ALBERTO LUIZ ALVES VIOTTI

POSSIBILIDADES DE USO DE SOFTWARE LIVRE COMO FERRAMENTAS DE ANÁLISE EM INVESTIGAÇÕES DIGITAIS

Monografia apresentada ao Departamento de Ciência da Computação da Universidade Federal de Lavras, como parte das exigências do curso de Pós- Graduação *Lato Sensu* em Administração de Redes Linux para obtenção do título de especialista em Redes Linux

Orientador
Prof. joaquim Quintero Uchôa

LAVRAS
MINAS GERAIS – BRASIL
2005

ALBERTO LUIZ ALVES VIOTTI

POSSIBILIDADES DE USO DE SOFTWARE LIVRE COMO FERRAMENTAS DE ANÁLISE EM INVESTIGAÇÕES DIGITAIS

Monografia apresentada ao Departamento de Ciência da Computação da Universidade Federal de Lavras, como parte das exigências do curso de Pós- Graduação *Lato Sensu* em Administração de Redes Linux, para obtenção do título de especialista em Redes Linux

APROVADA em 10 de setembro de 2005

Prof. _____

Prof. _____

Prof. Joaquim Quintero Uchôa
(Orientador)

LAVRAS
MINAS GERAIS – BRASIL

DEDICATÓRIA

Dedico esse trabalho a todos que acreditam ser o domínio tecnológico condição essencial para o desenvolvimento de uma nação qualquer.

Espero que este seja um trabalho que ajude a solidificar o modelo de prestação de serviços no mundo da computação, por meio do uso do GNU/Linux como plataforma para realização de investigações digitais.

A todos que dedicam parte de suas vidas à disseminação do conhecimento por meio do desenvolvimento, testes, aperfeiçoamento, documentação e ensino ligados ao software livre.

Optando pela franqueza em relação a meus sentimentos e assumindo o risco de ser rotulado de bajulador, dedico o trabalho ao professor Joaquim Quintero Uchoa por considera-lo um abnegado quando o assunto é promoção do software livre.

AGRADECIMENTOS

A Deus acima de tudo. Obrigado pela vida e tudo que há nela que nos permite crescer como seres divinos. Em Ti coloco toda a minha fé de um mundo fraterno.

A minha família, Sarita e a pequena Sarah. Pilares da minha existência e alegria diária que me conforta. Obrigado por estarem sempre comigo haja o que houver. Sem vocês pouca coisa valeria a pena.

Aos meus avós maternos, Sebastião (in memorian) e Maria das Neves (in memorian), aos meus pais, Luiz e Anna, à minha irmã Luciana, aos meus irmãos André Gustavo e Alexandre, com eles cresci e tenho aprendido a viver. Obrigado também aos meus avós paternos Ary (in memorian) e Aracy (in memorian), que não conheci, mas que também fazem parte dessa inabalável torre de sustentação.

Obrigado ao professor Joaquim Quintero Uchoa pela orientação e pelo empenho na manutenção da qualidade do curso de pós-graduação em Administração de Redes Linux, contribuindo para valorizar os profissionais formados pelo DCC/UFLA em parceria com a FAEPE.

Agradeço também àqueles que por meio do trabalho na área das investigações forenses digitais forneceram subsídios a esse trabalho.

RESUMO

As informações tem um papel estratégico nas organizações contemporâneas. Os processos organizacionais são cada vez mais informatizados e dependentes da tecnologia digital.

Os riscos inerentes as atividades das organizações devem ser gerenciados para permitir a continuidade das operações e a proteção dos ativos que sustentam essas atividades. Para isso, deve-se instituir um programa de gestão da segurança das informações que inclua controles tendentes a minimizar os impactos dos incidentes de segurança.

A forense digital insere-se aí como mecanísmo para identificar causas e responsáveis pelos incidentes de modo a indicar onde deve ser reforçada a segurança e obter evidências para punir os infratores.

O uso de ferramentas de código aberto para investigações digitais é indicado devido as suas características, destacando-se a flexibilidade e a transparência. Flexibilidade por permitir que soluções *ad hoc* sejam construídas sempre que faltem ferramentas para determinada tarefa, transparência devido à disponibilidade do código fonte conferir idoneidade às evidências encontradas e credibilidade às conclusões exaradas a partir delas.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

1 INTRODUÇÃO

Esse trabalho versa sobre o uso de ferramentas de código aberto baseadas no sistema operacional GNU/Linux para apoio à investigações e análises forenses em equipamentos digitais. Inicialmente são apresentados: a motivação que originou o presente estudo e as justificativas para conduzi-lo. A seguir é feita a delimitação do tema, incluindo o enfoque nos procedimentos e técnicas aplicáveis à forense computacional e ao GNU/Linux para obtenção, análise e registro das evidências presentes nos dispositivos digitais. Evidenciam-se então os objetivos visualizados e o problema elaborado para nortear a busca desses objetivos. Por fim, a hipótese relacionada ao problema e o caminho a ser trilhado para avaliação dela são apresentados.

São várias as motivações desse trabalho. A primeira refere-se a oportunidade de estudar com certa profundidade os mecanismos do GNU/Linux e, portanto, propiciar a consolidação do aprendizado. A segunda motivação é a relação do tema com a atividade de auditoria, profissão do autor, pois a aplicação do conteúdo do estudo ao trabalho diário favorece o envolvimento. Em terceiro, surgiu o interesse em contribuir para o aprimoramento da segurança dos sistemas digitais e, em conseqüência, com a segurança das informações. Essa motivação origina-se na percepção do crescimento ininterrupto do uso de sistemas computacionais e do comprometimento deliberado do funcionamento normal deles, com utilização dos mais diversos meios. Por último, visualiza-se a oportunidade de contribuir para o crescimento

do uso do GNU/Linux no segmento da segurança computacional. A seguir são apresentadas as razões que justificam o estudo.

