a um. Esse bit também pode ser encontrado em registros de diretórios com um propósito de manter um registro de quais clusters do buffer de alocação de índices estão em uso e quais não estão.

⁴Conjunto de registros que são continuamente modificados quando um arquivo ou diretório sofre alteração

⁵Dados que definem ou descrevem outra parte dos dados [Wyman Walt Scrivens 2012].

Já o registro **\$BADCLUS** é usado para manter uma lista dos clusters que são marcados como defeituosos pelo sistema operacional. Quando o sistema detecta um cluster com defeito ele registra esse cluster no **\$BADCLUS** e esse cluster passa a não ser mais utilizado pelo sistema.

A figura 2.4 demonstra a estrutura e formatação da *Master File Table*.

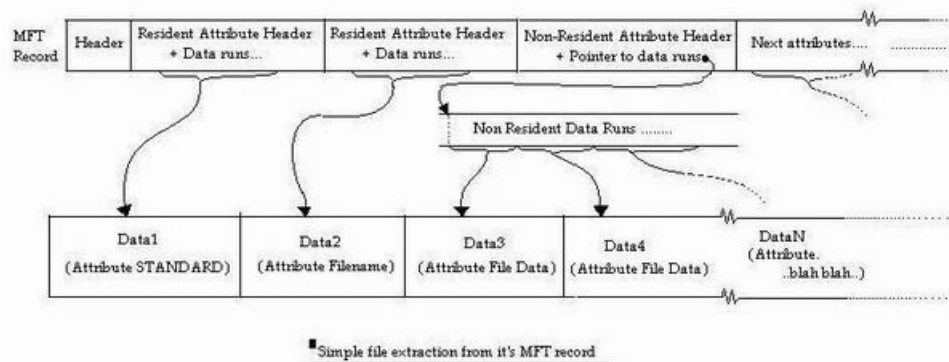


Figura 2.4: Registro MFT - Representação da estrutura da MFT [T. 2013].

2.4 Alocação de Arquivos

2.4.1 Registros da MFT

Os registros da MFT incluem inúmeros atributos e valores, sendo limitado em 1 KB fisicamente. Os atributos de registros são divididos em dois componentes lógicos - cabeçalho e área de dados - sendo que no cabeçalho são armazenados os tipos dos atributos, a localização e o tamanho do valor do atributo [Svensson 2005].

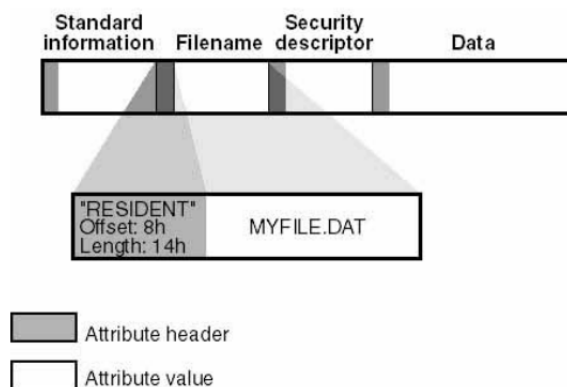


Figura 2.5: Registro MFT - Representação de um atributo e o tamanho de dados.

A tabela 2.3 define os inúmeros atributos do sistema de arquivos NTFS.

Tipo do Atributo	Descrição
\$VOLUME_INFORMATION	Versão do volume de dados.
\$VOLUME_NAME	Nome do volume de dados.
\$FILE_NAME	Nome do arquivo ou do diretório. Um arquivo pode ter múltiplos atributos de \$FILE_NAME.
\$STANDARD_INFORMATION	Atributos de arquivo (somente leitura, data de criação e de modificação, etc).
\$SECURITY_DESCRIPTOR	Presente para manter compatibilidade com outras versões do sistema de arquivos. Armazenava informações de segurança dos arquivos e diretórios.
\$DATA	Dados do arquivo.
\$INDEX_ROOT	Implementa a alocação do nome do arquivo.
\$INDEX_ALLOCATION	Implementa a alocação do nome do arquivo.
\$BITMAP	Índices bitmap para grandes diretórios (somente para diretórios).
\$OBJECT_ID	Usado pelo serviço de rastreamento de link distribuído.
\$REPARSE_POINT	Armazena um ponto de reanálise de dados.
\$ATTRIBUTE_LIST	Lista o local de todos os registros de atributos da MFT que não couberam no registro da MFT.
\$EA_INFORMATION, \$EA	Atributos estendidos de compatibilidade.
\$LOGGED_UTILITY_STREAM	A EFS (Encrypting File System - Essa tecnologia permite a encriptação de arquivos de forma transparente protegendo informações confidenciais de invasores com acesso físico ao computador [Wikipédia 2013]) armazena informações que são usadas na manipulação de arquivos encriptados.

