FERNANDO VILGINO SBAMPATO

**ESTUDO DA APLICABILIDADE DE TÉCNICAS DE SANITIZAÇÃO
DE DADOS EM DISCOS RÍGIDOS ATUAIS**

**São Paulo
2018**

FERNANDO VILGINO SBAMPATO

**ESTUDO DA APLICABILIDADE DE TÉCNICAS DE
SANITIZAÇÃO DE DADOS EM DISCOS RÍGIDOS
ATUAIS**

Dissertação apresentada à Escola
Politécnica da Universidade de São
Paulo para a obtenção do Título de
Mestre em Ciências.

**São Paulo
2018**

FERNANDO VILGINO SBAMPATO

**ESTUDO DA APLICABILIDADE DE TÉCNICAS DE
SANITIZAÇÃO DE DADOS EM DISCOS RÍGIDOS
ATUAIS**

Dissertação apresentada à Escola
Politécnica da Universidade de São
Paulo para obtenção do Título de Mestre
em Ciências.

Área de concentração:
Engenharia de computação

Orientadora: Dra. Tereza Cristina Melo
de Brito Carvalho

**São Paulo
2018**

Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 18 de Julho de 2018.

Assinatura do autor

Assinatura do orientador

FICHA CATALOGRÁFICA

Sbampato, Fernando Vilgino

Estudo da aplicabilidade de técnicas de sanitização de dados em discos rígidos atuais/ F.V. Sbampato. – versão corr. - São Paulo, 2018. 108 p.

Dissertação (Mestrado) – Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais (PCS).

1. Sanitização de dados 2. Disco rígido 3. Técnicas de sanitização de dados. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas (PCS). II. T.

AGRADECIMENTOS

À Deus em primeiro lugar, pela sua infinita graça sobre a minha vida e pelos milagres já realizados e que serão realizados em meu favor. À minha orientadora Tereza Cristina Melo de Brito Carvalho, pela sua dedicação e apoio neste projeto. Sou grato à sua dedicação e amor.

Aos meus pais, Flávio (em memória) e Janete, pelo admirável apoio em todos os projetos que participei e pelos conselhos. À minha esposa Márcia, por sempre estar ao meu lado em todos os momentos desde que estamos juntos e ao meu filho Fernando que me motiva e encoraja sempre nos novos desafios enfrentados na vida. Ao meu irmão Fabio pelo apoio.

Aos meus colegas de pós-graduação e do LARC-USP (Laboratório de Arquitetura e Rede de Computadores) em estima: Karen Langona, Marcos Wando, Fernando Frota Redígolo, Rosangela de Fátima Pereira Marquesone, Ricardo Wandré Dias Pedro, Ewerton Rodrigues Andrade, Bruno Medeiros, Cleber Okida, Marco Antonio Torrez, Nelson Mimura Gonzales e Bruno Bastos por me ajudarem nesta dissertação, tanto na troca de ideias, como pela amizade e pela troca de experiências que foram essenciais ao longo deste trabalho, ajudando a construir e solidificar o conhecimento que hoje partilhamos.

À Ericsson Telecomunicações do Brasil, pelo suporte financeiro realizado por meio da FDTE (Fundação para o Desenvolvimento Tecnológico da Engenharia) e à Ericsson Research (Canadá, Suécia e Finlândia) pelo apoio durante a realização de pesquisas e projetos em que participei.

Agradeço à Edilaine e ao Ricardo pelo suporte nas questões relacionadas a contratos de bolsa e em geral. Para terminar, agradeço à Universidade de São Paulo e à Escola Politécnica da Universidade de São Paulo pela enorme oportunidade concedida, pelo espaço cedido que tornou possível o desenvolvimento da pesquisa e por contar com grandes professores e profissionais que sempre auxiliam na construção do conhecimento.

DEDICATÓRIA

Agradeço a Deus primeiramente pela oportunidade, à minha esposa Márcia Sbampato, à minha mãe Janete Sbampato, ao meu filho Fernando Sbampato, e ao meu irmão Fabio Sbampato, por acreditar neste projeto e por estarem aqui em sua conclusão.

RESUMO

A sanitização de dados é um dos desafios que está em aberto quando tange a segurança de dados nos discos rígidos. Há duas formas de realizar este procedimento de sanitização de dados nos discos rígidos. A primeira é a utilização de técnicas físicas que visam à destruição do disco rígido por completo. A segunda é a utilização de técnicas lógicas que visam realizar a sanitização dos dados armazenados no disco, permitindo que este seja reutilizado. A proposta principal deste trabalho é a de verificar por meio da técnica de Microscopia de Força Magnética (MFM) a possibilidade de recuperação dos dados originais após o processo de sanitização ter ocorrido por meio de uma técnica lógica. Com este objetivo foram selecionadas oito técnicas lógicas (Gutmann, VSITR, RCMP TSSIP OPS-II, CSEC ITSG-06, DoD 5220.22-M, AR 280-19, GOST R 50739-95 e ISM 6.9.92), após esta seleção foi realizada uma avaliação lógica dessas técnicas com o intuito de selecionar duas técnicas para a avaliação experimental. Para realizar a avaliação experimental foram utilizados dois microscópios (Dimension Icon e o MultiMode 8) para aplicar a técnica de MFM em disco rígido. O objetivo foi comprovar a eficiência das técnicas lógicas na sanitização de dados armazenados nos discos rígidos.

Palavras-chave: Sanitização de dados, disco rígido, técnicas físicas de sanitização, técnicas lógicas de sanitização, Microscopia de Força Magnética.

ABSTRACT

Data sanitization is one of the challenges you face when it comes to data security on hard disk. There are two ways to perform this data sanitization procedure on hard disks. The first one is the deployment of physical techniques aimed at destroying the hard drive altogether. The second one is the use of logical techniques that aim to sanitize the data stored on the disk, allowing it to be used again. The main purpose of this work is to verify through the Magnetic Force Microscopy technique (MFM) the possibility of recovering of the original data after the sanitization process has occurred through of a logical technique. To this purpose, eight logics (Gutmann, VSITR, RCMP TSSIP OPS-II, CSEC ITSG-06, DoD 5220.22-M, AR 280-19, GOST R 50739-95 and ISM 6.9.92) were selected. After this selection, a logical evaluation of these techniques was carried out for selecting two techniques for the experimental evaluation. To perform the experimental evaluation, two microscopes (Dimension Icon and MultiMode 8) were used to apply the MFM technique to the hard disk. The objective was to verify the efficiency of the logical techniques in the sanitization of data stored in the hard disks.

Keywords: Data sanitization, hard disk, physical sanitization techniques, logical sanitization techniques, Magnetic Force Microscopy

SUMÁRIO

1	Introdução.....	17
1.1	Motivação e descrição do problema	18
1.2	Objetivos	20
1.3	Método adotado para a pesquisa	21
1.4	Trabalhos relacionados	24
1.5	Organização	26
2	Sanitização de Dados em Discos Rígidos.....	27
2.1	Modelos de disco rígido e capacidade de armazenamento	27
2.2	Processo de gravação e leitura dos dados	29
2.3	Mudanças na Forma de Armazenamento de dados dos consumidores.....	34
2.4	Técnicas de sanitização de dados.....	35
2.4.1	Técnicas Físicas.....	35
2.4.2	Técnicas lógicas	38
2.5	Ataques para recuperar dados em disco rígido.....	46
2.6	Recomendações e leis para sanitização de dados	47
2.7	Considerações finais do capítulo.....	49
3	Procedimento Experimental para Avaliação de Sanitização de Discos Rígidos	51
3.1	Avaliação Lógica das Técnicas de Sanitização de dados	51
3.2	Avaliação de Sanitização com o uso de Microscopia de Força Magnética	54
3.2.1	Microscopia de força magnética	54
3.2.2	Modelos de Microscópios utilizados para a avaliação Experimental.....	58
3.2.3	Procedimento Experimental para Análise por MFM.....	65
3.2.4	Tempo e Energia Elétrica desprendido para a Sanitização dos Dados nos Discos Rígidos	68
3.2.5	Preparação das amostras e Recursos utilizados	69
3.2.6	Resultados das Avaliações Lógica e Experimental	72
3.3	Resultados das avaliações realizadas	72
3.3.1	Resultados da avaliação lógica	72
3.3.2	Leitura das Amostras por meio da Técnica de MFM.....	75
3.4	Considerações Finais do Capítulo.....	87
4	Considerações Finais.....	89
4.1	Contribuições	89
4.2	Participação em Publicações e Publicações relacionadas a pesquisa	90
4.3	Trabalhos Futuros	91
5	Referências.....	93
	Apêndice A - Componentes do disco rígido	101
A.1	Componentes mecânicos.....	101
	Os principais componentes mecânicos são:	101
A.2	Componentes lógicos	104
	Os principais componentes lógicos são:	104
	Apêndice B – Preparação das amostras para o microscópio MultiMode 8.....	107

LISTA DE ILUSTRAÇÕES

Figura 1: Sequência de passos realizados na revisão bibliográfica.....	23
Figura 2: Capacidade de armazenamento de dados dos discos rígidos com o passar dos anos (<i>IEEE-SPECTRUM</i> , 2015).....	28
Figura 3: Apresentação da compressão de dados realizada nas modulações FM, MFM e RLL (<i>KOZIEROK</i> , 2001).....	31
Figura 4: Utilização dos setores com o novo padrão 4k (<i>ACER</i> , 2016).....	33
Figura 5: Degausser (<i>DATA</i> , 2008).....	36
Figura 6: Próton PPD-1 (<i>ESALINK</i> , 2018).....	37
Figura 7: 0300 Low/medium Volume Hard Drive Shredder (<i>ESALINK</i> , 2016).....	37
Figura 8: Utilização dos equipamentos Próton PPD - 1 e 0300 Low/medium Volume Hard Drive Shredder (<i>DATASHEETS</i> , 2018).....	38
Figura 9: Imagem resultante da aplicação da técnica MFM em disco rígido (<i>BELARMINO et al.</i> , 2004).....	47
Figura 10: <i>String</i> e sequência de bits gravados nos volumes lógicos.....	52
Figura 11: Passos realizados para a avaliação lógica nos volumes criados.	52
Figura 12: Interação magnética entre ponta e amostra (<i>Bruker, MultiMode 8</i> , 2016).....	56
Figura 13: Componentes do microscópio MultiMode 8 (<i>Bruker, MultiMode 8</i> , 2016).....	59
Figura 14: Tipos de scanners (<i>Bruker, MultiMode 8</i> , 2016).....	60
Figura 15: Scanners J.	60
Figura 16: Modelos de cantiléver utilizados no MultiMode 8 (<i>Bruker, MultiMode 8</i> , 2016).....	61
Figura 17: Porta amostra do microscópio MultiMode 8 (<i>Bruker, MultiMode 8</i> , 2016).....	62
Figura 18: Microscópio Dimension Icon e seus componentes.....	63
Figura 19: Porta amostra utilizado no Dimension Icon para colocação das amostras.	64
Figura 20: Visão geral dos componentes do porta amostra do Dimension Icon (<i>Bruker, Dimension Icon</i> , 2017).....	65
Figura 21: Preparação dos discos rígidos para análise por MFM.	67
Figura 22: Tempo gasto em segundo para sobrescrever o passo 0 em todos o disco rígido (Próprio autor).....	68
Figura 23: Amostra utilizada no MultiMode 8 e no Dimension Icon.	70
Figura 24: Exemplo de imagens geradas com MFM de uma amostra de um disco rígido.	76
Figura 25: Imagens das amostras afetada pelo toque da ponta do microscópio Multimode 8.....	78
Figura 26: Imagens das amostras com 0's armazenados – Microscópio Multimode 8.....	78
Figura 27: Imagens das Amostras com valores diferentes de 0 armazenados – Microscópio Multimode 8.....	79
Figura 28: Processo de preparação e colocação da amostra no porta amostra do Dimension Icon (Próprio autor).....	81
Figura 29: Imagens da Amostra com 0's gravados – Microscópio Dimension Icon.	81
Figura 30: Imagens em 2D - Amostra com 1's gravados – Microscópio Dimension Icon.	82
Figura 31: Imagens em 3D - Amostra com 1's gravados – Microscópio Dimension Icon.	82
Figura 32: Imagens em 2D - Amostra com Dados Originais gravados – Microscópio Dimension Icon.	83
Figura 33: Imagens em 3D - Amostra com Dados Originais gravados – Microscópio Dimension Icon.	83
Figura 34: Região utilizada para verificar a possibilidade de restauração dos dados gravados.....	84
Figura 35: Imagens em 2D - Amostra com Dados Sanitizados (<i>Guttmann</i>) – Microscópio Dimension Icon.....	85
Figura 36: Imagens em 3D - Amostra com Dados Sanitizados (<i>Guttmann</i>) – Microscópio Dimension Icon.....	85
Figura 37: Imagens em 2D - Amostra com Dados Sanitizados (<i>Guttmann</i>) nos discos 5 e 6 – Microscópio Dimension Icon.....	86
Figura 37: Imagens em 2D - Amostra com Dados Sanitizados (<i>RCMO TSSIT OPS-II</i>) nos discos 5 e 6 – Microscópio Dimension.....	87

Figura 39: Apresentação dos componentes mecânicos dentro de um disco rígido.	104
Figura 40: Apresentação dos componentes lógicos dentro de um setor (SEAGATE, 2013).	105
Figura 41: Apresentação dos componentes lógicos do disco rígido (PIROPO, 2007).	106
Figura 42: Processo para preparação da amostra.	107

LISTA DE TABELAS

Tabela 1: Passos da técnica de sanitização Gutmann.....	25
Tabela 2: Relação da capacidade de armazenamento, dos setores e da resolução dos setores.	32
Tabela 3: Passos da técnica de sanitização Gutmann.....	39
Tabela 3: Detalhes das técnicas de sanitização de dados.	45
Tabela 5: Distância de Hamming para os volumes sanitizados.	73
Tabela 6: Resultado da Análise das Técnicas Lógicas.....	74

LISTA DE ABREVIATURA E SIGLAS

ACM	<i>Association for Computing Machinery</i>
AF	<i>Advanced Format</i>
ATA	<i>Advanced Technology Attachment</i>
ATAPI	<i>Advanced Technology Attachment Packet Interface</i>
ARLL	<i>Advanced Run Length Limited</i>
AMI	<i>Amazon Machine Image</i>
AWS	<i>Amazon Web Services</i>
AFM	<i>Atomic Force Microscopy</i>
BSI	<i>Bundesamt für Sicherheit in der Informationstechnik</i>
CESG	<i>Communications - Electronic Security Group</i>
CPNI	<i>Center for the Protection of National Infrastructure</i>
CSEC	<i>Communications Electronic Security Group</i>
CSA	<i>Cloud Security Alliance</i>
CSCC	<i>Cloud Standard Customer Council</i>
DSP	<i>Digital Signal Processing</i>
DMTF	<i>Distributed Management Task Force</i>
DoD	<i>Department of Defense of UK</i>
DRAM	<i>Dynamic Random Access Memory</i>
ENISA	<i>European Network and Information Security Agency</i>
EPRML	<i>Extended Partial Response, Maximum Likelihood</i>
EIDE	<i>Extended Intelligent Drive Electronics</i>
EFM	<i>Electrical Force Microscopy</i>
FM	<i>Frequency Modulation</i>

GCHQ	<i>UK Government Communications Headquarters</i>
GNTAIA	<i>Government's National Authority for Information Assurance</i>
GLBA	<i>Gramm – Leach – Bliley</i>
HIPAA	<i>Health Insurance Portability and accountability Act</i>
IBM	<i>International Business Machines</i>
IDE	<i>Intelligent Drive Electronics</i>
IDEMA	<i>International Disk Equipment and Materials Association</i>
IEEE	<i>Institute of Electrical and Electronic Engineers</i>
ITU – T	<i>International Telecommunication Union</i>
ISACA	<i>Information Systems Audit and Control Association</i>
ISM	<i>Information Security Manual</i>
LARC	<i>Laboratório de Arquitetura e Redes de Computadores</i>
LBA	<i>Logical Block Address</i>
LFF	<i>Laboratório de Filmes Finos</i>
LNMS	<i>Laboratório de Novos Materiais Semicondutores</i>
MFM	<i>Modified Frequency Modulation</i>
MFM	<i>Magnetic Force Microscopy</i>
NASA	<i>National Security Agency</i>
NCSC	<i>National Computer Security Center</i>
NIST	<i>National Institute of Standards and Technology</i>
NCSC	<i>National Computer Security Center</i>
PATA	<i>Parallel Advanced Technology Attachment</i>
PRML	<i>Partial Response, Maximum Likelihood</i>
PIPEDA	<i>Personal Information Protection and Electronic Documents Act</i>

PCS	Departamento de Engenharia de Computação e Sistemas Digitais
RAID	<i>Redundant Array of Inexpensive Disks</i>
RCMP	<i>Royal Canadian Mounted Police</i>
RLL	<i>Run Length Limited</i>
RMS	<i>Root Main Square</i>
SATA	<i>Serial Advanced Technology Attachment</i>
SCSI	<i>Small Computer System Interface</i>
SPM	<i>Scanner Probe Microscopy</i>
SThM	<i>Scanner Thermal Microscopy</i>
USAF	<i>US Air Force</i>
UKGCHQ	<i>UK Government Communications Headquarter</i>
VSITR	<i>Verchlussachen – IT- Richtlinien</i>

1 Introdução

Desde o desenvolvimento dos primeiros computadores, como o modelo IBM 305 RAMAC Disk System desenvolvido pela IBM em 1956 (CHEN et al., 2006) até os atuais computadores, como por exemplo *desktops* e *notebooks*, é possível utilizar o dispositivo disco rígido para realizar o armazenamento final dos dados dos usuários (ABRAMOVITCH et al., 2002).

Os discos rígidos evoluíram consideravelmente ao longo dos anos, acompanhando os avanços tecnológicos do setor (WOOD, 2009). Esta evolução ocorreu tanto no tamanho físico, como na capacidade de armazenamento, passando de 60 cm de diâmetro do prato do disco rígido com capacidade de armazenamento de 5 MB, para os discos rígidos de 4 Terabytes de armazenamento e 3,8” de tamanho do prato do disco rígido, como por exemplo, os discos rígidos da empresa SEAGATE (SEAGATE, 2018).

Desta forma sua evolução possibilitou que seu custo/benefício se mantivesse favorável ao consumidor, mesmo quando comparados a novos dispositivos de armazenamento de dados, como o dispositivo de estado sólido ou SSD (*Solid State Disk*) (DURKEE, 2010; KASAVAJHALA, 2011). Atualmente os discos rígidos são amplamente utilizados em vários ambientes computacionais, como por exemplo, o ambiente de computação em nuvem, para armazenamento de dados dos consumidores disponibilizado pela Internet (DIKAIKOS et al., 2009), em *data centers*, entre outros.

Porém, desde o começo da evolução dos discos rígidos até o presente momento, um dos desafios em aberto é a **remoção definitiva dos dados do disco após o seu uso**, quando eles não são mais necessários ao seu proprietário, devido ao valor agregado que os dados podem receber e de potenciais implicações do ponto de vista de segurança e de privacidade caso os dados sejam recuperados por terceiros (GUTMANN, 1996; SIMSON, 2003; HUGHES et al., 2006; MCDONALD et al.; 2009, MARKOV, 2013; LU et al. 2014; ROJAS et al.; 2015).

A remoção de dados, também denominado **sanitização de dados** é um processo que visa a eliminação total dos dados armazenados em uma mídia, por exemplo, um disco rígido, garantindo que os dados sanitizados não sejam mais recuperáveis por nenhum indivíduo, tendo ele autorização ou não para acessar estes

dados (BARRETO, 2009). Este desafio de **sanitização de dados em discos rígidos** é o abordado neste trabalho.

1.1 Motivação e descrição do problema

Segundo CHEN et al. (2012) os dados contêm um ciclo de vida composto de sete fases:

- 1º) Criação dos dados;
- 2º) Transferência dos dados para o local definido pelo(s) seu(s) proprietários(s), que deve ocorrer de acordo com a política estabelecida pelos envolvidos;
- 3º) Uso dos dados, onde são estabelecidos os privilégios necessários para que os usuários tenham acesso a estes dados;
- 4º) Compartilhamento dos dados com todos os envolvidos;
- 5º) Utilização dos dados pelos envolvidos;
- 6º) Arquivamento dos dados após seu uso pelo envolvidos; e
- 7º) Sanitização dos dados.

Na sétima fase de sanitização dos dados, o usuário deseja que seus dados sejam removidos do disco rígido e deseja que estes não sejam mais utilizados por mais nenhum indivíduo (MARKOV, 2013). O NIST na publicação 800 - 88 (KISSEL et al.; 2014) definiu quatro maneiras para que a sanitização de dados ocorra na mídia que armazena estes dados.

