RBTI – Revista Brasileira em Tecnologia da Informação
Volume 5, número 2 – 2023
ISSN: 2675-1828
Editor Científico: Jaime Cazuhiro Ossada

Estudo de caso de análise forense de arquivos apagados em área não alocada da memória

Marcelo Cirilo de Souza
USP, marcelocirilo.mc@gmail.com

RESUMO

Em análise computacional forense é necessário buscar vestígios em todas as partições do disco analisado. Contudo, muitas vezes as evidências foram apagadas previamente ou por que já se passou muito tempo e o usuário não necessitava mais daquele arquivo ou na tentativa de ocultar provas. Sendo assim, este artigo demonstra na prática como é feito o estudo e análise de arquivos apagados e que se encontram na área não alocada do disco não volátil (HDD, SSD, etc.) dos dispositivos de memória. Para se fazer tal análise foi feita uma breve revisão sobre algoritmos hash, para buscas de arquivos através de valores de hashes conhecidos, assim como foi feita uma análise de cabeçalhos em hexadecimal para que fosse possível determinar qual o tipo de arquivo em questão, qual o tamanho do arquivo original e qual origem de tais arquivos antes da sua exclusão. Para que fosse feita uma pratica alcançável para todos os usuários foram utilizados softwares forenses gratuitos Guymager, para criação de imagens, e Autopsy para análise dos dados recuperados. Apesar de haver programas gratuitos e comerciais para análise forense, a interpretação dos dados se faz necessária e se mostrou capaz de recuperar e identificar arquivos que haviam sidos previamente apagados da memória, mesmo que tivessem sidos parcialmente sobrescritos.

Palavras-Chave: Computação forense, File Carving, Recuperação de Dados.

Data do recebimento do artigo: 15/01/2023
Data do aceite de publicação: 01/03/2023
Data da publicação: 31/12/2023

Case study of file carving in unallocated space

ABSTRACT

In computer forensic analysis is necessary to look for evidences in all partitions of the analyzed disk. However, many times the evidences were previously deleted in an attempt to hide proofs or because a long time have passed and the user did not need that file and then deleted it. Therefore, this article demonstrates how is done the analysis of deleted files, how deleted files are discovered in the unallocated area of the non-volatile memories (HDD, SSD, etc.) of the hard disk driver analyzed. In order to carry out such analysis, a brief review of hash algorithms was made, for files with previously known hash values, as well as an analysis of the file headers by viewing the files in hex so that it was possible to determine the file extension, its original and its directory of origin. In order to make the practice accessible to all users, free forensic softwares were used such as Guymager, for forensics acquisition propose, and Autopsy, in order to analyze the recovered data. Although, there are free and commercial programs for forensic analysis, the interpretation of the data is necessary and it has proved to be capable of recovering and identifying files that had been erased from memory even if they had been partially overwritten.

Key Words: Computer Forensics; File Carving; Data Recovery.

1 Introdução

Dentro das ciências forenses se encontra a ciência forense digital, sendo que este termo foi utilizado pela primeira vez 2001 para identificar o uso cientifico da cadeia de custódia para preservação, coleta, processamento, análise e interpretação de dados a partir de fontes digitais que pudessem conter vestígios de crimes (Gary, 2001). Por sua vez, dentro da ciência forense digital se encontra a computação forense, ou informática forense, que é um ramo específico destinado a buscas, aquisição e interpretação de dados na procura de vestígios, principalmente de atos delitivos, em mídias não voláteis (HDD, SSD, etc.) de computadores.

