Um Modelo de Triagem de Dados Digitais Aplicado à Perícia Criminal em Informática

Mateus de Castro Polastro[1] e Pedro Monteiro da Silva Eleuterio[2]

[1] polastro.mcp@dpf.gov.br [2] pedro.pmse@dpf.gov.br

Setor Técnico-Científico em Mato Grosso do Sul - Polícia Federal do Brasil

Resumo. A difusão do uso dos computadores nas práticas criminosas, associada ao vertiginoso crescimento da capacidade de armazenamento de dados dos dispositivos de armazenamento computacional, vem sobrecarregando as unidades de perícia criminal da área de Informática. Assim, a utilização de estratégias para processamento e triagem desses dados antes deles serem encaminhados à perícia criminal se faz necessária. Neste trabalho, os autores propõem um modelo de Triagem de Dados Digitais a ser aplicado aos dispositivos de armazenamento apreendidos em médias e grandes operações policiais. A triagem é realizada pelos investigadores utilizando os meios disponibilizados pela equipe pericial e, após a análise, somente os itens considerados relevantes pela equipe de investigação são encaminhados para a perícia realizar os devidos exames. Em um estudo de caso apresentado neste trabalho, o método proposto possibilitou a redução em 85% do número de itens e em 80% no volume de dados a analisar pelos peritos criminais, trazendo maior celeridade e melhores resultados aos casos. Dessa forma, o uso de um sistema de triagem provou ser eficiente e pode ser aplicado em casos similares para reduzir o volume de dados digitais a serem periciados e, consequentemente, reduzindo o tempo gasto pelos peritos criminais de computação forense.

Palavras-chaves: investigação, operação policial, triagem de dados, computação forense, exame pericial em informática.

1 Introdução

O uso dos computadores em empresas, residências e indústrias é, há muitos anos, muito comum no Brasil e no mundo. Atualmente, estima-se que, no Brasil, 49% domicílios possuem computadores (aproximadamente 30,6 milhões de domicílios) [1], número muito superior aos 25% dos domicílios no ano de 2005. Aumento significativo também pode ser observado nos domicílios com acesso à Internet, que atualmente está estimado em 27,2 milhões de domicílios [1], equivalente a 85,9 milhões de usuários somente no Brasil.

No dia-a-dia das pessoas é natural o uso da tecnologia da informação e comunicação (TIC), pois ela traz muitos benefícios para seus usuários. No entanto, é cada vez mais comum os usuários que utilizam esse aparato tecnológico para cometer atos ilícitos. A

adfa, p. 1, 2011.

utilização de equipamentos computacionais (computadores de mesa, *notebooks*, *tablets*, *smartphones*, entre outros) para a prática de crimes pode ser classificada de duas formas [2]: como ferramenta de apoio aos crimes convencionais e como meio para cometimento de crimes. No primeiro caso, o equipamento computacional auxilia as pessoas na prática de crimes que geralmente independem do uso de equipamentos computacionais, tais como tráfico de entorpecentes, compra de votos em eleições e sonegação fiscal. Na segunda modalidade de crimes, existem aqueles onde o equipamento computacional é essencial para o seu cometimento e que, sem ele, o crime não poderia ser praticado, como, por exemplo, os crimes de ataque a sítios de Internet, compartilhamento de arquivos de pornografia infantil pela Internet, difusão de programas maliciosos para roubo de senhas, entre outros [2].

Associado ao crescente uso de computadores para a prática dos mais variados tipos de crimes, vem ocorrendo há anos o crescimento da capacidade de armazenamento de dados desses dispositivos e atualmente é possível encontrar discos rígidos com capacidade de armazenamento de 3 TB (terabytes) de dados por preços bastante acessíveis. Em 1980 o preço por gigabyte de dados de armazenamento em discos rígidos era de US$ 193.000,00 enquanto atualmente está em torno de US$ 0,03 [3].

Como consequência desses fatores, a necessidade de análise de equipamentos computacionais por peritos criminais para encontrar evidências dos crimes cometidos por seus usuários é cada vez maior. Tal fator tem levado as unidades de perícia criminal a ter um acúmulo de trabalho muito grande, mesmo aumentando seus orçamentos [4]. Para ilustrar tal cenário, o gráfico exibido na Fig. 1 mostra a evolução do volume de dados processados pelos laboratórios de computação forense do FBI (*Federal Bureau of Investigation*), dos Estados Unidos da América, onde de 2003 até o ano de 2013, saltou de 82 TB para 5.973 TB [5], um aumento de 7.300%.

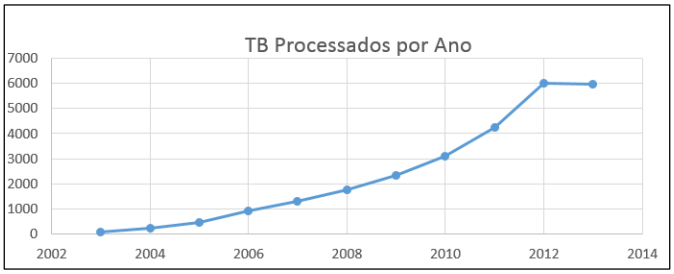

Fig. 1 – Dados processados (em terabytes) entre os anos de 2003 e 2013 pelos laboratórios de computação forense do FBI.

Outro problema enfrentado pelos órgãos públicos responsáveis por examinar tais evidências digitais é o custo. Com orçamento limitado, os gastos para processar os dados devem ser cuidadosamente empregados, a fim de conseguir minimizar o problema da crescente demanda juntamente com as prioridades inerentes dos crimes investigados. E essa equação geralmente resulta em aumento da carga de trabalho, deixando os órgãos responsáveis com maior atraso na entrega dos laudos periciais de equipamentos computacionais [4]. Há a necessidade de implantar mecanismos que possibilitem analisar:

(i) mais rapidamente o que não necessita de análise aprofundada, e; (ii) de forma completa o que pode conter informações relevantes relacionadas aos crimes investigados.