Na história recente da humanidade vêm sendo notadas alterações importantes na sociedade capitalista. Antes o capital era a maior força. A capacidade de gerar riqueza e de conquistar vantagem competitiva daqueles que conseguiam fornecer produtos e serviços a menores preços, pelo investimento do capital disponível, não produz atualmente diferencial tão acentuado. É a informação que está paulatinamente substituindo a posição estratégica do capital e construindo a assim chamada sociedade da informação[1].

Robins (apud Martineli [2]) pesquisou um dos fatores fundamentais para análise do quadro econômico: a evolução da distribuição na composição da mão-de-obra em países industrializados durante o século XX. Como pode ser verificado na Figura 1, houve redução da participação da mão-de-obra do setor primário, representado pelos homens do campo, ao longo de todo o século XX. Ao mesmo tempo, os trabalhadores industriais, que tiveram crescimento explosivo no século XIX com o advento da revolução industrial, registraram participação de 20% do total da mão-de-obra dos países industrializados no início e no final do período analisado, com variação positiva a partir de 1930, impulsionados pela indústria da guerra. Enquanto isso os chamados trabalhadores do conhecimento, considerados aqueles com formação científica ou técnica especializada tem um súbito aumento a partir dos anos 70, registrando participação próxima aos 30% no final do século XX.

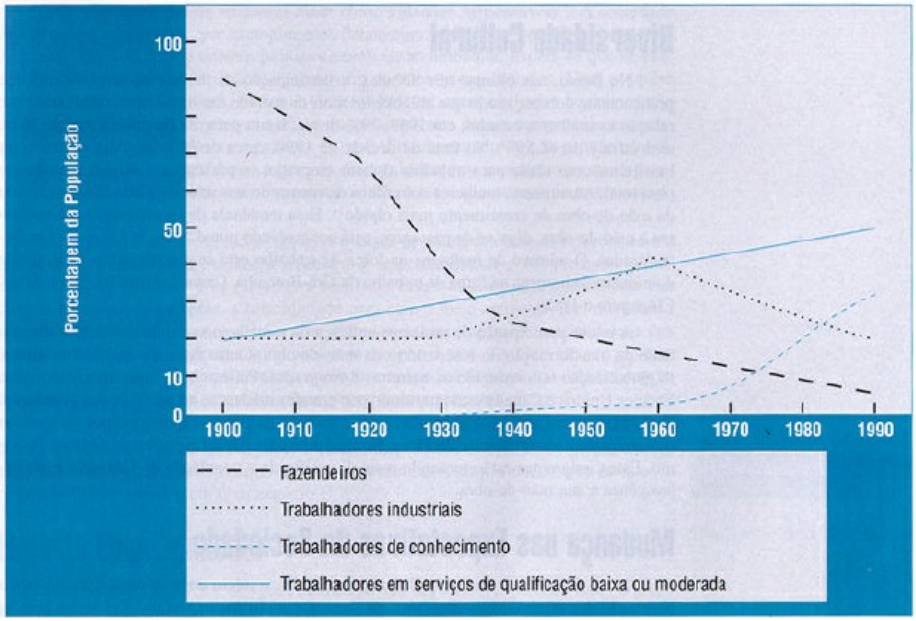

Figura 1 - Mudança na composição da força de trabalho nos países industrializados – extraída de [2]

Apesar da pesquisa ter sido realizada em países industrializados, demonstra uma tendência devida ao fenômeno denominado de globalização. A facilidade de interação entre os povos de todo mundo caracteriza essa nova ordem mundial cujos principais fatores são a unificação dos mercados financeiros e a disponibilidade de informações nos quatro cantos do globo. O desenvolvimento tecnológico viabiliza essa situação, pois basta ter acesso a um computador ou a um celular para interagir em escala global. Fica assim caracterizada a valorização das funções que lidam com maior quantidade de informações e que exigem capacidade analítica para processá-las e para produzir o conhecimento necessário às atividades exigidas no mundo contemporâneo.

Fatores técnicos, políticos e econômicos contribuíram para a construção desse cenário. Do lado político-econômico, um fator de destaque foi a convergência da regulamentação das telecomunicações nos países da comunidade européia, onde há desenvolvimento e uso intensivo das novas tecnologias, favorecendo a expansão do mercado das telecomunicações. Pelo lado técnico está a convergência da informática, da eletrônica e das telecomunicações. A face mais visível dessa transformação no mundo atual é a utilização em larga escala, tanto no âmbito das organizações quanto nos lares, de tecnologias digitais que, às vezes, dão suporte aos processos informacionais. Convive-se com computadores, telefones celulares, *handhelds*, câmeras fotográficas e toda sorte de equipamentos que fazem uso do sistema binário.

Nessa nova ordem, onde há predomínio da informação, o diferencial das organizações é construído sobre os alicerces da competência na gestão do conhecimento e no uso eficiente das informações para a tomada de decisões. A busca dos instrumentos para a gestão eficiente nesse cenário torna-se imprescindível para obter-se resultados satisfatórios. Essa necessidade é a causa do crescimento da dependência dos processos organizacionais em relação aos sistemas informatizados.

No mesmo passo do crescimento da informatização de processos organizacionais estão os incidentes de segurança. O CERT.br[1], compila dados sobre incidentes de segurança na Internet e fornece informações a respeito. Na Figura 2 pode ser visualizado o crescimento dos incidentes de segurança

1 http://www.nbso.nic.br/stats/incidentes/

relacionados a invasões de sistemas computacionais. Vale destacar que os dados de 2005 são parciais e correspondem ao período de janeiro a julho.