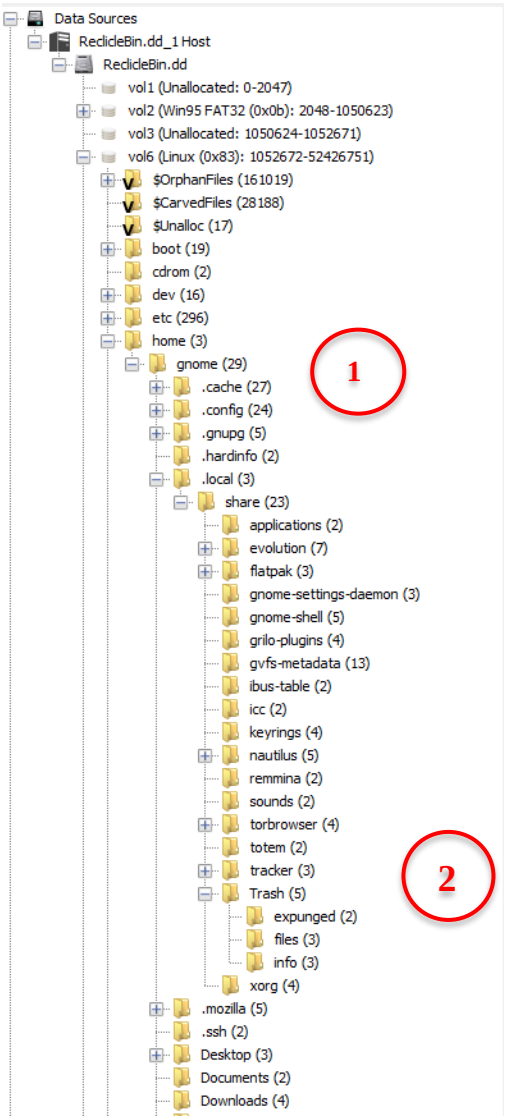
O ordenamento jurídico brasileiro entende que aparelhos eletrônicos, principalmente celulares e computadores, permitem acesso a um conjunto de informações pessoais, não se limitando a tratativas de práticas ilícitas em redes sociais, mas também a acessos e/ou produção de programas maliciosos, provas em documentos de texto ou planilhas. Sendo assim, os mandados de busca e apreensão são normalmente acompanhados de autorização de apreensão dos dispositivos eletrônicos e posteriormente análise pela equipe forense.

Na tentativa de ocultar as provas digitais, os usuários comumente deletam os arquivos de interesse e acreditam que desta maneira esses arquivos excluídos não poderão ser mais visualizados e que as informações de metadados (dados dos dados) não serão recuperados. No entanto, os arquivos deletados nos computadores são comumente enviados para a lixeira, onde unicamente ocorre a criação de uma nova pasta para onde vão os arquivos deletados.

Ao analisar a lixeira em sistemas operacionais Windows com softwares forenses, é possível observar que os arquivos deletados perdem o nome original de arquivo e assim são criados dois novos arquivo, sendo que este dois novos arquivos possuem 6 caracteres aleatórios idênticos, sendo diferenciados pelo prefixo $I, que possui informações como data de exclusão, tamanho do arquivo original e data da exclusão, e o prefixo $R, que contém o conteúdo do arquivo original (Llamas, 2019).

Já em sistemas operacionais Linux, os arquivos deletados não perdem seus nomes originais, mas o arquivo propriamente dito e seus metadados ficam alocados em pastas distintas sob a árvore da lixeira. Outra diferença entre os dois sistemas operacionais é que enquanto a lixeira aparece na raiz do sistema operacional Windows, ou seja, fora da árvore do usuário, o Linux mantém a lixeira dentro da árvore do usuário.

Figura 1 - Árvore Sistema Linux



Fonte: Elaborado pelo autor

Acima é exposto o sistema Linux visualizado com o Autopsy, sendo que 1 representa o nome de usuário (gnome) e 2 representa a pasta da lixeira (trash), em que files possui os arquivos deletados e info possui os metadados dos arquivos deletados. Caso se tratasse do sistema Windows, essa estrutura 2 estaria logo abaixo do Vol6, contendo os arquivos com prefixo $R e $I em uma única pasta.

Para analisar a lixeira pode-se visualizar os arquivos um a um, o que não é indicado quando se tem inúmeros arquivos dentro desta pasta. No entanto, quando se tem informações previas do arquivo procurado um dos artifícios utilizado é a busca através da

numeração hash, pois se o conteúdo do arquivo não foi alterado, mesmo que se altere seus metadados, os valores de hash permanecem o mesmo.

