UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
FACULDADE DE COMPUTAÇÃO
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

ERIC FELIPE DE OLIVEIRA PEREIRA

UM ESTUDO DE CASO EM RECUPERAÇÃO DE DADOS UTILIZANDO DATA CARVING APLICADO À COMPUTAÇÃO FORENSE

Belém
2017

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
FACULDADE DE COMPUTAÇÃO
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

ERIC FELIPE DE OLIVEIRA PEREIRA

UM ESTUDO DE CASO EM RECUPERAÇÃO DE DADOS UTILIZANDO DATA CARVING APLICADO À COMPUTAÇÃO FORENSE

Trabalho de Conclusão de Curso apresentado como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Msc. Inácio Leite Gorayeb

Coorientador: Prof. Dr. Josivaldo de Souza Araújo

Belém
2017

Eric Felipe de Oliveira Pereira
Um Estudo de Caso em Recuperação de Dados Utilizando Data Carving Aplicado à Computação Forense/ Eric Felipe de Oliveira Pereira. – Belém, 2017.
34 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Msc. Inácio Leite Gorayeb

Monografia – Universidade Federal do Pará
Instituto de Ciências Exatas e Naturais
Curso de Bacharelado em Ciência da Computação, 2017.

1. Recuperação de Dados. 2. Computação Forense. 3. Data Carving. 4. Sistema de arquivos. 5. NTFS. I. Título.

ERIC FELIPE DE OLIVEIRA PEREIRA

UM ESTUDO DE CASO EM RECUPERAÇÃO DE DADOS UTILIZANDO DATA CARVING APLICADO À COMPUTAÇÃO FORENSE

Trabalho de Conclusão de Curso apresentado como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Data da Defesa: 29 de Setembro de 2017
Conceito:

Banca Examinadora

Prof. Msc. Inácio Leite Gorayeb
Faculdade Metropolitana da Amazônia - FAMAZ
Orientador

Prof. Dr. Josivaldo de Souza Araújo
Faculdade de Computação - ICEN/UFPA
Coorientador

Prof. Dr. Roberto Samarone dos Santos Araújo
Faculdade de Computação - ICEN/UFPA
Membro da Banca

Msc. Deivison Pinheiro Franco
Analista Sênior de Segurança do Banco da Amazônia
Membro da Banca

Belém
2017

Dedico este trabalho aos meus avós, Francisco Candeira de Oliveira e Geulina dos Santos Oliveira, que sempre me incentivaram e acreditaram em mim.

AGRADECIMENTOS

Agradeço primeiramente a Deus, por tudo.

Agradeço aos meus pais, Cleide e Edivaldo, que sempre lutaram pra oferecer o melhor pra mim. Agradeço aos meus segundos pais, Vô Nego e Vô Zuzu (In Memorian), mesmo não estando mais aqui na terra eu nunca vou esquecer de tudo o que eles me ensinaram e fizeram por mim e sei que onde eles estão agora, estão felizes por mim.

Agradeço à minha namorada, Ludmila, que nos momentos que precisei ela esteve do meu lado me incentivando ou dando puxões de orelha.

Agradeço ao Professor Josivaldo, que mesmo diante de todas as minhas dificuldades não desistiu de mim.

Agradeço ao Professor Inácio, pela oportunidade oferecida a mim.

Agradeço a todos os meus amigos e companheiros de curso, Ronald, Paulo, Tiago, Brunelli, Hugo, Alexandre, Jeff, Gustavo, por todos os bons momentos, as bagunças, as zoações, as gargalhadas, os almoços e jantares no RU.

“O insucesso é apenas uma oportunidade
para recomeçar com mais inteligência”
(Henry Ford)

RESUMO

A área da Computação Forense está em crescimento constante, principalmente no Brasil, então são necessários estudos que abordem temas recorrentes no dia a dia de um perito forense. Além do mais, todos os dias, armazenamos vários e vários arquivos em meios digitais, alguns de suma importância, e por algum motivo esses dados podem ser danificados ou perdidos, de forma intencional ou não. Por esse motivo, o estudo na recuperação de dados se faz necessário, principalmente para os sistemas de arquivos NTFS, o mais usado por usuários finais. Este trabalho apresenta uma avaliação comparativa entre ferramentas de recuperação de dados, com o intuito de agregar conhecimento na área estudada. Para tal, foi feito um levantamento bibliográfico sobre os principais conceitos relativos à recuperação de dados, e subsequentemente um estudo de caso para analisar as ferramentas de recuperação. Desse modo, ao fim do estudo de caso, foi averiguado que parte das ferramentas testadas tiveram resultados positivos, conseguindo recuperar dados no sistema de arquivos NTFS.

Palavras-chave: Recuperação de dados. Computação Forense. Data Carving. Sistema de arquivos. NTFS.

ABSTRACT

The area of Computer Forensics is steadily growing, especially in Brazil, so it is necessary studies that order recurring topics in the day to day of a forensic expert. Moreover, every day, we store several and several files in digital media, some of paramount importance, and for some reason such data may be intentionally or unintentionally damaged or lost. For that reason, the study in data recovery is required, primarily for NTFS file systems, the most used by end-users. This work presents a comparative assessment of data recovery tools, in order to aggregate knowledge in the studied area. To do this, a bibliographical survey was made on the main concepts relating to data retrieval, and subsequently a case study to analyze the recovery tools. Thus, at the end of the case study, it was verified that part of the tested tools had positive results, retrieving data on the NTFS file system.

Keywords: Data recovery. Computer Forensics. Data Carving. File system. NTFS.

LISTA DE ILUSTRAÇÕES

LISTA DE QUADROS

LISTA DE ABREVIATURAS E SIGLAS

HD	*Hard Disk*
DFRWS	*Digital Forensic Research Workshop*
NTFS	*New Technology File System*
FAT	*File Allocation Table*
USB	*niversal Serial Bus*
EXT	*Extended File System*
GB	*Gigabyte*
RAID	*Redundant Array of Independent Disks*
CPF	Cadastro de Pessoas Físicas
JPG	*Joint Photographic Experts Group*
GNU/GPL	*GNU General Public License*

SUMÁRIO

1 INTRODUÇÃO

1.1 Considerações Iniciais

Na atualidade, o fluxo de geração e tráfego de dados está demasiadamente alto, segundo (TERRA, 2015), somente no ano de 2016 a previsão era de que o tráfego de dados ultrapassasse o volume de um zettabyte pela primeira vez, equivalente a um sextilhão de bytes. Com esse aumento, surgem, em conjunto, os ataques, os incidentes, além da preocupação de como manter seus dados digitais em segurança, e em casos de perda, como proceder. No Brasil, há o CERT.br (Centro de Estudos para Resposta e Tratamento de Incidentes em Computadores), esse grupo é responsável pelo tratamento de incidentes conectadas na rede da Internet brasileira. Além disso, o grupo faz a divulgação anualmente dos dados estatísticos dos incidentes registrados pelo CERT.br. Segundo (CERT.BR, 2017), somente em 2016 foram registrados um total de 647.112 incidentes, como pode-se ver na Figura 1.