Assim, uma triagem dos dados poderia ser realizada pela equipe de investigação antes dos materiais de informática apreendidos serem enviados à perícia criminal, a fim de otimizar os escassos recursos disponíveis. Seria um paralelo com que se faz na área médica, onde a triagem é comumente utilizada para identificar a necessidade e a prioridade de atendimento do paciente pelo médico [6].

Este artigo propõe um modelo de Triagem de Dados Digitais (TDD), a ser realizado pela equipe de investigação, que detém o maior conhecimento sobre o crime investigado. O objetivo do modelo proposto é evitar que equipamentos computacionais irrelevantes sejam periciados e que os relevantes sejam analisados de forma minuciosa, evitando, assim, que as unidades de processamento dessas evidências sejam sobrecarregadas de trabalho e não consigam dar resposta à população e à justiça em tempo hábil. Sugere-se a aplicação deste modelo em médias e grandes operações policiais, onde grande quantidade de equipamentos computacionais é apreendida.

O artigo está organizado da seguinte maneira: a Seção 2 apresenta alguns conceitos essenciais para o entendimento do fluxo de trabalho de peritos criminais que processam evidências digitais, a forma e os atores envolvidos na requisição de laudos periciais, bem como aspectos relacionados às operações policiais brasileiras e aos materiais por elas apreendidos. A Seção 3 apresenta o modelo de Triagem de Dados Digitais proposto e a Seção 4 mostra o resultado da aplicação do modelo em casos reais da Polícia Federal do Brasil, além de detalhar o parque tecnológico utilizado. Por fim, a Seção 5 apresenta a conclusão do trabalho.

2 Conceitos

Nesta seção serão descritos alguns aspectos relacionados a operações policiais em geral e, particularmente, na Polícia Federal do Brasil. Também serão revisados alguns critérios de aceitação e priorização de equipamentos computacionais para serem submetidos a exames periciais de computação forense. Além disso, serão apresentados alguns conceitos básicos relacionados ao modelo tradicional de trabalho em exames de dispositivos de armazenamento computacional, seguido da forma como tais exames periciais são solicitados e elaborados no âmbito criminal brasileiro.

2.1 Operações Policiais

Somente no ano de 2014 a Polícia Federal do Brasil realizou 336 operações policiais, resultando em 2.353 prisões [7]. As operações policiais são uma força-tarefa que acontece após um período de investigação (geralmente longo) e que visa a levantar mais provas dos crimes investigados com a apreensão de materiais, bem como efetuar a prisão das pessoas responsáveis. É comum em grandes operações a existência de dezenas de alvos, ou seja, dezenas de lugares (residências, empresas, órgãos públicos, entre outros) em que equipes de policiais comparecem para cumprir os mandados de busca e apreensão e, em alguns casos, até mesmo de prisão.

Embora o cumprimento desses mandados seja executado simultaneamente pelas forças policiais, eles não necessariamente estão interligados, ou seja, os alvos das operações policiais podem não ter conexão entre si. Um exemplo onde os alvos não estão interligados é em algumas operações de combate a pedofilia em que os suspeitos foram identificados por um monitoramento das atividades na Internet, onde são identificados suspeitos de posse e compartilhamento de arquivos. Nesses casos, tais suspeitos geralmente não se conhecem e a polícia os agrupa em uma determinada operação apenas para fins logísticos e sigilosos. Nesses casos, as provas levantadas durante as investigações e operações são analisadas individualmente, sem qualquer tipo de cruzamento ou vínculo entre elas.

Por outro lado, existem as operações onde os alvos estão interligados. Trata-se do tipo mais comum de operação policial na Polícia Federal do Brasil e que geralmente ocorre em grandes operações de combate à lavagem de dinheiro, desvio de recursos públicos, contrabando, tráfico de drogas, desmatamento, entre outras tipificações criminais. Nesses tipos de operações há extrema necessidade de se correlacionar as informações dos alvos, o que torna a análise simultânea essencial e muitas vezes traz como consequência a descoberta de novas práticas supostamente ilícitas e pessoas suspeitas.

2.2 Critérios de Aceitação e Priorização de Exames em Dispositivos de Armazenamento Computacional

Atualmente, principalmente como consequência de grandes operações policiais, tem-se observado grande quantidade de dispositivos de armazenamento computacionais apreendidos pelas forças policiais. Tal fato ocorre em todos os tipos de crimes, pois, como descrito na Seção 1, a maioria dos crimes utiliza direta ou indiretamente equipamentos computacionais para serem efetivados [2].

No entanto, nem todos os itens apreendidos possuem o mesmo valor para comprovar a prática delituosa. De uma forma geral, os dispositivos de armazenamento computacionais podem ser classificados em uma das três categorias a seguir [4], quanto à relação com o fato investigado:

- Direta: há a clara evidência de que o equipamento possui as provas do crime, como nos casos em que alguma testemunha viu determinada pessoa manipulando um arquivo de interesse no equipamento apreendido;
- Circunstancial: há maiores chances de ter arquivos de interesse no equipamento, como, por exemplo, no computador de um servidor público suspeito de desvio de recursos públicos;
- Nenhuma: o computador não tem relação direta com o crime praticado pela pessoa, como, por exemplo, quando uma pessoa é presa por racismo em uma briga de trânsito e apreende-se o computador dela em busca de outras provas relacionadas ao crime de racismo.