O algoritmo hash se trata de uma função que recebe arquivos em tamanhos variados e tem em sua saída uma string em tamanho fixo. Na entrada os arquivos são lidos em sua forma de bits e na saída do algoritmo é observado uma string, muitas vezes com valores em hexadecimal. Existem uma gama de algoritmos hashes que se diferenciam pelo tamanho da string de saída, tempo de execução e baixa probabilidade de colisão, que é quando dois arquivos possuem o mesmo valor de hash. Dentre os algoritmos de hashes mais comuns estão os algoritmos MD5 e SHA1 (Rountree, 2011).

Figura 2

Fonte: https://wallpapercave.com/w/ew8mE2V

Tabela 1 - Exemplos de hash da Erro! Fonte de referência não encontrada.

HASH	VALOR
MD5	ca7bcdd9d77b57e4b05fb9ec2a585cf7
SHA1	684e90d09666c84a59f277ff66db71546c7dbb5c

Fonte: Elaborado pelo autor

O hash MD5 produz um valor de hash de 128 bits expresso em 32 caracteres, enquanto o hash SHA1 produz um valor de hash de 160 bits expresso em 40 caracteres, sendo ambos expressos em hexadecimal.

Existem bibliotecas de valores de hashes de arquivos conhecidos, principalmente de arquivos maliciosos ou que contenham pedofilia. Desta maneira, através de comparações de valores de hashes se tornam simples a identificação de alguns arquivos. Como exemplo deste método existe o Project Vic (https://www.projectvic.org/) que

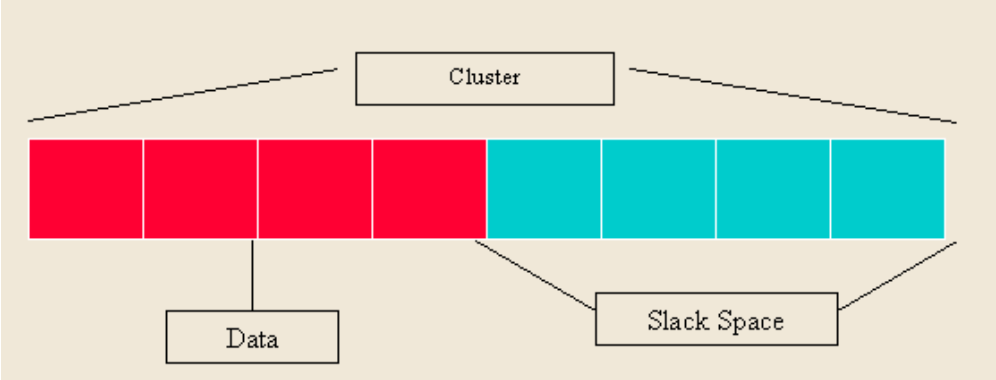
dentre várias atribuições detém e compartilha com autoridades os hashes de arquivos de vítimas de exploração sexual infantil, que são normalmente compartilhados pela internet.

Contudo, é possível deletar permanentemente os arquivos, seja utilizando o conjunto de teclas delete + shift ou esvaziando a lixeira. Estas ações podem levar alguns usuários a acreditarem que desta vez os arquivos foram apagados de forma definitiva, uma vez que não é mais possível visualizar o arquivo no sistema operacional, mas o que acontece é que os arquivos deletados permanentemente são designados para a área não alocada da memória.

A área não alocada corresponde ao espaço livre do disco ou o espaço que o arquivo deletado ocupava e passa a ser liberado para o usuário, podendo este espaço ser ocupado por outro arquivo, ou seja, um arquivo sobrescrevendo o outro. Mas o que ocorre é que a região física da memória do arquivo recém deletado ainda está ocupado pelo arquivo original deletado e permanecerá assim até que seja sobrescrito e neste ponto que os softwares forenses podem analisar e recuperar os dados nas áreas não alocadas.