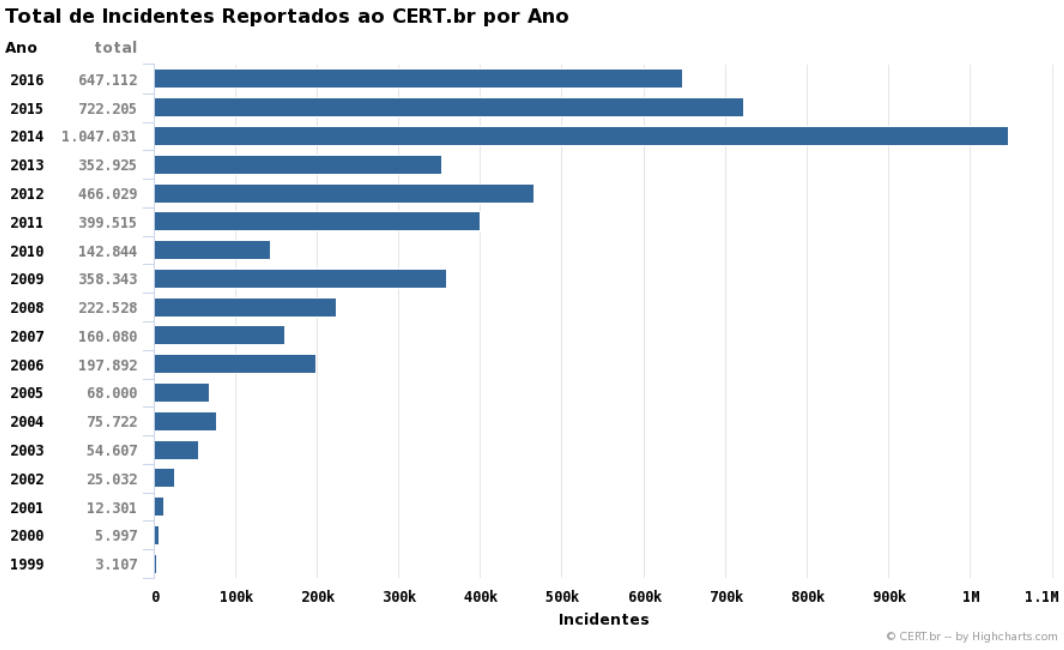

Figura 1 – Total de incidentes reportados ao CERT.br por ano (CERT.BR, 2017)

Desses incidentes, o que mais tem estado em destaque nas notícias ao redor do mundo é o Ransomware, um tipo de malware que sequestra e criptografa os arquivos da vítima, exigindo valores para o resgate e obtenção da chave de acesso de seus arquivos, segundo (NERY, 2016). Ele atua buscando as extensões dos arquivos mais importantes para o usuário, como .doc ou .jpg, e então coloca esses arquivos em uma pasta criptografada, sobrescrevendo os arquivos originais. De acordo com (FBI, 2015), as infecções por esse tipo de malware podem ser devastadoras e a recuperação dos dados pode ser um processo difícil ou até mesmo impossível, exigindo os serviços de um especialista em recuperação de dados bem conceituado.

Segundo (ROMER, 2015), o Brasil concentra 92,31% dos casos de Ransomware na América latina, uma porcentagem bastante alta considerando que a América latina possui um total de 20 países. No Brasil, os ataques estão mais direcionados em empresas, com criminosos se aproveitando principalmente da vulnerabilidade de redes corporativas para instalar seus Ransomwares e exigir resgates para devolução dos dados.

Visto que muitos incidentes causam a perda de dados, então a recuperação de dados é um estudo importante para a atualidade e este trabalho visa apresentar uma comparação de ferramentas de recuperação de dados, que utilizam a técnica de *Data Carving*, para dessa maneira auxiliar os interessados na área da Computação Forense.

1.2 Motivação

Cada vez mais as vidas das pessoas estão sendo documentadas e passadas em meios digitais, com isso, a importância da preservação e recuperação de dados digitais acompanham este crescimento. Por acidente, por culpa, ou ainda por ação de terceiros (humanos ou não) dados importantes correm risco de desaparecer. A recuperação de dados é essencial para o desenvolvimento da Computação Forense, requerendo então, a necessidade de novos estudos sobre essa prática.

1.3 Justificativa

Os cenários e dispositivos de armazenamento digital ao longo dos anos vem sofrendo evolução e mudanças, com isso, novas técnicas e ferramentas vem sendo desenvolvidas e apresentadas como solução para diversas situações. Portanto, a comparação e testes de ferramentas em alguns dos cenários atuais de armazenamento, torna-se importante para que quando uma resposta a um determinado incidente seja necessária, o responsável pela ação (perito, por exemplo) esteja seguro de como e qual técnica e ferramenta utilizar.

1.4 Objetivos

As subseções a seguir detalham as metas deste trabalho, levando em conta o tema exposto.

1.4.1 Objetivo Geral

Realizar uma pesquisa no âmbito da recuperação de dados, identificando técnicas e métodos, que possam auxiliar a Computação Forense em situações que exijam esta prática ou trabalhos futuros, bem como testar a efetividade de algumas ferramentas em diferentes cenários.

1.4.2 Objetivos Específicos

Dado o objetivo geral acima, tem-se os seguintes direcionamentos:

- Introduzir conceitos relacionados a Computação Forense;
- Apresentar procedimentos e técnicas de recuperação de dados;
- Apresentar estudo de caso em cenários controlados;
- Fazer comparações entre as ferramentas de recuperação de dados;

1.5 Metodologia

A pesquisa buscou, dentre as técnicas de recuperação de dados, quais as que melhor se ajustam dentro de um determinado cenário, dependendo do tipo de mídia utilizada, usando um modelo de estudo qualitativo.

Primeiramente, foi realizado um levantamento bibliográfico com o objetivo de compreender o objeto de estudo. Foram utilizados artigos, dissertações, livros e outros dados coletados na Internet.

Além disso, foi realizado um estudo de caso com avaliação de forma qualitativa, das ferramentas escolhidas e ao final, resultados foram coletados para amostra, observando quais benefícios elas trazem para o objetivo final que é a obtenção dos dados e recuperação de arquivos.

1.6 Trabalhos Relacionados

No desenvolvimento deste trabalho, buscou-se na literatura estudos sobre a recuperação de dados no âmbito da Computação Forense.

Os autores (NASCIMENTO; JERÔNIMO; SEGUNDO, 2010) buscaram abranger a quantidade de sistemas de arquivos estudados, que foram Ext3, Ext4, FAT32 e NTFS. Além do mais eles utilizaram outras ferramentas de recuperação de dados que são: Foremost, TSK/Autopsy e FTK Imager. Todas as ferramentas utilizadas nos testes estão com versões defasadas, pois se trata de um artigo do ano de 2010. Os autores não especificam quais tipos de mídias foram utilizadas nos testes.

O autor (CRUZ, 2015), estuda apenas os sistemas de arquivos Ext4, enfatizando os sistemas com núcleo Linux e os sistemas operacionais Android. Foram utilizadas 6 ferramentas de recuperação diferentes para fazer seu estudo. O autor não relata quais os tipos de mídias que foram utilizados nas experimentações.

O autor (POVAR; BHADRAN, 2010), aborda uma visão apenas teórica do assunto, não realizando estudo de caso, porém apresenta o funcionamento do *Data Carving* de uma maneira minuciosa, exemplificando com arquivos de extensão .jpeg e .pdf.

1.7 Estrutura do Trabalho

Este trabalho está organizado em 3 partes:

1 O capítulo 2 apresenta uma visão geral sobre computação forense, bem como, os conceitos básicos para a realização deste trabalho. No capítulo 3 é apresentado o estudo de caso bem como seus procedimentos e resultados. O capítulo 4 presenta algumas considerações sobre o trabalho desenvolvido e trabalhos futuros.

2 COMPUTAÇÃO FORENSE

2.1 Visão Geral

O autor (MELO, 2009) faz uma definição ramificada de Computação Forense, conforme segue: "A Computação Forense pode ser definida como uma área da Ciência da Computação que se desenvolve gradualmente para atender à demanda oriunda da Criminalística, e também como uma parte da Criminalística que se apropria de fundamentos da Ciência da Computação."