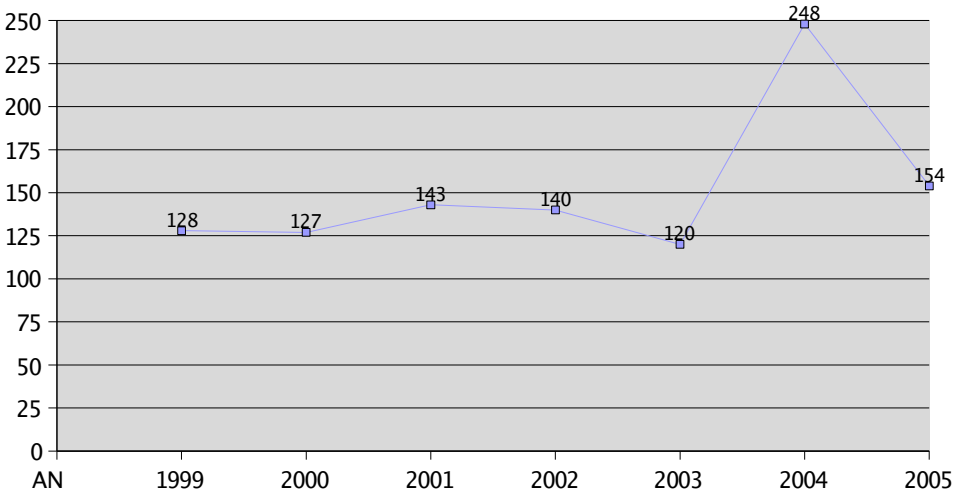

Figura 2 – Evolução das invasões reportadas ao CERT.br

Portanto, o crescente uso das novas tecnologias da informação e comunicação – NTIC para suporte aos processos informacionais nas organizações força a necessidade de desenvolvimento de controles nesses ambientes de forma a evitar o uso indevido dos ativos tecnológicos usados para coletar, armazenar, processar e distribuir informações.

Vários mecanismos têm sido desenvolvidos para propiciar o ajuste do grau da segurança dos sistemas informatizados a um nível adequado para uma realidade organizacional específica. Um desses mecanismos, a forense computacional, se insere nessa busca de aperfeiçoamento contínuo da segurança por meio da avaliação de sistemas computacionais atacados, com a premissa de não alterar qualquer registro e o objetivo de produzir evidências que

possibilitem acionar os infratores. Portanto, a justificativa desse trabalho é a necessidade de estabelecer procedimentos e técnicas para análise forense computacional de forma a utilizá-las em casos concretos para o aperfeiçoamento da segurança, seja pela identificação das vulnerabilidades dos sistemas, seja pela identificação e punição dos responsáveis pela invasão.

Feita a justificativa, importa especificar os aspectos da forense computacional tratados nesse trabalho. Assim, o enfoque é nos procedimentos e técnicas usados na investigação e análise de sistemas computacionais corrompidos, sem entrar em detalhes de arquiteturas específicas ou sistemas operacionais particulares. Procurar-se-á vislumbrar as características das ferramentas adequadas à atividade do investigador e avaliar a pertinência do uso das de código aberto baseadas no GNU/Linux, verificando possíveis restrições.

Outro limite estabelecido para o trabalho é a abrangência das análises forenses, qual seja, os equipamentos digitais que coletam, processam, armazenam e distribuem dados. Portanto, o trabalho não trata de investigações de redes de computadores, mas de computadores individuais que até podem estar conectados a uma rede.

O objetivo ao restringir dessa maneira o estudo é avaliar a adequação das ferramentas baseadas no GNU/Linux para conduzir investigações e análises forenses em sistemas digitais individuais.

Os objetivos específicos são:

- contextualizar a forense computacional nos processos de gestão da segurança das informações e este na gestão das informações;

- descrever os procedimentos e as técnicas da forense computacional estabelecendo um paralelo com o processo de análise forense tradicional;
- apresentar algumas ferramentas livres utilizadas para a análise forense computacional, disponíveis para GNU/Linux, e conhecer as características fundamentais.

O problema associado aos objetivos elencados é o de saber se há efetividade no uso de *software* livre para as análises forenses computacionais. A hipótese que se pretende avaliar, baseado na constatação da existência de distribuições voltadas para o mister, é que o *software* livre atende plenamente às necessidade dos peritos digitais.

No desenvolvimento do trabalho será realizada pesquisa bibliográfica sobre o tema, seguida de compilação das ferramentas livres voltadas para GNU/Linux e utilizadas para forense digital. Por fim, proceder-se-á a análise da adequação das características do *software* livre aos requisitos da investigação e análise forenses.

2 A ANÁLISE FORENSE NO CONTEXTO DA SEGURANÇA DAS INFORMAÇÕES

O objetivo desse capítulo é vislumbrar as atividades de análise forense num contexto mais amplo em que o objetivo é o gerenciamento e a proteção das informações. Procurar-se-á destacar a importância do processo de tratamento de incidentes para a gestão efetiva da segurança das informações referentes a determinado escopo organizacional. Dessa forma, a análise forense é apresentada como um recurso relevante para tratar os incidentes de segurança.

2.1 Gestão informacional

Conforme apresentado na *introdução*, a informação tem sido considerada como o recurso estratégico que, bem gerenciado, confere às organizações o diferencial dos vencedores na acirrada competição vigente no mundo atual. Assim, é preciso haver compreensão da gestão informacional antes de entrar na seara da segurança das informações. E, para isso, convém tratar de outros dois termos: dados e conhecimento.

Dados são atributos de determinados objetos ou fenômenos, características inerentes à própria existência deles. Correspondem a uma anotação direta da observação da realidade. É o que ensina Davenport [1]. Segundo ele, dados são:

> *"'observações sobre o estado do mundo'. Por exemplo: "existem 697 unidades no armazém". A observação desses fatos brutos, ou entidades quantificáveis, pode ser feita por pessoas ou por uma tecnologia apropriada. Da perspectiva do gerenciamento da informação, é fácil capturar, comunicar e armazenar os dados. Nada se*

perde quando representado em bits, o que certamente conforta o pessoal de TI."