Entende-se que os dispositivos de memória estão divididos em clusters, que são espaços menores da memória de tamanho fixo usados para armazenamento. O tamanho destes cluster podem ter tamanho variado de acordo com o sistema de arquivos padrão utilizado pelo sistema operacional, NTFS (New Technology File System) é o sistema de arquivos padrão do Windows, enquanto o ext4 é o do Linux. Sendo assim, caso um arquivo armazenado no disco de memória não ocupe todo um cluster, haverá a formação de Slack Espaço (slack space). Contudo, esse pequeno pedaço da memória desocupado pode conter vestígios de arquivos que já foram apagados. O conjunto de Slack Espaço é a memória não alocada, podendo conter cluster inteiros e/ou somente pedaços de clusters - Slack Espaço.

Figura 3 – Slack Space



Fonte: WU, 2023

O método de procurar e analisar dados que se encontram em áreas não alocadas recebe diversas denominações, sendo elas data recovery, data carving, file carving ou simplesmente carving. Há duas maneiras simples de visualizar e reconhecer os arquivos pelo método carving, sendo que uma das maneiras é através dos algoritmos hashes, desde que nenhum pedaço do conteúdo do arquivo tenha sido sobrescrito, ou seja, alterado.

Outra maneira de analisar os dados na área da memória não alocada é observar o cabeçalho dos arquivos em hexadecimal, pois desta maneira é possível identificar qual a extensão do arquivo em questão, independente se tenham trocado manualmente para forjarem ser de outro tipo de extensão. E desta maneira visualizar arquivo por arquivo que contenha a extensão do arquivo desejado (Darnowski & Chojnacki, 2015).

Na Tabela 2 é apresentada uma tabela com as assinaturas mais comuns para imagens, enquanto que a Figura 6 apresenta como essas assinaturas em hexadecimal são apresentadas nos arquivos.

Tabela 2 - Assinaturas de arquivos

Formato do arquivo	Assinatura (hexadecimal)
JFIF, JPE, JPEG, JPG	FF D8 FF
DOCX, PPTX, XLSX	50 4B 03 04
PDF	25 50 44 46

Fonte: KESSLER, 2022

2 Metodologia

O sistema operacional a ser analisado foi um Ubuntu versão 2022.4, sendo que para analisar a recuperação de dados foram feitos dois procedimentos. Primeiramente, a Figura 2, que se encontrava no Desktop do Ubuntu, foi deletado e desta forma foi analisado o sistema operacional com a imagem na lixeira, posteriormente, a lixeira foi esvaziada para que assim fosse analisado os dados recuperados do espaço da memória não alocada.

Para a criação da imagem do sistema operacional Ubuntu foi utilizado o programa Guymager versão 0.8.13-1, e para o processamento e análise da imagem foi utilizado o programa forense Autopsy versão 4.19.3. Ambos os softwares são gratuitos e nativos dos sistemas operacionais Linux Kali e Caine.

Frisa-se que imagem é a cópia bit a bit do disco a ser analisado, podendo esta imagem ser salva em algum diretório para que possa ser posteriormente analisada por algum software forense. O software que for utilizado para a criação da imagem precisa ser capaz de calcular pelo menos dois valores de hashes com algoritmos distintos para garantir que não ocorra colisão e desta forma garantir a cadeia de custódia.

Para as buscas por hash foi utilizado o algoritmo hash MD5, porém foi utilizado este unicamente por praticidade, já que o mesmo possui uma string menor em comparação com outros algoritmos, mas o recomendado é trabalhar com algoritmos hashes da família SHA, como SHA1, pois a probabilidade de colisão é menor, ou, preferencialmente, trabalhar com dois algoritmos hashes distintos, para que a probabilidade de colisão seja praticamente nula.