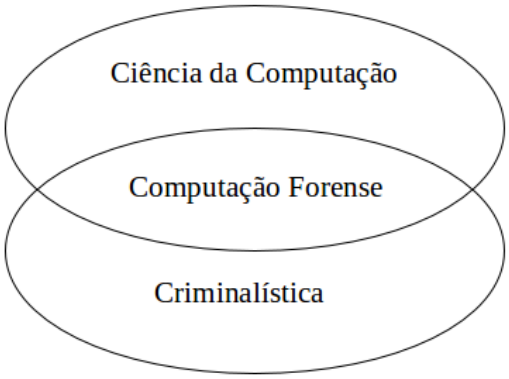

Figura 2 – Relação Ciência da Computação, Criminalística e Computação Forense (MELO, 2009)

Para se compreender melhor a Computação Forense, é necessário entender um pouco mais sobre a Ciência Forense e bem como da computação.

2.2 Conceitos

2.2.1 Ciência Forense

A Ciência Forense é uma área interdisciplinar que envolve Física, Biologia, Química, Matemática e várias outras ciências de fronteira, com o objetivo de dar suporte às investigações relativas à justiça civil e criminal (CHEMELLO, 2006).

De acordo com (SERRANO, 2002), a história da Ciência Forense teve início na China, na disnastia Tang, onde teve seu primeiro registro no século VII, pelo Ti Yen Chieh, ao fazer o uso dos vestígios de um crime para resolve-lo, além da lógica.

No início de sua estruturação, a Ciência Forense, utilizava-se de profissionais de formação genérica. Contudo, face ao progresso, alguns crimes passaram a ser executados com maior complexidade, pois aliados a utilização do conhecimento tecnológico, vieram a ser executados com maior grau de sofisticação. Por consequência, passou-se a exigir a colaboração de outras áreas da ciência, com a participação de profissionais com a especialização correspondente para

fazer frente às necessidades de conhecimento que devem ser aplicados em cada caso, visando realizar com mais eficácia uma investigação policial(CALAZANS; CALAZANS, 2005).

2.2.2 Computação

De acorco com o dicionário Michaelis(WEISZFLOG, 2015): sf.1 Ato ou efeito de computar; cálculo, contagem ou cômputo. 2 Inform V processamento de dados. 3 por ext Qualquer trabalho ou atividade que envolve o uso do computador.

2.3 Etapas da Computação Forense

A Computação Forense faz parte de um processo investigativo. Ao início de uma investigação um perito é acionado e o mesmo deve ter inúmeros cuidados e também necessita ser metódico para que todas as evidências permaneçam íntegras e assim não afetem o andamento da investigação e o laudo final. Esse processo é dividido em quatro etapas: Coleta de dados, exame dos dados, análise das informações e interpretação dos resultados (PEREIRA, 2010), como pode ser observado na Figura 3.

- Coleta de Dados: Essa etapa é vista como a mais importante durante o processo, pois é nela que é recolhida a massa crítica de todos os dados da investigação, portanto todo cuidado é pouco para manter a integridade dos dados e não afetar o resultado final. Além disso, outras atividades como coleta de equipamento, embalagem e identificação são realizadas nesta etapa;

- Exame dos Dados: Nessa segunda etapa pode-se dizer que é separado o "joio do trigo", pois o objetivo é filtrar as informações mais importantes para o caso e deixar as irrelevantes de lado, como por exemplo, arquivos do sistema. Ao início do processo é definida as ferramentas utilizadas, e essa escolha é relativa ao tipo de investigação e informações que estão sendo buscadas;

- Análise das Informações: Nesse terceiro momento, todas as informações extraídas anteriormente são analisadas com a intenção de encontrar dados significativos para a investigação do caso. Todos os dados pertinentes são relacionados com informações referentes à investigação, para que assim seja possível realizar a conclusão;

- Interpretação dos Resultados: Na última etapa, é apresentado pelo perito um relatório técnico que deve expressar com toda a veracidade e embasamento possível o que foi encontrado durante a investigação nos dados analisados. Todo o processo pericial deve estar presente nesse relatório, desde as ferramentas usadas para que seja comprovado a integridade das informações contidas no laudo .

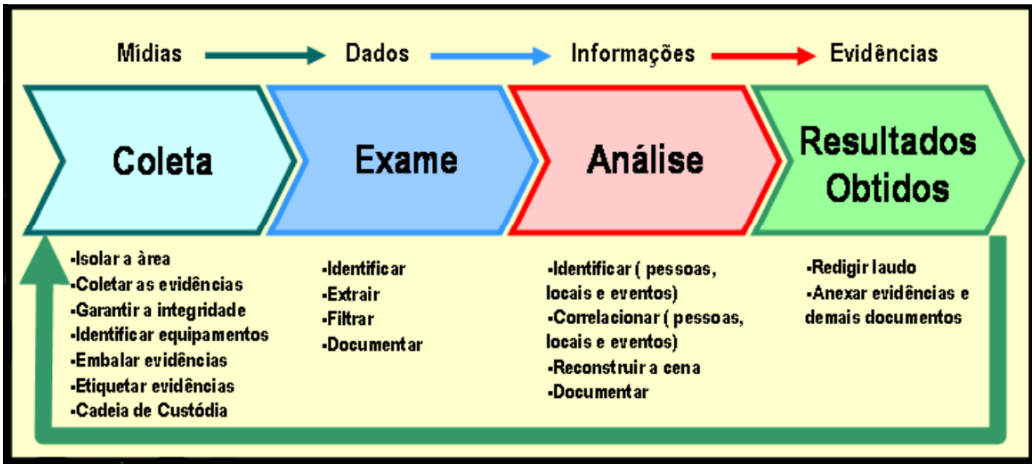

Figura 3 – Etapas da Computação Forense (PEREIRA, 2010)

2.4 Teorias da Computação Forense

A Computação Forense utiliza teorias que são herdadas da área do Direito, como por exemplo, a teoria de Locard.

2.4.1 Teoria de Locard

A Teoria de Locard ou Princípio da troca de Locard é um dos principais fundamentos da análise forense. Foi assim definida, em homenagem à Edmund Locard, que foi um criminalista francês pioneiro em ciência forense e criminologia. De acordo com esse princípio, qualquer um, ou qualquer coisa, que entra em um local de crime leva consigo algo do local e deixa alguma coisa para trás quando parte, segundo (LEOBONS, 2007).

Este princípio também é válido na Computação Forense, pois onde há intrusos também haverá vestígios, mesmo que esses sejam extremamentes difíceis de serem detectados eles estarão lá em algum lugar. Nessas situações a análise forense se tornará mais complicada e vagarosa, pois a dificuldade faz com que o tempo de análise seja maior.

Caso um perito fosse realizar uma análise em um HD previamente apreendido e o mesmo não tomasse as precauções devidas no seu manuseio como, por exemplo, manipular o conteúdo deste dispositivo com configuração de somente leitura, o perito correria o risco de alterar acidentalmente o conteúdo do dispositivo, deixando, com isso, algo de si e levando algo consigo, podendo causar tecnicamente a nulidade da prova.

2.4.2 Teoria dos frutos da árvore envenenada

A teoria dos frutos da árvore envenenada é um conceito da área do Direito Processual Penal, que se pode utilizar na Computação Forense também. De acordo com (OLIVEIRA, 2017) A teoria dos "*fruits of the poisonous tree*", ou teoria dos frutos da árvore envenenada, cuja origem

é atribuída à jurisprudência norte-americana, nada mais é que a simples consequência lógica da aplicação do princípio da inadmissibilidade das provas ilícitas, ou seja, se uma prova for obtida de forma ilícita, todas as outras provas ou conclusões obtidas a partir desta serão consideradas ilícitas também.