Tabela 2.3: Registros reservados do NTFS na Master File Table.

A figura 2.7 mostra essa indexação.

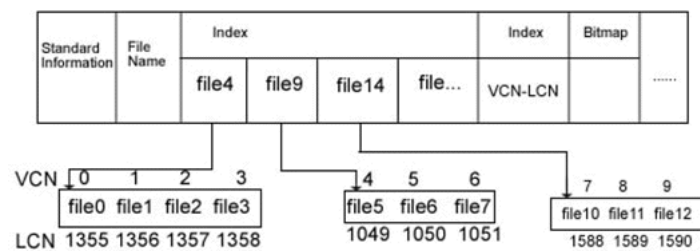


Figura 2.7: Registro de diretório .

As entradas de índice consistem do nome do arquivo e da cópia das informações padrões da entrada de diretório. Como o número de arquivos em diretórios pode crescer, o espaço pode não ser suficiente para armazenar todas as entradas de índice então, os índices adicionais são armazenados na área de alocação conhecida como “*Index Allocation Buffer*”. A área de bitmap também é usada para indicar qual VNCs presente no “*Index Allocation Buffer*” estão sendo usados.

A figura 2.8 demonstra o funcionamento desse processo.

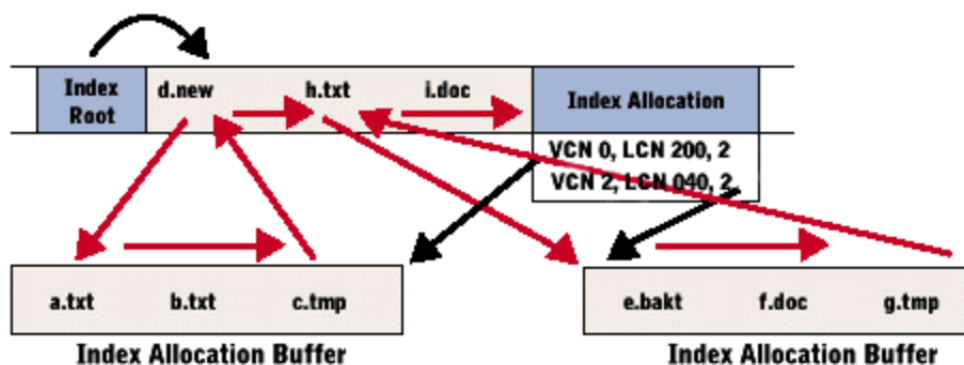


Figura 2.8: Registro de diretório e a “*Index Allocation Buffer*”.

2.4.4 Remoção de Arquivos

Quando um novo arquivo é criado em um sistema de arquivos NTFS, o arquivo bitmap do registro da MFT precisa ser atualizado pois os clusters anteriormente considerados *não alocados* agora estão ocupados por um arquivo. Assim, quando um arquivo é excluído o arquivo bitmap do registro da MFT também precisa ser atualizado. Os bits correspondentes ao cluster anteriormente alocado pelo arquivo removido são zerados (marcados como *não alocados*). O

arquivo foi então marcado para remoção porém seu conteúdo continua no disco no mesmo lugar que estava anteriormente alocado, até que o mesmo local seja utilizado para armazenar outro arquivo. A entrada do índice do arquivo é também marcada para exclusão porém, a entrada do índice é rapidamente sobrescrita, o que no caso do arquivo pode demorar dependendo do tamanho do volume de dados [Svensson 2005].