Baseado na experiência dos autores, observa-se que em grandes operações policiais a maioria dos dispositivos de armazenamento computacional apreendidos se enquadra

inicialmente na categoria "Circunstancial", existindo chances do mesmo ter as evidências procuradas. Em geral, trata-se de mandados de busca e apreensão cumpridos em residências ou empresas que possuem pessoas envolvidas no crime onde são apreendidos equipamentos computacionais pessoais e de trabalho dos suspeitos e, na maioria dos casos, não existe certeza de que nesses equipamentos há provas.

2.3 Modelo Tradicional de Exames Periciais em Computação Forense

Diversos autores [2, 8, 9] definem de maneira análoga o modelo tradicional de exames periciais em dispositivos de armazenamento computacional. A divergência entre os autores está na nomenclatura adotada em relação aos nomes e na quantidade das fases existentes em um exame tradicional de computação forense. Independente dos nomes adotados, é possível resumir o modelo tradicional de exames em dispositivos de armazenamento computacional em quatro principais fases, assim descritas (optamos em não adotar nomes das fases neste artigo):

- Fase 1: os dados digitais do dispositivo de armazenamento computacional são duplicados por meio de técnicas forenses. Em seguida, os exames periciais são sempre realizados na(s) cópia(s), garantindo a preservação do material original contra alterações;

- Fase 2: é uma fase automatizada com uso de diversos softwares forenses. Nesta fase, os dados digitais contidos no dispositivo de armazenamento computacional são processados, recuperados, extraídos, categorizados, indexados, bem como são realizados outros tipos de processamentos capazes de recuperar informações previamente apagadas no dispositivo;

- Fase 3: os dados contidos no dispositivo de armazenamento computacional, incluindo os recuperados na fase anterior, são analisados em buscas das evidências digitais solicitadas conforme tipo de investigação. Esta é a fase do exame pericial propriamente dito, onde o perito criminal irá buscar as evidências desejadas, identificando-as e separando-as para a escrita do laudo pericial;

- Fase 4: formaliza o trabalho desenvolvido durante os exames periciais com a elaboração do laudo pericial, que é um documento de texto que geralmente contém a descrição do material, os procedimentos de preservação e garantia da prova realizados, detalha a metodologia dos exames periciais e, principalmente, apresenta os resultados dos exames e suas conclusões.

A Fig. 2 ilustra o modelo tradicional de exames periciais em dispositivos de armazenamento computacional. Na esfera criminal brasileira, geralmente tais exames são realizados exclusivamente por um ou mais peritos criminais, sendo que, a partir da conclusão da Fase 1, as demais fases podem ser realizadas em paralelo, não sendo necessariamente sequenciais.

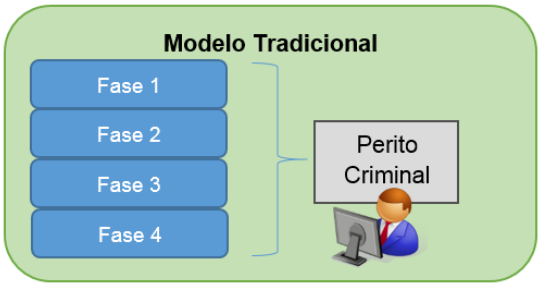

Fig. 2 – Modelo Tradicional de exames periciais em dispositivos de armazenamento computacional.

2.4 Formas e Atores Envolvidos na Requisição de Laudos Periciais no Âmbito Criminal Brasileiro

Para que o modelo proposto neste trabalho seja compreendido, é importante descrever como as requisições de exames periciais são solicitadas na esfera criminal brasileira, bem como quais os agentes envolvidos nesse trâmite.

Resumidamente, em operações policiais brasileiras, diversas equipes formadas por policiais (geralmente não incluem peritos criminais), cumprem mandados de busca e apreensão de equipamentos em casas, empresas, órgãos públicos e em qualquer lugar determinado pela Justiça. Após selecionar os materiais com interesse à investigação, os mesmos são apreendidos e descritos em documento específico, de forma a identificar unicamente o dispositivo, como por exemplo, no caso de discos rígidos, inclui a marca, o modelo e o número de série.

Tais materiais apreendidos fazem parte de um Inquérito Policial, que possui uma equipe de investigação conhecedora dos envolvidos e das suspeitas das atividades supostamente ilícitas praticadas por eles. Dessa forma, os materiais apreendidos são enviados para a Perícia Criminal, a fim de responder quesitos elaborados pelo presidente do Inquérito Policial, com auxílio da equipe de investigação. Após a realização dos exames periciais, os Laudos e seus anexos digitais [2], comuns em laudos periciais de informática, são enviados para a equipe de investigação, apresentando os resultados dos exames. A Fig. 3 ilustra esse processo, que possui apenas uma interação entre Perícia Criminal e equipe de investigação.

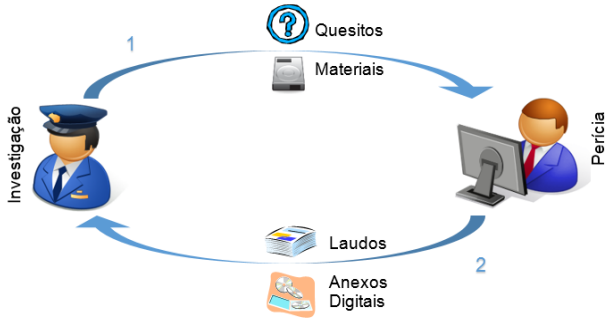

Fig. 3 – Funcionamento do Modelo Tradicional entre a perícia e a equipe de investigação.