Um dado só terá seu estado modificado, transformando-se em informação, quando alguém o puder utilizar para um determinado fim. É o que diz Peter Drucker (apud Davenport [1]) quando afirma que *"informação é o dado revestido de relevância e propósito"*. Importante nessa definição de Drucker é que propósito e relevância são características que só podem ser conferidas por seres humanos. Quando as pessoas, ao buscarem atingir determinado objetivo, se valem de determinados dados, utilizando-os como subsídio para a efetividade de suas ações, estão de fato fazendo uso de informação.

A literatura técnica sobre segurança das informações tem conceitado informação como um ativo, um bem utilizado pelas organizações na consecução dos objetivos estratégicos delas. Essa conceituação está inserida no próprio texto de introdução da norma brasileira sobre segurança das informações, NBR ISO/IEC 17799[1] [3]. *"A informação é um ativo que, como qualquer outro ativo importante para os negócios, tem um valor para a organização e conseqüentemente necessita ser adequadamente protegida"*.

Para Sêmola [4] *"a informação representa a inteligência competitiva dos negócios e é reconhecida como ativo crítico para atividade operacional e*

1 Observe que a norma brasileira NBR ISO/IEC 17799 consiste na tradução da norma ISO/IEC 17799 publicada em 2000. Já a norma internacional originou-se de uma adaptação para os padrões internacionais da primeira parte da norma britânica BS 7799. Em abril de 2005 foi publicada a última versão da norma internacional.

saúde da empresa". Percebe-se nessa definição o reconhecimento de que a informação permeia todos os níveis administrativos, do estratégico ao operacional, e, da mesma forma que a definição dada na norma brasileira mencionada, o foco patrimonial. Essa última característica pode ser resultado das primeiras pesquisas na área serem originadas no meio militar e da apropriação dos conceitos derivados delas para proteção do patrimônio das organizações. A importância de se considerar a informação como um bem consiste em reconhecer a necessidade da proteção segundo o valor avaliado.

Conhecimento é outro termo que ajuda a construir a grade conceitual necessária para tratar o tema da segurança das informações. Trata-se do uso eficaz da informação no desenvolvimento das atividades de determinada organização. Enquanto as informações são a base para resposta de cada porquê, quanto, como, quando, onde, o conhecimento é a reunião de todas essas respostas, é a consolidação da compreensão dos aspectos envolvidos com um tema específico. Assim, pode-se afirmar que conhecimento é a agregação de informações sobre determinada especialidade.

Enquanto a informação é formada a partir da relevância e do propósito que o ser humano confere aos dados, o conhecimento é construído pela experimentação e reunião de informações de origens variadas. Conversando com os outros, observando o que fazem, lendo, refletindo, tocando objetos, enfim vivenciando o ambiente é que as pessoas constroem conhecimento.

Assim, o desenvolvimento do conhecimento será tanto maior quanto maior for a liberdade para experimentar, mesmo que de forma sistemática. Esse

cenário de liberdade é típico de instituições de pesquisa como as universidades, mas cada vez mais buscado por outras organizações, como meio de transformar o conhecimento de seus empregados em inovação que resulte em diferencial competitivo.

A questão que se apresenta é a de como valorizar o conhecimento, incentivando a criatividade por meio da construção de um ambiente mais livre, vale dizer, com poucos controles, sem deixar de aplicar restrições para armazenar, transmitir, utilizar as informações. Surge aí uma tensão, um conflito que deve ser objeto das atenções das organizações quando do gerenciamento da segurança das informações. Escolher a política certa para cada ambiente é fator crítico de sucesso.

As tarefas de gerenciar os dados, a informação e o conhecimento são atribuições das organizações atuais, cada uma com objetivos específicos, cada uma utilizando métodos e recursos próprios. Aqui vale mencionar que os computadores são mecanismos com extraordinária capacidade para manusear dados, armazenando e recuperando cada vez maiores quantidades, a velocidades cada vez maiores. Assim, a tarefa de gerenciar dados tem na administração de recursos tecnológicos um componente essencial.

Por outro lado, o uso da tecnologia no apoio à gestão das informações e na gestão do conhecimento, apesar de necessário, não é suficiente para a efetiva gestão informacional e menos ainda do conhecimento. Davenport [1] apresenta elementos que apoiam a assertiva:

> *"Muitas pesquisas empíricas indicam que os administradores seniores preferem informações que não residem no computador. Vários estudos demonstraram que a informação computadorizada não oferece a variedade, a atualidade ou a relevância que esses executivos exigem. Como resultado, a maioria tem nas informações verbais suas fontes mais importantes. Uma comparação desses estudos sugere que essa preferência não mudou substancialmente desde que a pesquisa sobre o tópico começou, na década de 60".(p. 41)*

Portanto, gerenciar informações não se resume a administração das tecnologias da informação e comunicação. Refere-se a um processo em que são tomadas medidas gerenciais para identificar cada passo, as fontes e pessoas envolvidas, os problemas que surgem e as soluções aplicadas. Novamente Davenport [1] acrescenta elementos:

> *"Os administradores tendem a obter de fontes humanas dois terços da informação que usam. A maior parte dessa informação provém de contatos pessoais; o restante, de conversas telefônicas. No outro terço encontra-se a informação estruturada, que em grande parte vem de documentos sobre o ambiente externo, de pesquisas de mercado a revistas do setor industrial e o Wall Street Journal. Sempre que pergunto a meus clientes ou gerentes de pesquisas se obtêm as informações que necessitam no computador, quase todos dizem que não. Uma avaliação recente sobre planejamento e estratégias de administração também descobriu que 'um ceticismo substancial era expresso pelos entrevistados quando se lhes perguntava se os problemas informacionais da empresa podiam ser resolvidos por melhores sistemas de computadores'".(p. 41)*

O alerta é para a tendência de alguns em tratar a informação como se dado fosse, ou valorizar apenas aquelas informações que podem ser obtidas, ou melhor, formadas a partir de registros de sistemas informatizados. O resultado desse tratamento pouco rigoroso é a valorização do uso de recursos tecnológicos, eficazes no trato de dados, em detrimento dos procedimentos gerenciais, mais indicados para o gerenciamento das informações.