2.5 **Análise NTFS**

Svensson (2005, p.33) diz que registro da MFT pertencentes à arquivos já apagados podem ser recuperados pois, como vimos anteriormente, os arquivos não são excluídos do disco e sim deixam de ser referenciados, permanecendo no mesmo lugar até que outro arquivo sobrescreva seu conteúdo. As chances de se recuperar um arquivo excluído, no entanto, diminuem com o tempo uma vez que o sistema NTFS sobrescreve arquivos não mais alocados antes de alocar espaço adicional para a MFT. Informações padrões de arquivo, assim como seu nome, estão contidas nesses registros da MFT. Isso pode ser útil para auxiliar a encontrar as partes do arquivo não mais alocado uma vez que as informações de apontamento para as partes do arquivo não mais alocadas e ainda residentes no disco, estão presentes no registro de indexação do arquivo encontrado. Sem esse registro do arquivo ainda seria possível se recuperar um arquivo deletado fazendo uma procura física no disco desde que este arquivo não esteja fragmentado. Arquivos fragmentados são muito difíceis de se recuperar totalmente através de pesquisas físicas no disco.

Tendo isso em vista, Svensson (2005, p.33) relata que o índice pertencente à um arquivo removido pode não mais ser localizado visto que o mesmo pode ser rapidamente sobrescrito, o que já foi visto anteriormente. Essa entrada de arquivo (índice) pode ser muito importante na reconstituição do crime. Quando uma entrada de índice é marcada para remoção, todas as entradas seguintes são movidas para cima sobrescrevendo, portanto, o índice marcado para remoção. Se essa entrada de índice é a última entrada em um registro de diretório ou no buffer de alocação de índices (index allocation buffer) ela pode então não ser sobrescrita podendo assim ser recuperada.

3 *File Carving*

File Carving é o processo de recuperação de fragmentos de arquivos não mais indexados baseado no seu conteúdo e na ausência de metadados do sistema de arquivos [Hand Zhiqiang Lin 2012].

O processo de File Carving é útil na recuperação de arquivos e amplamente utilizado em investigações digitais (*em forense computacional*). Devido a isso, esse é um dos processos mais importantes e desafiadores da computação forense [Garfinkel 2007].

Um dos primeiros desafios do processo de file carving, segundo Memon (2008, p. S2), se encontra na tentativa de se recuperar arquivos fragmentados. Um ponto chave no processo de recuperação desses arquivos fragmentados é encontrar a relação de fragmentação do arquivo que pode beneficiar o processo de recuperação de dados. Técnicas tradicionais não conseguem recuperar arquivos quando o sistema de arquivos é ou está corrompido, quando a tabela de alocação não está presente ou possui endereçamentos errados ou incompletos. Assim, a técnica de File Carving apresenta sua capacidade de recuperar arquivos em espaços não alocados do disco (*área do disco não apontada pela tabela de partição - no caso do NTFS, a Master File Table*).

3.1 Assinatura de Arquivo

As ferramentas de file carving mais comuns analisam as informações de assinatura do arquivo, através do cabeçalho e do rodapé do arquivo, determinando assim o conteúdo do arquivo entre esses blocos. Infelizmente, ferramentas poderosas de file carving atuais ainda falham na recuperação de arquivos fragmentados.

3.2 Arquivos Fragmentados

Segundo Kloe (2007, p. 8), um arquivo fragmentado é um arquivo dividido em várias partes onde cada parte pode estar localizada em um lugar diferente em um mesmo conjunto de dados. Sistemas operacionais modernos, tais como o NTFS, tentam gravar os dados evitando a fragmentação porém, ainda existem algumas situações em que a fragmentação ocorre. Já Memon define que um arquivo é dito fragmentado quando o mesmo está armazenado de forma descontinuada nos clusters, e que o maior desafio para o processo de data carving é justamente a recuperação de arquivos quando esses estão fragmentados em duas ou mais partes.