Esse sistema tradicional tem funcionado satisfatoriamente em operações pequenas ou em casos isolados, quando o crime praticado é evidente para a equipe de investigação, que pode expressar na forma de quesitos específicos os questionamentos a serem respondidos pelos peritos criminais.

Entretanto, quando as dimensões da operação aumentam, geralmente os materiais de armazenamento computacional são apreendidos para tentar levantar as mais variadas evidências de crimes que, nesse momento, ainda não são tão claras para a equipe de investigação. Consequentemente, uma maior quantidade de dispositivos computacionais acaba sendo apreendida. Além disso, nesses casos, o presidente do Inquérito Policial geralmente é incapaz de sugerir quesitos específicos para a perícia, resultando na generalização dos quesitos, fazendo com que os peritos criminais acabem "investigando" o conteúdo dos dispositivos em busca de ilícitos. Entretanto, como os peritos criminais não têm conhecimento das investigações, esse processo se torna mais demorado, mais custoso e com menos resultados satisfatórios na prática.

3 Modelo Proposto: Triagem de Dados Digitais (TDD)

Esta seção descreve detalhadamente o modelo proposto neste trabalho, chamado de Triagem de Dados Digitais, ou simplesmente TDD, a ser aplicado em médias e/ou grandes operações policiais, que resultem em numerosa apreensão de dispositivos de armazenamento computacional.

O modelo proposto (TDD) surgiu após várias grandes operações realizadas pela Polícia Federal Brasileira entre os anos de 2006 a 2013, que resultou na apreensão de muitos dispositivos de armazenamento computacional que deveriam ser periciados nos moldes descritos na subseção anterior: sem quesitos específicos devido à ausência de clareza nos ilícitos praticados pelos suspeitos. Sendo assim, tais operações resultaram no atraso ao atendimento dessas solicitações de exame pericial, devido ao grande número de materiais a serem examinados, bem como na baixa taxa de resultados efetivos, devido à generalidade dos quesitos propostos e na própria imparcialidade dos peritos criminais, que não detinham o conhecimento das investigações.

Baseado em experiências práticas, o TDD foi proposto para resolver essas questões da seguinte forma: implantar um sistema de triagem dos materiais de informática apreendidos a ser realizado pela equipe de investigação, mas coordenada pelos peritos criminais que garantem, a partir da aplicação de técnicas forenses, a preservação do material apreendido. Uma vez realizada a triagem pela equipe de investigação, os ilícitos cometidos estarão mais claros e, consequentemente, os dispositivos de armazenamento computacional que possuírem indícios das provas dos ilícitos averiguados são então encaminhados para a perícia com quesitos específicos, servindo de entrada ao modelo tradicional, conforme modelo do TDD ilustrado na Fig. 4. O modelo proposto possui quatro fases, denominadas Cópia, Processamento, Análise Investigativa e Documentação, conforme detalhamento a seguir:

- Cópia: os peritos criminais recebem os materiais apreendidos da operação e copiam os arquivos ativos em um servidor de armazenamento com grande capacidade e redundância (RAID 5). Os arquivos ativos são os arquivos diretamente acessíveis nos

dispositivos, não incluindo arquivos apagados e recuperados, fragmentos de arquivos, entre outros. Os materiais são protegidos contra escrita com uso de técnicas e equipamentos forenses, a fim de manter a cadeia de custódia e a integridade dos dados.

- Processamento: com uso de programas forenses, alguns tipos de arquivos são descompactados e extraídos. Posteriormente, todos esses arquivos (de todos os materiais apreendidos na operação) são indexados e categorizados. Uma ferramenta remota, que permite buscas por palavras-chaves, navegação e visualização de diversos tipos de arquivos, é disponibilizada para a equipe de investigação. Esta fase é também realizada pelos peritos criminais.

- Análise Investigativa: utilizando a ferramenta remota disponibilizada pelos peritos criminais, a equipe de investigação analisa os dados contidos em todos os materiais apreendidos na operação de uma só vez, incluindo a realização de buscas por palavras-chaves, filtros de tipos de arquivos, entre outras funcionalidades. Como detém todo o conhecimento da investigação, os possíveis ilícitos cometidos são mais bem entendidos pela equipe.

- Documentação: a equipe de investigação faz um relatório textual dos materiais analisados, enviando-o ao presidente do Inquérito Policial. Somente os dispositivos de armazenamento computacional que possuírem indícios dos ilícitos investigados são reenviados para a peritos criminais, alimentando o Modelo Tradicional, com a diferença de que agora as atividades ilícitas dos investigados estão mais claras e os quesitos enviados aos peritos criminais são específicos e detalhados.

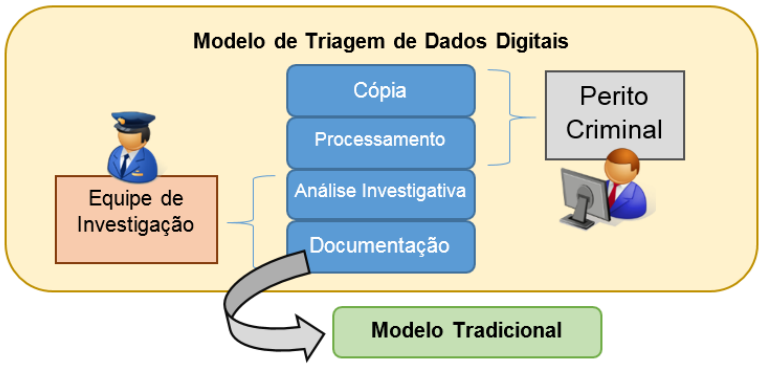

Fig. 4 – Modelo de Triagem de Dados Digitais proposto para grandes operações, que alimenta o Modelo Tradicional, envolvendo perícias em dispositivos de armazenamento computacional.