Um exemplo prático seria ao subtrair um HD de forma ilegal, realizar uma perícia no mesmo e achar provas de um crime qualquer. Porém, ao apresentar esses resultados em juízo, eles não terão valores algum, pois a prova é um fruto de uma árvore envenenada, visto que não houve autorização prévia para a perícia do HD.

2.5 Sistema de arquivos

A causa por trás de um sistema de arquivos é bem compreensível, os computadores necessitam de uma forma para armazenamento de arquivos. Os sistemas de arquivos oferecem uma técnica para que usuários armazenem dados em uma hierarquia de arquivos e diretórios (CARRIER, 2005). Um sistema de arquivos consiste em dados estruturais e de usuários de uma maneira que o computador saiba onde encontrá-los. Na maioria dos casos o sistema de arquivos independe de um computador específico.

De acordo com (HAGEN, 2001), cada sistema de arquivos pode ter tamanho de *cluster* (unidade mínima de armazenamento), modo de gerar inode[1] e capacidade de armazenamento distintos.

Para gravação de dados, um HD é decomposto em setores. Estes setores são agrupados em blocos ou *clusters* (a forma como é chamado depende do sistema) cujo tamanho é determinado pelo sistema. A Figura 4 exemplifica a segmentação de um HD.

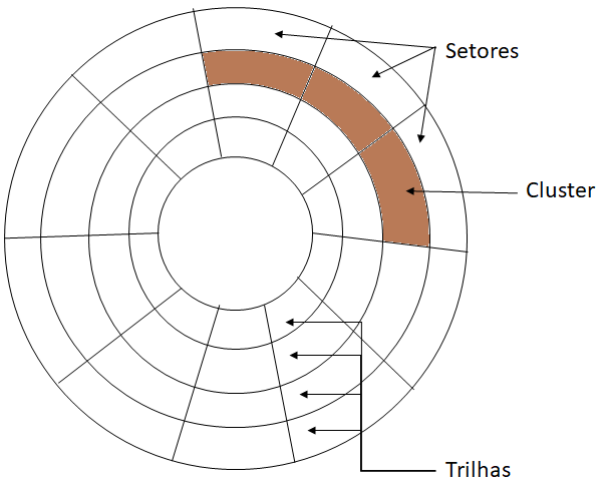

Figura 4 – Divisão do disco, adaptado de (NASCIMENTO; JERÔNIMO; SEGUNDO, 2010).

[1] Estrutura de dados usada para representar um objeto do sistema de arquivos

2.5.1 Características de um sistema de arquivos

Cada sistema de arquivos dispõe de suas próprias características. Porém, (CARRIER, 2005) define essas características em cinco categorias, que são: sistema de arquivos, conteúdo, metadados, nome de arquivo e aplicação. A categoria sistema de arquivos inclui as informações gerais, sua estrutura e tamanho. A categoria conteúdo diz respeito aos dados que compõe o conteúdo real de um arquivo. O maior volume dos dados de um sistema de arquivos pertence a esta categoria, e é normalmente organizado em um conjunto de recipientes de tamanho padrão. Cada sistema de arquivos atribui um nome diferente para os recipientes, como clusters no caso dos sistemas de arquivos FAT e NTFS e blocos para os sistemas de arquivos da família EXT.

A categoria de metadados contém informações, sobre onde o conteúdo do arquivo está armazenado, seu tamanho, a hora e data em que o arquivo foi lido, escrito ou acessado pela última vez. A categoria nome de arquivo contém os dados que atribui um nome a cada arquivo. Na maioria dos sistemas de arquivos, estes dados estão localizados no conteúdo de um diretório, compondo uma lista de nomes de arquivos e seus endereços. Por fim, a categoria de aplicativo contém os dados que não são necessários durante o processo de leitura ou escrita de um arquivo, mas agregam funcionalidades ao sistema de arquivos, como por exemplo, controle de cotas de usuário.

2.5.2 Sistema de Arquivos FAT

Segundo (CARRIER, 2005) "*File Allocation Table*"(FAT) é um dos sistemas de arquivos mais simples encontrados em sistemas operacionais comuns. Ele é suportado por todos os sistemas Windows e a maioria dos sistemas Unix. O FAT é comumente encontrado em cartões de memória, câmeras digitais e unidades USB.

O sistema de arquivos FAT não segue claramente o modelo de cinco categorias descritas anteriormente, por exemplo, o sistema não possui algum dado que se enquadre na categoria de aplicação. Com o surgimento de dispositivos de armazenamento mais aprimorados e com maior capacidade, o sistema FAT foi ganhando versões, identificadas pelos nomes FAT12 e FAT16, sendo o primeiro, quase um desconhecido e o último, padrão dos sistemas operacionais da Microsoft por muito tempo. As versões surgem com a intenção de corrigir determinadas limitações do sistema de arquivos anterior. O próprio FAT16, por exemplo, passou por isso: esta versão só trabalha com, no máximo, 2 GB, assim, para aplicá-lo em um disco de 5 GB, seria necessário dividi-lo em 3 partições (2 GB + 2 GB + 1 GB, por exemplo) para ser possível o aproveitamento de toda a capacidade da unidade (ALECRIM, 2011).

Diante deste e de outros problemas, a Microsoft lançou, em 1996, o FAT32, que se tornou o sistema de arquivos do Windows 95 e do Windows 98, sendo também compatível com versões lançadas posteriormente, como Windows 2000 e Windows XP, embora estes tenham um sistema de arquivos mais avançado, o NTFS.

2.5.3 Sistema de Arquivos NTFS

"*New Technologies File System*"(NTFS) foi criado pela Microsoft, e sua comercialização se deu a partir do Windows NT, sendo o sistema de arquivos padrão da plataforma Windows até os tempos atuais. Segundo (ALECRIM, 2011), uma das principais características do NTFS, refere-se ao quesito "recuperação": em caso de falhas, como o desligamento inesperado do computador, o NTFS é capaz de retroceder os dados à condição anterior ao imprevisto. Isso é viável, em parte, porque, durante o processo de boot, o sistema operacional consulta um arquivo de log que registra todas as operações efetuadas e entra em ação ao identificar nele os pontos problemáticos. Ainda neste aspecto, o NTFS também tolera redundância de dados, isto é, replicação, como o que é feito por sistemas RAID (*Redundant Array of Independent Disks*)[1], por exemplo.

Outra peculiaridade do NTFS é seu modo de permissões de acessos. O Unix sempre foi considerado um sistema operacional seguro por trabalhar com o princípio de que todos os arquivos precisam ter variados níveis de permissões de acesso. O NTFS também é capaz de permitir que o usuário defina quem e como acessar pastas ou arquivos.

Nos sistemas de arquivos podem ocorrer diferentes modos de realizar a exclusão de dados, dentre eles estão:

- Deleção: O disco rígido é formado por estruturas magnéticas que podem ser ativadas ou desativadas. Enquanto fisicamente existem as informações ativas ou inativas, logicamente mostra-se às informações “0” e “1”. Um arquivo vai ocupar, por padrão, o espaço “1” nos HDs. Após um arquivo ser deletado, apesar do computador não enxergar mais as informações naquele local, o sistema apenas reconhecerá aquele espaço como espaço livre. Com isso, usuários podem armazenar novos dados naquele local, sendo que as informações serão sobrescritas.