Garfinkel (2007, p. S4) defendia a hipótese de que diferentes tipos de arquivos poderiam apresentar diferentes padrões de fragmentação, o que poderia determinar uma fragmentação diferenciada para um tipo de arquivo específico no disco. Arquivos comumente de sistemas, instalados juntamente com a parte do sistema operacional apresentariam um baixo fator de fragmentação enquanto arquivos comuns e altamente importantes na análise forense, por se tratarem de informações do dia a dia tais como documentos (.doc), mídias (.avi, .jpeg), arquivos de correio eletrônico armazenados (.pst), planilhas com cálculos e fórmulas (.xls), arquivos de log e arquivos texto (.txt), aparentam ter uma tendência a maior fragmentação se comparados à arquivos menos significativos para o processo forense. Garfinkel (2007, p. S2) então chegou a uma das principais conclusões, que até 42% de arquivos de e-mail (.pst), 17% de arquivos microsoft word (.doc) e 16% de arquivos de imagem (.jpeg) são fragmentados, deixando bem claro que a recuperação de arquivos fragmentados é um problema crítico para a análise forense. Outro fator importante que tem influencia a fragmentação, que pode ser percebido por Garfinkel (2007, p. S6), é que em dispositivos com maior capacidade de armazenamento o fator de fragmentação é menor se comparado a dispositivos de menor capacidade de armazenamento.

- Quando não há mais espaço de mídia suficiente na sequência física para a gravação de um arquivo, assim, para ser alocado no disco ele tem que ser dividido em dois ou mais fragmentos.
- Na sequência de um arquivo já alocado, o espaço restante no cluster não é suficiente para a gravação do arquivo sequencialmente, sendo necessária a divisão do arquivo em dois ou mais fragmentos.

3.2.1 Fragmentação Linear

Kloe referencia a fragmentação linear a arquivos que estão fragmentados seguindo a sequência original no disco. A figura 3.1 demonstra a fragmentação linear na ordem original do disco.



Figura 3.1: Exemplo de Fragmentação Linear

3.2.2 Fragmentação Não Linear

Conforme Kloe, a fragmentação não linear representa os arquivos fragmentados fora da ordem normal de sequência do disco. A figura 3.2 demonstra como os fragmentos do arquivo F1 estão fora da sequência natural do disco.

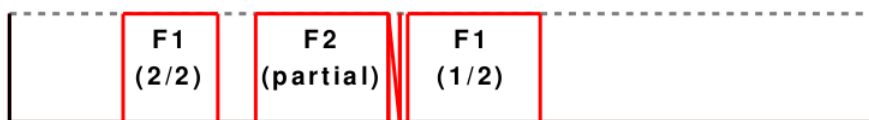


Figura 3.2: Exemplo de Fragmentação Não Linear

3.2.3 Arquivo Parcial

A figura 3.2 demonstra não somente a fragmentação não linear mas também a sobreposição de parte do arquivo. Isso demonstra um arquivo não mais alocado que fora parcialmente sobrescrito. O arquivo pode nunca mais ser totalmente recuperado porém, parte da informação útil do arquivo ainda está presente na área antes alocada para o mesmo. Kloe diz que para o processo de carving não há diferença entre uma informação parcial e um arquivo fragmentado que ainda não tenha sido totalmente recuperado. Um primeiro detalhe sobre a afirmação de Kloe é que um dos arquivos pode ser totalmente recuperado dependendo do tempo e da profundidade de busca da ferramenta de carving (em algum momento conseguiria encontrar as demais partes do arquivo), enquanto um arquivo parcial (sobrescrito) não possui mais suas partes perdidas (por terem sido sobrescritas) no disco e que a ferramenta de carving em algum momento terá que definir se tratar de um arquivo parcial.

3.2.4 Recuperação de Arquivos Fragmentados

Para recuperar arquivos fragmentados, qualquer ferramenta de file carving precisa ser capaz de identificar o ponto de início do arquivo e os blocos necessários para reconstruir o arquivo. Segundo Memon (2008, p. S4), há basicamente três pontos a se seguir no processo de carving:

1. Identificar o ponto inicial do arquivo.
2. Identificar os blocos pertencentes ao arquivo.
3. Organizar os blocos de forma correta para que se possa reconstruir o arquivo.

[PAREI AQUI!!!!!!!!]