Nesse sistema, inicialmente somente os materiais são enviados para os peritos criminais, que copiam os dados desses dispositivos no servidor de armazenamento e realizam o processamento dos arquivos, incluindo a indexação, disponibilizando-os de forma categorizada, em ferramenta remota a ser utilizada pela equipe de investigação. Posteriormente, após a triagem, são enviados quesitos específicos somente para os materiais considerados relevantes e que necessitam de exames periciais detalhados. Nesses casos, os exames periciais tradicionais são realizados e os laudos periciais, bem como seus anexos digitais, são enviados para a equipe de investigação. Tal sistema é ilustrado na Fig. 5, que acrescenta uma interação entre perícia e equipe de investigação, quando

comparado ao modelo ilustrado na Fig.3. Os dispositivos computacionais que retornam a Criminalística são examinados no Modelo Tradicional, segundo as técnicas forenses vigentes, incluindo a recuperação de arquivos apagados, quebra de senhas e outras atividades técnicas complexas, que não são realizadas durante a Triagem.

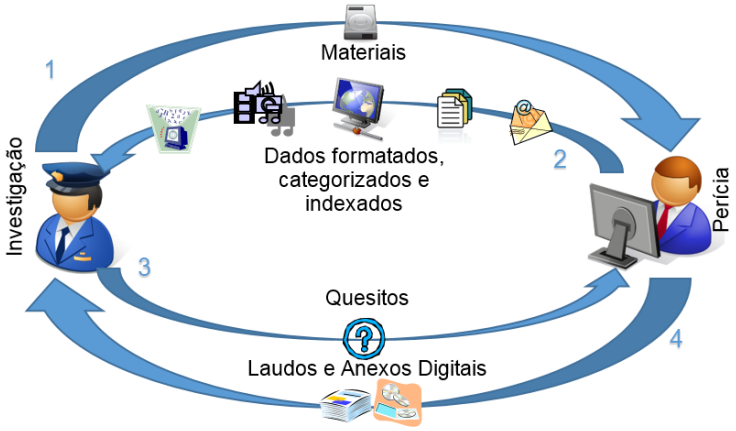

Figura 5 – Funcionamento do Modelo de Triagem de Dados Digitais para médias e grandes operações, aliado ao Modelo Tradicional, entre a perícia e a equipe de investigação.

A Tabela 1 faz um resumo comparativo entre os principais pontos sobre os exames periciais realizados pelo Modelo Tradicional simples e também com a inclusão do TDD. Primeiramente, somente os dispositivos que foram classificados como relevantes pela equipe de investigação, ou seja, que possuem algum indício do(s) crime(s) investigado(s), serão submetidos a exame pericial, diferentemente de todos os dispositivos apreendidos no Modelo Tradicional. Tal fato permite a otimização do tempo de trabalho do perito criminal, que é um recurso escasso no sistema criminal brasileiro, uma vez que somente irá periciar materiais considerados relevantes à investigação, ou seja, das categorias “Direto” ou “Circunstancial” da classificação descrita na Seção 2.

Outro ponto importante é a qualidade dos quesitos elaborados pela equipe de investigação em grandes operações. No Modelo Tradicional, geralmente são propostos quesitos genéricos, sem apontar diretamente quais são as suspeitas dos crimes cometidos, pois essa equipe não tem tamanha clareza nesse momento. Já com a triagem, a equipe de investigação entenderá a relação dos materiais apreendidos com os suspeitos, bem como as possíveis práticas ilícitas, de forma a elaborar, para cada dispositivo de armazenamento computacional, quesitos específicos e relevantes que irão dar suporte técnico e material às investigações.

Uma das grandes vantagens do sistema proposto diz respeito ao tempo de acesso aos dados contidos nos materiais apreendidos pela equipe de investigação. Enquanto no Modelo Tradicional a equipe de investigação deve aguardar a conclusão dos exames periciais para ter acesso aos dados por meio das mídias anexas aos laudos periciais (e de todos os exames para pode unir os resultados), no Modelo Proposto o acesso aos dados dos materiais apreendidos é muito mais rápido, uma vez que somente as cópias dos arquivos são realizadas juntamente com um breve conjunto de processamentos,

permitindo o acesso dos dados categorizados e indexados, através de ferramenta remota. Isso é fundamental para não "esfriar" a investigação devido à natural demora da realização dos muitos laudos a serem emitidos.

No Modelo Tradicional, geralmente cada dispositivo de armazenamento computacional é examinado pelo perito criminal de forma independente, elaborando um laudo pericial para cada material examinado. Em laudos de informática, é comum a apresentação das evidências encontradas gravadas em mídias ópticas anexas ao laudo [2]. Imaginando uma grande operação com 50 discos rígidos, serão elaborados 50 laudos com seus anexos digitais. Quando a equipe de investigação receber tais laudos e anexos, ela terá que juntar as evidências que estão gravadas em pelo menos cinquenta mídias ópticas. Se houver necessidade de buscar por determinada palavra-chave nas mídias, terá que fazer isso em cada uma dessas mídias, ou seja, repetindo a operação por no mínimo cinquenta vezes. Já no modelo proposto, as buscas realizadas na fase de Triagem são simultâneas em todos os 50 materiais, permitindo unir as evidências e entender o significado de cada material e de cada suspeito na investigação. Assim, nesse modelo, ao verificar os materiais que possuem os indícios de ilícitos, reenviando-os para a Perícia, e recebendo os laudos com seus anexos, o cruzamento desses dados já estará muito facilitado pela fase de Triagem. Um ponto importante que deve ser ressaltado é que, por experiência prática dos autores, a grande maioria dos dispositivos de armazenamento computacional apreendidos em grandes operações não retornam para a Perícia Criminal após a triagem, ou seja, grande parte dos materiais inicialmente classificados [4] nas categorias "Direta" ou "Circunstancial" são rebaixados pela equipe de investigação para a categoria "Nenhuma", evitando-se a realização de exames periciais desnecessários. Esse ponto será detalhado na Seção 4 deste artigo, que apresenta dados de casos reais da aplicação do modelo proposto em uma grande operação policial brasileira.