Para a busca pelo método carving foram separados os arquivos que continham no cabeçalho informações que indicassem se tratarem de imagens, sendo estas mesmas visualizadas um a um. Contudo, também foram feitas pesquisas por valores de hashes em arquivos que se encontravam no espaço da memória não .

3 Resultados

O software Guymager a priori calcula o hash do disco a ser analisado, porém, o software Autopsy calcula novamente o hash do disco. Desta maneira, com os hashes

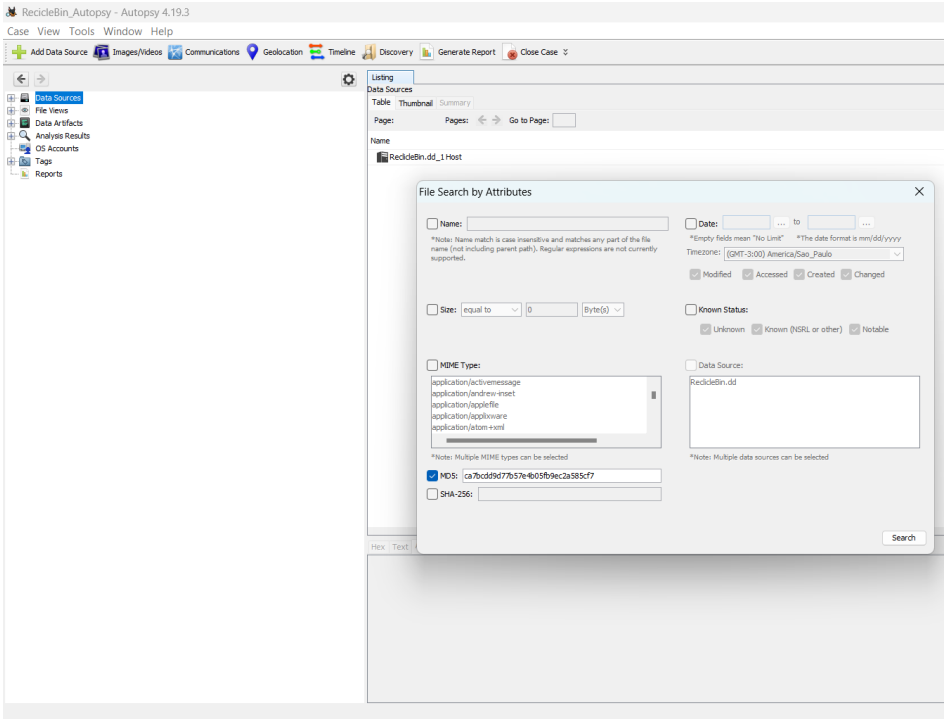
coincidindo a partir de softwares distintos é possível garantir a cadeia de custódia do disco analisado.

Foi utilizado para o processamento da imagem um Intel(R) Core(TM) i7-1065G7 CPU @ 1.30GHz 1.50 GHz e RAM de 20 GB. O sistema operacional Ubuntu analisado possuía o tamanho de 26 GB. Levando o tempo de processamento pelo Autopsy de 15 minutos, cerca de 1,7 GB/min. Frisa-se que para este primeiro processamento não foi feito o método carving, ou seja, não foi observado o espaço não alocado da memória.

Isso demostra que para discos acima de 500 GB levaria mais de 4 horas de processamento, o que demonstra que para discos de maior capacidade é necessário um computador mais robusto, caso contrário o tempo de processamento pode não viável para um laboratório forense.