2.2 Gestão da segurança das informações

O gerenciamento da segurança da informação é parte do processo organizacional de gestão das informações e, portanto, envolve não apenas a administração de recursos tecnológicos, mas também dos processos e das pessoas envolvidas neles.

Não obstante, gerenciar a segurança das informações nos dias atuais requer que sejam administrados também a infraestrutura tecnológica, pois a utilização dela em todos os processos de negócio das organizações é cada vez maior. Krause (apud Oliveira [5]) assim apresenta a questão:

> *"A segurança da informação dedica-se, primordialmente, com a segurança dos processos de registro, tráfego, processamento, armazenamento e recuperação da informação, independentemente do suporte no qual ela se encontra. Em vista da modernidade tecnológica, com a capacidade de armazenamento da informação em discos flexíveis ou locais em máquinas do tipo PC como em grandes servidores de arquivos, o suporte priorizado como alvo da segurança na atualidade veio a ser o digital, eletrônico, o qual necessita de tecnologias também*

> *modernas e atualizadas que garantam a segurança da informação neles contida".*

A segurança das informações trata de garantir que a informação, esse bem produzido pelas pessoas e cada vez mais sendo obtido, processado, transmitido e armazenado nos equipamentos digitais, do qual depende cada vez mais a sociedade, seja confiável.

Para que seja possível confiar numa determinada informação é preciso verificar alguns de seus atributos. Quando uma informação, mantida por determinado sistema é necessária e o acesso a ela no momento e local oportunos não é obtido, perde-se a confiança de consegui-lo em outra oportunidade. Quando determinada informação, armazenada por alguém ou transmitida em um meio específico, não é recuperada ou recebida como havia sido produzida originalmente, também há prejuízo para a confiança. Por fim, quando é esperado que determinada informação só seja acessível a pessoas específicas e outras pessoas acabam tendo acesso a elas, a confiança também fica abalada.

Os aspectos apresentados, a disponibilidade, a integridade e a confiabilidade da informação são por muitos apresentados como aqueles atributos que permitem qualificar quão segura é uma informação. O nível de proteção a ser conferido a determinada informação é então definido de acordo com a necessidade que se tem do período de tempo e local para mantê-la acessível, apenas pelas pessoas credenciadas e como fora produzida ou transmitida.

Também existem aqueles que consideram outros atributos, como a legalidade, a autenticidade e o não-repúdio, na qualificação da segurança das informações. Não obstante esses entendimentos, nesse trabalho serão considerados apenas a confidencialidade, a integridade e a disponibilidade como atributos que caracterizam a segurança. Essa escolha é feita considerando que o principal quadro de referência teórico adotado para avaliar a gestão da segurança das informações, a norma NBR/ISO 17799 [3], apresenta esse conceito.

Caracterizada pela disponibilidade, confidencialidade e integridade, a segurança da informação determina não só a qualidade das decisões, mas também a própria continuidade das atividades. Com o crescimento da importância das informações para os processos de negócio torna-se necessário o desenvolvimento de controles que proporcionem a proteção adequada às informações.

Mas, instituir controles tem um custo, interfere na produtividade e, com freqüência, exige gastos com treinamentos e equipamentos. Para saber quais controles implementar, sem prejuízo para a atividade-fim e sem despender recursos maiores do que o valor da informação, deve-se proceder a uma análise que compare os custos e benefícios associados. Isso faz parte do processo de gestão dos riscos.

Krause [6] afirma que gerenciar os riscos às informações é uma tarefa dinâmica e de complexidade crescente. Desde o início da era da informação, a tecnologia de armazenamento, processamento, acesso e transferência das

informações evoluiu consideravelmente fazendo com que os esforços para proteger efetivamente a informação fossem continuamente exigidos. Portanto, para que os riscos associados à informação e à tecnologia da informação sejam identificados e gerenciados eficientemente, é essencial que, além do processo de análise e mensuração dos riscos seja bem compreendido por todas as partes, ele seja executado periodicamente.

Essa dinâmica de contínua análise e mensuração dos riscos, exigência da complexidade crescente e mudança constante no ambiente tecnológico e de negócios, é a essência do Sistema de Gerenciamento da Segurança das Informações – SGSI proposto pela BS7799 – parte 2 [7], norma inglesa sobre gerenciamento da segurança das informações.

Conforme explicitado na BS7799 – parte 2 [7], no item referente ao escopo, a norma especifica requisitos para definição, implementação e manutenção de um SGSI documentado dentro do contexto dos riscos globais de negócio de uma organização. Ela especifica os requisitos para implementação de controles de segurança, personalizados de acordo com as necessidades de uma organização em particular ou de parte dela.

A última revisão da BS7799 – parte 2 publicada em 2002 é o resultado da adequação da norma para compatibilizá-la com outros sistemas de gestão, a exemplo das normas da série 9000 e 14000. Nesse processo foi introduzido o modelo PDCA (do inglês Plan-Do-Check-Act) como parte da abordagem do sistema de gerenciamento para planejar, implementar e aperfeiçoar a efetividade de um sistema de gerenciamento da segurança das informações.

Conforme explicado no anexo B da BS7799 – parte 2 [7], estabelecer e administrar um Sistema de Gerenciamento de Segurança das Informações requer a mesma abordagem necessária em outros sistemas de gestão. O modelo de processo descrito segue um ciclo de atividades contínuo: PLANEJAR, DESENVOLVER, CHECAR e AGIR. Esse ciclo de atividades pode ser descrito como um círculo virtuoso porque o propósito dele é garantir que as melhores práticas da organização sejam documentadas, incentivadas e aprimoradas com o tempo.

Essa última característica é importante para mostrar que a segurança das informações não é um projeto, com início, meio e fim, mas sim um programa da organização. Por isso, o desejo de administradores em conferir segurança às informações deve ser acompanhado de disposição para fornecer apoio, tempo, dinheiro e pessoas que produzam os resultados esperados.