- Formatação: Assumindo que o sistema operacional Windows é o segundo mais utilizado entre os usuários, de acordo com (ZURIARRAIN, 2017), nele existem as opções de formatação rápida e formatação completa. Segundo (VILAR, 2016), na formatação rápida os dados gravados em disco não são excluídos, na verdade apenas os apontamentos para o conteúdo presente na unidade de armazenamento são removidos da tabela de arquivos, isto é, nenhum dado é apagado de fato, apenas a sobrescrita fica disponível, para armazenamento de novos dados.

  A formatação completa, por outro lado, é antagônica ao pensamento comum, e da mesma forma que a formatação rápida, não apaga absolutamente nada dos dados, apenas a tabela de arquivos. A demora no processo é justificada pela verificação do disco em busca de setores defeituosos.

[1] Subsistema de armazenamento composto por vários discos individuais

- Wipe: Wipe em tradução para português significa limpar, e é exatamente esse o seu propósito, fazer uma limpeza no dispositivo de armazenamento desejado. Dois métodos de wipe são mais conhecidos, eles são o wipe zero e wipe randômico.

  De acordo com (FISHER, 2017), wipe zero também pode ser chamado de Zero-fill, ele funciona de uma forma bem simples: ele percorre todo o disco substituindo tudo por zeros. Se for utilizado zero-fill em uma unidade e verifica se todos os dados foram substituídos, pode-se ter certeza de que a informação é menos provável de ser recuperada.

  Então há o wipe randômico, que funciona de maneira semelhante, porém ele utiliza de caracteres aleatórios para fazer o preenchimento da unidade de armazenamento, sendo assim mais improvável a recuperação dos dados.

2.6 Recuperação de Dados

Na atualidade, grande parte dos nossos dados são criados e arquivados de forma digital, desde lembretes de lista de compras até artigos acadêmicos. De acordo com (CALDAS; SILVA, 2016), todos os dias uma grande massa de dados são geradas seguidamente em vasta medida, fugacidade, pluralidade e trafegam pela rede mundial de computadores. A facilidade para a troca desses dados aliados com a grande massa de equipamentos geradores de dados, como smartphones e câmeras digitais, tornam os dados digitais mais acessíveis. Então, de uma forma ou outra, algo de errado pode ocorrer e pode-se perder alguns desses dados. Para reverter quadros como esse utiliza-se das técnicas de recuperação de dados.

2.6.1 Técnicas de Recuperação

Uma das técnicas utilizadas para a recuperação é a chamada recuperação via inode, e para se entender essa técnica precisamos saber o que é um inode. Como analogia para melhor entendimento, todo cidadão brasileiro possui um número de CPF, onde esse número é único e identifica cada pessoa, então no sistema de arquivos Linux e Unix cada objeto possui um identificador único, um número de inode.

De acordo com (MONTES, 2014), cada número de inode aponta para um inode que é uma estrutura de dados que possui informações do arquivo, como por exemplo: tipo de arquivo, tamanho de arquivo, permissões, entre outras informações. Ao apagar um arquivo em um sistema ext2 o sistema irá proceder da seguinte maneira: marcar como livre o *inode number* alocado ao objeto, incrementar o número de inodes livres e os blocos de discos utilizados pelo objeto são recolocados na lista de blocos livres. Então, para realizar a recuperação de dados basta "reconstruir" o inode referente ao arquivo que se deseja recuperar, pois o sistema não apaga o conteúdo de um inode, ele apenas transfere o mesmo para a lista de blocos livres. Se não for conhecido o número inode, talvez será recuperado apenas partes do arquivo que se deseja recuperar.

Outra técnica conhecida é a recuperação via *Pattern*, segundo (FILHO, 2016), onde a maioria dos arquivos possuem patterns que são padrões de identificação. Um exemplo disso são os arquivos no formato .JPG que sempre inicia com 0xFFD8FF. Para a recuperação desse tipo é feito um "esculpimento" no espaço não alocado, usando a extração de cabeçalho e rodapé para a localização e recuperação de arquivos. Essa técnica também é conhecida como *Data Carving* ou *File Carving*, e independe do sistema de arquivos.

3 ESTUDO DE CASO

Nesse capítulo será mostrado a descrição do ambiente de testes, bem com a realização dos testes e os resultados encontrados.

3.1 Ambiente de Teste

Para o estudo de caso foi montado um ambiente de teste composto por uma mídia digital (pendrive) da marca Sandisk Cruzer Blade com capacidade de 32GB e uma mídia magnética (HD) da marca Samsung HD161HJ com capacidade de 160GB. O sistema de arquivos escolhido foi o NTFS, pois ele é compatível com a maioria dos sistemas e além disso é o mais usado por usuários finais. Foram colocados nas duas mídias exatamente os mesmos arquivos, um total de 1042 arquivos ocupando 16,8GB contendo arquivos de formatos variados como documentos, músicas, vídeos e executáveis de programas.

Para um melhor entendimento, o estudo de caso foi dividido em cenários.

3.2 Cenários e Procedimentos

Foram definidos 5 níveis dentro dos cenários, como visto anteriormente, que são:

- Deleção simples;
- Formatação Rápida;
- Formatação Lenta;
- Wipe zero;
- Wipe random;

Os procedimentos iniciais realizados nos cenários foram semelhantes, porém cada cenário possui suas particularidades. Inicialmente, as mídias foram formatadas em NTFS, logo em seguida foram postos rigorosamente os mesmos arquivos nas duas mídias.

- Deleção simples: Após, todos os arquivos, tanto no Pendrive quanto no HD, foram deletados utilizando o atalho de tecla Shift + delete, para que dessa maneira os arquivos não fossem para a lixeira.
- Formatação Rápida: Posteriormente, utilizando o assistente gráfico do Ubuntu, a formatação rápida foi executada nas duas mídias.

- Formatação Lenta: Seguidamente, a formatação lenta foi executada tanto no HD como no Pendrive, utilizando o assistente gráfico do Ubuntu. O processo de formatação lenta no sistema operacional utilizado faz uma sobrescrita completa com zeros no armazenamento, sendo assim análogo ao processo de wipe zero.

- Wipe zero: Logo após, com o uso do comando dd na plataforma Linux, foi realizado Zero-fill (ou wipe zero) nas duas mídias. Um exemplo do uso desse comando pode ser visualizado na Figura 5.

```
labiocad@labiocad-PC08:~$ sudo dd if=/dev/zero of=/dev/sda1
[sudo] senha para labiocad:
dd: escrevendo em '/dev/sda1': Não há espaço disponível no dispositivo
312579761+0 registros de entrada
312579760+0 registros de saída
160040837120 bytes (160 GB, 149 GiB) copied, 14677,3 s, 10,9 MB/s
```

Figura 5 – Exemplo do comando DD para wipe zero.

- Wipe random: Prontamente, o HD e o Pendrive passaram pela operação de wipe random, utilizando o comando *dd if=/dev/urandom of=/dev/sdX bs=4096* no terminal Linux, onde 4096 é o tamanho padrão da unidade de alocação.

Logo em seguida, em todos os cenários foram realizadas cópias forenses[1] com o auxílio do comando DD, para que assim fossem realizadas as operações com as ferramentas de recuperação de dados.

```
labiocad@labiocad-PC08:~$ sudo dd if=/dev/sda1 | gzip > imagem_hd.img.gz
[sudo] senha para labiocad:
312579760+0 registros de entrada
312579760+0 registros de saída
160040837120 bytes (160 GB, 149 GiB) copied, 2624,14 s, 61,0 MB/s
```

Figura 6 – Exemplo do comando DD para cópia forense.