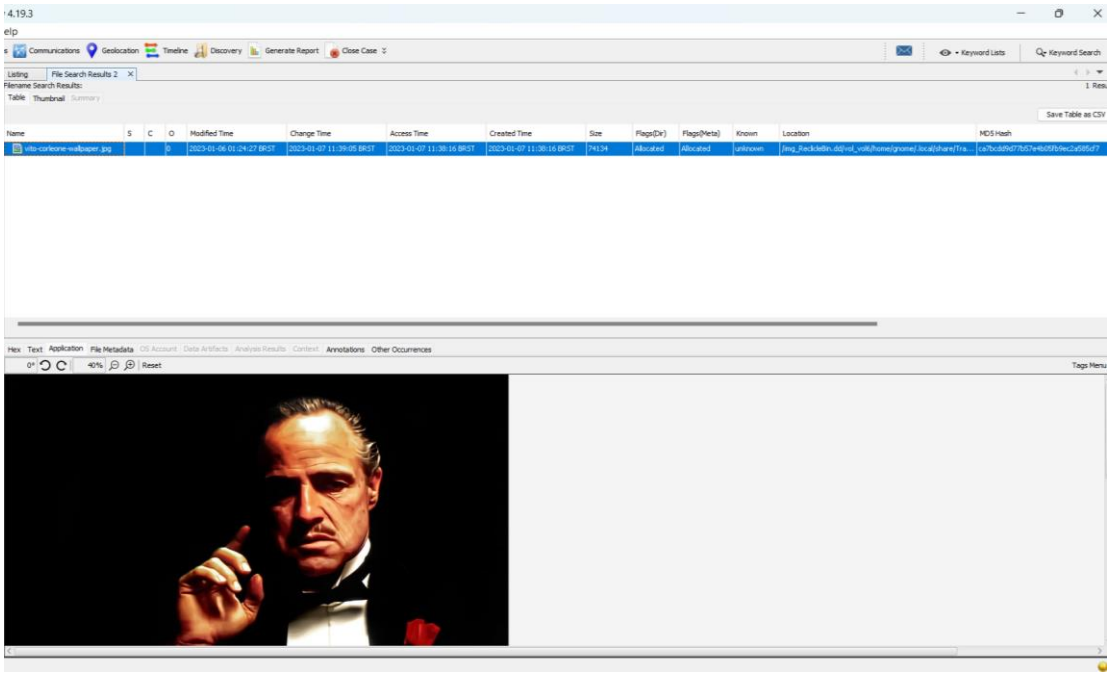
Primeiramente foi processado o sistema operacional com o arquivo Figura 2 na lixeira, sendo que a busca pelo hash MD5, da Tabela 1, levou menos de meio minuto para mostrar o resultado, Figura 4 e Figura 5, o que demostra que a indexação de dados no processamento foi eficaz, assim como a revelação do arquivo foi rápida

Figura 4: Busca por hash



Fonte: Elaborado pelo autor

Figura 5: Busca por hash



Fonte: Elaborado pelo autor

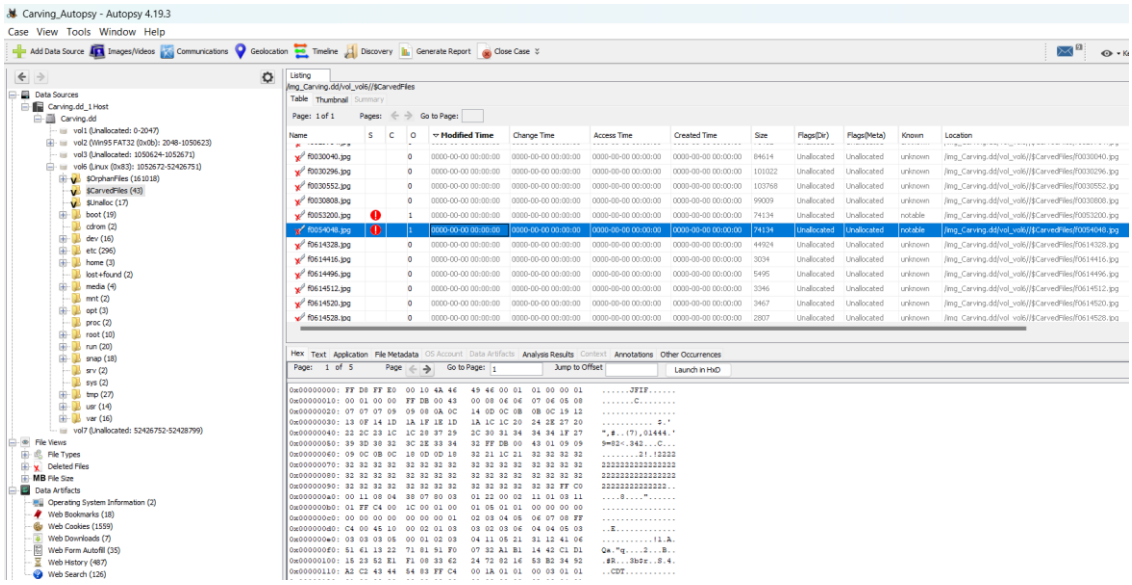
Na Figura 4 é demonstrada a busca por um hash já conhecido, enquanto a Figura 5 demonstra o resultado desta busca. Ainda na Figura 5 é possível observar que mesmo que o arquivo se encontre na lixeira (coluna location), o mesmo se encontra em um espaço da memória alocado (coluna flag), sendo assim, não poderia ser sobrescrito, a menos que a lixeira esteja programada para ser esvaziada após determinado dias de exclusão do arquivo. Ademais, é carregada a data de criação na pasta no diretório, neste caso corresponde com a de exclusão.