- 1º) Descarte: descartar a mídia de maneira simples, sem a aplicação de qualquer método de sanitização, devido aos dados armazenados não terem nenhum valor ao usuário, por exemplo, o descarte de uma folha de papel impressa, que pode ser destinada a reciclagem;
- 2º) Limpeza: nível de sanitização de dados mais elevado, pois prestigia a confiabilidade e a privacidade dos dados, devido à importância que o mesmo possa ter ao usuário. Neste nível as técnicas de sanitização podem atuar como garantia que estes dados originais sobrescritos por

uma técnica lógica não poderão ser mais recuperados após o processo de sobrescrita ter ocorrido com êxito;

3º) Purgar: nível de sanitização mais alto que o nível de Limpeza, pois seu objetivo é o de proteger os dados impedindo que os mesmos sejam recuperados por um ataque até mesmo em nível de laboratório. Neste cenário também o uso de técnica lógica de sanitização de pode contribuir, uma vez que estes dados originais sobrescritos por uma técnica lógica não poderão ser mais recuperados; e

4º) Destruição: destruição da mídia, por exemplo, um disco rígido que armazena os dados considerados sensíveis ao usuário. Esta destruição do disco rígido pode ser evitada se o dispositivo estiver em condições de uso novamente por meio da utilização das técnicas lógicas de sanitização de dados para garantir que estes dados não possam ser recuperados.

Esta sanitização dos dados deve ocorrer de forma segura, visando garantir que não haja mais nenhum resquício dos dados originais armazenados anteriormente no disco rígido, não importando o ambiente no qual o disco rígido está inserido, tais como um ambiente corporativo ou acadêmico (BOHN et al., 2011). Entretanto, no trabalho de Balduzzi (BALDUZZI et al., 2012) foi analisada a segurança do serviço de computação em nuvem da *Amazon Elastic Compute Cloud* EC2 da empresa *Amazon*, sendo que 98% dos dados removidos das AMIs (*Amazon Machine Image*) **que estavam armazenados nos discos rígidos em uso no ambiente puderam ser recuperados neste experimento**, causando insegurança nos consumidores em relação a privacidade de seus dados armazenados nos discos rígidos utilizados neste ambiente.

A ***Amazon*** utiliza técnicas de sanitização de dados ao **final da vida útil do disco rígido**, como por exemplo a técnica DoD 5220.20.22-M do Departamento de Defesa Norte Americano (DOD, 1991), ou seja, quando o disco rígido utilizado por ela no ambiente chega ao final da sua vida útil, quando o disco rígido será removido e não será mais usado no ambiente computacional, sendo totalmente destruído. Uma possível abordagem para evitar que atacantes não tenham acesso a estes dados armazenados em discos rígidos que estão em uso em um ambiente é a utilização das técnicas lógicas de sanitização de dados. O objetivo das técnicas lógicas é sobrescrever os dados originais que estão no disco rígido, de maneira que a *string*

binária final resultante da operação seja a mais diferente possível da *string* binária original dos dados.

Porém, trabalhos que utilizam a técnica de Microscopia de Força Magnética (*Magnetic Force Microscopy* - MFM) em **modelos de discos rígidos antigos** conseguiram recuperar dados armazenados após o processo de sanitização de dados ter ocorrido (GUTMANN, 1996). O objetivo deste estudo é verificar a aplicabilidade destas técnicas lógicas de sanitização de dados nos discos rígidos analisando a possibilidade de recuperação dos dados após a execução do processo de sanitização dos dados ter ocorrido no mesmo. Se a recuperação dos dados originais não for possível, permitirá que o disco rígido seja reutilizado e não destruído como é feito em muitas empresas após sua remoção do ambiente ao qual ele está inserido, mesmo tendo seu estado de uso aceitável para uma nova utilização (YAN; XUE; XU, 2013).

Tais abordagens podem gerar benefícios financeiros para os consumidores, empresas, entre outros, na medida em que os recursos computacionais são reutilizados. Traz também benefícios do ponto de vista da sustentabilidade, evitando o consumo de matérias-primas para a fabricação de novos discos rígidos (THIBADEAU, 2005) e diminuição de poluentes que seriam lançados na natureza.

1.2 Objetivos

Este trabalho propõe um estudo da aplicabilidade de técnicas de sanitização de dados em discos rígidos atuais. Seu principal objetivo é avaliar se é possível a recuperação dos dados após a aplicação de uma técnica lógica de sanitização de dados em um disco rígido, usando métodos físicos de acesso ao disco tradicionalmente associados à recuperação de dados em discos rígidos, baseados em MFM. Como objetivos secundários deste trabalho são propostos:

- Avaliação das técnicas lógicas de sanitização de dados selecionadas neste trabalho com o objetivo de verificar como as técnicas lógicas são eficazes para sanitizar os dados armazenados;
- Seleção de duas técnicas lógicas para a avaliação experimental;
- Obtenção de imagens que estabeleçam os padrões dos dados gravados nos discos rígidos durante o processo de sanitização de dados; e

- Avaliação experimental que visa verificar a possibilidade de restauração dos dados que foram submetidos às técnicas de sanitização de dados aplicadas nos discos rígidos utilizados nesta pesquisa, verificando assim, a eficiência das técnicas lógicas na sanitização dos dados.

Para atender e realizar as etapas propostas neste trabalho se faz necessário a adoção de um método científico. Este método vai auxiliar na pesquisa, no estudo e no desenvolvimento deste trabalho até a sua conclusão.

1.3 Método adotado para a pesquisa

Para a realização deste trabalho é utilizado o método hipotético-dedutivo (PRODANOV; FREITAS, 2013). Este método será usado para a realização da validação da proposta de conceito, especificando uma hipótese de solução e avaliando tal hipótese. O método utilizado para realizar a pesquisa dos artigos científicos relacionados nesta pesquisa foi o da revisão sistemática (RF, 2007).

Para isto foi definida uma pergunta chave para a busca de artigos: "como realizar a sanitização de dados em discos rígidos?".

Para efetuar o mapeamento das referências da área foram realizadas as seguintes etapas:

Identificação das bases de dados consultadas: na pesquisa realizada foram consultadas as seguintes bases de dados eletrônicas de publicações científicas, IEEE (*Institute of Electrical and Electronic Engineers*), ACM (*Association of Computing Machinery*), Springer, WebScience e SciPress. Também foram consultados padronizações e documentos de boas práticas de órgãos, tais como: NIST (*National Institute of Standards and Technology*), ENISA (*European Network and Information Security Agency*), CPNI (*Center for the Protection of National Infrastructure*), CSA (*Cloud Security Alliance*), DTMF (*Distributed Management Task Force*), CSCC (*Cloud Standard Customer Council*), ITU-T (*International Telecommunication Union*) e ISACA (*Information Systems Audit and Control Association*);

Palavras-chaves para a consulta: na pesquisa as seguintes palavras-chaves foram adotadas: eliminação de dados (*data elimination*), remoção de dados (*data*

removal), destruição de dados (*data destruction*), apagamento de dados (*data erasure*), sanitização de dados (*data sanitization*), estas palavras-chaves também foram combinadas com outras possibilidades como: sanitização de dados em disco rígido (*data sanitization in hard disk*), eliminação de dados em disco rígido (*data elimination in hard disk*), remoção de dados em disco rígido (*data removal in hard disk*), destruição de dados em disco rígido (*data destruction in hard disks*), apagamento de dados em disco rígido (*data erasure in hard disk*). Com o objetivo de abranger o tema mais ricamente foram utilizadas *strings* de busca com combinações como sanitização de dados em computação em nuvem (*data sanitization in cloud computing*), método para eliminação de dados em computação em nuvem entre outras, com o objetivo de verificar em diferentes ambientes como ocorre o processo de sanitização de dados nos discos rígidos utilizados;

Estratégia de busca: para a pesquisa não se limitar aos artigos em português, as palavras-chaves também foram empregadas em inglês. A busca efetuou-se a partir da consulta aos sítios das bases identificadas, e, a partir do buscador do Google (*Google Academic*);

Revisão e seleção dos artigos: durante a seleção dos artigos para o estudo, a avaliação dos títulos e dos resumos (*abstracts*) foram realizadas, quando o resumo não era suficiente para a avaliação do artigo a leitura rápida foi necessária, objetivando não excluir um artigo importante da revisão sistemática. Após a realização desta etapa foram identificados 261 (duzentos e sessenta e um artigos);

Análise crítica dos artigos selecionados: o objetivo dessa etapa foi analisar os artigos identificados que envolvem o contexto de sanitização de dados em geral. Foi realizada a leitura mais detalhada dos artigos para classificar quais artigos iriam compor a quadro de artigo que norteariam a pesquisa realizada neste trabalho. Ao final da pesquisa foram identificados e selecionados quatro (04) artigos que estão diretamente alinhados com o trabalho realizado nesta dissertação.

A Figura 1 apresenta a sequência de passos realizados durante a revisão sistemática para selecionar os artigos mais relevantes para o tema de pesquisa abordado neste trabalho.

Figura 1: Sequência de passos realizados na revisão bibliográfica.

Ao todo foram realizadas 5 etapas que culminaram em quatro (04) artigos considerados relevantes para a pesquisa da aplicabilidade de técnicas de sanitização de dados em discos rígidos. Estes artigos foram analisados mais detalhadamente para verificar quais seriam suas contribuições com o tema pesquisado neste trabalho.

O objetivo foi selecionar trabalhos que também realizam uma avaliação experimental em disco rígido utilizando a técnica de MFM como a proposta neste trabalho, porém, durante o desenvolvimento desta pesquisa foi verificado que poucos trabalhos utilizaram a técnica de MFM na prática para a recuperação de dados, por isso foram selecionados apenas quatro artigos que aplicaram a técnica de MFM em disco rígido.

O objetivo foi selecionar trabalhos que também realizam uma avaliação experimental em disco rígido utilizando a técnica de MFM como a proposta neste trabalho, porém durante o desenvolvimento desta pesquisa foi verificado que poucos trabalhos aplicaram a técnica de MFM na prática para recuperação de dados, por isso foram selecionados apenas 4 artigos que aplicaram a técnica.

1.4 Trabalhos relacionados

O trabalho de (GUTMANN, 1996) apresenta um estudo sobre a aplicação de técnicas de sanitização de dados em discos rígidos. O autor aponta fragilidades inerentes ao disco rígido, como por exemplo, "**transbordamento**" de dados nas bordas das trilhas que permitem que segmentos de dados sejam reconstruídos mesmo após a aplicação de uma técnica lógica de sanitização de dados no disco rígido. Outro fato está na fragilidade da modulação de dados usada para escrever os dados no prato do disco rígido que facilita sua reconstrução por meio da utilização da técnica de MFM.

Entretanto estes argumentos foram elaborados em 1996 quando foi realizado este experimento com a tecnologia utilizada na época, porém, a tecnologia empregada nos discos rígidos evoluiu, corrigindo problemas que eram anteriormente encontrados. A cabeça de leitura e gravação eletromagnética é mais precisa na escrita dos dados, ou seja, não há mais este fator de "transbordamento" de dados nas bordas das trilhas (KANG; LEE, 2010) impedindo que trechos de dados possam ser recuperados por este motivo de "transbordamento" de dados para fora da trilha do disco rígido.

A capacidade de armazenamento dos discos rígidos aumentou consideravelmente desde 1996, com isto o dispositivo teve uma redução do tamanho físico de todos os materiais envolvidos no processo de gravação magnética. A espessura de cada subcamada magnética e de revestimento de carbono protetor é de apenas vários nanômetros e a cabeça de leitura e gravação eletromagnética paira sobre a superfície do disco rígido a alturas inferiores a 50 nanômetros que é em torno de $5,0 \times 10^{-6}$ cm, obrigando que as modulações utilizadas evoluíssem.

A atual modulação EPRML (*Extended Partial Response, Maximum Likelihood*) comprime mais os dados que as antigas modulações utilizadas nos discos rígidos que foram realizados os testes, tornando mais difícil sua releitura por meio da técnica de MFM (SHIROISHI, 2007; BELARMINO et al., 2004), fazendo necessário que novas avaliações experimentais devam ser realizadas, para comprovar se nos discos rígidos é possível a recuperação de dados após o processo de sanitização de dados ser realizado.

O trabalho (BAUER; PRIYANTHA, 2001) criou um arquivo de 100 KB em um disco rígido e em seguida sanitizou o espaço no disco rígido destinado ao

armazenamento dos dados. Após a realização do processo de sanitização os autores entraram em contato com 20 empresas especializadas em recuperar dados "perdidos". Destas 20 empresas, apenas uma empresa se interessou em tentar recuperar estes dados que foram sanitizados. **Porém, os autores não realizaram esta tentativa de recuperação do arquivo**, o que mantém o desafio de verificar a eficácia das técnicas lógicas em sanitizar os dados armazenados em disco rígido em aberto e necessitando que uma avaliação experimental seja realizada para comprovar seu uso.

O trabalho (WRIGHT; KLEIMAN, 2008) apresentou um estudo teórico sobre a utilização da técnica de MFM para recuperar dados em discos rígidos. Os autores apresentaram fatos que se contrapõem a recuperação de dados, como a modulação utilizada atualmente, a velocidade de rotação utilizada nos discos rígidos que aumentaram (atualmente a velocidade de rotação está em torno de 7200 RPM há 15000 RPM) e gravam os dados mais superficialmente no prato do disco rígido. Entretanto, não foi realizada nenhuma avaliação experimental com um microscópio que aplica a técnica de MFM, mantendo o desafio ainda em aberto e necessitando que soluções sejam apresentadas para comprovar a aplicabilidade das técnicas de sanitização de dados em discos rígidos.

O trabalho (KANEKAL, 2013) apresentou a recuperação de dados armazenados em disco rígido utilizando a técnica de MFM e técnicas de processamento digital de imagem. Porém, o disco rígido que foi alvo desta pesquisa não tinha sido sanitizado por uma técnica lógica antes do processo de aquisição das imagens pelo microscópio utilizado para aplicar a técnica de MFM. Mantendo o desafio de recuperação de dados após o processo de sanitização de dados ter ocorrido em aberto.

A Tabela 1 apresenta onde cada trabalho relacionado atuou na questão da proteção dos dados.

Tabela 1: Passos da técnica de sanitização Gutmann.

Trabalhos relacionados	Sanitização	Recuperação de dados	Aval. Experimental (MFM)
(Gutmann, 1996)	Sim	Sim	Sim
(Bauer; Priyantha, 2001)	Sim	Não	Não
(Wright; Kleima, 2008)	Não	Não	Não

(Kanekeal, 2013)	Não	Sim	Não
------------------	-----	-----	-----

Como apresentado na Tabela 1 nenhum dos trabalhos relacionados realizou as três etapas de sanitização por completo, mantendo a desafio de sanitização de dados em discos rígidos em aberto.

1.5 Organização

Este trabalho está organizado da seguinte forma:

No Capítulo 2 é apresentado como os discos rígidos evoluíram com o passar dos anos até chegar aos discos rígidos utilizados recentemente. Também é apresentado como ocorre o processo de gravação e leitura de dados nesta mídia. Apresentam-se possíveis ataques que podem ser realizados para recuperar dados em disco rígido, destacando o ataque por meio da MFM. São também relacionadas as técnicas de sanitização de dados encontradas na literatura para os discos rígidos que serão alvos desta pesquisa e por fim, são apresentadas as recomendações e leis sobre sanitização de dados adotadas por países e órgãos de referências.

No Capítulo 3 são apresentados os detalhes da avaliação lógica, bem como o resultado relacionado à aplicação das técnicas lógicas nos volumes/espço que foram utilizados na pesquisa. São apresentados conceitos sobre MFM e os microscópios MultiMode 8 e o Dimension Icon, que foram utilizados para a aplicação da técnica de MFM. Por fim, são apresentados os resultados da avaliação experimental e os detalhes da análise que comprovam a eficiência das técnicas lógicas em sanitizar dados armazenados em discos rígidos.

O Capítulo 4 apresenta as considerações finais deste trabalho, e a participação em projetos e as publicações relevantes ligadas ao tema proposto deste trabalho.

2 Sanitização de Dados em Discos Rígidos

Neste capítulo são apresentados modelos que antecederam os discos rígidos. O objetivo é contextualizar a evolução que ocorreu neste dispositivo em relação às suas características e propriedades, fazendo com que ele seja utilizado até o presente momento para realizar o armazenamento dos dados dos consumidores (SEAGATE, 2018). É apresentada a mudança ocorrida na forma de como os dados são armazenados ao longo dos anos até o presente momento, bem como possíveis ataques para tentar recuperar dados em disco rígido (com destaque ao ataque que foi utilizado neste trabalho para tentar recuperar dados). São apresentadas as técnicas lógicas de sanitização de dados encontradas na literatura destinada a realizar a sanitização de dados que estão armazenados no disco rígido, e por fim, são apresentados padrões, leis, políticas e atores utilizados para realizar a sanitização de dados em diferentes países.

2.1 Modelos de disco rígido e capacidade de armazenamento

O armazenamento de dados em discos rígidos teve início com o modelo 305 RAMAC Disk System desenvolvido pela IBM em 1956, visando o mercado corporativo, tendo sua capacidade máxima de armazenamento de 5 MB em um disco rígido de 50 pratos com 60 cm de diâmetro cada (CHEN et al., 2002). Com o passar dos anos a IBM continuou investindo em equipamentos para armazenamento de dados, como os modelos IBM Model 1301 *Disk Storage Unit*, lançado em 1961, com a capacidade de 28 MB de armazenamento, o IBM 1311 *Disk Storage Drive*, lançado em 1962, com capacidade de 2,6 MB, porém, com um tamanho inferior quando comparado ao IBM RAMAC Disk, com 6 pratos de 14 polegadas (cerca de 35 cm), e o IBM 3340 *Direct Access Storage Facility* popularmente conhecido como "*Winchester*" devido ao fato de utilizar dois conjuntos de disco com capacidade de 30 MB cada (DANIEL; MEE; CLARK, 1999).

A grande contribuição do modelo IBM 3340 foi o desenvolvimento da cabeça de leitura e gravação eletromagnética leve e discos lubrificados instalados em um

compartimento a vácuo (STEVENS, 1981), técnica utilizada até os dias atuais nos discos rígidos. Em 1980 a IBM rompeu a barreira do gigabyte com o modelo IBM 3380 *Direct Access Storage Device* com a capacidade de 2,52 Gigabytes e velocidade de transferência de dados de 3 MB/s. Também em 1980 a empresa *Shugart Technology*, hoje denominada *Seagate Technology*, lançou o modelo ST-506 com capacidade de armazenamento de 5 MB (MONETT, 1985). A principal contribuição deste modelo foi a placa controladora que era instalada diretamente no computador permitindo executar o sistema operacional, aplicativos e acessar arquivos salvos diretamente no disco rígido tornando-se um componente ligado diretamente ao computador (GROCHOWSKI; HOYT, 1996). O Apêndice A deste trabalho apresenta os componentes mecânicos e lógicos que compõem o disco rígido.

Com o passar dos anos o disco rígido tornou-se cada vez mais viável, tendo seu tamanho físico reduzido, chegando até o tamanho de 3,8 polegadas, 2,5 polegadas por exemplo. Em 1996, a *Seagate Technology* lançou o modelo *Seagate Barracuda* com capacidade de 2,5 Gigabytes com a velocidade de 7200 RPM que se tornou padrão para os discos rígidos modernos (ARPACI-DUSSEAU et al., 1997). A tecnologia aplicada no disco rígido evoluiu com o passar dos anos possibilitando que a capacidade de armazenamento de dados aumentasse consideravelmente (TANENBAUM; ZUCCHI, 2009). Neste sentido a Figura 2 apresenta a evolução da capacidade do disco rígido em armazenar dados com o passar dos anos desde 1980 até 2015.

Figura 2: Capacidade de armazenamento de dados dos discos rígidos com o passar dos anos (*IEEE-SPECTRUM*, 2015).

Como apresentado na Figura 2, até meados de 1995 a capacidade de armazenamento estava em Gigabyte. Entretanto, no começo da década de 2000 até o seu final, a capacidade de armazenamento aumentou consideravelmente, saindo da casa de um dígito, indo até a casa de quatro dígitos, com 1000 gigabytes de capacidade de armazenamento. É possível notar que os avanços na capacidade de armazenamento chegam à casa de Terabytes de armazenamento de dados nos discos rígidos (SEAGATE, 2018a). Estes discos rígidos são utilizados em vários tipos de serviços oferecidos ao consumidor, como serviços de armazenamento de dados como Dropbox, Google Drive, Onedrive (Microsoft), por meio da Internet, permitindo que vários usuários compartilhem o mesmo disco rígido ao mesmo tempo, por meio do paradigma de multi-inquilino (TAKABI; JOSHI; AHN, 2010).

O tamanho físico do disco varia entre 3,5", 2,5" e 1.8" e sua velocidade de rotação entre 5400 RPM há 15000 RPM, variando a interface utilizada para conexão entre IDE/ATA (*Advanced Technology Attachment*), SATA (*Serial Advanced Technology Attachment*) SCSI (*Small Computer System Interface*), *FIBRE CHANNEL*, e SAS (*Serial Attached SCSI*). É possível encontrar no mercado modelos híbridos que permitem a combinação da tecnologia do disco rígido com a tecnologia do SSD¹. Todos os discos rígidos utilizados nesta pesquisa são de 3,5" de 7500 RPM sendo seu padrão de interface SATA.

Permanece assim, a necessidade que possíveis soluções ao desafio de sanitização de dados em discos rígidos que está em aberto sejam apresentadas para tranquilizar os consumidores em relação a segurança e privacidade dos dados armazenados nos discos rígidos utilizados nestes ambientes como dispositivo final de armazenamento de dados.