2.3 Administração de incidentes

O processo mais abrangente de gestão da segurança inclui as tarefas relacionadas à reação aos incidentes de segurança. Após planejar e implementar mecanismos para controle dos riscos às informações, o ambiente deve ser permanentemente monitorado e ações tomadas para sanar as deficiências identificadas. Os incidentes de segurança constituem uma forma de apontar essas deficiências.

Como o processo de implantação dos controles é baseado na avaliação dos custos e benefícios relacionados aos riscos existentes, pode ocorrer de não

haver justificativa para eliminar alguns riscos, apenas reduzindo-os a valores menores. Também pode ocorrer de novas vulnerabilidades surgirem no ambiente e, nesse caso, não haverem sido previstos mecanismos para tratar os riscos associados.

Portanto, uma estrutura para lidar com os incidentes deve ser prevista para compor o Sistema de Gerenciamento da Segurança das Informações. Essa constatação tem apoio na própria norma internacional sobre segurança das informações [8], recentemente revista e entre cujas alterações está a inclusão de um conjunto de controles sob o tópico de tratamento de incidentes.

Essencialmente, dois aspectos devem ser considerados no processo de tratamento de incidentes: a detecção e a resposta. A detecção cuida da análise do ambiente a procura de indícios que justifiquem a execução dos procedimentos para resposta aos incidentes. A resposta, por seu turno, consiste na adoção de políticas e execução de procedimentos visando a minimizar os impactos do incidente.

Os métodos para a detecção de incidentes incluem a análise de registros do sistema monitorado. Segundo Bernstein et alli [9] os registro de eventos dos sistemas, conhecidos como *logs* constituem *"a fonte de informação mais útil (embora não totalmente confiável)"*. Partindo dessa premissa, algumas indicações presentes nos arquivos de *log* podem consistir em incidentes de segurança, a exemplo de:

- horário de utilização atípico;
- padrões de utilização e erros atípicos;

- utilização de uma conta padrão;
- presença de uma conta nova, desconhecida;
- utilização de uma conta anteriormente inativa;
- modificações inexplicadas em arquivos;
- lacunas nos *logs* do sistema;
- a descoberta de utilitários *hackers*;
- conexões a partir de endereços suspeitos da Internet;

Não obstante a importância dos arquivos de *log,* não devem ser esquecidos outras fontes de informações indicativas de comprometimento do sistema. Mesmo por que, invasores experientes encobrem seus rastros com habilidade. Além disso, quanto mais rápido for detectada uma invasão menos tempo o invasor terá e menos estragos poderá fazer.

As ferramentas para detecção de invasões aos sistemas computacionais são outro recurso para a identificação de incidentes de segurança. Essas ferramentas vêm sendo constantemente aprimoradas, incorporando métodos derivados do sistema imunológico humano [10] que tentam reduzir os falso-positivos, indicação de que houve uma invasão sem que essa tenha ocorrido. Não obstante esses esforços, os falso-positivos ainda existem.

Hatch et al [11] relaciona diversas outras maneiras de identificar o comprometimento de um sistema, a exemplo da rápida diminuição do espaço em disco, da atividade de rede anormal, da existência de interfaces de redes promíscuas, da existência de usuários desconhecidos no sistema, de processos estranhos em execução, do uso excessivo de CPU, além de contatos de outros administradores avisando que máquinas dele estão sendo alvo de tentativas de invasão por sua máquina.

Após o incidente ter sido identificado, ações pré-estabelecidas devem ser postas em prática rapidamente, de modo a minimizar os impactos do incidente. Porém, responder a um incidente exige dedicação e pode ocorrer outros incidentes enquanto um está sob investigação. Nessa circunstância é preciso estabelecer prioridades e, para tanto, deve haver uma política que defina com clareza o que é mais relevante. Podem ser considerados o potencial prejuízo financeiro de um ataque, os danos que a invasão pode causar a imagem, a extensão do comprometimento da rede ou a possibilidade de comprometimento de sistemas que suportam processos relacionados a vida humana.

Por exemplo, a equipe de respostas a incidentes de um hospital enfrenta três incidentes. O primeiro referente a alteração da *home page* institucional, comprometendo a imagem, o segundo afetando a rede administrativa que suporta os sistemas de faturamento de todos os hospitais participantes de uma organização inter-hospitalar e o terceiro que compromete os prontuários e receituários dos pacientes. Qual deve ser a prioridade da equipe? A política para tratamento de incidentes dirá.

Ainda, devido a conexão à Internet, característica da maioria das redes de computadores existentes no mundo, deve existir procedimentos documentados para troca de informações sobre incidentes com outras equipes de resposta a incidentes. Assim, torna-se mais factível rastrear os ataques. O CERT da Carnegie-Mellon University é o caminho mais natural. O seu representante no Brasil é o antigo NBSO, agora CERT.br, responsável por

receber, analisar e responder a incidentes de segurança em computadores, envolvendo redes conectadas à Internet brasileira.

Claro que algumas organizações podem não querer divulgar que seus sistemas foram comprometidos pensando que essa ação preservaria sua imagem. No entanto, a divulgação pode ajudar na medida em que profissionais gabaritados ajudarão na solução dos problemas. Além do mais, pode-se alegar que não serão divulgados detalhes devido a existência de investigação para apurar responsabilidades.