Posteriormente, a lixeira do sistema Ubuntu foi esvaziada e refeito todos os passos para criação de imagem e processamento. Foi utilizada a mesma configuração de computador de processamento, mas neste processo foi feito o carving, ou seja, analisado o espaço da memória não alocado, levando o tempo de processamento de 25 minutos, logo, cerca de 1 GB/min. O que significou uma redução de 40% da velocidade de processamento. Portanto, foi possível perceber que o processamento carving demanda mais capacidade do sistema.

Para a segunda etapa foi necessário observar os arquivos que se encontravam na pasta CarvedFiles, presente na Figura 6, e observar o cabeçalho dos arquivos, conforme destacado na figura abaixo. Neste caso de interesse, os arquivos com as iniciais

FFD8FFE0, conforme destacado em vermelho na figura. Contudo, o software forense Autopsy automatiza a leitura dos cabeçalhos e separa os arquivos que possuem em seu cabeçalho assinaturas de imagens e vídeos, além disso o software cria esses arquivos identificado os thumbnails (miniaturas), não sendo necessário verificar arquivos que possuam em sua assinatura identificações de arquivos de texto, por exemplo, e podendo focar somente em arquivos de imagem que contenham o cabeçalho de interesse.

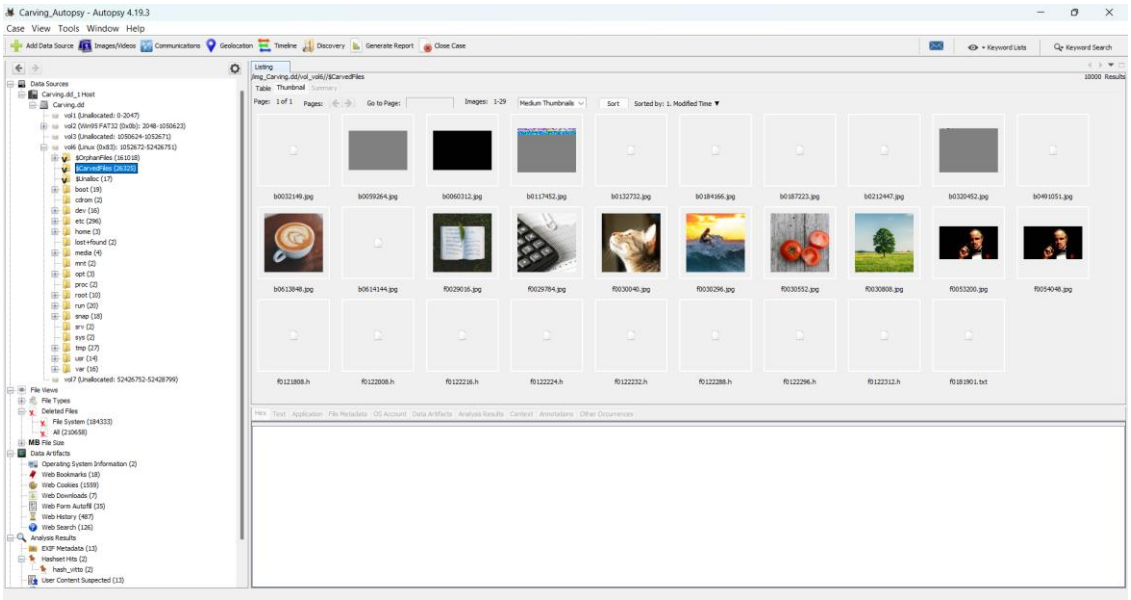
Figura 6: Arquivos recuperados por carving