2.2 Processo de gravação e leitura dos dados

Para realizar a leitura e gravação dos dados nos pratos do disco rígido é utilizada a cabeça de leitura e gravação eletromagnética. Esta cabeça de leitura e gravação eletromagnética contém uma bobina que utiliza impulsos magnéticos para manipular as moléculas da superfície do prato para gravar os dados. Quando está positiva atrai o pólo negativo e quando está negativa, atrai o pólo positivo. De acordo com esta

polaridade são gravados os bits (0's e 1's) no prato. Para o processo de leitura dos dados a cabeça de leitura e gravação eletromagnética "lê" o campo magnético gerado pelas moléculas e gera uma corrente elétrica correspondente. A placa lógica/controladora do disco rígido analisa a variação do sinal ocorrida no prato para determinar o bit lido (0 ou 1) (FARIAS, 2006). Para cada prato há duas cabeças de leitura e gravação eletromagnética, uma superior e outra inferior (CARLOMAGNO; SCHERER, 2000).

Para ocorrer esta gravação ou leitura dos dados foi desenvolvida a **modulação**. A primeira modulação utilizada para a conversão de números digitais (bits) em informações magnéticas em disco rígido foi a Modulação em Frequência Modificada (*Modified Frequency Modulated* - MFM) derivada da Modulação de Frequência (*Frequency Modulation* - FM). Foi utilizada primeiramente em discos flexíveis, sendo utilizada na família de discos rígidos Seagate ST - 225 de 20 MB até o ST-4096 de 80 MB, conhecida também como especificação de interface ST-506 ou ST-412. Neste esquema de modulação, quando há um bit 1, o sinal é invertido, quando há um bit 0 precedido de um bit 1, o sinal não é invertido e quando há um bit 0 precedido de outro sinal 0, o sinal é invertido (PORCHER; GARCIA, 2000).

A modulação MFM foi rapidamente substituída pela Modulação de Comprimento de Percurso Limitado (*Run Length Limited* - RLL). Essa modulação, ao invés de modular um bit por vez, como ocorre no método FM e MFM, modula vários bits por vez. Nesta modulação os padrões dos bits que representam os dados são convertidos em códigos, ao invés de serem armazenados bit a bit. A modulação RLL usa um esquema de codificação no qual as alterações de fluxo magnético se baseiam na ocorrência do sinal magnético permitindo que os dados sejam gravados com menos alterações do fluxo magnético do que seria necessário para o mesmo número de bits de dados (CHALLA; GANNAGE, 1994).

A modulação tem dois parâmetros definidores: conhecidos como RLL (X, Y): sendo que X é o comprimento do percurso, e Y é o limite do percurso, o qual se refere a sequência de espaço no dado modulado sem a inversão do sinal (PATRICK, 2002). Como, por exemplo, a modulação RLL (2, 7), que define o espaço mínimo entre inversões do sinal de 2 espaços. Enquanto o tamanho máximo entre as inversões do sinal é de 7 espaços, fazendo com que os dados ocupem 2/3 do espaço que ocupariam com a modulação MFM (LEE, 2003). A Figura 3 apresenta a comparação

entre as compressões FM, MFM e RLL. É possível notar que a mesma sequência de bits (10001111) ocupa tamanhos diferentes para o armazenamento no prato do disco rígido após a aplicação das diferentes modulações.

Figura 3: Apresentação da compressão de dados realizada nas modulações FM, MFM e RLL (KOZIEROK, 2001).

Como apresentado na Figura 3, com os avanços das modulações foi possível diminuir o espaço físico de gravação dos dados nos discos rígidos, possibilitando que mais dados fossem gravados em um mesmo espaço físico. Com o passar dos anos foi necessário aumentar a densidade (medida da quantidade de dados (bits) que podem ser armazenados em um dispositivo de armazenamento) dos discos rígidos para permitir maior capacidade de armazenamento de dados. Porém, as primeiras cabeças de leitura e gravação eletromagnéticas "varriam" a superfície dos pratos detectando as mudanças de fluxo que ao atingirem um valor limite conhecido como **Threshold**, indicavam a presença da reversão do sinal. Esta técnica recebe o nome do **Peak Detection** justamente por perceber esta variação do sinal (STEYAERT et al., 1995). Porém, o aumento da densidade causou também o aumento do "ruído" que é a tendência do sinal de uma reversão influenciar em outra reversão dificultando a interpretação da reversão de sinal.

Para evitar esta tendência do sinal de uma reversão influenciar em outra reversão, foi desenvolvida a modulação (*Partial Response, Maximum Likelihood - PRML*) que utiliza algoritmo de correção de erros, que visa "varrer" os sinais de entrada conforme recebido (*Partial Response*) e indica a maior possibilidade vislumbrada de correção (*Maximum Likelihood*) (ABBOTT et al., 1994). Os discos rígidos utilizam modulação *Extended Partial Response, Maximum Likelihood - EPRML*, versão melhorada da PRML (JIANG; LO, 2005).

Outro componente lógico que sofreu modificações ao passar dos anos foi o setor, que inicialmente armazenava 512 bytes. Com o aumento da capacidade de armazenamento dos discos rígidos o tamanho do setor passou a ser um elemento limitante (SEAGATE, 2018), sendo necessário o desenvolvimento de novas técnicas para que a capacidade de armazenamento dos discos rígidos possa continuar a crescer. A tabela 2 apresenta a relação dos setores totais baseado na capacidade de armazenamento e nas suas resoluções correspondentes.

Outro componente lógico que sofreu modificações ao passar dos anos foi o setor, que inicialmente armazena 512 bytes. Com o aumento da capacidade de armazenamento dos discos rígidos o tamanho do setor passou a ser um elemento limitante (SEAGATE, 2016), sendo necessário o desenvolvimento de novas técnicas para que a capacidade de armazenamento dos discos rígidos possa continuar a crescer. A Tabela 2 apresenta a relação de setores totais baseado na capacidade de armazenamento e suas resoluções¹ correspondentes.

Tabela 2: Relação da capacidade de armazenamento, dos setores e da resolução dos setores.

Capacidade	Número de setores totais	Resolução do setor
40 MB	80,000	0,001%
400 GB	800,000,000	0,0000001%

Como pode ser visto na Tabela 2, com o desenvolvimento da tecnologia de armazenamento de dados, o prato do disco rígido em que são armazenados os dados aumentou sua capacidade de armazenamento, passando de 40 MB para 400 GB por exemplo. Onde era possível ter 80,000 setores em um disco rígido de 40 MB, um disco de 400 GB tem 800,000,000 setores, ou seja, 10 mil vezes mais setores. Se a resolução de um setor era de 0,001% em um disco rígido de 40 MB, em um disco rígido de 400 GB a resolução do setor passou a ser de 0,0000001%, ou seja, com o aumento da capacidade de armazenamento no prato dos discos rígidos o tamanho físico dos setores vem diminuindo.

Com o aumento da densidade nos discos rígidos o setor consome um espaço físico menor, causando um aumento na taxa de erros e tornando a correção dos erros mais difícil (STENFORT, 2003). Um novo padrão de setor com maior capacidade de

¹ Resolução = Tamanho que um setor ocupa no disco rígido de acordo com a capacidade que ele possui em armazenar dados.

armazenamento de dados foi desenvolvido para solucionar esta questão; este padrão foi denominado de Formato Avançado (*Advanced Format* - AF) (IDEMA, 2013).

O AF, também conhecido como setores de 4K (por poder armazenar 4096 bytes), foi aceito em janeiro de 2011 pela comunidade (SEAGATE, 2013; SEAGATE, 2016). A primeira linha de discos rígidos a utilizar este padrão foi o da Seagate, o Barracuda Green 2 TB. A Figura 4 apresenta como o padrão 4K utiliza os setores de 512 bytes.

Figura 4: Utilização dos setores com o novo padrão 4k (ACER, 2016).

Diversos sistemas de arquivos utilizados pelos sistemas computacionais modernos continuam a assumir que os setores são compostos por 512 bytes. Devido a este fato ocorre uma emulação de setor de 512 bytes, até que a indústria programe o padrão de 4K nativamente nos discos rígidos. O processo de emulação para a gravação inicializa-se a partir da requisição de escrita dos dados no disco rígido. O disco "lê" a entrada em formato 4K do disco rígido e carrega-se para a DRAM (*Dynamic random access memory*), onde o bloco é dividido em sub-setores de 512 bytes. A seguir o setor emulado de 512 bytes modificado é alterado na DRAM e todo o setor de 4K é reescrito no disco. Este processo é conhecido como Leitura - Modificação - Gravação e acrescenta tarefas mecânicas a mais, afetando o desempenho do disco rígido.

Para esta comunicação (gravação e/ou leitura) entre o sistema de arquivo e o disco rígido é utilizado o Endereçamento de Bloco lógico (*Logical Block Address* - LBA). Cada setor de 512 bytes recebe um endereço a partir do zero. Por meio da solicitação do computador para gravação dos dados é devolvido um endereço de LBA no final da gravação e é informado ao computador em quais trilhas e setores os dados estão gravados.

Por padrão o endereço LBA 0 é alinhado com o primeiro setor de 512 bytes no setor físico de 4K. A este padrão de alinhamento é dado o nome de Alinhamento 0. Este padrão de alinhamento opera muito bem no AF, pois permite facilmente mapear oito setores contínuos em um único setor de 4K. Os atuais sistemas operacionais da Microsoft a partir do MS - Windows Vista já aceitam este padrão (MICROSOFT, 2016), o GNU/Linux a partir da versão do núcleo 2.6.031 já aceita também este padrão (LINUX, 2016).

2.3 Mudanças na Forma de Armazenamento de dados dos consumidores

A maneira como os dados são armazenados mudou drasticamente com o passar dos anos (WANG et al., 2011). Antes o consumidor que armazenava os dados detinha geralmente o dispositivo em que os dados estavam armazenados (MAGGIANI, 2009). Com o advento de novos ambientes computacionais é possível que o consumidor tenha seus dados fora de seus domínios, tanto físico, como lógico (HAMLEN et al., 2012), passando todo o gerenciamento dos seus dados para ser realizado no ambiente administrado pelo provedor do serviço oferecido a ele, por exemplo, em um serviço de armazenamento de dados como o oferecido pela empresa Dropbox (MARTURANA; ME; TACCONI, 2012), e não mais no ambiente gerenciado pelo consumidor.

O provedor deste novo modelo de serviço define as políticas de segurança vigentes e os procedimentos que devem ser realizados em cada ciclo de vida dos dados mantidos em seu ambiente (ENISA, 2016; CHAWKI, 2016; MELL; GRANCE et al., 2011). Com estes novos cenários e modelos de serviços, o consumidor, proprietário dos dados, não está mais diretamente envolvido na operação de sanitização de dados, porque a sanitização dos dados ocorre agora no ambiente do provedor do serviço e não no ambiente gerenciado pelo consumidor.

Nestes novos ambientes o consumidor não tem mais acesso ao disco rígido, tão pouco à infraestrutura utilizada (LOMBARDI; PIETRO, 2011), o que impossibilita que ele realize seus procedimentos para sanitizar os dados armazenados nos discos rígidos. Tais procedimentos ficam a cargo do provedor do serviço (BISHOP et al.,

2006), o que pode causar uma insegurança ao consumidor, devido ao fato de depender de agentes externos ao seu ambiente para realizar a sanitização de seus dados.

Nestes novos tipos de serviços de armazenamento de dados a sanitização dos dados é um desafio que está em aberto, e necessita que soluções sejam apresentadas (FERNANDES et al., 2014) o que evidencia ainda mais a importância desta pesquisa da aplicabilidade de técnicas lógicas de sanitização de dados em discos rígidos, para garantir que os dados dos consumidores não serão recuperados após a realização do procedimento de sanitização de dados ocorrido no disco rígido, garantindo assim mais segurança a eles.

2.4 Técnicas de sanitização de dados

Independente da técnica de sanitização de dados utilizada para remover os dados do disco rígido, seu principal objetivo é impossibilitar que qualquer indivíduo tenha acesso aos dados removidos novamente (DONG; KUN; YU, 2009). O disco rígido que foi submetido a esta operação não deve mais armazenar nenhum vestígio dos dados originais (DHAMDHERE; PATIL, 1993), garantindo que os dados armazenados na mídia não estejam mais acessíveis e que qualquer tentativa para recuperar estes dados não terá êxito em sua execução. O objetivo desta seção é relacionar as técnicas de sanitização pesquisadas na literatura destinada a realizar a sanitização de dados que serão alvo do estudo neste trabalho.

2.4.1 Técnicas Físicas

As técnicas físicas de sanitização visam a alteração das características físicas do disco rígido, inviabilizando-o para novos usos. Porém, como já mencionado o objetivo deste trabalho é provar que as técnicas lógicas para sanitização de dados são eficientes, permitindo que o mesmo seja utilizado novamente até o final da sua

vida útil, por isso as técnicas físicas serão apresentadas brevemente, apenas para conhecimento.

2.4.1.1 Desmagnetização

Esta técnica age sobre a forma magnética do disco rígido, alterando seu funcionamento e impedindo a recuperação dos dados no disco (DATA, 2008). A Figura 5 apresenta um equipamento desmagnetizador conhecido como Próton T-1.5, utilizado para realizar a desmagnetização em disco rígido.

Figura 5: Degausser (DATA, 2008).

2.4.1.2 Destruição do disco rígido

Existem equipamentos no mercado que visam a destruição física do disco rígido. A Figura 6 apresenta o modelo Próton PPD-1, que pressiona o disco rígido, dobrando-o em um ângulo de aproximadamente 90°. Após o disco magnético é descolado para fora do equipamento. A amassadura inutiliza o disco rígido entortando o prato e danificando toda a sua estrutura física (PRODUCTS, 2016) sendo que sua destruição é realizada manualmente pelo usuário. A Figura 6 apresenta o equipamento 0300 *Low/Medium Volume Hard Drive Shredder*, que permite a possibilidade de triturar todo o disco rígido.

Figura 6: Próton PPD-1 (ESALINK, 2018).

Figura 7: 0300 Low/medium Volume Hard Drive Shredder (ESALINK, 2016).

A Figura 8 A, apresenta como é "dobrado" o disco rígido dentro do equipamento Próton PPD - 1 e o seu estado após o término da operação que dura em torno de 20

segundos. A Figura 8 B apresenta o antes e depois de um disco rígido que foi submetido ao equipamento 0300 *Low/medium Volume Hard Drive Shredder*.

Figura 8: Utilização dos equipamentos Próton PPD - 1 e 0300 *Low/medium Volume Hard Drive Shredder* (DATASHEETS, 2018).

Como apresentado na Figura 8 em ambos os casos o disco rígido torna-se inutilizável após o final da operação realizada, como já mencionado anteriormente estas técnicas físicas foram mencionadas para conhecimento, não sendo alvo desta pesquisa.

2.4.2 Técnicas lógicas

Diferente das técnicas físicas, as técnicas lógicas permitem o reuso do disco rígido (GUTMANN, 1996). Estas técnicas devem garantir que os dados não possam ser recuperados dos discos rígidos (ABRAMOVITCH; FRANKLIN, 2002). Estas técnicas de sanitização de dados serão as técnicas estudadas para o estudo proposto neste trabalho de recuperação de dados em discos rígidos por meio da técnica de MFM.

2.4.2.1 Gutmann

Técnica desenvolvida pelo pesquisador Peter Gutmann (GUTMANN, 1996), do Departamento de Ciência da Computação da Universidade de Auckland/Nova Zelândia para remoção de dados. Esta técnica especifica 35 padrões para sobrescrever os dados armazenados em cada setor do disco rígido; estes passos são apresentados na Tabela 3.

Tabela 3: Passos da técnica de sanitização Gutmann.

Passo	Notação binária do padrão	Notação hexadecimal	Modulação RLL (1,7)	Modulação RLL (2,7)	Modulação MFM
1	Aleatório	Aleatório			
2	Aleatório	Aleatório			
3	Aleatório	Aleatório			
4	Aleatório	Aleatório			
5	01010101 01010101 01010101	55 55 55	RLL (1,7)		MFM
6	10101010 10101010 10101010	AA AA AA	RLL (1,7)		MFM
7	10010010 01001001 00100100	92 49 24		RLL (2,7)	MFM
8	01001001 00100100 10010010	49 24 92		RLL (2,7)	MFM
9	00100100 10010010 01001001	24 92 49		RLL (2,7)	MFM
10	00000000 00000000 00000000	00 00 00	RLL (1,7)	RLL (2,7)	
11	00010001 00010001 00010001	11 11 11	RLL (1,7)		
12	00100010 00100010 00100010	22 22 22	RLL (1,7)		
13	00110011 00110011 00110011	33 33 33	RLL (1,7)	RLL (2,7)	
14	01000100 01000100 01000100	44 44 44	RLL (1,7)		
15	01010101 01010101 01010101	55 55 55	RLL (1,7)		MFM
16	01100110 01100110 01100110	66 66 66	RLL (1,7)	RLL (2,7)	
17	01110111 01110111 01110111	77 77 77	RLL (1,7)		
18	10001000 10001000 10001000	88 88 88	RLL (1,7)		
19	10011001 10011001 10011001	99 99 99	RLL (1,7)	RLL (2,7)	

20	10101010 10101010 10101010	AA AA AA	RLL (1,7)		MFM
21	10111011 10111011 10111011	BB BB BB	RLL (1,7)		
22	11001100 11001100 11001100	CC CC CC	RLL (1,7)	RLL (2,7)	
23	11011101 11011101 11011101	DD DD DD	RLL (1,7)		
24	11101110 11101110 11101110	EE EE EE	RLL (1,7)		
25	11111111 11111111 11111111	FF FF FF	RLL (1,7)	RLL (2,7)	
26	10010010 01001001 00100100	92 49 24		RLL (2,7)	MFM
27	01001001 00100100 10010010	49 24 92		RLL (2,7)	MFM
28	00100100 10010010 01001001	24 92 49		RLL (2,7)	MFM
29	01101101 10110110 11011011	6D B6 DB		RLL (2,7)	
30	10110110 11011011 01101101	B6 DB 6D		RLL (2,7)	
31	11011011 01101101 10110110	DB 6D B6		RLL (2,7)	
32	Aleatório	Aleatório			
33	Aleatório	Aleatório			
34	Aleatório	Aleatório			
35	Aleatório	Aleatório			

Os quatro primeiros passos sobrescrevem os dados aleatoriamente com 0's e 1's. Do quinto até o trigésimo primeiro passo, cada bloco recebe dados com padrões específicos de 0s e 1s de acordo com o método de modulação usado. Como apresentado na Tabela 2, do passo trinta e dois ao passo trinta e cinco que encerra o método, os dados são novamente sobrescritos aleatoriamente com 0s e 1s. Os padrões determinísticos entre as gravações aleatórias são permutados antes da escrita a ser realizada, com o objetivo de dificultar a recuperação dos dados.

2.4.2.2 VSITR

Criado pelo Departamento Alemão de Segurança da Informação - *Bundesamt für Sicherheit in der Informationstechnik* - (BSI), a técnica VSITR (*Verschlusssachen - IT - Richtlinien*) propõe que, para apagar os dados do disco rígido de forma segura é

necessário sobrescrever cada setor do disco rígido de forma segura sete vezes (SBS, 2018).

- 1º passo – sobrescreve os dados com 0s;
- 2º passo – sobrescreve os dados com 1s;
- 3º passo – sobrescreve os dados com 0s;
- 4º passo – sobrescreve os dados com 1s;
- 5º passo – sobrescreve os dados com 0s;
- 6º passo – sobrescreve os dados com 1s; e
- 7º passo – sobrescreve os dados com 0s;

Esta técnica sobrescrever o padrão 0 e 1 no disco rígido, sendo que o padrão 0 é sobrescrito quatro vezes e o padrão 1 é sobrescrito 3 vezes. Nesta técnica não está presente o padrão de sobrescrita aleatório.

2.4.2.3 RCMP TSSIT OPS-II

A técnica RCMP TSSIT OPS-II é apresentada no documento publicado pela *Royal Canadian Mounted Police - RCMP (POLICE, 2003)*. Esta técnica emprega sete passos para sobrescrever os dados nos setores do disco rígido. O RCMP TSSIT OPS-II não é mais utilizado pelo governo canadense, tendo sido substituído pela técnica CSEC ITSG-06.

- 1º passo – sobrescreve os dados com 0s;
- 2º passo – sobrescreve os dados com 1s;
- 3º passo – sobrescreve os dados com 0s;
- 4º passo – sobrescreve os dados com 1s;
- 5º passo – sobrescreve os dados com 0s;
- 6º passo – sobrescrever os dados com 1s; e
- 7º passo – sobrescreve os dados aleatoriamente com 0s e 1s e verifica a operação;

A verificação da operação de sobrescrever no sétimo passo ocorre como se fosse um passo a mais, pois consiste em checar qual é o valor escrito em cada bloco

que foi escrito no sétimo passo. Para isso são guardados os valores das sequências escritas em cada bloco em uma *string* binária. Se o valor que estiver sendo escrito não corresponder com o valor escrito no bloco, é cancelada a operação e é realizado o sétimo passo novamente até que os valores sejam os mesmos.