Outro aspecto a ser considerado está relacionado a abordagem da resposta ao incidente em relação ao agente da invasão após essa ter sido identificada. São duas as possibilidades: manter o sistema funcionado e observar as ações do invasor e, eventualmente, bloquear algumas ações, ou interromper as ações do infrator desconectando ou mesmo desligando o equipamento em questão. Nenhuma das estratégias é a ideal e a escolha de uma ou de outra é dependente da criticidade do equipamento comprometido para as operações da organização em questão e da política de resposta a incidentes, que pode privilegiar a punição dos atacantes ou o rápido retorno das atividades normais. Para Hatch [11]:

> *"Embora alguma evidência em sua máquina pudesse incriminar a parte responsável, é muito melhor (de um ponto vista legal) 'apanhar o perpetrador no ato' - e isso significa manter sua máquina em atividade e acessível ao cracker enquanto você chama as autoridades responsáveis para ajudar a localizar o cracker. A maioria dos crackers fugirá, se achar que foi*

> *ou está sendo descoberta – significando que você pode não ser capaz de reunir informações suficientes para rastreá-los"(Hatch, p.691).*

De uma forma ou outra será necessário retomar o controle do sistema, colocando-o em funcionamento normal e, de preferência, com as correções das falhas exploradas no ataque que acabou de ocorrer. Antes de realizar a reinstalação de todo o sistema ou apenas consertar as falhas encontradas deve-se produzir uma imagem do sistema invadido para possibilitar posteriores investigações. Essa imagem consiste numa cópia *bit* a *bit* das unidades invadidas e outras informações extraídas do sistema.

Independente da abordagem adotada é preciso documentar as ações a serem executadas e instruir os responsáveis sobre como executar as tarefas que lhes competem. A documentação é fundamental por diversos motivos, vale destacar a garantia que o problema foi identificado e devidamente detalhado de modo a poder ser revisto em um momento posterior, a cientificação a todos os envolvidos no processo de negócio afetado e a possibilidade de uso dos documentos como evidência em um processo judicial ou administrativo.

O gerenciamento da documentação relacionada aos incidentes de segurança deve garantir que ela esteja disponível a quem de direito assim que necessário e que não haja corrupção do conteúdo. Esse último aspecto é fundamental no caso de existir intenção de usar parte da documentação como evidência em um processo juducial. Berstein et alli [9] assevera:

> *"Algumas provas mais valiosas para a acusação será o que você e outras pessoas da sua equipe de respostas a*

incidentes tiverem anotado em seus livros de registros. Sua empresa deverá, então, estabelecer um recurso para o tratamento das evidências. Você e outras pessoas deverão entregar os livros de registro a alguém encarregado dessa função ao final de cada dia no qual tenham feito pelo menos uma anotação. A pessoa encarregada de administrar as evidências deverá, então, tirar fotocópias da(s) pagina(s) relevante(s), assinar e datar cada página e, depois, guardar cada fotocópia em um cofre até que elas sejam necessárias. Você deve utilizar procedimentos semelhantes em relação aos logs de sistema e às listagens impressas de arquivos alterados – submeta-os ao seu funcionário encarregado da administração de evidências para garantir sua verificação e armazenamento seguro. Talvez você ainda precise gravar em video-tape seus logs de acesso e de impressão, de modo a assegurar que eles não foram simplesmente fabricados por você" Berstein et alli (p. 326)

3 ANALISE FORENSE DIGITAL

Este capítulo trata dos aspectos da forense digital, destacando a conotação legal associada à palavra forense, as diferentes abordagem das investigações digitais, a influência da arquitetura dos equipamentos eletrônicos na investigação digital e a validade das informações sobre eventos passados existentes em sistemas computacionais.

Vários crimes podem envolver o uso de sistemas digitais, tanto diretamente quanto indiretamente. Exemplos do primeiro tipo são o acesso e alteração de dados sem a devida autorização, o envio de e-mails com ameaças ou a disponibilização de material de pedofilia. A pesquisa de informações na Internet para cometer um sequestro é um exemplo do segundo tipo. Qualquer que seja o motivo, quando uma violação é detectada pode-se iniciar uma investigação para responder a questões como: qual a causa da violação, quem ou o que a produziu, etc.

Após a detecção do incidente e o encaminhamento da decisão de desligar o sistema ou mantê-lo em funcionamento monitorado, deve ser conduzida uma investigação que identifique o que ocorreu, quais foram os responsáveis, por que a invasão foi possível. Só então pode-se providenciar ações para punição dos responsáveis e melhoria das condições de segurança.

A necessidade de investigação de sistemas digitais usados em crimes proporcionou o estabelecimento da forense computacional. O conceito dessa recente área de pesquisa vem evoluindo. Algumas definições iniciais valorizavam os procedimentos utilizados pelos investigadores, como:

> *"Ela é a ciência que estuda a aquisição, preservação, recuperação e análise de dados que estão em formato eletrônico e armazenados em algum tipo de mídia computacional." (apud [12])*

Por outro lado, alguns consideravam o objetivo do trabalho como fundamental na conceituação, a exemplo de Farmer & Venema:

> *"Coletar e analisar dados de uma maneira tão livre de distorção ou ajuste quanto possível para reconstruir dados ou o que ocorreu no passado com um sistema."*[1] [13]

Farmer & Venema [14] revisaram essa definição retirando a expressão *"in a maner as free from distorsion or bias as possible"* por acreditarem ser mais provável extrair informações adicionais relevantes quando procedimentos mais intrusivos são executados. Por outro lado, esses procedimentos podem prejudicar a admissão das informações em uma corte judicial.

Percebe-se que existe uma tensão entre preservação do estado do sistema e o conhecimento mais profundo dos fatos ocorridos no sistema digital sob investigação. É nesse ponto que a palavra forense alcança conotações diferentes. Usada para designar as investigações digitais sem preocupação com os requisitos legais necessários para admissão nos tribunais ou como referência aos procedimentos e técnicas desenvolvidos e testados no sistema penal.

Seja para apurar os responsáveis por um crime ou para compreender como as ações criminosas foram levadas a cabo, é necessário o mesmo rigor nos procedimentos de forma a conferir maior credibilidade às conclusões. Mas,

1 *"Gathering and analyzing data in a manner as free from distortion or bias as possible to reconstruct data or what has happened in the past on a system."*

se a intenção é levar os responsáveis a um tribunal deve-se abrir mão de alguns procedimentos, já que é essencial garantir que as evidências sejam irrefutáveis.

