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

Lista de Tabelas

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.

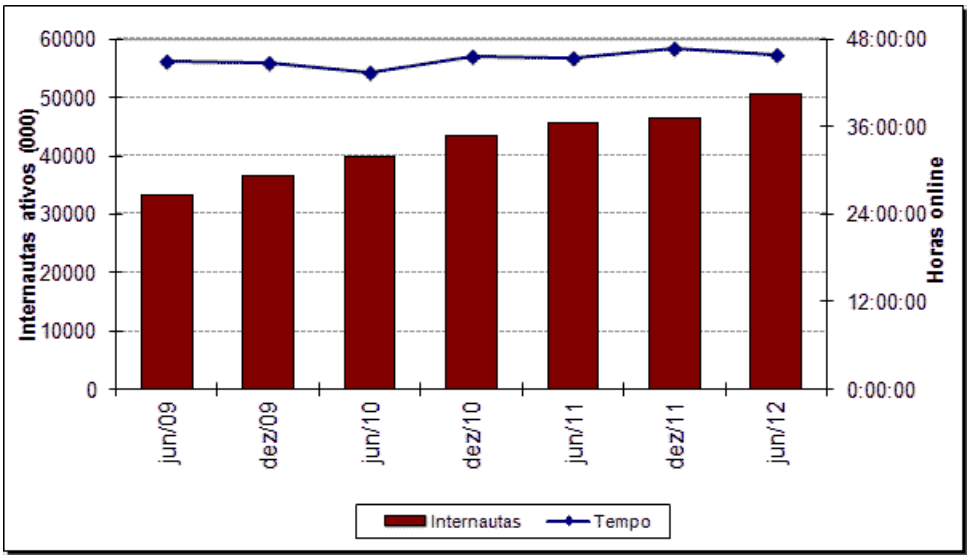

Figura 1.1: Internautas ativos em residências e no trabalho e horas navegadas - 2012 (IBOPE//NETRATINGS, 2012)

Nesse meio, existem usuários que promovem o cibercrime[1] ou atividades ilegais na rede[2]. As informações computacionais são armazenadas em discos rígidos[3] 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

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 possiblidade 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 clusters[4]. 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 space[5]. 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.

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.

"In forensic practice, file carving can recover files that have been deleted and have had their directory entries reallocated to other files, but for which the data sectors themselves have not yet been overwritten (GARFINKEL, 2007)."[6]

[4]Unidade 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.

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

[6]Na prática forense, file carving pode recuperar arquivos que tenham sido apagados e tenham suas entradas de diretório realocadas para outros arquivos, desde que seus setores de dados ainda não tenham sido sobrescritos.

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.

> During a digital forensic investigation many diferent pieces of data are preserved for investigation, of which bit-copy images of hard drives are the most common. These images contain the data allocated to files as well as the unallocated data. The unallocated data may still contain information that is relevant to an investigation, in the form of (parts of) intentionally deleted or automatically removed temporary files. Unfortunately, this data is not always easily accessible: a string search on the raw data might recover (parts of) interesting text documents, but it won't help to get to information present in for example images or compressed files. Besides that, the exact strings to look for may not be known beforehand. To get to this information, the deleted files have to be recovered (KLOET, 2007).[7]

1.2 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.

[7] 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.

1.3 Hipótese(s)

Pretende-se com a análise dos metadados[8] 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:

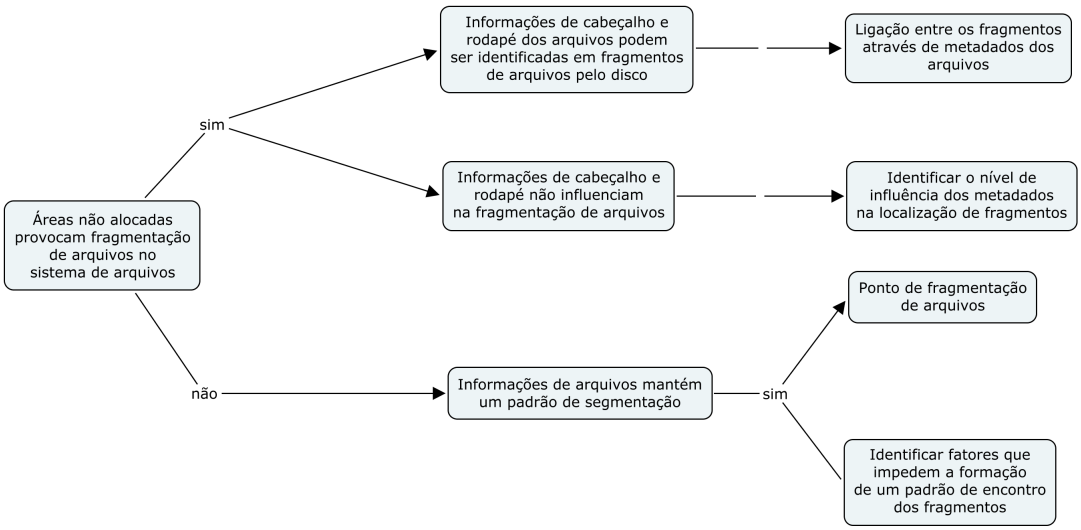

Figura 1.2: Hipótese de pesquisa

1.4 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.

1.5 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:

8

- 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; e
- Identificar formas de localização de fragmentos dos arquivos não alocados.

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.

1.6 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 área não alocada, 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.

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 NTFS

Projetado pela Microsoft, 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). Sendo o predecessor do antigo sistema FAT, é o sistema mais presente em investigações de sistemas Windows (CARRIER, 2005).

2.1 Conceito

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 bytes[1] são alocados em arquivos no sistema (CARRIER, 2005).

2.2 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 (MFT), 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

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

- Espaço não utilizado.

A figura 2.1 demonstra um registro de entrada na MFT.

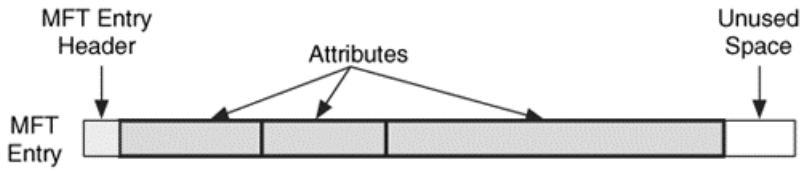

Figura 2.1: Entrada na MFT - Cabeçalho e espaço reservado aos diferentes tipos de atributos. Para o exemplo, a entrada possui 3 atributos.

2.2.1 Cluster

De acordo com Svensson (2005, p. 25), um disco rídigo 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 se apresenta na ordem dos clusters, 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.

2.2.2 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)[2]. Os primeiros 16 registros da MFT são

[2]Conjunto de registros que são continuamente modificados quando um arquivo ou diretório sofre alteração

reservados pelo sistema de arquivos para manter metadados[3] 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[4].
$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[5], 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.

2.3 Alocação de Arquivos

...

2.4 Análise NTFS

...

[3] Dados que definem ou descrevem outra parte dos dados (WYMAN WALT SCRIVENS, 2012).

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 dos arquivos fragmentados. 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 ainda 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 Número Mágico

...

3.3 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. S2) tinha 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 para a análise forense, por se tratarem de informações do dia a dia tais como documentos (.DOC), arquivos de e-mail armazenados localmente (.PST), planilhas com cálculos e fórmulas (.XLS) e arquivos de log e arquivos texto (.TXT), esses arquivos aparentam ter uma tendência a maior fragmentação.

- 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.

3.3.1 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

3.3.2 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.

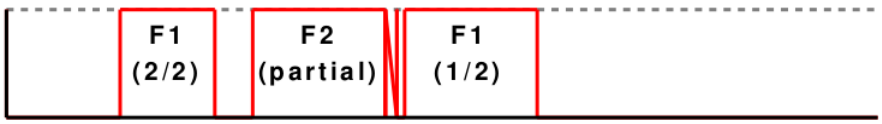

Figura 3.2: Exemplo de Fragmentação Não Linear

3.3.3 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.

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

CALOYANNIDES, M. A. *Privacy Protection and Computer Forensics*. 2. ed. [S.l.]: Artech House, 2004.

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>.

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.

PAL HUSREV T. SENCAR, N. M. A. Detecting file fragmentation point using sequential hypothesis testing. *ELSEVIER*, n. 5, p. S2–S13, 2008.

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

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

Anexos

...