2.4.2.4 CSEC ITSG-06

A técnica CSEC ITSG - 06 é apresentada no *IT security Guidance 06: Clearing and Declassifying Electronic Data Storage Device*, publicado pelo *Communication Security Establishment Canada* - (CESC). Esta técnica emprega três passos para sobrescrever os dados nos setores do disco rígido;

- 1º passo – sobrescreve os dados aleatoriamente com 0s e 1s;
- 2º passo – sobrescreve os dados com o complemento do valor escrito no passo anterior (1 substitui por 0 e vice-versa); e
- 3º passo – sobrescreve os dados aleatoriamente com 0s e 1s e verifica a operação;

Essa técnica também é utilizada pelo *Navy Staff Office Publication 5239 Module 26: Information Systems Security Program Guidelines* publicado pela marinha Norte Americana. Também aplicada em discos rígidos antigos com modulações MFM e RLL. A verificação no terceiro passo ocorre de forma análoga a do RCMP TSSIT OPS- II, com a comparação dos dados aleatórios escritos e a posterior leitura destes dados.

2.4.2.5 DoD 5220.22-M

Técnica desenvolvida pelo Departamento de Defesa dos Estados Unidos (*Department of Defense of the United States of America*) - DoD. Esta técnica sobrescreve cada setor do disco rígido três vezes (DOD, 1991). Entretanto, esta técnica nunca foi aprovada como padrão de técnica lógica de sanitização de dados a ser utilizada pelo Departamento de Defesa dos Estados Unidos (DATA, 2018). O

departamento a utilizou como referência em seus documentos sobre sanitização de dados.

Porém, esta técnica tem sido utilizada por outras empresas com a *Amazon* em seus serviços de armazenamento de dados (AMAZON, 2015), no departamento da força aérea Americana (*US Air Force*) - USAF por meio da *Force System Security Instruction 5020* - AFSSI - 5020. No *National Computer Security Center* - NCSC, um grupo que está ligado à Agência Espacial Americana (*National Aeronautics and Space Administration*) - NASA. A técnica também foi apresentada no *Forest Green Book* como NCSC - TG - 025, no *Communications Electronic Security Group* - CESG que está ligado ao *UK Government Communications Headquarters* - CCH com o nome de HMG IS5 (CENTER, 2018) para sanitizar dados.

- 1º passo - sobrescreve os dados com 0s e verifica a operação;
- 2º passo - sobrescreve os dados com 1s e verifica a operação; e
- 3º passo – sobrescreve os dados aleatoriamente com 0s e 1s e verifica a operação;

Semelhante à técnica RCMP TSSIT OPS-II que verificar a execução da sobrescrita dos dados no 7º passo a técnica DoD 5220.22-M verifica a sobrescrita dos dados no 3º passo.

2.4.2.6 AR 380-19

Desenvolvida pelo exército dos Estados Unidos - *US Army* - e publicada na *Army Regulation 380 - 19* (DHILLON, 2007), emprega três passos para sobrescrever os dados nos setores do disco rígido:

- 1º passo - sobrescreve os dados aleatoriamente com 0s e 1s;
- 2º passo – sobrescreve os dados aleatoriamente com 0s e 1s; e
- 3º passo – sobrescreve os dados com o complemento do valor sobrescrito no passo anterior, (1 substitui por 0 e vice-versa);

Semelhante à técnica RCMP TSSIT OPS-II que verificar a execução da sobrescrita dos dados no 7º passo e a técnica DoD 5220.22-M que verifica a

sobrescrita dos dados no 3º passo a técnica AR 380-19 também verifica a sobrescrita dos dados no 3º passo.

2.4.2.7 Técnica GOST R 50739-95

Técnica utilizada pelo governo russo, definida no documento GOST R 50739-95 (RUSSO, 2016). Utiliza três passos:

- 1º passo - sobrescreve os dados com 0s;
- 2º passo – sobrescreve os dados aleatoriamente com 0s e 1s; e
- 3º passo – sobrescreve os dados aleatoriamente com 0s e 1s

Em nenhum dos seus passos é realizado a verificação para verificar se o processo de sobrescrita ocorreu com êxito quando aplicado no disco rígido.

2.4.2.8 ISM 6.9.92

Técnica aplicada pelo governo australiano, apresentada no a e definida no *Australian Department of Defense: Intelligence & Security* Esta técnica usualmente é implementada em 1 passo. Para discos rígidos com capacidade de 15 GB ou menos este passo é executado três vezes.

- 1º passo - sobrescreve os dados aleatoriamente com 0s e 1s; e verifica a operação;

É também utilizada pelo governo da Nova Zelândia, tendo sido publicada no *New Zealand Information and Communications Technology Security Manual* e também referenciado no NZSIT 402 do *Government Communications Security Bureau*.

2.4.2.9 Considerações sobre Técnicas Lógicas

Durante a pesquisa foram encontradas mais técnicas para sanitizar dados, porém, como elas eram semelhantes a alguma técnica já mencionada nesta pesquisa, elas foram citadas juntamente com a técnica já apresentada na pesquisa. No decorrer da pesquisa foi observado que mesmo com o desenvolvimento de uma técnica por órgão de referência, ela não passou a ser adotada por outros órgãos do mesmo país como um padrão para sanitizar seus dados, como no caso da técnica DOD 5220.22-M do departamento de defesa americano e a técnica AR 38-19 desenvolvida pelo exército dos Estados Unidos, demonstrando que há muitas divergências na escolha sobre qual técnica lógica é mais eficaz para sanitizar dados em disco rígido.

É possível observar que as técnicas são semelhantes, diferenciando-se em alguns dos passos por utilizar um padrão diferenciados dos demais ou mesmo uma quantidade de passos diferenciada. Como exemplo pode-se verificar que as técnicas CSEC ITSG-06, DOD 5220.22- M, AR 38-19 e GOST R 50739-95 tem o mesmo número de passos, porém os passos que as compõem são diferentes, o que ocasiona no processo final da sanitização de dados um resultado diferente (ROEBUCK, 2012).

A Tabela 4 apresenta detalhes ligados às técnicas de sanitização de dados relacionadas nesta pesquisa, sendo destacado o passo que é considerado o diferencial da técnica, pois possibilita que o resultado do processo de sanitização de dados seja diferente das demais técnicas, o país de origem e o número total de passos que compõem a técnica.

Tabela 4: Detalhes das técnicas de sanitização de dados.

Técnica	País	Número de passos	Passo considerado diferencial
Gutmann	Nova Zelândia	35	Do passo 5 ao passo 31 onde são sobrescritos padrões diferentes sobre os dados originais.
VSITR	Alemanha	7	Não tem um passo principal, ocorre uma variação entre sobrescrever 0s e 1s nos passos.
RCMP TSSIT OPS-II	Canadá	7	7º passo, onde os dados são sobrescritos com o valor do complemento armazenado na operação anterior.
CSEC ITSG-06	Canadá	3	2º passo, onde os dados são sobrescritos com o valor do complemento armazenado na operação anterior.

DOD 5220.22-M	Estados Unidos	3	3º passo, sobrescrever os dados com valores aleatórios.
AR 38-19	Estados Unidos	3	3º passo, onde os dados são sobrescritos com o valor do complemento armazenado na operação anterior.
GOST R 50739-95	Rússia	3	2º e 3º passos sobrescrevem os dados com valores aleatórios.
ISM 6.9.92	Austrália	1	A técnica tem apenas um passo

Como apresentado na Tabela 3 ao total foram relacionadas oito técnicas lógicas que foram desenvolvidas tanto por pesquisadores, como por órgãos, com o objetivo de realizar a sanitização de dados nos discos rígidos (IVANKA MENKER, 2013).

2.5 Ataques para recuperar dados em disco rígido

Um ataque é uma exploração de uma falha de um sistema informático, cujos motivos podem ser os mais variados possíveis, como obter acesso ao sistema, roubar informações, como senhas, dados bancários, segredos industriais, entre outros (PASQUALETT, ET. AL., 2013).

Do ponto de vista da recuperação de dados em discos rígidos, uma classe de ataque ocorre por meio de códigos maliciosos, executados sobre o disco rígido ligado em um computador, seja acessando o conteúdo diretamente ou, para caso dos dados removi50 dos, por meio do acesso à tabela de alocação de arquivos e dos blocos de dados do disco rígido.

Isso ocorre porque em uma simples remoção dos dados eles não são removidos definitivamente do dispositivo, eles são removidos apenas da tabela de alocação. Com isso, os dados antigos permanecem armazenados no disco rígido, até que a área reservada aos dados antigos seja regravada com novos dados, e durante este intervalo de tempo em que não ocorre a regravação de novos dados armazenados neste local (DORION, 2010; GONZALEZ et al., 2012).

Há casos em que o disco rígido não pode mais ser ligado a um computador para esta recuperação, como por exemplo, para discos com problemas ou término de sua vida útil. Caso a mídia magnética esteja intacta, um ataque que pode ser realizado em posse do disco rígido é por meio da utilização da técnica de Microscopia de Força

Magnética (MFM) para recuperar dados (RUGAR et al., 1990). A técnica permite analisar a estabilidade de padrões de bits em mídias de alta densidade, como as utilizadas nos discos rígidos (KOBLSCHKA; WEI; HARTMANN, 2007) por meio de um microscópio de varredura. A Figura 9 apresenta uma imagem resultante da aplicação da técnica de MFM em disco rígido onde foram gravados dados.

Figura 9: Imagem resultante da aplicação da técnica MFM em disco rígido (BELARMINO et al., 2004).

2.6 Recomendações e leis para sanitização de dados

Atualmente, há vários manuais para realizar a sanitização de dados disponíveis aos usuários (POLICE, 2003; DEFENSE, U S. Department, 2004; GILL, et. al., 2006; JEANNETTE, 2016). Nos Estados Unidos o laboratório do Departamento de Defesa americano avalia todos os processos e ferramentas relacionados com a sanitização de dados em disco rígido por meio da agência federal e as aprovações são dadas somente após a devida diligência, devido à tamanha importância dada à questão da segurança dos dados governamentais.

No Reino Unido, desde 1998 é exigido que as organizações estabelecessem procedimentos para o tratamento dos dados, para evitar a divulgação accidental ou não autorizada dos dados e a destruição do disco rígido, sujeitas a Autorizada Técnica Nacional para Garantia da Informação do Reino Unido (*Government's National Technical Authority for Information Assurance*) - GNTAIA, e estão sobre orientações

da saúde pública, de portabilidade e de responsabilidade (*Health Insurance Portability and Accountability Act*) - HIPAA, proteção de informação pessoal e documentos eletrônicos (*Personal Information Protection and Electronic Documents Act*) - PIPEDA, e pela proteção financeira (Lei *Gramm - Leach - Bliley*) - GLBA, que visam proteger os usuários, os dados e o processo de sanitização de dados utilizados (HUGHES; COUGHLINCOMMINS, 2009). A Lei de proteção de privacidade on-line infantil (*Children's Online Privacy Protection Act*) - COPPA que regula as questões relativas a coleta e tratamento de dados de crianças menores de 13 anos, garantindo assim privacidade, confiabilidade e a integridade dos dados que estão armazenados nos discos rígidos utilizados pelos usuários.

Na comunidade Europeia existe a Diretiva 95/46 buscando regular as relações das empresas nas operações que visem o tratamento dos dados pessoais, bem como busca uniformizar a questão do tratamento de dados pessoais entre os países membros da União Europeia, sendo que a questão da proteção das pessoas singulares em relação ao tratamento de dados pessoais é considerada um direito fundamental, segundo o artigo 8.0 n° 1, da Carta dos Direitos Fundamentais da União Europeia. A diretiva busca tutelar as relações onde existam as empresas de um lado ou grandes grupos frente ao indivíduo, sendo que esses dados pessoais podem ser tratados por meios automatizados ou não automatizados, bastando apenas que seja realizado sob esse prisma, porém quando as relações de tratamento de dados pessoais forem tratadas entre particulares, exclusivamente, ou estejam envolvidos os interesses nacionais, como a segurança do Estado, por exemplo, ficam excluídas da proteção da referida norma.

Os países europeus têm fortes leis para a proteção e sanitização de dados. No Reino Unido e nos Países Baixos os discos rígidos comercializados têm um comando de sanitizar dados "*in - drive*", isto é, embutido na configuração dos discos rígidos. Sendo que este método não requer nenhum software adicional para realizar a sanitização dos dados armazenados no disco rígido, e atende aos requisitos legais da HIPAA, PIPEDA e GLBA, tendo sido aprovado pelos Estados Unidos (RAMAN; PRAMOD, 2013).

No Brasil não existe leis específicas que trate a respeito da proteção de dados pessoais dos usuários, nem uma maneira efetiva no seu tratamento, existindo brechas jurídicas nesse assunto. Existem, atualmente, vários projetos de lei que tramitam no Congresso Nacional, que tratam especificamente de uma política nacional para a

proteção dos dados pessoais, como os Projeto de Lei no 5276/16, apensado ao projeto 4060/12 e os outros Projetos de Lei que tramitam no Senado Federal, como o 131/14 (dispõe sobre o fornecimento de dados de cidadãos ou empresas brasileiros a organismos estrangeiros), 181/14 (estabelece princípios, garantias, direitos e obrigações referentes à proteção de dados pessoais) e o 330/2014 (D sobre a proteção, o tratamento e o uso dos dados pessoais, e dá outras providências).

Entretanto, a maioria dos discos rígidos comercializados ao redor do mundo não tem este mecanismo de sanitização automático como padrão de fábrica habilitado para o usuário final, impossibilitando que o usuário realize o processo de sanitização de dados diretamente no disco rígido, o que mantém o desafio da aplicabilidade de técnicas lógicas de sanitização de dados em disco rígido em aberto.

2.7 Considerações finais do capítulo

Neste capítulo foi apresentado como os discos rígidos evoluíram ao longo do tempo, desde o primeiro protótipo lançado pela IBM, até os discos rígidos utilizados comercialmente para armazenamento de dados. Foi apresentado o processo de leitura e gravação de dados nos discos rígidos, e o novo padrão de leitura e gravação utilizando setores de 4K. Foram apresentadas as oito técnicas de sanitização de dados pesquisadas na literatura para aplicação em disco rígido.

Foi apresentada a mudança ocorrida na maneira de armazenamento de dados dos consumidores, antes o consumidor detinha o controle do ambiente, agora com novas possibilidades de serviços de armazenamento de dados oferecidas o consumidor delega ao provedor dos serviços a função de sanitização de seus dados do ambiente que ele está ligado.

Foram apresentados os ataques que podem ser utilizados para recuperar dados em disco rígido, sendo destacado o ataque que utiliza a técnica de MFM para recuperar dados, que será o ataque abordado neste trabalho. Por fim, foram apresentadas as recomendações de órgãos de referência da área de segurança de dados, relacionando alguns exemplos de recomendações sugeridas por eles, juntamente com leis específicas para sanitização de dados que foram elaborados por

países com o objetivo de apresentar a importância da sanitização de dados em discos rígidos no atual cenário da computação.

3 Procedimento Experimental para Avaliação de Sanitização de Discos Rígidos

Neste capítulo são apresentados detalhes ligados às duas avaliações que foram realizadas para verificar a aplicabilidade de técnicas lógicas de sanitização de dados em discos rígidos atuais. O objetivo é contextualizar como ocorreu a avaliação lógica nas técnicas de sanitização de dados selecionadas e como foram criadas as amostras utilizadas na avaliação experimental, além de mostrar detalhes de como foi realizada a avaliação experimental por meio dos microscópios selecionados que aplicam a técnica de MFM.

3.1 Avaliação Lógica das Técnicas de Sanitização de dados

O principal objetivo da avaliação lógica é classificar entre as oito técnicas lógicas de sanitização de dados apresentadas no capítulo anterior, quais são as técnicas que sanitizam os dados com mais eficiência, ou seja, a técnica que realizar a sanitização dos dados permitindo que a relação de proximidade com os dados originais seja a mais diferente possível. Diferente neste contexto significa que a *string* binária A que compõem os dados originais seja a mais diferente possível da *string* binária A' resultante da operação de sanitização dos dados.

Para realizar a avaliação lógica foi criado inicialmente um **volume de 512 bytes** contendo uma ***string* de mesmo tamanho**, que ocupava todo o volume; o tamanho foi selecionado para simular o tamanho de um setor físico utilizado nos discos rígidos atuais para armazenamento de dados. A Figura 10 apresenta a ***string* de caracteres de 512 bytes** e a sequência de bits correspondente à *string* de caracteres que armazenada em cada disco rígido utilizado na pesquisa.

[illegible]

Figura 10: *String* e sequência de bits gravados nos volumes lógicos.

Tal volume foi replicado 80 vezes, de maneira a se aplicar cada uma das 8 diferentes técnicas de sanitização para cada 10 volumes replicados. O volume original foi mantido sem alteração (por exemplo, sem a aplicação de uma técnica de sanitização), para permitir a comparação da sequência binária apresentada na Figura 10 com as sequências binárias correspondentes de cada técnica lógica gerada no processo de sanitização de dados dos volumes replicados. A Figura 11 apresenta os passos realizados para a avaliação lógica nos volumes criados.

Figura 11: Passos realizados para a avaliação lógica nos volumes criados.

Cada técnica lógica foi aplicada dez vezes em volumes diferentes, ao final de cada rodada o valor binário da *string* resultante do processo foi armazenado para

posterior comparação. No seguinte endereço eletrônico estão as imagens relacionadas à aplicação de cada técnica lógica nos volumes lógicos alvo da avaliação <http://www.larc.usp.br/users/fsbampato/msc.html>.

Para realizar o processo de sanitização de dados nos volumes criados foi utilizado o comando “*dd*”, nativo em sistemas operacionais Linux, que permite escrever diretamente nos blocos dos volumes, sem o uso de um sistema de arquivos. Para cada técnica lógica foi criado um *script* correspondente na linguagem de programação Bash com o objetivo de realizar o procedimento o mais automatizado possível. Tanto as sequências utilizadas no *dd* como os scripts usados para sanitizar os volumes que foram utilizados nesta pesquisa estão disponíveis no endereço eletrônico <http://www.larc.usp.br/users/fsbampato/msc.html>.

Outros *softwares* proprietários prometem realizar a sanitização de dados e foram pesquisados neste trabalho. Não foram utilizados por diferentes motivos, seja porque não eram nativos das distribuições Linux (como o *dd*) ou porque não permitiam que alterações fossem realizadas em seu código-fonte e ou ainda porque não permitiam a aplicação das oito técnicas lógicas de sanitização de dados selecionadas nesta pesquisa.

A comparação foi realizada utilizando dois arquivos de texto, onde o primeiro armazenava a sequência de bits do volume separado com a *string* original de dados e o segundo arquivo continha a sequência de dados sanitizado por uma técnica lógica. A comparação foi realizada verificando posição a posição de cada valor armazenado nos arquivos, e contabilizando o número de posições iguais e diferentes do processo de sanitização de dados.

Para realizar a avaliação lógica foi utilizada a métrica da distância de Hamming (HAMMING, 1950). A distância de Hamming permite verificar o número de posições que uma *string* binária A difere de uma *string* binária A', sendo ambas as *strings* binárias do mesmo tamanho em relação ao número de elementos que as compõem. Para fornecer o número de posições diferentes nas *strings* binárias ela compara posição a posição das duas *strings* binárias e, para cada comparação, o algoritmo soma o valor 1 à distância se e somente se ambas as posições foram iguais, fazendo este procedimento até o último *bit* que corresponde à sequência da *string* binária.

Durante a pesquisa para verificar qual seria a métrica mais adequada para realizar avaliação experimental foram relacionadas mais métricas como a distância de Levenshtein (HASHISUME, K. et. al.; 2004), a distância de Jaro Winkler (WINKLER,

1990), a distância de Monge Elkan (MONGE, 1996), coeficiente de Dice's (DICE, 1945) e similaridade de Jaccard (COHEN, et, al., 2003), utilizadas para recuperação de dados. Porém estas métricas são normalmente utilizadas para *strings* não binárias (*string* de caracteres). Como o alvo deste trabalho é avaliar *strings* binárias, essas métricas não foram selecionadas para a avaliação lógica.

Para realizar a comparação entre a *string* binária original e a *string* binária sanitizada foi criado um programa na linguagem de programação Python que aplica a distância de Hamming. Sendo que seu código está disponível no seguinte endereço eletrônico <http://www.larc.usp.br/users/fsbampato/msc.html>. Este programa recebe as duas *strings* binárias, a *string* binária com dados originais e a *string* binária com dados sanitizados e retorna o valor da distância de Hamming correspondente entre as *strings* binárias.

3.2 Avaliação de Sanitização com o uso de Microscopia de Força Magnética

A partir da avaliação lógica das técnicas de sanitização, duas das oito técnicas lógicas foram selecionadas para a avaliação de sua aplicabilidade por meio da técnica de Microscopia de Força Magnética. Na próxima seção são descritos os detalhes das técnicas e os microscópios utilizados.

3.2.1 Microscopia de força magnética

A técnica de MFM foi desenvolvida em 1980, e faz parte da família de microscópios do tipo *Scanning Probe Microscopy* - SPM, sendo que sonda/ponta varre a superfície da amostra, registrando ponto a ponto algum tipo de interação que ocorra entre a ponta/sonda e amostra. Dependendo do tipo de interação, a microscopia recebe um nome específico: por exemplo, quando a interação consiste em uma passagem de corrente de tunelamento é denominado de microscópio de Tunelamento.