Fonte: Elaborado pelo autor

Conforme figura 7 observa-se que essa filtragem de cabeçalho fez com que a análise saísse de um cenário de 26.325 arquivos recuperados para um cenário com somente 29 arquivos de interesse, somente utilizando a técnica de cabeçalho.

Figura 7 - Thumbnail de arquivos de carving



Fonte: Elaborado pelo autor

Como observado na figura acima, esta técnica de cabeçalho permitiu a observação do arquivo de interesse que se encontrava no espaço da memória não alocada visualizando os thumbnails. Frisa-se que o quarto thumbnail da figura acima é um exemplo clássico de imagem no espaço da memória não alocada que já teve boa parte do seu conteúdo sobrescrita, apresentando somente alguns pixels da imagem original e apresentado um borrão escuro em boa parte da imagem, que é a parte sobrescrita

4 Considerações finais

Observou-se que a busca por espaço da memória não alocada faz o tempo de processamento aumentar, ou seja, a velocidade de processamento cair, em cerca de 40%, por isso, o recomendado em discos de tamanho considerável é fazer o processamento sem carving e caso não se encontre o arquivo de interesse, deve-se fazer outro processamento somente com carving.

Ressalta-se que no segundo caso, como a imagem de interesse não estava sobrescrita, a técnica de busca de hash se mostrou eficaz, mas raramente arquivos de interesse possuem hashes conhecidos, sendo necessário o uso da técnica de busca por arquivos através do cabeçalho em hexadecimal. Os experimentos mostraram eficácia nas

buscas por imagens, separando os arquivos de imagens dos demais arquivos, como arquivos de texto, por exemplo.

Mesmos os programas utilizados sendo gratuitos e livres, ambos cumpriram com a função e foram capazes de recuperar arquivos em espaços não alocados da memória, mas o tempo de processamento é dependente da capacidade de processamento, pois ao utilizar um computador mais robusto, como o Processador Intel(R) Xeon(R) Silver 4214R CPU @ 2.40GHz 2.39 GHz RAM 64,0 GB, o tempo de processamento do Autopsy caiu quase pela metade, mostrando que para laboratórios forenses é necessário o uso equipamentos adequados.

Referências Bibliográficas

Darnowski, F., & Chojnacki, A. (2015). Selected Methods of File Carving and Analysis of Digital Storage Media in Computer Forensics. TELEINFORMATICS REVIEW, pp. 26-27.

Gary, P. (2001). A Road Map for Digital Forensic Research. Utica, NY: DFRWS.

Kessler, G. (09 de dezembro de 2022). File signatures table. Fonte: GCK'S FILE SIGNATURES TABLE: https://www.garykessler.net/library/file_sigs.html

Llamas, J. M. (2019). Analisis and Design of Digital Forensics and Incident Response Procedure. Madri: Universidad Politécnica de Madrid.

Rountree, D. (23 de setembro de 2011). 2 - Cryptography. Security for Microsoft Windows System Administrators, pp. 29-69.

Wu, W. (05 de janeiro de 2023). CISSP PRACTICE QUESTIONS – 20211124. Fonte: https://wentzwu.com/2021/11/24/cissp-practice-questions-20211124/

.