3.3 Aspectos de Carving

3.3.1 Cabeçalho e Rodapé

Atualmente a simples comparação byte por byte é uma operação muito rápida que computadores atuais podem processar com facilidade. Sendo assim, a verificação de cabeçalhos e rodapés estáticos de tipos de arquivos se torna um dos primeiros passos que um algoritmo de recuperação pode seguir. Com base nas informações contidas no cabeçalho e rodapé de um arquivo pode-se determinar o tipo do arquivo em questão e até mesmo, em alguns casos, informações mais específicas referentes ao arquivo (metadados¹) [Garfinkel 2007].

A análise de cabeçalho e rodapé é a técnica mais básica de carving. A técnica consiste em procurar no conjunto de dados por padrões que definem o início do bloco de um arquivo, como por exemplo o cabeçalho de um arquivo de imagem (.png) demonstrado pela figura 3.3 [Kloet 2007].

O valor hexadecimal “89 50 4e 47 0d 0a 1a 0a” representa os 8 bytes do cabeçalho de arquivos PNG. Cada arquivo PNG é marcado a partir desse cabeçalho. A marca final do arquivo ou o rodapé do arquivo delimita o final do arquivo. O conteúdo entre o cabeçalho do arquivo e o final do arquivo seria a representação da informação do arquivo 3.4.

Segundo Kloet (2007, p. 9), a análise de cabeçalho e rodapé é um método que apresenta alguns problemas relevantes:

¹ Informação descritiva sobre um dado ou um item, metainformação

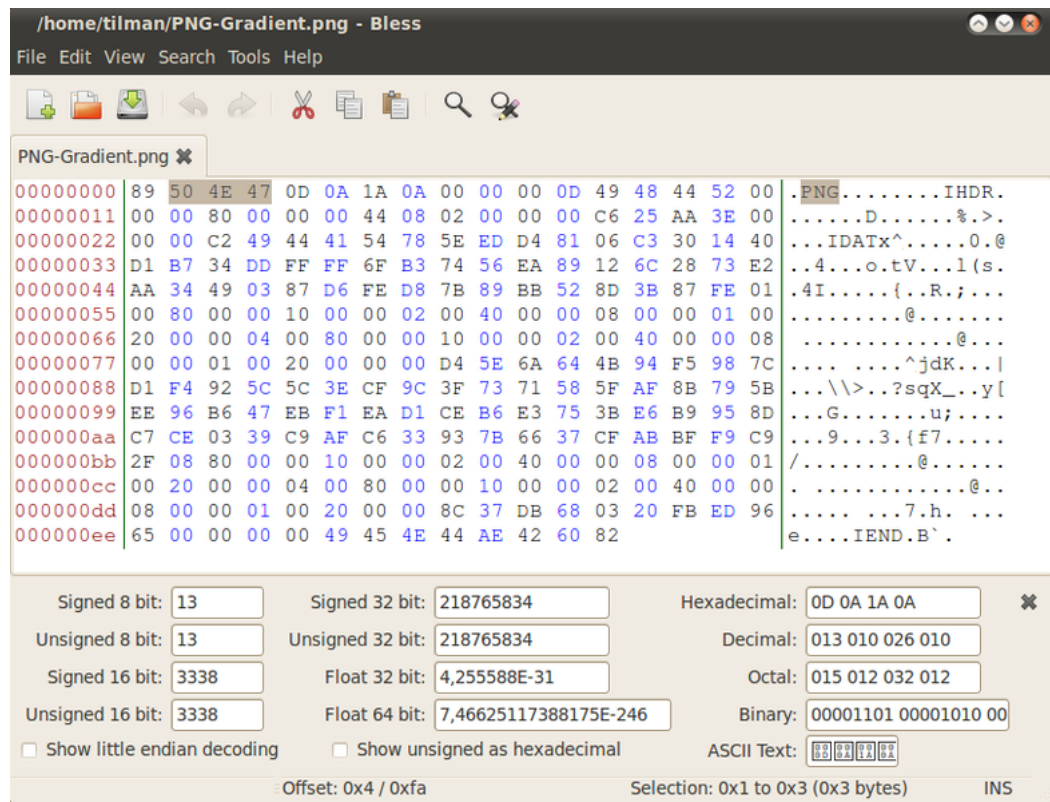


Figura 3.3: Estrutura de cabeçalho de um arquivo do tipo PNG [FileFormat 2013]



Figura 3.4: Área de dados de um arquivo do tipo PNG [Kloet 2007]

- Muitos falsos positivos.
- Incapacidade de detectar arquivos fragmentados e arquivos parciais.
- Alguns arquivos não apresentam cabeçalho fixo, assim como alguns arquivos também apresentam rodapé variável, o que impossibilita sua identificação por esse método.