Gerd Binnig e Heinrich Rohrer ganharam o prêmio Nobel em 1986 por inventarem o microscópio de Tunelamento, que foi a primeira modalidade de SPM a ser desenvolvida. Atualmente existe uma infinidade de modos de operação para o SPM, sendo alguns exemplos, a Microscopia de Força Elétrica - (*Electrical Force Microscopy*) - EFM, a Microscopia de Scanner Térmico - (*Scanner Thermal Microscopy*) - SThM, a Microscopia de Força Atômica de Condutividade e a Microscopia de Força Magnética empregada nesta pesquisa.

As varreduras nas amostras podem ocorrer de três modos descritos a seguir, sendo que estes modos podem ser aplicados de acordo com o tipo de técnica selecionada para realizar a varredura:

- **AFM² de contato:** neste modo de contato a sonda consiste em uma ponta presa a um cantiléver (peça utilizada para prender a sonda/ponta que realiza a varredura na amostra), que se mantém em contato com a superfície da amostra durante toda a varredura da amostra. Durante a varredura da amostra a ponta passa por diferentes alturas (devido a amostra não ter uma superfície totalmente plana), gerando variações na deflexão do cantiléver que são emitidas até um fotodiodo de quatro quadrantes utilizado para posicionar a ponta. Por meio de um sistema de feedback de um computador, a deflexão é mantida constante durante toda a varredura, movendo a amostra verticalmente, a cada ponto.
- **AFM de contato intermitente:** neste modo de contato a sonda consiste em uma ponta presa a um cantiléver oscilante que durante a varredura toca a superfície da amostra no ponto de máxima amplitude de oscilação. Sendo que o cantiléver neste modo de operação oscila em sua frequência de ressonância em uma amplitude típica entre 20nm a 10nm. No período que ocorre a varredura, a ponta oscilante passa por diferentes alturas da amostra gerando variação na amplitude de oscilação da ponta e; por meio de um sistema de feedback de um computador a amplitude de oscilação da ponta é mantida constante.
- **AFM de não contato:** neste modo de contato a sonda consiste em uma ponta presa a um cantiléver oscilante que durante a varredura não toca a

² AFM- *Atomic Force Microscopy* = técnica que utiliza uma ponta muito fina como sonda para varrer a superfície da amostra, esta ponta fica na extremidade livre de um cantiléver que irá vibrar sobre a superfície da amostra ou simplesmente tocar a superfície.

amostra. A ponta é mantida a distâncias de unidades ou dezenas de nanômetros da superfície da amostra, sendo utilizado a força de Van Der Waals atrativa. A força total de interação entre a ponta e a amostra neste caso de operação é extremamente baixa, na ordem de 10^{-12} quando comparada às forças agentes no modo de contato (10^{-6} a 10^{-7}), sendo indicado para a análise de amostras delicadas, **como no caso de disco rígido, ou adesivas.**

No caso da técnica de MFM o modo indicado é o modo de AFM de não contato, que foi usado neste trabalho. A sonda/ponta utilizada é recoberta com um filme fino de um material ferromagnético, com imantação permanente. A sonda/ponta oscila próximo à superfície da amostra, mas sem tocá-la, detectando: a) a mudança na frequência de ressonância do cantiléver, ou b) na diferença de fase entre a força atuante no cantiléver e o movimento oscilatório do cantiléver, ou c) a mudança da amplitude de oscilação do cantiléver denominada *Root Mean Square* - RMS. Esta alteração induzida na oscilação do cantiléver é gerada pela interação do campo magnético da amostra com o campo magnético da sonda/ponta. A Figura 12 apresenta a interação magnética entre a ponta e a amostra ocorrida durante uma varredura de uma amostra.

Figura 12: Interação magnética entre ponta e amostra (Bruker, MultiMode 8, 2016)

Na técnica de MFM os três diferentes métodos podem ser usados para se criar as imagens que apresentam o campo magnético das amostras. No primeiro método a frequência de ressonância do cantiléver é mantida fixa enquanto a amplitude de

oscilação é medida. Isto é feito por meio do sensor de deflexão ótico do AFM. A variação na amplitude de oscilação é então detectada a cada ponto e armazenada como a coordenada Z, juntamente com a posição (X, Y) da varredura, formando a imagem do campo magnético da superfície da amostra.

O segundo método é com a frequência de ressonância do cantiléver mantida fixa, enquanto a diferença de fase, entre a força do transdutor e a oscilação do cantiléver é mantida, sendo também feito pelo sensor de deflexão ótico do AFM. É registrada, então a diferença de fase a cada ponto armazenado como coordenada Z, juntamente com a posição (X, Y) da varredura, formando a imagem do campo magnético da superfície da amostra.

O terceiro método é o de detecção por frequência que consiste em fixar a fase em 90°, o que corresponde à ressonância do sistema acoplado. É medida a variação de frequência com relação à ressonância do sistema não acoplado, por meio do sensor de deflexão ótico do AFM. A cada ponto é medida a variação da frequência e é armazenada junto com a posição (X, Y) da varredura, formando a imagem do campo magnético da superfície da amostra.

Resumidamente, a formação da imagem pode ser feita das seguintes formas: quando a ponta oscilante varre a superfície da amostra, passa por pontos de diferentes valores do componente vertical do gradiente de força magnética entre ponta e amostra, gerando assim variações na frequência de oscilação. Estas variações são medidas por meio da amplitude, fase ou frequência. Juntamente com a posição (X, Y) correspondente armazenada no computador formam a imagem de campo magnético ao longo da superfície da amostra. São fornecidas imagens a respeito de estrutura de domínio magnético que ocorram naturalmente ou que tenham sido gravadas na superfície da amostra.

Para a criação das imagens foi utilizado um recurso denominado de *interleave*, que consiste na obtenção de duas modalidades de imagens simultaneamente em uma mesma área de varredura. No primeiro modo sem o uso do recurso *interleave*, o microscópio varre a ida e a volta na direção de varredura rápida, sendo obtido o mesmo tipo de informação para ida e volta. É realizada uma varredura lenta na direção perpendicular e são efetuados passos cujos avanços correspondem ao tamanho de pixel da imagem gerada.

Quando o recurso *interleave* é acionado, inicialmente o sistema realiza uma varredura de ida e volta, como padrão, mas avançando apenas metade do tamanho

do pixel na direção da varredura lenta. É realizada em seguida uma varredura de ida e volta, obtendo-se novas informações da varredura de ida e volta. Devido a esta interação é possível obter duas imagens em modos diferentes, na mesma região da amostra. Por padrão durante a varredura é possível ter uma imagem e **MFM**, sendo mantida uma distância fixa da superfície da amostra em relação a ponta utilizada na varredura, para o modo de AFM de não contato.

3.2.2 Modelos de Microscópios utilizados para a avaliação Experimental

Foram utilizados dois modelos de microscópios que aplicam a técnica de MFM para realizar a avaliação experimental. A seguir são apresentados os componentes e as principais diferenças entre os dois modelos de microscópios utilizados.

3.2.2.1 Microscópio MultiMode 8

O microscópio MultiMode 8 da empresa Bruker (BRUKER, 2018) é composto por um conjunto de componentes que fazem parte de sua estrutura física.

Seus principais são componentes:

- **Computador:** principal controlador do sistema integrado ao MultiMode 8, nele são incluídos uma placa *frame grabber*, e uma interface de Processamento de Sinal Digital (*Digital Signal Processing* - DSP). Por meio de um software, o usuário envia os comandos para a controladora do Microscópio via uma porta de entrada/saída (I/O) utilizados para a comunicação de dados entre o computador e o microscópio.
- **Controladora:** gerencia o microscópio e o escaneamento realizado por ele por meio de um micro cabo de 68 pinos – SCSI conectado entre o computador e a controladora. A conexão no microscópio é feita por um cabo de 37 pinos conectado na frente da controladora.

- **Microscópio:** pode ser configurado para utilizar diversos scanners (sistema ou estágio de varredura que têm a propriedade de realizar deslocamentos com alta precisão), dependendo do tipo de técnica que deseja utilizar para a aquisição de imagens. A Figura 13 apresenta os componentes do microscópio MultiMode 8

Figura 13: Componentes do microscópio MultiMode 8 (Bruker, MultiMode 8, 2016).

- **Scanner:** é um sistema ou estágio de varredura que tem a propriedade de realizar deslocamentos com alta precisão. Conforme o microscópio, a varredura pode ser realizada com a sonda (*probe*) sobre a amostra ou com a amostra sobre a sonda. Há diferentes tipos de scanners, scanners grandes, tipo J, que permitem grandes *Scan Sizes* (medida de “janela” utilizada para criar as imagens) e scanners pequenos como o A, limitados, que geram pequenas imagens, e fornecem melhores resoluções de imagem. Os *scanners* pequenos tendem a serem mais estáveis em relação a ruído, scanners maiores oferecem varreduras mais amplas, necessitando de um sistema de amortecimento adicional contra ruído e precauções em tamanhos de varredura menores de alta resolução. A

Figura 14 apresenta dois tipos de modelos de *scanners* diferentes utilizados no MultiMode 8 para aquisição de imagens.

Figura 14: Tipos de *scanners* (Bruker, MultiMode 8, 2016).

O *scanner* é um equipamento que pode ser trocado, ou seja, para cada tipo de técnica de microscopia que é aplicada pelo microscópio pode ser utilizado um scanner diferente. O modelo de *scanner* utilizado para a microscopia de força magnética é o *scanner* J. A Figura 15 apresenta o *scanner* J utilizado na avaliação experimental pelo MultiMode 8.

Figura 15: Scanners J.

Este equipamento é responsável por realizar as varreduras nas amostras e criar as imagens. Quanto menor for a amplitude de varredura, maior será a ampliação apresentada na imagem. O maior *Scan Size* (Tamanho da “janela” que será apresentada na imagem obtida pela varredura) permitido por este scanner é de 120 μm .

- **Suporte da sonda:** também chamado de porta pontas ou cantiléver, é responsável por prender a ponta/sonda que realiza a varredura. De acordo com a técnica de microscopia utilizada é selecionado um cantiléver diferente. A Figura 16 apresenta alguns tipos de cantiléver que podem ser utilizados no microscópio MultiMode 8. Os demais componentes serão mencionados ao longo do capítulo onde será apresentada uma aplicação direta do mesmo.

Figura 16: Modelos de cantiléver utilizados no MultiMode 8 (Bruker, *MultiMode 8*, 2016).

Para cada cantiléver é utilizado um material diferente em sua fabricação, uma vez que cada técnica de microscopia apresenta características diferentes (por exemplo, o cantiléver que é utilizado na MFM é imantado permanentemente para que seja possível verificar a presença do(s) campo(s) magnético(s) na(s) amostra(s) utilizada(s) para aquisição de imagens).

Uma questão importante com o MultiMode 8 é o tamanho da amostra. A amostra a ser utilizada no microscópio deve ser de 15 milímetros de diâmetro, uma vez que seu porta amostra limita o tamanho físico da amostra. A Figura 17 apresenta a localização do porta amostra junto ao microscópio MultiMode 8.

Local onde é colocada a amostra

Figura 17: Porta amostra do microscópio MultiMode 8 (Bruker, MultiMode 8, 2016).

A amostra deve ser preparada em um formato geométrico de um quadrado para que seja colocada no centro do porta amostra. A amostra é então presa ao porta amostra por meio de uma fita dupla face comum e este é preso ao porta amostra do microscópio por meio de uma corrente de vácuo criada entre o suporte e o porta amostra, não interferindo nos demais componentes e na leitura da amostra pelo microscópio para criar as imagens.

3.2.2.2 Dimension Icon

Comparado ao MultiMode 8 o Dimension Icon é um microscópio mais robusto e sofisticado. A Figura 18 apresenta o microscópio Dimension Icon e seus componentes.

Figura 18: Microscópio Dimension Icon e seus componentes.

Entre as principais diferenças entre os dois microscópios, está o porta amostra do Dimension Icon, que permite colocar todo o prato do disco rígido em seu interior, evitando que o mesmo tenha de ser cortado, como no caso do porta amostra do MultiMode 8. Outro fato importante é que seu porta amostra não está preso ao microscópio diretamente como no caso do MultiMode 8, mas está instalado fisicamente em uma caixa que serve de suporte e proteção. A Figura 19 apresenta o porta amostra utilizado pelo microscópio Dimension Icon onde é possível verificar o tamanho físico do porta amostra, comprovando a possibilidade de colocação do prato do disco rígido em sua superfície.

Figura 19: Porta amostra utilizado no Dimension Icon para colocação das amostras.

O porta amostra têm um tamanho total de 6". Para fixar a amostra no centro do porta amostra é utilizado uma corrente de ar que prende a amostra ao porta amostra durante a varredura.

O Dimension Icon não permite a troca de scanner para a aplicação das diferentes técnicas de varredura, possuindo um único scanner com capacidade de 90 μm de *Scan Size*. Por meio da esfera central do controle é possível deslocar o porta amostra na direção desejada. Os botões auxiliares têm as funções de ajustar tanto o foco, como o zoom apresentado na tela do computador ao usuário para orientar a escolha da área de varredura do microscópio desejada pelo controlador. A Figura 20 apresenta uma visão geral dos componentes do porta amostra do Dimension Icon.

Figura 20: Visão geral dos componentes do porta amostra do Dimension Icon (Bruker, Dimension Icon, 2017).

O **LCD display** apresenta os valores dos ganhos relacionados aos eixos verticais e horizontais. O **óptico** é um sistema que ajuda na localização do cantiléver em relação à amostra. O sistema foca automaticamente na maioria das amostras ajustando a altura do *scanner* no local desejado. E essa é uma vantagem em comparação ao microscópio MultiMode 8, que faz este ajuste manualmente. A **chave on/off** para prender a amostra cria um "canal" de vácuo no centro do porta amostra. O **sistema de estágios** é dividido em dois sistemas, o primeiro é o (X, Y) e o segundo é o Z, proporcionando uma repetitividade de posicionamentos de 2 μm , e permitindo o posicionamento micrométrico de amostras sob a ponta com uma direção de 154 mm (direção Z) por 185 mm (direção Y). O estágio Z fornece imagens de baixo ruído e alta precisão. Um parafuso de chumbo motorizado aciona o estágio Z para a aproximação do *scanner* com a amostra para que ocorra a varredura e a criação das imagens.

3.2.3 Procedimento Experimental para Análise por MFM

O teste experimental consiste em verificar se é possível a recuperação dos dados nos discos rígidos atuais, que foram submetidos às técnicas de sanitização de dados, por meio da técnica de MFM.

Os microscópios modelos Dimension Icon e MultiMode 8 foram utilizados em forma de parceria com o Instituto de Física da Universidade de São Paulo - IFUSP, existentes no Laboratório de Novos Materiais Semicondutores (LNMS), coordenado

pelo professor Alain André Quivy, e no Laboratório de Filmes Finos (LFF), coordenado pela professora Maria Cecilia Salvadori. O conhecimento necessário para a utilização dos dois equipamentos foi adquirido, por meio da disciplina de Microscopia de Força Atômica e Tunelamento (PGF5205) que o candidato cursou durante o período da realização desta pesquisa, a disciplina foi ministrada no IFUSP do campus da capital. O conteúdo do programa da disciplina visa apresentar tanto a parte teórica de operação dos microscópios e de técnicas empregadas por eles, quanto a parte prática da sua utilização.

Para realizar a avaliação experimental foram utilizados sete discos rígidos, sendo quatro discos rígidos novos e três usados, formatados da seguinte maneira:

- **1º Disco rígido novo:** foram gravados 0's em todo o disco rígido. O objetivo foi verificar a imagem do padrão de gravação do *bit* 0 na superfície do disco;
- **2º Disco rígido novo:** foram gravados 1's em todo o disco rígido. O objetivo foi verificar a imagem do padrão de gravação do *bit* 1 na superfície do disco;
- **3º Disco rígido novo:** foi gravada a mesma *string* de 512 bytes usada na avaliação lógica apresentada na Figura 10 em todo o disco rígido, com o objetivo de se ter uma imagem de referência dos dados que foram gravados no disco rígido bem como o padrão de gravação destes dados;
- **4º Disco rígido novo:** foi gravada a *string* de 512 bytes em todo o disco rígido. Após isso, foi aplicada a técnica lógica de Gutmann selecionada, para fins de avaliação lógica
- **Discos Rígidos Usados (5º, 6º e 7º discos rígidos):** foram realizados os mesmos passos do quarto disco rígido, porém, em discos em diferentes estágios de sua vida útil. Nos casos dos 5º e 6º discos rígidos utilizam-se da mesma técnica de sanitização selecionada para o 4º disco rígido, que é a técnica de Gutmann. No 7º disco rígido foi utilizada a técnica RCMP TSSIT OPS - II de sanitização de dados para verificar a exatidão do passo aleatório utilizado no último passo pela técnica de Gutmann para sobrescrever os dados no disco rígido.

A Figura 21 apresenta os passos que foram realizados nos discos rígidos utilizados para a aquisição das imagens.

Figura 21: Preparação dos discos rígidos para análise por MFM.

O objetivo dos procedimentos executados nos primeiros discos rígidos (novos) foi verificar os padrões de gravação dos dados sem que haja “resquícios” de gravação de dados antigos nos quatro discos rígidos, e nos três últimos é verificar a eficiência das técnicas lógicas selecionadas em discos rígidos que estavam no ambiente de produção, “sofrendo” constantes mudanças nas regiões de gravação disponível no prato do disco rígido. Cada disco rígido tem um tempo diferente de vida útil no ambiente que estão inseridos, porém, ambos estão inseridos no mesmo ambiente computacional, o primeiro disco rígido usado tem dois anos de uso, enquanto o segundo tem três anos e meio de uso e o terceiro têm dois anos e meio de uso.

Os quatro primeiros discos rígidos novos são da marca Seagate de 500 GB de armazenamento do padrão SATA II com Firmware C414 do lote ST3500414CS. O quinto, sexto e sétimo discos rígidos são da marca Maxtor, com 500 GB de armazenamento do padrão SATA II com Firmware MX15 do lote 9GT156 – 325.

Os discos foram preparados com os diferentes padrões de dados, sendo os discos 4-7 submetidos a um processo de sanitização. Para a sanitização foram utilizados dois computadores com configurações diferentes, de maneira a verificar o tempo de duração do processo de sanitização de dados em discos rígidos atuais: um Desktop Intel Core 2 Duo E6400 (com processador de 2.13 Ghz e 2.5 GB de memória

RAM DDR2) e um Servidor Intel Xeon E3-1240v2 (com processador de 3.40 Ghz e 16 Gb de memória Ram DDR3).

3.2.4 Tempo e Energia Elétrica despendido para a Sanitização dos Dados nos Discos Rígidos

No primeiro computador (Desktop Intel Core 2 Duo E6400) foram necessárias 28 horas para sobrescrever cada padrão (0,1 e aleatório) nos três primeiros discos rígidos. No quarto disco rígido foram necessários 32 dias e 12 horas para sobrescrever o disco rígido, devido ao fato que a técnica de Gutmann para sobrescrever os dados originais tem 35 passos de sobrescrita. Este mesmo valor de tempo também foi gasto nos quinto e sexto discos rígidos. No sétimo disco rígido onde foi aplicada a técnica de RCMP TSSIT OPS - II foram necessários 8 dias devido ao fato da técnica possuir 7 para passos sobrescrever os dados originais armazenados no disco rígido. Ao total foram gastos 101 dias para preparar todos os discos rígidos.

A Figura 22 apresenta o tempo gasto para sobrescrever o passo 0 no primeiro disco rígido por meio do comando “dd” utilizado para sobrescrever os dados originais.

```
fsbampato@fsbampato-desktop:~$ sudo su
[sudo] password for fsbampato:
root@fsbampato-desktop:/home/fsbampato# dd if=/dev/zero of=/dev/sda1 bs=512 count=976773168 iflag=nocache oflag=direct
dd: error writing '/dev/sda1': No space left on device
976766977+0 records in
976766976+0 records out
500104691712 bytes (500 GB) copied, 89887,8 s, 5,6 MB/s
root@fsbampato-desktop:/home/fsbampato#
```

Figura 22: Tempo gasto em segundo para sobrescrever o passo 0 em todos o disco rígido (Próprio autor).

Sendo considerado que um disco rígido tem uma potência em torno de 8 W e que foram necessários 101 (2424 horas) dias para preparar os discos rígidos, foram gastos 58 KWh neste período.

No segundo computador (Servidor Intel (R) Xeon (R)) utilizado, foram necessárias 16 horas para sobrescrever cada padrão (0.1 e a *string* de 512 bytes) nos três primeiros discos rígidos. No quarto disco rígido foram necessários 18 dias e 7 horas para sobrescrever os dados devido à técnica de Gutmann ter 35 passos de sobrescrita. Este tempo também foi gasto nos quinto e sexto discos rígidos usados.

No sétimo disco rígido onde foi aplicada a técnica RCMP TSSIT OPS - II foram necessários 8 dias devido ao fato de a técnica ter 7 passos para sobrescrever os dados originais. Ao total foram gastos 56 dias e 21 horas para preparar todos os dados originais. Ao total foram gastos 11 KWh neste período.