É importante frisar que a preservação dos materiais de informática apreendidos não é afetada com o novo modelo, pois os procedimentos de preservação e garantia de integridade, como cópia dos dados sem alteração do conteúdo do dispositivo original, cálculos de valores *hash*, entre outros procedimentos recomendados [2], são também realizados em todos os materiais pelos peritos criminais, seja no Modelo Tradicional ou no Modelo TDD proposto. Sendo assim, o novo modelo não permite questionamentos jurídicos sobre a correta preservação do material e sua integridade.

Tabela 1 – Comparação das principais características do Modelo Tradicional e do Modelo de Triagem de Dados Digitais proposto.

#	Descrição	Modelo Tradicional simples	Triagem de Dados Digitais com Modelo Tradicional
1	Quantidade de materiais a serem examinados pela perícia criminal	Todos apreendidos na operação	Somente os que contém indícios de ilícitos levantados pela Triagem, que é realizada pela equipe de investigação
2	Tipo dos quesitos recebidos pela perícia criminal	Genéricos, solicitando "investigação"	Específicos, focando no crime a ser provado

#	Descrição	Modelo Tradicional simples	Triagem de Dados Digitais com Modelo Tradicional
3	Tempo de acesso aos dados dos materiais apreendidos pela equipe de investigação	Demorado, pois deve aguardar a elaboração dos laudos periciais (todos para acesso completo do que foi apreendido)	Rápido. Aguarda apenas a cópia e processamento dos dados no servidor de armazenamento e disponibilização da ferramenta de acesso remoto
4	Resultado efetivo dos laudos periciais	Médio, pois o perito criminal não detém o conhecimento das investigações prévias	Ótimo, pois o perito criminal responde aos quesitos específicos, identificando as provas existentes
5	Uso do tempo do perito criminal	Irracional, pois irá fazer exames periciais completos em diversos materiais que não contêm itens relevantes	Racional, só irá examinar materiais relevantes, produzindo resultados muito mais importantes e efetivos
6	Análise e uso do laudo pericial e seus anexos pela equipe de investigação	Demorado. Cada um dos laudos periciais elaborados, incluindo seus anexos digitais, devem ser unidos e pesquisados pela equipe de investigação, a fim de conhecer o conjunto de ilícitos apontados em cada um dos materiais	Mais rápido. As buscas realizadas em todos os materiais durante a triagem já mapearam os ilícitos e a ligação entre os materiais.
7	Preservação e garantia da integridade dos materiais	Sim, o processo é realizado pelos peritos durante os exames periciais	Sim, o processo é realizado pelos peritos durante a cópia dos dados para a Triagem e também durante os exames periciais posteriores

4 Aplicação e Validação do Modelo Proposto (TDD) em Grandes Operações Policiais

Esta seção mostra a aplicação prática do modelo proposto de Triagem de Dados Digitais, aplicado em casos reais de grandes operações policiais da Polícia Federal Brasileira. Primeiramente, são detalhadas as soluções técnicas, incluindo o parque tecnológico utilizado e alguns detalhes técnicos do processo de extração, categorização e indexação. Em seguida, um estudo de caso de uma grande operação policial brasileira é exibido, onde são detalhados os números, quantidade de dados analisados e os ganhos conseguidos com a aplicação do modelo proposto, validando-o. Finalmente, são exibidos alguns dados de novas operações que foram inseridas no TDD e ainda aguardam por resultados futuros.

4.1 Parque Tecnológico e Ferramentas Computacionais utilizadas no TDD

Para a implementação do TDD, foi utilizado um servidor computacional, com alta capacidade de armazenamento, composto de 6 discos rígidos com capacidade de 3 TB cada, agrupados em RAID 5, totalizando 15 TB de espaço útil com redundância, além de um disco rígido independente para o sistema operacional e ferramentas utilizadas. Tal servidor está conectado na rede interna do Departamento de Polícia Federal em Mato Grosso do Sul, unidade onde os autores desempenham seus papéis de peritos criminais federais. O servidor computacional possui políticas de segurança de acesso aos dados, incluindo controle de usuários e auditoria das ações realizadas por eles. Isso é

necessário, pois para cada caso inserido no TDD, a equipe de investigação relacionada ao caso terá os privilégios necessários para acessar o conteúdo do caso de forma remota pela rede interna.

Para a cópia dos arquivos ativos dos materiais apreendidos para dentro do servidor de armazenamento do TDD, são utilizados equipamentos com bloqueio de escrita, o que garante a preservação e a integridade do material questionado (Fase de Cópia do TDD).

Em seguida, para a realização da Fase de Processamento do TDD, é realizado um pré-processamento dos dados com o objetivo de expandir alguns tipos de arquivos que funcionam como container de outros, tais como PST, DBX, EML, MBOX, ZIP e RAR. Utilizando a ferramenta "Apache Tika", a parte textual dos arquivos suportados é extraída. Para os não suportados (áudio e imagens, por exemplo), são extraídos apenas os metadados. Esse conteúdo textual extraído é indexado por uma ferramenta desenvolvida internamente, baseada na "Apache Lucene", que é uma biblioteca para busca de textos escrita em Java.