O mesmo ocorre para o método de cabeçalho com tamanho máximo possível de dados, técnica que visa recuperar os dados pelo comprimento máximo disposto para o tipo de arquivo para sua área de dados. Além desses os problemas da técnica de cabeçalho e rodapé, esta técnica ainda tem outros dois problemas ligados à área de dados. Tendo em vista que a área máxima não é exatamente a área máxima estipulada pela definição do tipo de arquivo e sim uma questão de referência, esses dados podem ser primeiramente coletados além do tamanho real do arquivo no dispositivo de origem como também poderá ser menor que o próprio tamanho dos dados do arquivo no dispositivo de origem.

3.3.2 Carving em Estrutura de Arquivo

...

3.4 Número Mágico

...

4 File Carving Avançado

...

4.1 Diferenciação

...

4.2 Fragmentação

...

4.3 Ponto de Fragmentação

...aqui serão incluídos também algoritmos de detecção de pontos de fragmentação como listado no mind map (Sequential Hypothesis Testing)...

5 *Ferramentas de File Carving*

Scalpel Foremost

5.1 Scalpel

...

5.2 Foremost

...

6 *Considerações Finais*

...em desenvolvimento...

Referências Bibliográficas

CARRIER, B. *File System Forensic Analysis*. [S.l.]: Addison Wesley Professional, 2005.

CYBERCITIZEN. *What is Cyber Crime?* cybercitizenship, 2012. Disponível em: <<http://www.cybercitizenship.org/crime/crime.html>>.

FILEFORMAT. *The PNG Image File Format*. 2013. Disponível em: <<http://www.fileformat.info/format/png/corion.htm>>.

GARFINKEL, S. L. Carving contiguous and fragmented files with fast object validation. *ELSEVIER*, v. 4S, p. S2–S12, 2007.

HAND ZHIQIANG LIN, G. G. B. T. S. Bin-carver: Automatic recovery of binary executable files. *ELSEVIER*, n. 9, p. S108–S117, 2012.

IBOPE//NETRATINGS, N. *Painel IBOPE/NetRatings*. cetic.br, 2012. Disponível em: <<http://www.cetic.br/usuarios/ibope/w-tab02-01-cons.htm>>.

KLOET, S. *Measuring and Improving the Quality of File Carving Methods*. Dissertação (Mestrado) — Eindhoven University of Technology, 10 2007.

MAHANT, B. S. H. Ntfs deleted files recovery - forensics view. *IRACST - International Journal of Computer Science and Information Technology and Security (IJCSITS)*, v. 2, n. 3, p. 491–497, 2012.

SVENSSON, A. *Computer Forensics Applied to Windows NTFS Computers*. Dissertação (Mestrado) — Stockholm's University / Royal Institute of Technology, 04 2005.

T., R. *Undelete a file in NTFS*. CodeProject, 2013. Disponível em: <<http://www.codeproject.com/Articles/9293/Undelete-a-file-in-NTFS>>.

WIKIPÉDIA. *Byte*. Wikipédia, 2012. Disponível em: <<http://pt.wikipedia.org/wiki/Bytes>>.

WIKIPÉDIA. *Disco Rígido*. Wikipédia, 2013. Disponível em: <<http://goo.gl/RXODS>>.

WIKIPÉDIA. *Encrypting File System*. 2013. Disponível em: <http://en.wikipedia.org/wiki/Encrypting_File_System>.

WYMAN WALT SCRIVENS, P. H. L. S. B. Metadados. *OUCH!*, 2012.

Anexos

...