Houve uma diminuição de 45 dias quando comparado ao primeiro computador utilizado, sendo que ambos os computadores estavam dedicados a realizar somente esta tarefa de sanitização de dados nos sete discos rígidos. Não tinha nenhuma tarefa adicional rodando em segundo plano no computador.

Como comprovado é necessário que sejam verificadas as configurações dos computadores utilizados no processo de sanitização de dados nos discos rígidos. Computadores mais robustos, do ponto de vista de configurações de hardware oferecem mais desempenho computacional para realizar este processo de sanitização de dados em vários discos rígidos. Devem contabilizar previamente o número de horas que podem ser gastas para sanitizar os discos rígidos e a quantidade de energia elétrica gasta no processo de sanitização, para evitar eventuais problemas com relação ao tempo e energia elétrica gastas no processo.

3.2.5 Preparação das amostras e Recursos utilizados

Após a etapa de sanitização ter ocorrido foram preparadas as amostras para o uso nos dois microscópios (Dimension Icon e MultiMode 8). As amostras utilizadas no Dimension Icon foram as mesmas do MultiMode 8: como o corte realizado no disco rígido para criar as amostras para o MultiMode 8 têm 15 milímetros de comprimento, os restantes dos pratos dos discos rígidos cortados foram utilizados como amostras para o Dimension Icon. A Figura 23 apresenta uma imagem de um dos pratos dos discos rígidos cortados que foram utilizados como amostra para o Dimension Icon e o MultiMode8.

Figura 23: Amostra utilizada no MultitMode 8 e no Dimension Icon.

Para garantir que, independente da região ou lado do prato que fosse utilizado no porta amostra do microscópio Dimesion Icon sempre estariam armazenados dados conhecidos na região da varredura, os discos foram preenchidos totalmente com dados desejados e totalmente sanitizados quando aplicável. Dada a necessidade de abrir os discos e cortar os pratos para a criação das amostras para o Multimode 8, foi necessária também uma etapa de aprendizado do procedimento, esta etapa é descrita no Apêndice B.

Por fim, após a preparação das amostras, elas foram examinadas nos dois microscópios para criar as imagens para verificar a aplicabilidade de técnicas lógicas em sanitizar os dados armazenados em discos rígidos atuais.

Para a análise das imagens geradas pelos microscópios foram utilizados três ambientes com monitores diferentes (Figura 24):

- Um computador do tipo desktop com um monitor de 22" e resolução 1920x1080;
- Um computador do tipo desktop com um monitor de 47" e resolução 1920x1080; e
- Um ambiente de visualização de alta resolução composto por um servidor Intel com um painel de visualização composto por 4 monitores de 47" polegadas, gerando uma tela de aproximadamente 94" e resolução 3840x2160.

A) Monitor LG de 22 polegadas

B) Monitor AOC de 47 polegadas

C) Painel com 4 monitores de 47 polegadas

Figura 24: Monitores utilizados nas análises das imagens de MFM (Próprio autor).

O objetivo principal de utilizar vários monitores com tamanhos diferentes foi o de observar mais detalhes das imagens geradas pelas varreduras realizadas nas amostras, o que demandou tanto o aumento da imagem (por isso o uso das telas gradativamente maiores) como o da resolução, culminando em um ambiente de visualização de alta resolução. O recurso de zoom nas imagens também foi utilizado, mas, com a variação do tamanho do monitor foi possível observar mais detalhes do que com monitores menores, devido ao fato das imagens "maiores", favorecerem a análise manual das imagens de MFM resultantes das varreduras das amostras que foi realizada por meio da análise visual. O recurso de zoom nas imagens também foi utilizado, mas, com a variação do tamanho do monitor foi possível observar mais detalhes do que com monitores menores, devido ao fato das imagens ficarem "maiores" nos monitores com mais polegadas, favorecendo a análise das imagens de MFM resultante das varreduras das amostras.

3.2.6 Resultados das Avaliações Lógica e Experimental

Nesta seção, são apresentados os resultados relacionadas as avaliações lógica e experimental nos microscópios Dimension Icon e MultiMode 8, para verificar a possibilidade de recuperação de dados após a aplicação de uma técnica lógica sobre os dados originais armazenados no disco rígido.

3.3 Resultados das avaliações realizadas

Nesta seção são apresentados os resultados relacionados às avaliações lógicas e experimental realizada nos Dimension Icon e no MultiMode 8 para verificar a possibilidade de recuperação de dados após a aplicação de uma técnica lógica.

3.3.1 Resultados da avaliação lógica

Devido ao fato das técnicas lógicas serem muito semelhantes entre si, ou seja, aplicarem os mesmos padrões (0, 1 e aleatório) para sobrescrever os dados que estão armazenados nos discos, foi mudada somente a quantidade de passos que são aplicados pelas técnicas para sobrescrever os dados. A única técnica que muda este padrão de sobrescrita (0, 1 e aleatório) é a técnica de Gutmann que sobrescreve os dados com padrões específicos. Como o objetivo desta avaliação lógica é avaliar a *string* binária final resultante do processo de sanitização, o tempo gasto para criar cada imagem de MFM é de aproximadamente meia hora cada, o acesso aos microscópios (Dimension Icon e MultiMode 8) foi restrito ao período de realização da disciplina a técnica de MFM, foram selecionadas somente duas técnicas lógicas de sanitização. Ao total foram concedidas duas aulas práticas para a aquisição das imagens totalizando 8 horas, sendo que neste tempo foram realizadas varreduras para

aquisição das imagens nas amostras com os dados 0's, 1's, aleatório, dados originais, dados sanitizados pela técnica de Gutmann e pela técnica RCMP TSSIT OPS-II.

Porém, como já mencionado anteriormente, como as técnicas lógicas aplicam os mesmos passos entre si, mudando o número de execução desses passos e a quantidade de passos utilizados por cada técnica (0, 1 e aleatório), sendo que o objetivo é verificar a *string* binária resultante do processo de sanitização de dados, o resultado final desta pesquisa não foi prejudicado.

A Tabela 5 apresenta o resultado da avaliação por meio da métrica da distância de Hamming para cada um dos 10 volumes sanitizados. Nesta tabela pode-se constatar que a técnica lógica VSITR não sofreu alteração no valor de nenhuma das 10 rodadas, uma vez que em seu último passo ela sobrescreve os dados originais com o padrão 0 em todo o volume, gerando o mesmo resultado final para todas as execuções.

Para verificar qual foi a técnica lógica que teve o melhor desempenho para sanitizar os dados originais foram calculados a média aritmética, o desvio padrão, e o intervalo de confiança para cada técnica selecionada. O objetivo é apresentar como as técnicas lógicas sanitizaram os dados ao final das 10 rodadas realizadas. A Tabela 6 apresenta a média aritmética, o desvio padrão e o intervalo de confiança para cada técnica lógica utilizada na avaliação lógica.

Tabela 5: Distância de Hamming para os volumes sanitizados.

Técnica	Rodada	Resultado	Técnica	Rodada	Resultado
Gutmann	1°	2159	VSITR	1°	1958
	2°	2122		2°	1958
	3°	2105		3°	1958
	4°	2122		4°	1958
	5°	2300		5°	1958
	6°	2340		6°	1958
	7°	2290		7°	1958
	8°	2400		8°	1958
	9°	2222		9°	1958
	10°	2335		10°	1958
RCMP TSSIT OPS-II	1°	2115	CSEG ITSG-06	1°	2104
	2°	2194		2°	2100
	3°	2119		3°	2139
	4°	2153		4°	2050
	5°	2009		5°	2121
	6°	2206		6°	2002
	7°	2006		7°	2042
	8°	2456		8°	2130

	9°	2134		9°	2563
	10°	2231		10°	2056
DoD 5220.22-M	1°	2118	AR 38-19	1°	2091
	2°	2082		2°	2110
	3°	2042		3°	2669
	4°	2116		4°	2169
	5°	2099		5°	2110
	6°	2015		6°	2110
	7°	2006		7°	2115
	8°	2156		8°	2312
	9°	2119		9°	2144
	10°	2116		10°	2156
GOST R 50739-95	1°	2041	ISM 6.9.92	1°	2125
	2°	2676		2°	2212
	3°	2584		3°	2112
	4°	2159		4°	2101
	5°	2000		5°	2100
	6°	2100		6°	2110
	7°	2225		7°	2123
	8°	2153		8°	2253
	9°	2130		9°	2100
	10°	2220		10°	2111

Tabela 6: Resultado da Análise das Técnicas Lógicas.

	Técnica de sanitização	Média	Desvio padrão	Intervalo de confiança 95%
1°	Gutmann	2239,5	107,332	57,5438452
2°	GOST R50739 95	2228,8	223,814	119,9937084
3°	AR 380 19	2198,6	176,989	94,88917483
4°	RCMP TSSIT OPS-II	2162,3	127,725	68,47767576
5°	CSEC ITSG 06	2131,2	158,193	84,81246698
6°	ISM 6 9 92	2124,7	53,1581	28,49977615
7°	DOD 5220 M	2086,9	49,8519	26,72718597
8°	VSITR	1958	0	0

De acordo com a Tabela 6 a técnica lógica de sanitização Gutmann é a primeira colocada, com a maior média em relação às demais técnicas lógicas, ou seja, a técnica que mais necessita de operações (inserção, deleção ou substituição) para que a *string* binária resultante da sanitização de dados volte a ser semelhante à *string* original, por isso a técnica de Gutmann foi a 1° selecionada para a avaliação

experimental. A técnica RCMP TSSIT OPS-II foi selecionada para ser aplicada no 7º disco rígido devido ao fato de aplicar no último passo o mesmo passo da técnica de Gutmann que é o passo aleatório. O objetivo foi verificar se o resultado de uma técnica diferente que aplica o último passo semelhante à técnica de Gutmann apresentaria imagens diferentes das capturadas utilizando a técnica de Gutmann anteriormente.

3.3.2 Leitura das Amostras por meio da Técnica de MFM

As amostras preparadas foram submetidas aos microscópios, sendo geradas imagens como resultado das varreduras, que foram analisadas por meio de um software especializado, o NanoScope Analysis 64 bits da própria empresa Bruker (BRUKER, 2018e) que é utilizado pelos dois microscópios.

Na varredura realizada por um microscópio que aplica a técnica de MFM são geradas 5 imagens.

- A primeira imagem apresenta o estado da superfície da amostra.
- A segunda imagem da taxa de erro usada para criar a imagem da amplitude, que mede a oscilação entre a ponta e a amostra.
- A terceira imagem da taxa de erro da fase, usada para criar a imagem da fase que mede a diferença de altura entre os pontos presentes na amostra.
- A quarta imagem da amplitude, imagem obtida na direção Z do campo tridimensional, onde é possível verificar a presença dos campos magnéticos contidos na amostra utilizada na varredura.
- A quinta imagem da fase, que é obtida por meio dos pontos X, Y, onde também se verifica a presença dos campos magnéticos da amostra utilizada na varredura

A Figura 24 apresenta exemplos destas cinco imagens.

Figura 24: Exemplo de imagens geradas com MFM de uma amostra de um disco rígido.

Nesta pesquisa, o enfoque foi dado às imagens de amplitude e fase (D e E na Figura 24), devido ao fato delas apresentarem a variação do campo magnético contido nas amostras dos discos rígidos, que representam as gravações dos dados no disco rígido. Para realizar a análise com mais precisão foi necessário estimar o tamanho do *bit* gravado no prato do disco rígido. Para realizar este cálculo foi estimado inicialmente o tamanho da área do prato do disco rígido, igual a 38,4336 cm². Como os discos rígidos utilizados na avaliação experimental têm somente um prato para gravar 500 GB (250 GB ou 2.147.483.648.000 *bits* para cada lado do prato), o tamanho do *bit* analisado no caso corresponde a aproximadamente 1,79 μm².

Esta medida é necessária para configurar o tamanho da imagem gerada na varredura da amostra, denominado *Scan Size*, bem como para quantificar a quantidade de bits presentes nas imagens analisadas. A *string* de 512 bytes (4096 bits) tem uma área total de 7331,84 μm², sendo necessário um *Scan Size* de no mínimo 86 μm². Entretanto, imagens com esta medida de *Scan Size* para este tipo de análise não são recomendados, pois este valor é considerado alto, não favorecendo a análise: quanto maior o *Scan Size* menos detalhes são possíveis de serem visualizados, ou seja, é mais difícil a visualização dos bits gravados na amostra. Nesta pesquisa foram realizadas varreduras com *Scan Size* menores, como 20, 10, e 5 μm,

onde é possível verificar trechos da *string* de 512 bytes gravada no terceiro disco rígido.

O tempo gasto para criar cada imagem foi aproximadamente meia hora cada. Ao total foram criadas 29 imagens por meio das amostras dos discos rígidos utilizados, ou seja, o tempo total para gerar as imagens foi de aproximadamente **14 horas**. A aquisição das imagens não foi realizada em uma única sessão e foram necessárias **4 sessões ao total**, sendo que a cada sessão eram confirmadas as configurações ajustadas para a aquisição das imagens pelo microscópio Dimension Icon.

As imagens criadas correspondem as amostras com os dados 0, com os dados 1, com os dados originais, com os dados sanitizados pela técnica de Gutmann e com os dados sanitizados pela técnica RCMP TSSIT OPS - II. Para criar as imagens de cada técnica lógica selecionada seriam necessárias mais 16 horas, porém a disciplina PGF5205 durante a qual o candidato realizou as varreduras das imagens já tinha se encerrado e só seria reaberta no próximo semestre. Optou-se por dar continuidade ao trabalho, empregando-se as imagens coletadas. As imagens relacionadas a cada varredura das amostras estão disponíveis no seguinte endereço eletrônico <http://www.larc.usp.br/users/fsbampato/msc.html>.

3.3.2.1 Análise das Imagens geradas pelo Microscópio MultiMode 8

Na amostra com os dados 0's armazenados em todo o prato do disco rígido foi configurado o *Scan Size* de 15 μm , porém, ao final do processo de varredura com o microscópio MultiMode 8 a ponta utilizada para aplicar a técnica de MFM tocou na amostra, prejudicando a varredura. Este toque deve-se à necessidade de se cortar a amostra (para cabê-la no porta amostra limitado em tamanho), que resultou em rebarbas em suas bordas, imperceptíveis a olho nu, mesmo efetuando-se um acabamento pós-corte para evitá-las.

A Figura 25 apresenta a imagem da amplitude da amostra com os dados 0's armazenados e a região que foi afetada pelo toque da ponta na amostra.

Figura 25: Imagens das amostras afetada pelo toque da ponta do microscópio Multimode 8.

Por meio do *software* utilizado na análise é possível eliminar esta área afetada da imagem e analisar a região onde a varredura ocorreu com sucesso. A Figura 26 apresenta as novas imagens geradas após o processo de eliminação da área afetada da imagem.

Figura 26: Imagens das amostras com 0's armazenados – Microscópio Multimode 8.

Entretanto, não foi possível verificar a presença de campo magnético nas imagens, somente “ruídos”: a técnica de MFM mede a variação do campo magnético contido na amostra, porém neste caso não houve a variação magnética presente na amostra com os dados zeros armazenados. Nas demais amostras foi possível verificar a presença da variação do campo magnético; a Figura 27 apresenta as imagens da amplitude das outras 5 amostras dos discos rígidos utilizados nesta pesquisa (o 7º disco não foi analisado com este microscópio).

Figura 27: Imagens das Amostras com valores diferentes de 0 armazenados – Microscópio Multimode 8.

Nestas amostras é possível apenas observar a “superfície” das trilhas onde estão armazenados os bits que correspondem aos dados armazenados nos pratos dos discos rígidos. Após esta verificação foi notado que possivelmente este fato ocorreu por desgaste da ponta utilizada na varredura: as pontas utilizadas para a varredura por meio da técnica de MFM se desgastam rapidamente devido a sua alta sensibilidade e à corrente elétrica utilizada para realizar a varredura, porém, este fato só é perceptível após a aquisição das imagens.

Após esta constatação foi realizado um novo contato com o laboratório, porém não foi possível realizar novas varreduras por não haver pontas novas em estoque. A aquisição teria um custo alto, ao redor de 652 dólares a caixa com 10 unidades (Bruker, AFM PROBES, 2017), aproximadamente 2.100,00 reais na época que estava sendo realizada a avaliação experimental. Além disso estas pontas são importadas dos Estados Unidos, incorrendo em custos adicionais para a importação, tais como taxas e frete, e o tempo para o transporte e desembaraço alfandegário. Dados os prazos e verbas disponíveis para o experimento, tal aquisição foi inviabilizada.

Esta ocorrência foi documentada para ilustrar que o uso da técnica de MFM para recuperação de dados em discos rígidos, conforme propagado na literatura, não é trivial, envolvendo fatores que estão além da posse do disco rígido a ser analisado. São necessárias pessoas com conhecimento teórico e prático na área de forense computacional e no uso de microscópios de MFM para recuperação de dados.

Além disso, os microscópios que aplicam a técnica de MFM são caros (da ordem de centenas de milhares de dólares) e estão localizados em sua grande maioria em instituições de pesquisa especializadas, como no caso da Universidade de São Paulo. Não é um simples usuário que aplica a técnica de MFM para recuperar dados em disco rígido. Todo este processo da utilização de MFM para recuperar dados torna-se assim complexo e caro, sendo necessário que a equipe envolvida analise se o valor financeiro gasto compensa a tentativa de recuperação dos dados armazenados nos discos rígidos.

3.3.2.2 Análise das Imagens geradas pelo Microscópio Dimension Icon

Esta sessão apresenta as imagens resultantes do processo de varredura das sete amostras utilizadas na avaliação experimental para verificar a aplicabilidade de técnicas lógicas em discos rígidos atuais. Antes de realizar a varredura das amostras para registrar o campo magnético é necessário preparar a mesma para que sua superfície esteja a mais limpa possível para ser colocada no porta amostra, para isso é utilizado um solvente específico (neste caso foi utilizado o álcool isopropílico (isopropanol)). A Figura 28 ilustra o processo de limpeza da amostra por meio deste solvente e a colocação da amostra no porta amostra do Dimension Icon para a realização da varredura para gerar as imagens que serão analisadas.

Figura 28: Processo de preparação e colocação da amostra no porta amostra do Dimension Icon (Próprio autor).

Após a colocação da amostra no porta amostra é necessário ajustar a ponta na posição desejada para a varredura. Isto é possível por meio de uma tela do computador que apresenta a posição da ponta e a amostra em um ângulo de 90° em relação a visão que o usuário tem na tela apresentada.

De maneira análoga ao caso anterior, não foi constatada a presença de campo magnético na amostra com dados 0's gravados, sendo registrados apenas "ruídos" na imagem. A Figura 29 apresenta as imagens relacionadas a amostra com os dados 0's gravados em todo o disco rígido.

Figura 29: Imagens da Amostra com 0's gravados – Microscópio Dimension Icon.

Na amostra com dados 1's gravados foi possível verificar a presença de campo magnético, possibilitando ter a imagem do padrão do bit 1 gravado na superfície do prato do disco rígido. A Figura 30 apresenta as imagens da amplitude e fase criadas na superfície do prato do disco rígido onde foram gravados 1's em todo o prato do disco rígido, com *Scan Sizes* de 5, 10 e 20 μm .

Figura 30: Imagens em 2D - Amostra com 1's gravados – Microscópio Dimension Icon.

Foram também criadas imagens em 3D das amostras, de maneira a permitir a visualização de detalhes adicionais. A Figura 31 apresenta as imagens em 3D com *Scan Size* de 5, 10 e 20 μm para a amostra com dados 1's gravados, onde é possível distinguir os valores "1" armazenados.

Figura 31: Imagens em 3D - Amostra com 1's gravados – Microscópio Dimension Icon.

A seguir foi analisada a amostra com os dados originais armazenados, de maneira a se recuperar uma sequência de dados possível nas imagens de MFM. As imagens resultantes da amostra com os dados originais gravados, com *Scan Size* de 5, 10 e 20 μm , são apresentadas na Figura 32 (imagens 2D) e Figura 33 (imagens 3D).

Figura 32: Imagens em 2D - Amostra com Dados Originais gravados – Microscópio Dimension Icon.

Figura 33: Imagens em 3D - Amostra com Dados Originais gravados – Microscópio Dimension Icon.

Na Figura 32 é possível verificar fisicamente a separação entre as trilhas presente no prato do disco rígido, e os trechos dos dados originais gravados dentro dos setores que compõem as trilhas do disco rígido. Na sendo que na Figura 32-G é destacado um segmento que contém informações da *string* original de dados e que foi utilizado para verificar a possibilidade de recuperação dos dados. Foram verificados trechos de dados gravados diferentes entre as trilhas, bem como trechos semelhantes. Como foi gravado uma sequência de dados de 512 bytes em cada setor do disco rígido e como as trilhas que compõem o prato do disco rígido são círculos concêntricos, já era esperado esta possibilidade que trechos conforme apresentados na Figura 32-G ocorressem durante a gravação dos dados. A Figura 34 apresenta a região da imagem de 20 µm que foi alvo da recuperação de dados para verificar a possibilidade de recuperação de dados após a sanitização do mesmo.

Figura 34: Região utilizada para verificar a possibilidade de restauração dos dados gravados.