A equipe de investigação passa a ter acesso a uma interface Web que permite a realização de diversas ações sobre os dados indexados, tais como busca por palavras-chaves e expressões regulares, pré-visualização de arquivos, seleção e exportação de arquivos suspeitos, entre diversas outras funcionalidades (Fase de Análise Investigativa do TDD).

Essa interface permite aos investigadores acesso rápido e preciso aos dados armazenados nas mídias apreendidas, podendo, dessa forma, identificar em quais delas realmente estavam os dados de interesse, permitindo a identificação dos materiais relevantes, descrito em relatórios (Fase de Documentação do TDD). Uma vez identificados os materiais relevantes, os mesmos são reenviados para serem examinados pela perícia Criminal, alimentando o Modelo Tradicional de exames periciais em dispositivos de armazenamento computacional.

4.2 Estudo de Caso de uma Grande Operação Policial Brasileira

No ano de 2013, a Polícia Federal do Brasil realizou uma grande operação para combater uma quadrilha que desviava dinheiro de programas de saúde. Nessa operação, diversos equipamentos computacionais foram apreendidos em empresas prestadoras de serviço, escritórios de contabilidade e um hospital filantrópico. No total foram apreendidos 91 itens que possuíam alguma forma de armazenamento de dados digitais, tais como discos rígidos internos e externos, *pen drives*, cartões de memória, *tablets* e mídias ópticas, que juntos continham aproximadamente 29 TB de capacidade total de armazenamento.

O processamento, análise e elaboração de laudo de todos esses itens iria demandar uma enorme quantidade de trabalho dos peritos criminais. Dessa forma, foi aplicado o Modelo de TDD proposto, em parceria com a equipe de investigação, para agilizar a forma com que a análise desses equipamentos seria realizada. As quatro fases (Cópia, Processamento, Análise Investigativa e Documentação) foram realizadas com os equipamentos apreendidos nesta operação, resultando em aproximadamente 6 TB de arquivos ativos copiados para o servidor de armazenamento da TDD. Essa diferença entre

os 29 TB de capacidade total para 6 TB de espaço utilizado no servidor é facilmente entendida, pois a maioria dos discos rígidos, por exemplo, não estavam completamente cheios de informações. Afinal, na fase de triagem somente o que está diretamente acessível no disco é copiado para o servidor de armazenamento, não incluindo arquivos apagados, fragmentos de arquivos, entre outras peças forenses complexas que são analisadas durante a aplicação do Modelo Tradicional, no caso, somente nos materiais relevantes que retornarem à Perícia.

Ao final da análise da equipe de investigação, os peritos criminais receberam para efetiva realização de perícia apenas 13 materiais, que possuíam uma capacidade total de 5,7 GB de dados. Assim, o modelo de triagem descrito possibilitou a redução em aproximadamente 85% do número de materiais analisados e em 80% no volume de dados a analisar pela perícia criminal.

Ademais, com essa análise prévia, a equipe de investigação pôde direcionar quesitos específicos aos peritos criminais para cada material de forma individual, diminuindo o tempo de exames nos materiais reencaminhados. Dessa forma, o uso de um sistema de triagem provou ser eficiente e pode ser aplicado em casos similares para reduzir o volume de dados digitais a serem analisados e, consequentemente, reduzindo o tempo gasto pelos peritos criminais na realização de exames em computação forense.

A **Tabela 2** mostra uma comparação dos números obtidos no estudo de caso realizado, permitindo observar diretamente diversas vantagens da aplicação da Modelo Proposto em detrimento ao Modelo Tradicional simples. Pode-se observar que o tempo gasto por um perito criminal para a realização dos exames de uma grande operação com o novo modelo proposto foi de 11 semanas no total, sendo 3 semanas de preparação da triagem e 8 semanas de exames periciais no material reenviado à perícia. Os autores estimam que, no Modelo Tradicional, seriam gastas 80 semanas para um perito criminal realizar os exames periciais. Sendo assim, o tempo gasto pelo perito criminal no Modelo de TDD foi de apenas 13,75% em comparação ao Modelo Tradicional.

Tabela 2 – Números do estudo de caso realizado em uma grande operação da Polícia Federal do Brasil, comparando-se o Modelo Tradicional e ao Modelo Proposto de TDD.

#	Descrição	Modelo Tradicional simples	Triagem de Dados Digitais com Modelo Tradicional
1	Número de materiais a serem examinados pela Perícia Criminal	91 itens, sendo 67 discos rígidos	13 discos rígidos
2	Capacidade total do material a ser examinado pela perícia criminal	29 TB	5,7 TB
3	Tempo estimado para a realização dos exames periciais, considerando a atuação de um único perito criminal	80 semanas (estimativa de acordo com a experiência dos autores em casos similares)	8 semanas (tempo gasto na realidade, pois os quesitos específicos auxiliaram na redução do tempo dos exames periciais)
4	Tempo gasto pela perícia para a realização das Fases de Cópia e de Processamento do modelo de TDD	-	3 semanas
5	Tempo gasto pela equipe de investigação	18 semanas (estimativa de acordo com a própria equipe de investigação)	18 semanas

Em consulta à equipe de investigação, responsável pela triagem, foi informado que a rapidez na disponibilização dos dados fez com que a análise investigativa fosse mais

efetiva, pois não houve o "esfriamento" das investigações. Além disso, a realização de buscas em todos os materiais de uma só vez foi muito elogiada, permitindo maior abrangência nas pesquisas e maior facilidade na interligação dos suspeitos e dos indícios. Sobre o tempo da análise investigativa, a equipe de investigação gastou cerca de 18 semanas e salientou que demoraria o mesmo tempo para analisar todos os 91 laudos que iriam ser emitidos no Modelo Tradicional. Portanto, a implantação do novo modelo não aumentou o gasto de tempo da equipe de investigação e, pelo contrário, a ferramenta Web disponibilizada forneceu facilidades inéditas para o melhor entendimento das investigações. Além disso, o presidente do Inquérito Policial deste caso relatou que a aplicação do novo modelo não trouxe nenhum prejuízo à investigação no tocante aos excelentes resultados obtidos.