3.2.1 Ferramentas

A primeira ferramenta escolhida para testes foi o Scalpel. Scalpel é uma ferramenta de código livre para recuperação de dados originalmente baseada no Foremost. Desenvolvida por Golden G. Richard III e apresentada na DFRWS (*Digital Forensic Research Workshop*) em 2005, onde sua versão mais atual foi lançada em 2014, utiliza uma base de dados contendo cabeçalhos e rodapés de vários tipos de arquivos conhecidos, para assim fazer o "esculpimento". Seu uso é simples, primeiramente precisa-se editar seu arquivo de configuração para especificar que tipos de extensões ele irá esculpir. Logo em seguida usou-se o comando para acionar a ferramenta e como parâmetro foi colocado o local onde ele irá salvar os arquivos recuperados, como podemos ver na Figura 7.

[1] Cópia fiel do armazenamento, realizada de forma *bit* a *bit*

```
labiocad@labiocad-PC08:~$ sudo nano /etc/scalpel/scalpel.conf
labiocad@labiocad-PC08:~$ sudo scalpel /home/labiocad/backup_pendrive.img -o /ho
me/labiocad/recupPEN
Scalpel version 1.60
Written by Golden G. Richard III, based on Foremost 0.69.

Opening target "/home/labiocad/backup_pendrive.img"

Image file pass 1/2.
/home/labiocad/backup_pendrive.img:  13.7% |*           |    4.1 GB    16:43 ETA
```

Figura 7 – Exemplo de uso da ferramenta Scalpel.

Nas imagens da mídia magnética o procedimento custou o dobro do tempo das imagens da mídia digital, que levaram em média de 45 minutos, pois seu tamanho é maior.

A segunda ferramenta utilizada é PhotoRec. É uma ferramenta gratuita, de código-livre e distribuída sob a licença GNU/GPL. O PhotoRec ignora o sistema de arquivos e segue os dados subjacentes, por isso ainda funcionará, mesmo que o sistema de arquivos da sua mídia tenha sido gravemente danificado ou reformatado. Para mais segurança, o PhotoRec usa o acesso somente leitura para lidar com a unidade ou o cartão de memória do qual está prestes a recuperar dados perdidos.

O seu uso é mais fácil comparado ao Scalpel, pois há uma interface gráfica com mais opções. Basta colocar o comando que aciona a ferramenta junto com o parâmetro da imagem a ser usada, que então a interface é carregada, como pode-se ver na Figura 8.

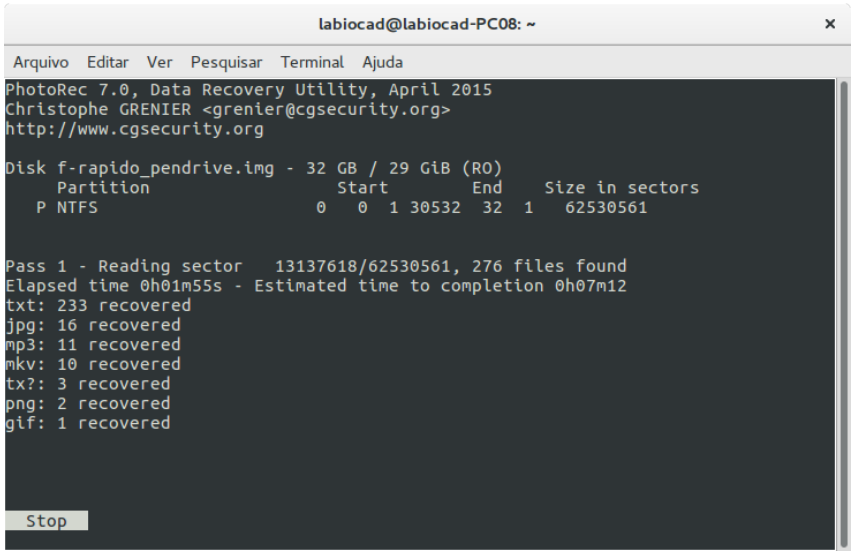

Figura 8 – Exemplo da interface da ferramenta PhotoRec.

A terceira ferramenta escolhida para testes foi o Recuva. É uma ferramenta exclusiva para o sistema operacional Windows, possui uma versão grátis e uma versão paga com mais recursos disponíveis. Sua interface é de fácil uso, onde na tela inicial a ferramenta pergunta quais os tipos de arquivos que desejamos procurar, como é visto na figura.

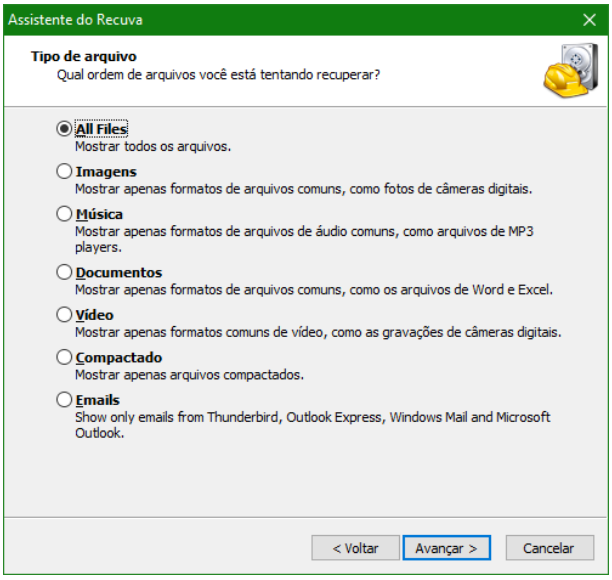

Figura 9 – Exemplo da tela inicial da ferramenta Recuva.

3.3 Resultados

- Deleção simples: O programa Scalpel conseguiu recuperar 13.276 arquivos, porém muito desses arquivos foram apenas parcialmente recuperados[1] e vários outros são falsos positivos[2]. A quantidade recuperada para o HD foi a mesma para o Pendrive.

  A ferramenta Photorec conseguiu recuperar vários arquivos, porém nenhum dos arquivos recuperados estão com seu nome original e alguns foram parcialmente recuperados e estão fragmentados, como por exemplo arquivos de legenda, que possuem a extensão .srt, foram recuperados como arquivo .txt e em mais de um arquivo muita das vezes. Alguns arquivos de vídeos foram totalmente recuperados mantendo a qualidade original. Arquivos .pdf também foram totalmente recuperados.

  A ferramenta Recuva conseguiu recuperar um total de 1024 arquivos no HD e 948 arquivos no Pendrive, quase todos com seus nomes e caminhos originais nos testes executados.

- Formatação Rápida: O Scalpel recuperou precisamente os mesmos arquivos recuperados no cenário de delação, assim como no HD e no Pendrive.

  Tanto no cenário de deleção quanto no de formatação rápida, assim como no HD e no Pendrive, a quantidade de dados recuperados foi exatamente a mesma utilizando a ferramenta PhotoRec.

  O Recuva conseguiu recuperar exatamente a mesma quantidade de arquivos no HD e no Pendrive, e grande parte desses arquivos estavam com seus nomes originais e foram totalmente recuperados, apenas 1% foi recuperado pacialmente e outros 1% não estavam com seu nome original. A quantidade recuperada foi semelhante à do cenário de deleção.

[1] Arquivos que após recuperados não estão completos
[2] Quando a ferramenta indica que há um arquivo recuperado mas não existe o mesmo

- Formatação Lenta: O Scalpel não foi capaz de recuperar dados nesse cenário.

  A ferramenta PhotoRec não conseguiu recuperar dado algum.

  Não foi possível a recuperação com o Recuva.

- Wipe zero: Nesse cenário, o Scalpel não conseguiu recuperar os dados da mídia digital. Ao final do processamento não foi apresentado nenhum arquivo.

  A ferramenta PhotoRec não foi capaz de recuperar dados.