De acordo com (CASEY, 2011) o valor do bit 1 é maior que o valor do bit 0, ou seja, os picos nas imagens correspondem ao bit 1, e os vales correspondem ao bit 0. A sequência de bits mapeada da Figura 33 foi portanto de 70 bits: 010110101100101011011011010110100110001001011100011010101010110101010. Utilizando a tabela ASCII (WESSEL, SMITH, 1991) para o mapeamento de bits realizado a *string* correspondente encontrada foi a “**1,5d**” da *string* original de dados apresentado na Figura 10.

É interessante ressaltar que, neste tipo de análise, em alguns casos há mais de uma possibilidade de recuperação de caractere. Por exemplo, no caso acima, para a última letra foi possível identificar duas possibilidades de caracteres, uma vez que apenas o último conjunto de bits possuía apenas 7 *bits* (0110010): o caractere “**d**” (sequência de bits 01100100) e o caractere “**e**” (sequência de bits ‘01100101’). Neste

caso, porém, como a sequência dos dados gravados é conhecida a identificação do caractere correspondente ocorreu mais facilmente. Porém, um atacante que não conhece a sequência dos dados originais gravada provavelmente despenderia um tempo considerável na tentativa de identificar qual seria a sequência correspondente, isso sem contar que o tipo do arquivo impacta nesta análise (ex.: se os bits pertencessem a um arquivo de vídeo, estes bits teriam um significado completamente diferente).

A Figura 35 e 36 apresentam as imagens em 2D e 3D relacionadas a amostra com os dados sanitizados pela técnica de Gutmann.

Figura 35: Imagens em 2D - Amostra com Dados Sanitizados (Guttmann) – Microscópio Dimension Icon.

Figura 36: Imagens em 3D - Amostra com Dados Sanitizados (Guttmann) – Microscópio Dimension Icon.

Como observado na Figura 36 não é possível estabelecer qualquer relação com a sequência de dados recuperada (**em 1,5 d**) anteriormente na Figura 35. As imagens de MFM para a amostra com os dados sanitizados não apresentam qualquer padrão de escrita, apenas são apresentados campos magnéticos que não estão ordenados logicamente. Como o último passo da técnica de Gutmann sobrescreve o padrão aleatório, é possível que em algum momento da sobrescrita sejam gravadas sequências lógicas curtas de dados que formem palavra na língua portuguesa, porém, estas palavras não expressam qualquer relação com os dados originais gravados antes da sanitização ocorrer no disco rígido. As imagens da Figura 35 evidenciam que não é possível a "reconstrução" dos dados originais após os mesmos terem sido sanitizados.

As imagens das amostras 5 e 6 (Figura 37), correspondentes a discos rígidos usados, foram semelhantes às da Figura 34. Isto evidencia que a varredura por MFM não separa as “camadas” de dados gravados anteriormente no disco rígido, apenas verifica as variações dos campos magnéticos da última gravação realizada no prato do dispositivo; em outras palavras, não é possível observar “camadas” anteriores de gravações de dados.

Figura 37: Imagens em 2D - Amostra com Dados Sanitizados (Guttmann) nos discos 5 e 6 – Microscópio Dimension Icon.

Nesta etapa dos trabalhos foi obtido o 7_ disco rígido, que foi preparado da mesma forma que os discos rígidos 4, 5 e 6. Foi sanitizado então com a técnica lógica RCMP TSSIT OPS-II, apresentada na Tabela 6 como sendo a quarta melhor técnica

lógica da análise lógica, sendo selecionada devido aplicar o passo aleatório como último passo, semelhante à técnica de Gutmann utilizada nos outros discos rígidos. Foi, então realizada mais uma nova sessão de varredura no microscópio Dimension Icon, para verificar a possibilidade da recuperação dos dados utilizando esta técnica lógica para sanitizar a *string* com os dados originais.

A Figura 38 apresenta as imagens de MFM geradas pela sétima amostra, para os dados que foram sanitizados pela técnica de RCMP TSSIT OPS-II.

Figura 38: Imagens em 2D - Amostra com Dados Sanitizados (RCMO TSSIT OPS-II) nos discos 5 e 6 – Microscópio Dimension.

De maneira análoga à da técnica lógica de Gutmann a técnica lógica RCMP TSSIT OPS-II removeu os dados de forma eficaz, impossibilitando sua recuperação e dando indícios que as técnicas lógicas para sanitização de dados em geral são eficazes para se evitar a recuperação de dados utilizado o ataque de MFM para recuperar dados.

3.4 Considerações Finais do Capítulo

Neste capítulo foram apresentados os procedimentos experimentais e os resultados referentes as avaliações lógica e experimental. Na avaliação lógica as quatro técnicas lógicas melhores colocadas foram as Gutmann, GOST R50739 95, AR 380 19, e RCMP TSSIT OPS - II, sendo utilizada na avaliação experimental as técnicas de Gutmann e a técnica RCMP TSSIT OPS - II.

Foram apresentados detalhes dos dois microscópios (MultiMode 8, Dimension Icon) utilizados para realizar a avaliação experimental, bem como detalhes de seus mecanismos, e de como a técnica de MFM é aplicada nas amostras de discos rígidos e as imagens resultantes da varredura realizada na amostra. Foram apresentados, ainda os recursos utilizados para a análise das imagens de MFM resultante das amostras dos discos rígidos alvo desta pesquisa.

Por fim, foram mostradas as imagens resultantes da aplicação da técnica de MFM nos discos rígidos utilizados como amostra. No microscópio MultiMode 8 não foi possível obter um resultado satisfatório devido à limitação encontrada no uso da ponta utilizada para realizar a varredura das amostras. No microscópio Dimension Icon as varreduras realizadas nas amostras ocorreram com sucesso, gerando as imagens de MFM das amostras utilizadas, que foram analisadas para se comprovar a eficiência de técnicas lógicas em sanitizar os dados que estão armazenados no disco rígido para evitar a sua recuperação.

4 Considerações Finais

O objetivo proposto nesta dissertação foi o de verificar a aplicabilidade de técnicas lógicas para sanitização de dados em discos rígidos atuais, sendo este objetivo atingido por meios dos resultados apresentados. Após a especificação dos requisitos para realizar tanto a avaliação lógica como a experimental, tais avaliações foram realizadas com sucesso. A avaliação lógica permitiu verificar como as técnicas lógicas removem os dados originais do disco rígido. A avaliação experimental permitiu verificar a eficácia das técnicas lógicas em sanitizar os dados em discos rígidos atuais.

Para a avaliação lógica foi utilizada a distância de Hamming, enquanto que para a avaliação experimental foi utilizada a técnica MFM para verificar a presença de campo magnético gravados nas amostras de discos rígidos utilizadas. Ao final da avaliação foi comprovado que não é possível a recuperação de dados após a aplicação de uma técnica lógica para sanitizar os dados.

4.1 Contribuições

A principal contribuição desta pesquisa é a verificação da aplicabilidade de técnicas lógicas em discos rígidos atuais. Apresentando assim uma possível solução ao desafio sobre sanitização de dados que estava em aberto. Assim os usuários podem sanitizar seus dados de discos rígidos de forma segura empregando técnicas lógicas, garantindo-se que os dados sanitizados não serão mais recuperados por um atacante que esteja de posse do disco rígido desse usuário.

É interessante ressaltar que neste tipo de análise em alguns casos há mais de uma possibilidade de recuperação de caractere. Por exemplo, no caso mencionado no capítulo 3, para a última letra foi possível identificar duas possibilidades de caracteres, uma vez que apenas o último conjunto de *bits* possuía apenas 7 *bits* (0110010): o caractere “d” (sequência de *bits* “01100100”) e o caractere “e” (sequência de bits “01100101”). Neste caso, como a sequência dos dados gravados é conhecida, a identificação do caractere correspondente ocorreu facilmente, porém, um atacante que não conhece a sequência dos dados originais gravada provavelmente

despenderia um tempo considerável na tentativa de identificar qual seria a sequência correspondente, isso sem contar que o tipo do arquivo impacta nesta análise (ex.: se os bits pertencessem a um arquivo de vídeo estes *bits* teriam um significado completamente diferente).

Outro ponto que deve ser destacado é que o quarto disco rígido utilizado onde foram recuperados dados possui dados homogêneos: (uma única *string* gravada repetidas vezes), porém, em um ambiente onde o disco rígido alvo tenha sido usado no dia a dia com diferentes aplicações, provavelmente existiriam vários tipos de arquivos armazenados com tamanhos e estruturas diferentes entre si, estando eles espalhados pelo disco rígido, e não em uma sequência ordenada nos setores, dificultando sua reconstrução. Isso inviabiliza a recuperação dos dados

Os resultados obtidos mostram que as técnicas lógicas oferecem uma maneira segura para que a sanitização de dados ocorra nos discos rígidos atuais e que os mesmos possam ser reutilizados novamente em seu ambiente ou em outro ambiente ou até mesmo, por exemplo, em um processo de reuso de equipamentos de informática onde estejam envolvidos discos rígidos, garantido-se que os dados não serão recuperados. Isso favorece ações de sustentabilidade voltadas ao reuso de discos rígidos em computadores remanufaturados, como tem sido praticado pelo CEDIR (Centro de Reuso e Descarte de Resíduos de Informática) da USP.

4.2 Participação em Publicações e Publicações relacionadas a pesquisa

O objetivo desta sessão é apresentar as atividades realizadas pelo candidato dentro do programa de mestrado. Estas atividades são pertinentes à participação em projeto de pesquisa e publicação de artigo. Dentre o período do programa de mestrado (Abril/2015 a Fevereiro/2017), o candidato participou de projetos como:

- **Acordo de Nível de Serviço de Segurança em Computação em Nuvem:** O principal objetivo deste projeto de pesquisa foi propor um arcabouço e mecanismo de execução para acordo de nível de serviço (SLA) de segurança em ambiente de computação em nuvem. Projeto de

pesquisa com parceira internacional (Ericsson Research). (Finalizado em 30/09/2015).

- **Uma estrutura tolerante a falhas adaptado a aplicativos baseados em nuvem:** seu principal objetivo foi desenvolver uma estrutura tolerante a falhas sob medida para aplicações baseadas na nuvem para que as métricas tolerantes a falhas possam ser adaptáveis de acordo com demandas de negócios e os recursos disponíveis (Ericsson Research). (Finalizado em 31/05/2016).

Além da participação em projetos, o autor tem co-autoria em três publicações. A primeira com o título **"Inclusion of security requirements in SLA lifecycle management for cloud computing"** (ROJAS et al., 2015), a segunda com o título **"A framework to orchestrate security SLA lifecycle in cloud computing"** (ROJAS et al., 2016) e a terceira com o título **"Managing the Lifecycle of Security SLA Requirements in Cloud Computing"** (ROJAS et al., 2018).

Os resultados prévios desta pesquisa foram apresentados em: V Workshop de pós-graduação da Escola Politécnica da Universidade de São Paulo, intitulado **"A study of the applicability of data sanitization techniques in a cloud computing environment"** (SBAMPATO, 2016).

O artigo **"Data Sanitization in Current Hard Disks"** foi aceito no Congresso Internacional de Ciências de Computação e Sistema de Informação - CICC SI (SBAMPATO et al., 2016).

Um artigo sobre o tema **"Data Destruction"** foi publicado em fevereiro de 2018 na revista **"Digital Forensics"**.

E por fim, está em processo de submissão um artigo contendo toda descrição do projeto ao journal **"The International Journal of Digital Forensics Incident Response"** mantido pela ELSEVIER.

4.3 Trabalhos Futuros

Em termos de publicações planeja-se escrever um artigo final para periódico com os resultados da pesquisa e detalhes sobre o método teórico e sobre as avaliações

realizadas nas amostras dos discos rígidos por meio da técnica de MFM. Outro ponto importante é a continuação do desenvolvimento desta pesquisa sobre a aplicabilidade de técnicas lógicas para a sanitização de dados em discos rígidos atuais.

É pretendido utilizar "*machine learning*" para o reconhecimento de padrões, ou seja, "treinar" um algoritmo para que este possa reconhecer o padrão do bit 0 e do bit 1 gravado nas imagens criadas por meio da técnica de MFM. Deste modo será possível determinar sequências maiores de dados, possibilitando até determinar possíveis frases que façam sentido para entendimento dos dados armazenados nas amostras.

5 Referências

- ABBOTT, W. L.; NGUYEN, H.; KUO, B.; OVENS, K.; WONG, Y.; CASASANTA, J. A digital chip with adaptive equalizer for prml detection in hard-disk drives. Solid-State Circuits Conference, 1994. Digest of Technical Papers. 41st ISSCC. 1994 IEEE International. [S.l.], 1994. p. 284–285.
- ABRAMOVITCH, D.; FRANKLIN, G. Disk drive control: The early years. Annual Reviews in Control, Elsevier, v. 26, n. 2, p. 229–242, 2002.
- ACER. Advanced Format. 2013.
- https://en.wikipedia.org/wiki/AdvancedFormat#cite_note-IDEMA2-4. Acessado em: 16/02/2018
- AMAZON, A. Amazon Web Services Overview of Security Processes. 2015.
- ARPACI-DUSSEAU, A. C.; ARPACI-DUSSEAU, R. H.; CULLER, D. E.; HELLERSTEIN, J. M.; PATTERSON, D. A. High-performance sorting on networks of workstations. ACM SIGMOD Record. [S.l.], 1997. v. 26, n. 2, p. 243–254.
- BALDUZZI, M.; ZADDACH, J.; BALZAROTTI, D.; KIRDA, E.; LOUREIRO, S. A security analysis of amazon's elastic compute cloud service. Proceedings of the 27th Annual ACM Symposium on Applied Computing. [S.l.], 2012. p. 1427–1434.
- BARRETO, G. L. Utilização de técnicas anti-forenses para garantir a confidencialidade. Curitiba, PUC-PR. Disponível em: <http://www.ppgia.pucpr.br/jamhour/Download/pub/RSS/MTC/referencias/TCC>. 2009.
- BAUER, S.; PRIYANTHA, N. B. Secure data deletion for linux file systems. Usenix Security Symposium. [S.l.: s.n.], 2001. v. 174.
- BELARMINO, A. T. d. N.; ROMERO, Á. L.; MELLO, G. S. de; FORTES, M. de A.; CAMPELLO, R. S. Levantamento sobre a utilização de técnicas de microscopia na recuperação de dados em discos rígidos. ICCyberâ2004, p. 52, 2004.
- BISHOP, M.; CRAWFORD, R.; BHUMIRATANA, B.; CLARK, L.; LEVITT, K. Some problems in sanitizing network data. Enabling Technologies: Infrastructure for Collaborative Enterprises, 2006. WETICE'06. 15th IEEE International Workshops on. [S.l.], 2006. p. 307–312.
- BOHN, R. B.; MESSINA, J.; LIU, F.; TONG, J.; MAO, J. Nist cloud computing reference architecture. Services (SERVICES), 2011 IEEE World Congress on. [S.l.], 2011. p. 594–596.
- BRENDER, N.; MARKOV, I. Risk perception and risk management in cloud computing: Results from a case study of swiss companies. International journal of information management, Elsevier, v. 33, n. 5, p. 726–733, 2013.

BRUKER, A. P. AFM PROBER. 2018. <http://www.brukerafmprobes.com/p-3309-mesp.aspx>. Acessado em 15/04/2018.

BRUKER, C. Cantilever. 2018. <http://nano.boisestate.edu/wpcontent/uploads/2010/12/NANOSCOPE-CONTROLLER-USING-NANOSCOPE-8-SOFTWARE-D-004-1030-000.pdf>. Acessado em 11/04/2018.

BRUKER, D. I. Dimension Icon. 2018. <http://www.nanophys.kth.se/nanophys/facilities/nfl/afm/icon/brukerhelp/Iconwebhelp.html>. Acessado em 17/04/2018.

BRUKER, M. MultiMode 8. 2018. <http://www.brukerafmprobes.com/p-3309-mesp.aspx>. Acessado em 09/04/2018.

BRUKER, N. A. NanoScope Analysis. 2018. <ftp://Release:Z3Analysis91@sboftp.bruker-nano.com>. Acessado em 15/04/2018.

CARLOMAGNO, M. S.; SCHERER, F. O. Práticas Dos Inovadores: Tudo Que Você Precisa Saber Para Começar a Inovar. [S.l.]: Editora Atlas SA, 2000.

CASEY, E. Digital evidence and computer crime: Forensic science, computers, and the internet. [S.l.]: Academic press, 2011.

CENTER, C. S. R. Cryptographic Module Validation Program Conference 2002. 2018. <https://csrc.nist.gov/events/2002/cryptographic-module-validation-programconference>. Acessado em 26/02/2018.

CHALLA, N.; GANNAGE, M. E. Methods and apparatus for hard disk emulation. [S.l.]: Google Patents mar. 1 1994. US Patent 5,291,584.

CHAWKI, J. Cloud computing standards: overview and itu-t positioning. International Telecommunications Union. 2016. <https://www.itu.int/dms pub/itu-t/oth/06/5B/T065B00001C0043PDFE.pdf>. Acessado em 19/02/2018.

CHEN, B. M.; LEE, T. H.; VENKATARAMANAN, V.; CHEN, B. M. Hard disk drive servo systems. [S.l.]: Springer Berlin, 2002.

CHEN, D.; ZHAO, H. Data security and privacy protection issues in cloud computing. Computer Science and Electronics Engineering (ICCSEE), 2012 International Conference on. [S.l.], 2012. v. 1, p. 647–651.

COHEN, W.; RAVIKUMAR, P.; FIENBERG, S. A comparison of string metrics for matching names and records. Kdd workshop on data cleaning and object consolidation. [S.l.: s.n.], 2003. v. 3, p. 73–78.

COSTA, P. R. Caracterização de Materiais por Microscopia de Força Atômica. 2018. <http://www.ebah.com.br/content/ABAAAEi4QAD/afm-caracterizacao-materiaispor-microscopia-forca-atmica-por-rodrigo-pasiani-costa?part=2>. Acessado em 17/04/2018.

DANIEL, E. D.; MEE, C. D.; CLARK, M. H. Magnetic recording: the first 100 years. [S.I.]: John Wiley & Sons, 1999.

DATA, D. Media Sanitization Solutions. 2018. <https://www.destructdata.com/home/>. Acessado em 26/02/2018.

DATA, P. Security. 2008. <http://www.protondata.com/portuguese/products.htm#degaussers>. Acessado em 20/02/2018.

DATASHEETS. Model 0300 Low/Medium Volume Hard Drive Shredder. 2018. <https://datasheets.globalspec.com/ps/5351/SecurityEngineeredMachinery/5162661F-3E4F-4CE9-BFE0-3B51FE3F1F14>. Acessado em 20/02/2018.

DHAMDHARE, D. M.; PATIL, H. An elimination algorithm for bidirectional data flow problems using edge placement. ACM Transactions on Programming Languages and Systems (TOPLAS), ACM, v. 15, n. 2, p. 312–336, 1993.

DHILLON, G. Principles of Information Systems Security: text and cases. [S.I.]: Wiley Hoboken, NJ, 2007.

DICE, L. R. Measures of the amount of ecologic association between species. Ecology, Wiley Online Library, v. 26, n. 3, p. 297–302, 1945.

DIKAIKOS, M. D.; KATSAROS, D.; MEHRA, P.; PALLIS, G.; VAKALI, A. Cloud computing: Distributed internet computing for it and scientific research. IEEE Internet computing, IEEE, v. 13, n. 5, 2009.

DOD, J. Industrial Security Manual for Safeguarding Classified Information. [S.I.]: DoD, 1991.

DONG, H.; KUN, S.; YU, C. Research on secure destruction of digital information. Apperceiving Computing and Intelligence Analysis, 2009. ICACIA 2009. International Conference on. [S.I.], 2009. p. 356–359.

DORION, P. Data destruction services: When data deletion is not enough. SearchDataBackup.com, 2010.

DURKEE, D. Queue, ACM, v. 8, n. 4, p. 20, 2010. ENISA. Cloud computing security strategy. 2016. <https://www.terena.org/activities/tfcsirt/meeting30/hogbencloudcomputing.pdf>. Acessado em 16/02/2018.

ESALINK. Proton PPD-1. 2018. <http://www.esalink.com.br/produtosdetalhes10.php>. Acessado em 20/02/2018:

FARIAS, P. C. B. Treinamento profissional em redes wireless. [S.I.]: Universo dos Livros Editora, 2006.

FERNANDES, D. A.; SOARES, L. F.; GOMES, J. V.; FREIRE, M. M.; INÁCIO, P. R. Security issues in cloud environments: a survey. *International Journal of Information Security*, Springer, v. 13, n. 2, p. 113–170, 2014.

GARFINKEL, S. L.; SHELAT, A. Remembrance of data passed: A study of disk sanitization practices. *IEEE Security & Privacy*, IEEE, v. 99, n. 1, p. 17–27, 2003.
GILL, P.; PHYTHIAN, M. *Intelligence in an insecure world*. [S.l.]: Polity, 2006.

GONZALEZ, N.; MIERS, C.; REDIGOLO, F.; SIMPLICIO, M.; CARVALHO, T.; NÄSLUND, M.; POURZANDI, M. A quantitative analysis of current security concerns and solutions for cloud computing. *Journal of Cloud Computing: Advances Systems and Applications*, Springer, v. 1, n. 1, p. 11, 2012.