4.3 Outras Operações Brasileiras Inseridas no Modelo Proposto

Uma vez que a aplicação do TDD em caso real foi um sucesso, tanto para a perícia, quanto para o presidente do Inquérito Policial, novas operações foram inseridas nesse modelo.

Atualmente (março/2015), o servidor de dados utilizado pelo Modelo TDD no estado de Mato Grosso do Sul conta com duas operações de porte médio em fase de análise da equipe de investigação. Pode-se observar que, mantida a mesma proporção de ganho de tempo dos peritos criminais, o Modelo de TDD reduzirá em cerca de 32 semanas o trabalho que um perito criminal gastaria na elaboração dos laudos periciais relacionados a essas duas operações.

Tabela 3 – Números de outras operações atualmente inseridas no Modelo de TDD.

#	Descrição	Número de materiais	Capacidade Total dos materiais	Espaço utilizado no Servidor TDD
1	Operação A	27 itens, sendo 19 discos rígidos	21 TB	5,3 TB
2	Operação B	10 itens, sendo 7 discos rígidos	3,1 TB	320 GB

5 Conclusão

Este artigo propõe um modelo de Triagem de Dados Digitais aplicado à perícia criminal na área de Informática. O modelo introduz um procedimento de triagem definido em quatro fases, com a participação conjunta entre perícia criminal e equipe de investigação, antes do modelo tradicional de exames periciais em dispositivos de armazenamento computacional. Com o novo modelo proposto, somente dispositivos computacionais relevantes, definidos pela triagem, são objeto de exames periciais.

O modelo proposto foi aplicado em um caso real, que consistia em uma grande operação da Polícia Federal do Brasil que apreendeu quase uma centena de dispositivos de armazenamento computacional. Entre as vantagens observadas do modelo proposto, pode-se destacar que a equipe de investigação teve acesso mais rapidamente aos dados armazenados nos dispositivos apreendidos. Além disso, a equipe de investigação, que detém o conhecimento do fato investigado, realiza a Análise Investigativa em todo o material apreendido de forma simultânea, trazendo maiores facilidades para o entendimento dos ilícitos e da interligação das evidências. Ademais, quando o material chega

para a perícia após o processo de triagem, os quesitos elaborados são muito mais específicos, tornando o laudo pericial ainda mais robusto como meio de prova material.

Para o sucesso do modelo de triagem proposto esperava-se que a quantidade de material examinado pelos peritos criminais e o tempo gasto nos exames fosse significativamente menor do que àquele que seria gasto usando o modelo tradicional. Isso foi comprovado no caso real em que o modelo proposto foi implantado e validado, existindo redução de aproximadamente 85% no número de materiais e em 80% no volume de dados encaminhados à perícia criminal. Ademais, a aplicação do modelo proposto nesse caso, de acordo com as estimativas realizadas, trouxe uma economia de 69 semanas de trabalho de um perito criminal, ou seja, aproximadamente 1 ano e quatro meses.

Além disso, os autores consultaram a equipe de investigação que participou da análise dos dados da operação onde o Modelo de TDD foi aplicado e constatou que diversos ganhos também foram observados do ponto de vista investigativo, conforme já descritos na Seção 4.

Portanto, o modelo proposto foi capaz de atender às necessidades da perícia criminal e das pessoas envolvidas na investigação, comprovando que ele pode ser aplicado a outros casos similares, principalmente quando houver necessidade de análise de dados digitais contidos em grande quantidade de dispositivos de armazenamento computacional apreendidos em médias e grandes operações policiais.

Agradecimentos

Os autores agradecem à Polícia Federal do Brasil, principalmente a Superintendência Regional no Estado de Mato Grosso do Sul. Também agradecem a todos os Peritos Criminais Federais envolvidos indiretamente de alguma forma neste trabalho, como consultores, autores de scripts e de programas utilizados nesta solução técnica.

Referências

[1] Comitê Gestor da Internet no Brasil. Pesquisa Sobre Uso das Tecnologias de Informação e Comunicação no Brasil - 2013. http://www.cetic.br/media/docs/publicacoes/2/TIC_DOM_EMP_2013_livro_eletronico.pdf. 2014.
[2] P. M. S. Eleutério; M. P. Machado. Desvendando a Computação Forense. Novatev Editora. ISBN 978-85-7522-260-7. 2011.
[3] M. Komorowski. A History of Storage Cost. http://www.mkomo.com/cost-per-gigabyte-update. Acessado em março de 2015.
[4] H. Parsonage. Computer Forensics Case Assessment and Triage - some ideas for discussion. 2009.
[5] RCFL. RCFL Annual Reports FY2003-2013. http://www.rcfl.gov/downloads. 2014. Digital Forensic Research Workshop (DFRWS). 2001.
[6] M. Gielen, D. Bolzoni. Prioritizing Computer Forensics Using Triage Techniques. University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science, Holanda. 2014.
[7] http://www.dpf.gov.br/agencia/estatisticas/2014. Acessado em março de 2015.
[8] Y. Yuso, R. Ismail, and Z. Hassan. Common phases of computer forensics investigation models. International journal of advanced computer science and information technology, 3(3):17–31, 2011.
[9] Digital Forensic Research Workshop (2001) a Road Map for Digital Forensic Research. In Report from the First Digital Forensic Research Workshop (DFRWS), (New York, NY, USA).