  Recuva não recuperou dados no Pendrive, já no HD a ferramenta não conseguiu nem iniciar o processo de busca, pois apresentou um erro de tamanho inválido com a imagem usada.

- Wipe random: O Scalpel apresentou 19710 arquivos recuperados da mídia magnética, porém nenhum desses arquivos estava legível, 3890 arquivos ao final do processo da mídia digital, entretanto nenhum arquivo estava totalmente recuperado.

  A ferramenta PhotoRec foi incapaz de recuperar dados nas duas mídias.

  O Recuva não conseguiu realizar operações pois o mesmo apresentou um erro de tamanho inválido na leitura das duas imagens, como pode ser visualizado na Figura 10.

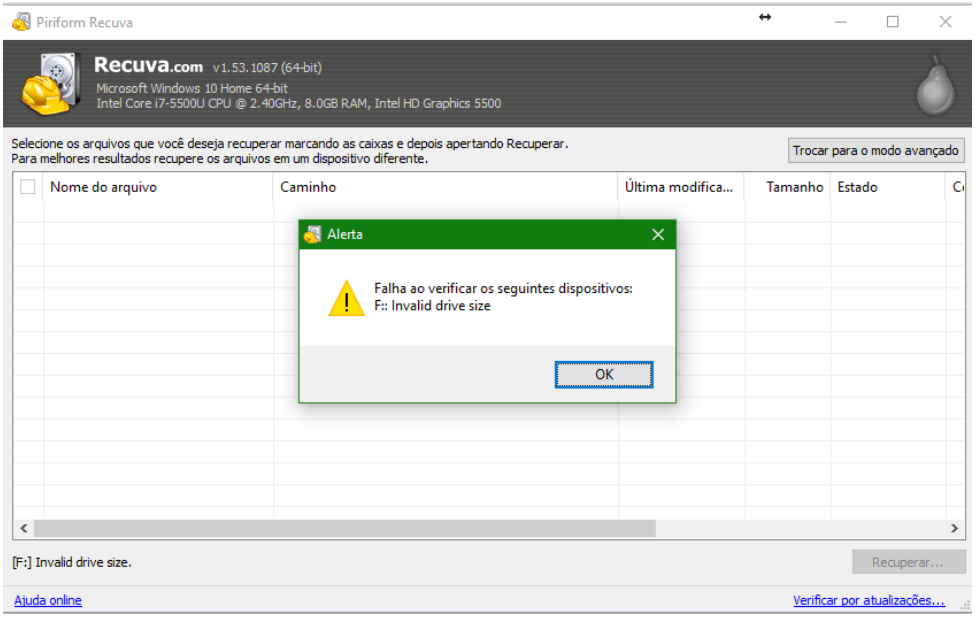

Figura 10 – Tela de erro da ferramenta Recuva.

Ao final dos testes foi observado que a quantidade de arquivos recuperados no cenário de deleção e no cenário de formatação rápida foram bem semelhantes, em todas as ferramentas. Pode-se notar do mesmo modo, que nos cenários de formatação lenta e wipe zero, a quantidade de arquivos recuperados foi nula, em todas as ferramentas.

Foi visto que os resultados obtidos nos cenários de deleção e formatação rápida são bem semelhantes, como visto anteriormente, a deleção e a formatação rápida funcionam da mesma maneira, por esse motivo os resultados também são semelhantes. A Figura 11 mostra o comparativo.

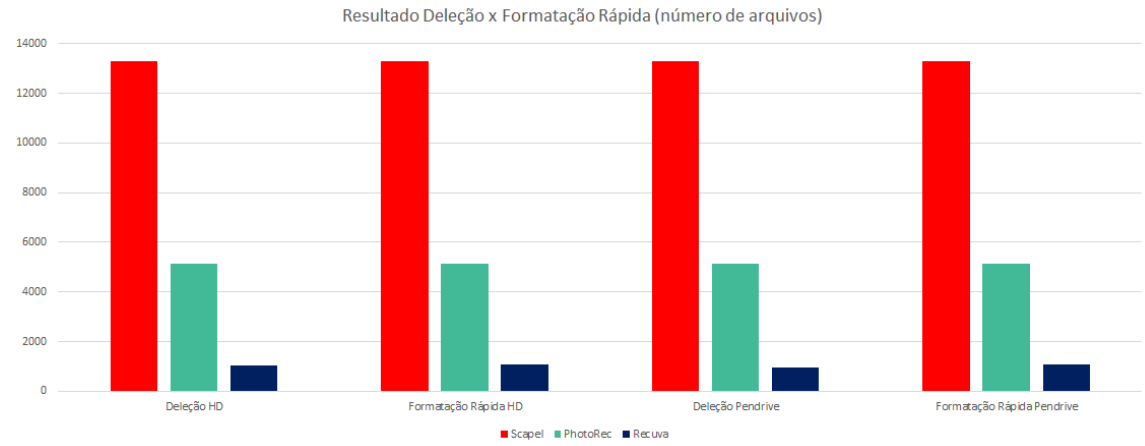

Figura 11 – Resultado Deleção x Formatação Rápida.

Verifica-se que a ferramenta Scalpel gerou o maior número de falso positivos, e a ferramenta que recuperou a quantidade de arquivos mais próximo ao original foi o Recuva, como pode-se constatar no Quadro 1.

Quadro 1 – Total de arquivos recuperados por cenário e ferramenta.

Cenário	Scalpel	PhotoRec	Recuva
Deleção HD	13276 arquivos	5119 arquivos	1024 arquivos
Deleção Pendrive	13276 arquivos	5130 arquivos	948 arquivos
Formatação Rápida HD	13276 arquivos	5119 arquivos	1074 arquivos
Formatação Rápida Pendrive	13276 arquivos	5130 arquivos	1074 arquivos
Formatação Lenta HD	0 arquivos	0 arquivos	0 arquivos
Formatação Lenta Pendrive	0 arquivos	0 arquivos	0 arquivos
Wipe Zero HD	0 arquivos	0 arquivos	Não foi possível
Wipe Zero Pendrive	0 arquivos	0 arquivos	0 arquivos
Wipe Random HD	19710 arquivos	3 arquivos	Não foi possível
Wipe Random Pendrive	3890 arquivos	0 arquivos	Não foi possível

Foram no total 1042 arquivos originais e em determinados cenários foram recuperados mais arquivos do que existiam, como por exemplo no uso da ferramenta Scalpel que chegou a mostrar um total de 19710 arquivos ao final de seu processamento. Então a ferramenta Scalpel foi a menos adequada nos testes, pois foi a que mais se distanciou do número de arquivos originais.

A ferramenta Photorec foi a intermediária dentre as escolhidas, visto que após o processo de recuperação, apresentou um número de arquivos recuperados não muito próximo ao exato, porém não se distanciou tanto quanto o Scalpel.

E a ferramenta que mais se adequou foi o Recuva, dado que sua assertividade foi a melhor dentre as escolhidas. O Recuva apontou um total de arquivos bem próximo ao total original, além do mais apresentou nomes e caminhos idênticos aos arquivos originais em grande parte dos seus resultados.

Dentro da computação forense é de suma importância que ao realizar uma recuperação de dados, mantenha-se o nome original do arquivo, como foi apresentado na ferramenta Recuva. Dado que, para o perito isso facilitará o seu trabalho além de que aumenta a assertividade de seu laudo final. Imagine que em um processo, um juiz realiza uma ação de busca de um arquivo de planilhas específico chamado "finanças", logo o perito realiza a cópia forense no local para realizar o processo de recuperação em seu laboratório. Então ao final do processamento a ferramenta entrega 1000 planilhas sem os nomes originais, sendo assim o perito deverá analisar todos os arquivos para encontrar apenas o arquivo que estava sendo buscado, aumentando a força de trabalho realizada, caso a ferramenta utilizada não seja a mais adequada para a situação.