GROCHOWSKI, E.; HOYT, R. F. Future trends in hard disk drives. *IEEE transactions on magnetics*, IEEE, v. 32, n. 3, p. 1850–1854, 1996.

GUTMANN, P. Secure deletion of data from magnetic and solid-state memory. *Proceedings of the Sixth USENIX Security Symposium*, San Jose, CA. [S.l.: s.n.], 1996. v. 14, p. 77–89.

HAMLEN, K.; KANTARCIOGLU, M.; KHAN, L.; THURASINGHAM, B. Security issues for cloud computing. *Optimizing information security and advancing privacy assurance: new technologies: new technologies*, IGI Global, v. 150, 2012.

HAMMING, R. W. Error detecting and error correcting codes. *Bell Labs Technical Journal*, Wiley Online Library, v. 29, n. 2, p. 147–160, 1950.

HUGHES, G.; COUGHLIN, T. Tutorial on disk drive data sanitization. <http://ucsd.edu/people/Hughes/DataSanitizationTutorial.pdf>. Acessado em 03/07 2018.

HUGHES, G. F.; COUGHLIN, T.; COMMINS, D. M. Disposal of disk and tape data by secure sanitization. *IEEE Security & Privacy*, IEEE, v. 7, n. 4, 2009.

IDEMA. The advent of advanced format. 2013. [Http://idema.org/?pageid=2369](http://idema.org/?pageid=2369). Acessado em 16/02/2018.

IEEE-SPECTRUM. Capacidade do armazenamento de dados ao longo dos anos segundo a lei de Moore. 2015.

JIANG, S.-H.; LO, F.-H. Prml process of multilevel run-length-limited modulation recording on optical disc. *IEEE transactions on magnetics*, IEEE, v. 41, n. 2, p. 1070–1072, 2005.

KANEKAL, V. *Data Reconstruction from a Hard Disk Drive using Magnetic Force Microscopy*. [S.l.]: University of California, San Diego, 2013.

KANG, C.-I.; LEE, B.-J. Method and apparatus improving prevention of track writing in a hard disk drive. [S.l.]: Google Patents, jul. 20 2010. US Patent 7,760,455.

KASAVAJHALA, V. Solid state drive vs. hard disk drive price and performance study. Proc. Dell Tech. White Paper, p. 8–9, 2011.

KISSEL, R.; REGENSCHEID, A.; SCHOLL, M.; STINE, K. Guidelines for media sanitization. [S.l.]: US Department of Commerce, National Institute of Standards and Technology, 2014.

KOBLISCHKA, M.; WEI, J.; HARTMANN, U. Optimization of the hf-mfm technique. Journal of Physics: Conference Series. [S.l.], 2007. v. 61, n. 1, p. 591.

KOZIEROK. Run length limited (RLL). 2001.
<http://www.pcguide.com/ref/hdd/geom/dataRLL-c.html>. Acessado em 16/02/2018.

LEE, P. J. Disk drive comprising a multiple-input sequence detector selectively based by bits of a decoded ECC coded word. [S.l.]: Google Patents, ago. 5 2003. US Patent 6,604,220.

LINUX. Linux on 4 kb sector disks: Practical advice. 2016.
<https://www.ibm.com/developerworks/library/l-linux-on-4kb-sector-disks/>. Acessado em 16/02/2018.

LOMBARDI, F.; PIETRO, R. D. Secure virtualization for cloud computing. Journal of Network and Computer Applications, Elsevier, v. 34, n. 4, p. 1113–1122, 2011.

MAGGIANI, R. Cloud computing is changing how we communicate. Professional communication conference, 2009. IPCC 2009. IEEE international. [S.l.], 2009. p. 1–4.

MARTURANA, F.; ME, G.; TACCONI, S. A case study on digital forensics in the cloud. Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2012 International Conference on. [S.l.], 2012. p. 111–116.

MCDONALD, O. C.; WANG, X.; BARROS, M. M. D.; BONILLA, R. K.; KE, Q. Strategies for sanitizing data items. [S.l.]: Google Patents mar. 24 2009. US Patent 7,509,684.

MELL, P.; GRANCE, T. et al. The nist definition of cloud computing. Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology Gaithersburg, 2011.

MICROSOFT. Política de suporte para discos rígidos com setor 4k no Windows. 2016. <https://support.microsoft.com/pt-br/kb/2510009>. Acessado em 16/02/2018.

MONETT, M. R. Rapid starting variable frequency oscillator with minimum startup perturbations. [S.l.]: Google Patents, ago. 6 1985. US Patent 4,533,881.

MONGE, A. E.; ELKAN, C. et al. The field-matching problem: Algorithms and applications. KDD. [S.l.: s.n.], 1996. p. 267–270.

PASQUALETTI, F.; DÖRFLER, F.; BULLO, F. Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, IEEE, v. 58, n. 11, p. 2715–2729, 2013.

PATRICK, L. Disk drive employing a multiple-input sequence detector responsive to reliability metrics to improve a retry operation. [S.l.]: Google Patents, jun. 11 2002. US Patent 6,405,342.

POLICE, R. C. M. Hard drive secure information removal and destruction guidelines. Retrieved September, v. 25, p. 2007, 2003.

PORCHER, V. G.; GARCIA, M. L. F. F. Network storage: Estudo de caso e implantação de uma san. 2000.

PRODANOV, C. C.; FREITAS, E. C. de. Metodologia do Trabalho Científico: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico-2a Edição. [S.l.]: Editora Feevale, 2013.

PRODUCTS, D. D. Services. 2016.

<http://www.semshred.com/0300harddriveshredder>. Acessado em 20/02/2018.

RAMAN, R.; PRAMOD, D. A study on data privacy, protection& sanitization practices during disk disposal by Indian educational institutes. *International Journal of Computer Science Issues (IJCSI)*, v. 2, 2013.

RF, S. Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. *Revista brasileira de fisioterapia*, SciELO Brasil, v. 11, n. 1, p. 83–89, 2007.

ROEBUCK, K. Data Wiping and Destruction: High-impact Strategies-What You Need to Know: Definitions, Adoptions, Impact, Benefits, Maturity, and Vendors. [S.l.]: Emereo Publishing, 2012.

ROJAS, M. A.; GONZALEZ, N. M.; SBAMPATO, F. V.; REDÍGOLO, F. F.; CARVALHO, T.; ULLAH, K. W.; NÄSLUND, M.; AHMED, A. S. A framework to orchestrate security sla lifecycle in cloud computing. *Information Systems and Technologies (CISTI)*, 2016 11th Iberian Conference on. [S.l.], 2016. p. 1–7.

ROJAS, M. A. T.; GONZALEZ, N. M.; SBAMPATO, F.; REDIGOLO, F.; CARVALHO, T. C. M. de B.; NGUYEN, K. K.; CHERIET, M. Inclusion of security requirements in sla lifecycle management for cloud computing. *Evolving Security and Privacy Requirements Engineering (ESPRE)*, 2015 IEEE 2nd Workshop on. [S.l.], 2015. p. 7–12.

ROJAS, M. A. T.; REDÍGOLO, F. F.; GONZALEZ, N. M.; SBAMPATO, F. V.; CARVALHO, T. C. M. de B.; ULLAH, K. W.; NÄSLUND, M.; AHMED, A. S. Managing the lifecycle of security sla requirements in cloud computing. In: *Developments and Advances in Intelligent Systems and Applications*. [S.l.]: Springer, 2018. p. 119–140.

RUGAR, D.; MAMIN, H.; GUETHNER, P.; LAMBERT, S.; STERN, J.; MCFADYEN, I.; YOGI, T. Magnetic force microscopy: General principles and application to longitudinal recording media. *Journal of Applied Physics*, AIP, v. 68, n. 3, p. 1169–1183, 1990.

RUSSO, G. Gost. 2018. <http://protect.gost.ru/v.aspx?control=8baseC=-1page=0month=1year=1search=RegNum=1DocOnPageCount=15id=126373pageK=BB9EFA6B-6634-47AD-969D-552AB6AD2127>. Acessado em 26/02/2018.

SBAMPATO, F. V. A study of the applicability of data sanitization techniques in a cloud-computing environment. 2016.

SBAMPATO, F. V.; ROJAS, M. A. T.; REDÍGOLO, F. F.; CARVALHO, T. C. M. de B. A study of the applicability of data sanitization techniques in a cloud computing environment. 2016.

SBS. Secure Data Sanitization. 2018. <http://www.securedatasanitization.com/Lobby>. Acessado em 22/02/2018.

SEAGATE. Disco rígido de 4 Terabytes. 2018. <https://www.seagate.com/br/pt/internal-hard-drives/hdd/barracuda>. Acessado em 31/01/2018.

SEAGATE, B. Transição para discos rígidos de formato avançado com setores de 4K. 2018. <http://www.seagate.com/br/pt/tech-insights/advancedformat-4k-sector-harddrives-master-ti>. Acessado em 07/02/2018.

SHERRILL-MIX, S. Latency, expression and splicing during infection with HIV. [S.l.]: University of Pennsylvania, 2015.

SHIROISHI, Y. Magnetic recording and reading device. [S.l.]: Google Patents, fev. 13 2007. US Patent 7,177,115.

STENFORT, R. J. Method and device for performing error correction on ECC data sectors. [S.l.]: Google Patents, dez. 9 2003. US Patent 6,662,334.

STEVENS, L. D. The evolution of magnetic storage. *IBM Journal of Research and Development*, IBM, v. 25, n. 5, p. 663–676, 1981.

STEYAERT, M. S.; DEHAENE, W.; CRANINCKX, J.; WALSH, M.; REAL, P. A cmos rectifier-integrator for amplitude detection in hard disk servo loops. *IEEE Journal of solid-state circuits*, IEEE, v. 30, n. 7, p. 743–751, 1995.

TAKABI, H.; JOSHI, J. B.; AHN, G.-J. Security and privacy challenges in cloud computing environments. *IEEE Security & Privacy*, IEEE, v. 8, n. 6, p. 24–31, 2010.

TANENBAUM, A. S.; ZUCCHI, W. L. Organização estruturada de computadores. [S.l.]: Pearson Prentice Hall, 2009.

THIBADEAU, R. System and method for magnetic storage disposal. [S.l.]: Google Patents, nov. 24 2005. US Patent App. 10/852,710.

WANG, Q.; WANG, C.; REN, K.; LOU, W.; LI, J. Enabling public auditability and data dynamics for storage security in cloud computing. *IEEE transactions on parallel and distributed systems*, IEEE, v. 22, n. 5, p. 847–859, 2011.

WESSEL, P.; SMITH, W. H. Free software helps map and display data. *Eos, Transactions American Geophysical Union*, Wiley Online Library, v. 72, n. 41, p. 441–446, 1991.

WINKLER, W. E. String comparator metrics and enhanced decision rules in the fellegi sunter model of record linkage. ERIC, 1990.

WOOD, R. Future hard disk drive systems. *Journal of magnetism and magnetic materials*, Elsevier, v. 321, n. 6, p. 555–561, 2009.

WRIGHT, C.; KLEIMAN, D. Overwriting hard drive data: The great wiping controversy. *International Conference on Information Systems Security*. [S.l.], 2008. p. 243–257.

YAN, G.; XUE, M.; XU, Z. Disposal of waste computer hard disk drive: data destruction and resources recycling. *Waste Management & Research*, Sage Publications Sage UK: London, England, v. 31, n. 6, p. 559–567, 2013.

Apêndice A - Componentes do disco rígido

Os discos rígidos atuais podem ser divididos em duas partes: A parte mecânica dos componentes do disco rígido e a parte lógica responsável pelo gerenciamento dos dados.

A.1 Componentes mecânicos

Os principais componentes mecânicos são:

- **Placa lógica / controladora:** responsável por tarefas como: movimentação dos pratos que compõem a estrutura do disco rígido e da cabeça de leitura e gravação eletromagnética; envio e recebimento dos dados entre o disco rígido e o computador e rotinas de segurança ligada à gravação dos dados;
- **Prato:** disco, onde os dados são armazenados fisicamente é feito de alumínio com uma cobertura de material magnético e por uma camada de material protetor. Quanto mais denso for o material magnético, maior capacidade de armazenamento de dados o disco rígido possui. Os discos rígidos atuais contêm mais de um prato na sua estrutura possibilitando maior espaço de armazenamento de dados (CHEN et al., 2006);
- **Buffer:** chip de memória responsável pelo armazenamento de trechos de dados durante a comunicação com o computador. Facilita a transferência de dados por lidar com dados de maneira mais rápida do que os discos rígidos;
- **Atuador:** tem a função de mover o braço móvel sobre a superfície dos pratos permitindo que a cabeça de leitura e gravação eletromagnética realize suas tarefas. A proximidade da cabeça de leitura e gravação eletromagnética com o prato é tão extrema que chegar a ser ter a percepção que ambas se tocam. Porém esta operação acontece em um ângulo de 90° e nunca há o toque entre a cabeça de leitura e gravação eletromagnética e o prato (MEYERS, 2004);

- **Porta de comunicação/interfaces:** conectada à cabeça de leitura e gravação eletromagnética, permite a troca de dados entre o disco rígido e a placa mãe (motherboard) do computador.

Os principais padrões de interfaces são:

Padrão IDE:

Primeiro padrão a ser utilizado foi o padrão *Intelligent Drive Electronics* ou *Integrated Drive Electronics* - IDE. Os discos rígidos que utilizaram este padrão têm um limite de armazenamento de 504 MB. O padrão IDE com o passar dos anos evoluiu para o padrão *Extended IDE* - EIDE o que possibilitou um avanço no armazenamento de dados para 128 Gigabytes.

Padrão ATA:

O próximo padrão a ser aceito foi o ATA - 3 com o tempo de acesso aos dados mais rápido que seu antecessor. Para aumentar ainda mais a velocidade de transmissão de dados foi desenvolvido o padrão ATA *Packet Interface*, sendo que este padrão de porta de comunicação/interface também foi utilizado em gravadores de CD's. Posteriormente, surgiu o ATAPI - 5, porém, foi o ATAPI - 6 que mudou o limite do tamanho de armazenamento de dados do disco rígido, passando de 128 Gigabytes para 128 Petas bytes. A taxa de transferência dos dados chega a 100 MB/s. Do padrão IDE até o padrão ATAPI - 6 é denominado na literatura como padrão Parallel ATA - PATA.

Padrão SATA:

No ano de 2000 houve uma ruptura no padrão de comunicação paralela fornecida pelo padrão PATA por meio do padrão *Serial Advanced Technology Attachment* - SATA. Na primeira versão do SATA a taxa de transmissão dos dados chegou à velocidade de 150 MB/s. Na segunda versão do SATA a taxa de transmissão chegou a 300 MB/s, sendo este padrão o utilizado nos discos rígidos atuais. A modificação do padrão de transmissão (do paralelo para o serial) possibilitou uma considerável perda de ruído (falha na transmissão de dados), permitindo mais eficiência na transmissão de dados.

A versão SATA III lançada em 2009 permite uma taxa de transferência de até 600 MB/s, A organização responsável pelo gerenciamento do padrão SATA é o *Serial ATA International Organization*, composta por empresas do setor de produção de discos rígidos, tendo sido formada em julho de 2004.

Padrão SCSI:

Outro padrão de porta de comunicação/interfaces que está conectada diretamente à placa lógica/controladora, porém, não está ligada diretamente ao disco rígido é o *Small Computer System Interface* - SCSI. Neste modelo tanto a placa lógica/controladora e a porta de comunicação/interfaces são adquiridas separadamente do disco rígido. Os discos rígidos do padrão SCSI são indicados para servidores devido à alta taxa de transmissão alcançada e à tecnologia de *Redundant Array of Inexpensive Disk* - RAID 1 implementada. Esta tecnologia permite o paralelismo em nível de disco, aumentando a eficiência dos servidores de arquivos.

O primeiro modelo SCSI foi lançado em 1986 com uma taxa de transferência de 5 MB/s permitindo 8 discos rígidos acoplados à porta de comunicação/interfaces. Dentre as melhorias alcançadas no padrão SCSI II, temos a possibilidade de acoplar 16 discos rígidos e a velocidade de transferência dos dados passou para 10 MB/s. Atualmente é utilizada a transferência *Serial Attached SCSI*, compatível com a SATA, podendo controlar até 128 discos rígidos. Trata-se da tecnologia empregada no ambiente de computação em nuvem para prover garantias e segurança no armazenamento dos dados do provedor.

Padrão SAS:

A sigla SAS significa *Serial Attached SCSI*, a transmissão é serial, *full-duplex*, sendo as taxas de transmissão máxima de 375 MB/s, com uma frequência máxima de 3.0 GHz. A SAS tem conexão dedicada ou exclusiva para evitar concorrência, podendo ser compartilhada por meio do uso de um expensor. Seu objetivo é fornecer maior agilidade na transmissão de dados, juntando confiabilidade, escalabilidade e gerenciamento e alto desempenho.

A Figura 39 apresenta a localização de cada componente mecânico dentro do disco rígido.

Figura 39: Apresentação dos componentes mecânicos dentro de um disco rígido.

Como é possível notar por meio da Figura 39 a placa controladora é o único componente mecânico tem contato no disco rígido, os demais componentes estão localizados dentro do disco rígido. É necessário abrir o disco rígido para ter contato diretamente a estes componentes mecânicos, porém, ao abrir o disco rígido seu funcionamento pode ser comprometido, devido ao fator da interação do ambiente com seus componentes.

A.2 Componentes lógicos

Os principais componentes lógicos são:

- **Trilhas:** são círculos concêntricos que se iniciam no final do prato e vão se tornando menores conforme se aproximam do centro do prato. Cada Trilha que compõe o prato recebe um endereço lógico que permite sua localização. A trilha mais externa recebe o endereço lógico 0, a próxima trilha recebe o endereço lógico 1 e assim sucessivamente.
- **Setores:** são pequenos trechos dentro das trilhas que armazenam os dados, sendo que cada setor pode armazenar 512 bytes de dados. Para que fosse possível definir o limite entre as trilhas e os setores foram criadas “marcas” que são sinais magnéticos especiais que orientam a cabeça de leitura e gravação eletromagnética, possibilitando que a placa logica/controladora tenha acesso aos dados. Cada setor é composto por:
 - Seção de intervalo (GAP) que separa cada setor do disco rígido;

- Seção de sincronização (SYNC), sendo que a marca de sincronização indica o início do setor e fornece o alinhamento de tempo;
- Seção de marca de endereço (*Address Mark*), responsável por marcar logicamente o endereço, identificar e localizar o número do setor
- Seção de dados (Data), que são os dados pertencentes aos usuários; e
- Seção de correção de erros (ECC), que contém o algoritmo de correção de erros que visam reparar e recuperar dados que são corrompidos durante o processo de leitura e gravação.

A Figura 40 apresenta os componentes lógicos que compõem o disco rígido.

Figura 40: Apresentação dos componentes lógicos dentro de um setor (SEAGATE, 2013).

Cada componente é organizado de forma que ele possa atuar ordenadamente, ou seja, que o dado seja recebido, verificado, corrigido e armazenado corretamente. Toda a organização lógica do disco rígido ocorre no prato, sendo que o prato gira em sentido horário. A Figura 41 apresenta uma ilustração da separação lógica de cada componente lógico.

Figura 41: Apresentação dos componentes lógicos do disco rígido (PIROPO, 2007).

Como apresenta a Figura 41, um disco rígido pode ter mais de um prato para armazenar dados, sendo que ambas as faces do prato são utilizadas para armazenar dados. Toda a organização lógica do disco rígido ocorre no prato, sendo que o prato gira em sentido horário.

Apêndice B – Preparação das amostras para o microscópio MultiMode 8

Para adquirir o conhecimento de como criar e preparar as amostras para o microscópio Multimode 8 foram utilizados três discos rígidos que já não estavam mais em uso, uma vez que seria necessário abrir e cortar o disco rígido para o tamanho do porta amostra. A Figura 42 apresenta o passo a passo utilizado para adquirir o conhecimento necessário para criar as amostras e o suporte utilizado para armazenar as amostras empregadas na avaliação experimental.

Figura 42: Processo para preparação da amostra.

Como apresentado na Figura 42, os discos rígidos devem ser abertos para se ter acesso ao seu interior, devem ser removidos os componentes ao redor do disco e, em seguida deve ser removido o prato onde estão gravados os dados. O disco rígido utilizado nesta figura não foi utilizado como amostra; este disco foi utilizado na fase de aprendizagem para criação das amostras. Para realizar o corte no disco rígido é utilizada uma serra com diamante para que o corte seja mais preciso e menos danoso ao prato do disco rígido. O prato do disco rígido deve ser preso em um local onde ele tenha estabilidade para evitar sua movimentação, desta forma o corte é mais preciso e é evitado que outras regiões do prato possam ser danificadas no processo do corte.

Após este processo o disco rígido passa a ser inutilizado, pois o corte do prato impede que ele seja utilizado novamente.

É recomendado que a sala utilizada para este processo de criação das amostras esteja o mais limpa possível para evitar contaminação da amostra por elementos estranhos dispersos no ar que possam prejudicar a leitura da amostra por meio da técnica de MFM no microscópio, por isso é recomendado o uso de "Salas Limpas". Salas Limpas são locais onde tecnicamente o ambiente é controlado, evitando que elementos estranhos estejam presentes no ar, sendo utilizadas para testes ou manufatura de produtos.