Outra situação que possa ocorrer são com os arquivos fragmentados, como a ferramenta PhotoRec apresentou. Isso ocorre quando ao ser realizado o procedimento de recuperação, um arquivo original é apresentado em mais de uma parte no final do processo. Imagine a hipótese de que um perito esteja buscando um documento de texto bastante extenso, porém a ferramenta utilizada apresenta esse documento fragmentado em 5 diferentes partes. Em juízo, ao ser apresentado o laudo final contendo os 5 arquivos, o perito pode ser questionado, dado que a busca foi apenas para um arquivo, essa situação pode até se enquadrar na teoria dos frutos da árvore envenenada.

4 CONCLUSÃO

O estudo do sistema de arquivos NTFS é de suma importância na atividade de um acadêmico ou profissional da área de Computação Forense. Visto que o uso do NTFS é bastante difundido entre os usuários finais graças ao sistema operacional Windows por ser seu sistema de arquivos padrão.

Desse modo, a possibilidade de um profissional forense necessitar realizar uma recuperação de dados nesse tipo de sistema de arquivos é algo tangível. Nesse sentido foi imprescindível um estudo comparativo das ferramentas e tipos de mídias utilizadas para a recuperação de dados em um sistema de arquivos NTFS.

Por conseguinte, o estudo realizado neste trabalho expôs as eficácias e carências, comparando as ferramentes e técnicas de recuperação de dados em sistema de arquivos NTFS.

Desta maneira, conclui-se que o uso de mídias digitais e mídias magnéticas não causam grandes diferenças no resultado final. Já no âmbito das ferramentas de recuperação, a que mais se mostrou eficiente foi o Recuva, pois recuperou totalmente a maior parte dos arquivos, com seus nomes e caminhos originais.

Todavia, os resultados não menosprezam a importância de qualquer ferramenta utilizada neste trabalho, dado que os fatores envolvidas em uma perícia são incontáveis e nem sempre a ferramenta mais adequada é a que possui os melhores resultados.

Por fim, este trabalho não tem a intenção de declarar a melhor ferramenta, mas sim acrescentar conhecimento para a área de computação forense no âmbito da recuperação de dados.

4.1 Trabalhos Futuros

Para trabalhos futuros é proposto a utilização de dispositivos de armazenamentos SSD, pois o uso deste tipo de dispositivo vem aumentando com a diminuição de preços e as vantagens que ele traz consigo, como por exemplo, o tempo de acesso reduzido e largura de banda superior, além disso alguns dispositivos possuem implementado em hardware uma funcionalidade de wipe automático, assim dificultando a recuperação de dados. Bem como é proposta a utilização de outras ferramentas para testes, como por exemplo o R-Studio. Além disso, é visto de suma importância uma ferramenta que ajude a filtrar os arquivos falso-positivos de uma forma automatizada, diminuindo assim o tempo necessário para a análise dos arquivos recuperados.

REFERÊNCIAS

ALECRIM, E. **Sistema de arquivos NTFS**. 2011. Disponível em: <https://www.infowester.com/ntfs.php>.

CALAZANS, C. H.; CALAZANS, S. M. Ciência forense: das origens à ciência forense computacional. **São Paulo, SP: USP**, 2005.

CALDAS, M. S.; SILVA, E. C. C. Fundamentos e aplicação do big data: como tratar informações em uma sociedade de yottabytes. **Bibliotecas Universitárias: pesquisas, experiências e perpectivas**, v. 3, n. 1, 2016.

CARRIER, B. **File system forensic analysis**. [S.l.]: Addison-Wesley Professional, 2005.

CERT.BR. **Incidentes Reportados ao CERT.br – janeiro a dezembro de 2016**. 2017. Disponível em: <https://www.cert.br/stats/incidentes/2016-jan-dec/analise.html>.

CHEMELLO, E. Ciência forense: impressões digitais. **Química Virtual**, p. 1–11, 2006.

CRUZ, D. d. O. Recuperação de dados em sistema de arquivos ext4. 2015.

FBI. **Ransomware on the Rise**. 2015. Disponível em: <https://www.fbi.gov/news/stories/ransomware-on-the-rise>.

FILHO, J. E. M. **Técnicas forenses para a recuperação para a recuperação de arquivos**. 2016. Disponível em: <http://eriberto.pro.br/palestras/recuperacao_arquivos.pdf>.

FISHER, T. **What is the Write Zero Method?** 2017. Disponível em: <https://www.lifewire.com/what-is-the-write-zero-method-2626052>.

HAGEN, W. v. **Linux Filesystems**. [S.l.]: Sams, 2001.

LEOBONS, M. H. e R. **Análise Forense**. 2007. Disponível em: <https://www.gta.ufrj.br/grad/07_1/forense/reconhecimento.html>.

MELO, S. **Computação Forense com Software Livre**. [S.l.]: Rio de Janeiro: Alta Books, 2009.

MONTES, A. E. **Entenda o básico sobre INODES em Linux/Unix com exemplos**. 2014. Disponível em: <http://www.escher86.com.br/entenda-o-basico-sobre-inodes-em-linuxunix-com-exemplos/>.

NASCIMENTO, J. dos S.; JERÔNIMO, K. de S.; SEGUNDO, P. C. de S. Análise de ferramentas forenses de recuperação de dados. 2010.

NERY, A. **Ransomware: saiba sobre o malware sequestrador de dados**. 2016. Disponível em: <https://www.security.ufrj.br/2016/04/ransomware-sabia-sobre-o-malware-sequestrador-de-dados/>.

OLIVEIRA, E. P. d. Curso de processo penal. Atlas, 2017.

PEREIRA, E. D. V. Investigação digital: conceitos, ferramentas e estudos de caso. In: **III Congresso Tecnológico TI e Telecom InfoBrasil**. [S.l.: s.n.], 2010. p. 43–44.

POVAR, D.; BHADRAN, V. Forensic data carving. In: SPRINGER. **International Conference on Digital Forensics and Cyber Crime**. [S.l.], 2010. p. 137–148.

ROMER, R. **Brasil concentra 92% dos casos de ransomware na América Latina**. 2015. Disponível em: <https://canaltech.com.br/seguranca/brasil-concentra-92-dos-casos-de-ransomware-na-america-latina-48259/>.

SERRANO, B. L. **Ciencia Forense: ¿cómo usar la ciencia y la tecnología para desvelar lo ocurrido?** Todo-Ciencia.com, 2002. Disponível em: <http://matap.dmae.upm.es/WebpersonalBartolo/articulosdivulgacion/crimenes_3.htm>.

TERRA. **Tráfego de dados na internet ultrapassará o "zettabyte"em 2016**. 2015. Disponível em: <https://www.terra.com.br/noticias/tecnologia/internet/trafego-de-dados-na-internet-ultrapassara-o-zettabyte-em-2016,024bfe32cdbda310VgnCLD200000bbcceb0aRCRD.html>.

VILAR, G. P. **Formatação rápida ou completa?** 2016. Disponível em: <http://www.itnerante.com.br/profiles/blogs/formatacao-rapida-ou-completa>.

WEISZFLOG, W. **Michaelis Moderno Dicionário da Língua Portuguesa**. [S.l.]: Editora Melhoramentos, 2015.

ZURIARRAIN, J. M. **Android já é o sistema operacional mais usado do mundo**. 2017. Disponível em: <https://brasil.elpais.com/brasil/2017/04/04/tecnologia/1491296467_396232.html>.