Assim, a forense digital trata de responder perguntas sobre fatos ocorridos nos equipamentos eletrônicos cujas respostas tem sustentação em evidências passíveis de serem aceitas em uma corte judicial como prova do fato ocorrido.

Deixando de lado as questões legais Carrier [15] define a investigação digital como um processo científico em que uma hipótese é elaborada e testes são feitos para confirmar ou refutar a hipótese. Nas palavras dele:

> *"Uma investigação digital é um processo em que é desenvolvida e testadas hipóteses de resposta a questões sobre eventos digitais. Isso é feito usando o método científico que consiste no desenvolvimento da hipótese, a partir das evidências encontradas, e no teste da hipótese, pesquisando evidências adicionais que mostram ser a hipótese absurda."*[1]

Carrier [15] também apresenta a definição de evidência digital, considerando-a como um objeto que contenha informação confiável que suporta ou nega uma hipótese. Assim, dados sobre o tráfego de rede registrados no *firewall* podem constituir evidência que nega indícios de um ataque específico que foram obtidos em servidor comprometido.

1 *"A digital investigation is a process where we develop an test hypoteses that answers questions about digital events. This is done using a scientific method where we develop a hypotesis using evidence that we find and then test the hypothesis by looking for additional evidence that shows the hypothesis is impossible"*

Da mesma forma que Carrier [15] este trabalho usa o termo evidência desvinculado dos aspectos legais aplicáveis às investigações digitais, uma vez que os requisitos jurídicos necessários para validade das evidências digitais variam de um Estado para outro, além do objetivo do trabalho não estar vinculado à área jurídica. O termo investigação forense é evitado, uma vez que a palavra forense remete a requisitos legais.

3.1 Influência da arquitetura dos sistemas digitais

Os sistemas digitais e, em particular, os computadores têm estruturas complexas. Para facilitar seu desenvolvimento, esses sistemas são projetados em uma estrutura em camadas com interfaces mais ou menos padronizadas. Assim, o desenvolvimento de uma camada pode abstrair os detalhes de implementação específicos das camadas adjacentes.

Isso é necessário, por exemplo, para que os mesmos dispositivos sejam usados pelas aplicações, independente da organização dos dados, para transferir e armazenar *bits*. Para isso, as aplicações precisam traduzir suas estruturas para a estrutura comum do computador.

A estrutura em camadas é espelhada nos sistemas de arquivos. Esses armazenam os arquivos em *bytes* e os organizam em uma estrutura de árvore invertida e, além de armazenar os nomes e conteúdo, também mantém atributos tais como propriedade, permissões de acesso, hora de última alteração dentre outros. Assim, a percepção dos arquivos, diretórios e seus atributos é uma das ilusões existentes nos sistemas computacionais, assim como também é ilusório

vê-los como blocos de dados e blocos de metadados (*inodes*). Na realidade os sistemas de arquivos usam vetores, com tamanho igual ao dos blocos definidos para o sistema de arquivos, como estruturas de dados e separam uma parte desses para uso próprio. Não obstante, para aplicações e seus usuários é mais útil enxergar arquivos e diretórios da maneira ilusória. Mas, até mesmo a estrutura de dados em *array*, apesar de facilitar a implementação do armazenamento de massa, é ilusória pois os discos rígidos armazenam os *bits* em domínios magnéticos.

Um exemplo simples de camada de abstração é o código ASCII em que cada caracter do alfabeto é representado por um número de 32 a 127. Quando um arquivo texto é salvo no computador, cada caracter é convertido em seu respectivo código, cuja representação binária é gravada no disco rígido. Se o conteúdo do arquivo é aberto na forma como foi gravado será visualizado uma série de uns e zeros, mas aplicando o filtro do código ASCII permitirá ver o texto como fora produzido. Os editores de texto são um exemplo de ferramenta operando na camada de abstração correspondente ao código ASCII.

Formalmente, cada camada de abstração pode ser descrita como uma função de entradas e saídas (figura 3). A entrada da camada são os dados e o conjunto de regras de tradução. As regras descrevem como a entrada deve ser processada e, em vários casos, consiste na especificação do objeto. As saídas são os dados derivados da entrada e uma margem de erro. No exemplo anterior, as entradas são os dados binários e o mapa de caracteres ASCII e a saída é a representação alfanumérica da entrada [16].

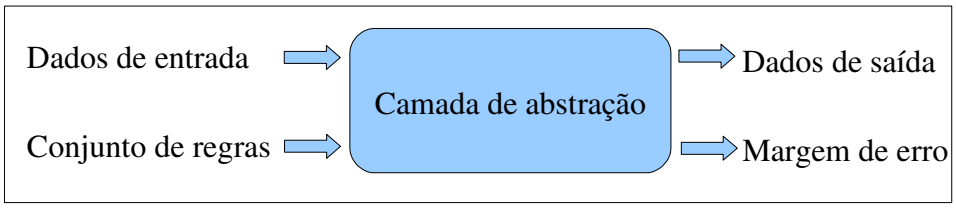

Figura 3 – Modelo das camadas de abstração – extraída de [16]

A saída de uma camada pode ser a entrada de outra, tanto como os dados a serem traduzidos quanto como o conjunto de regras a ser usado na próxima tradução.

No exemplo do código ASCII se o arquivo estivesse no formato HTML então a saída da primeira camada seria usada como entrada para a camada de conversão do conjunto de caracteres em um documento formatado. Os navegadores são ferramentas típicas dessa camada e, tipicamente, da primeira também.

Os *inodes* são um exemplo de camada de abstração que produz em sua saída metadados que serão usados como regras para a próxima camada. A estrutura de um *inode* inclui um descritor que indica o tipo do objeto, que pode ser um arquivo, diretório ou um tipo especial, e um ponteiro para o endereço do objeto no sistema de arquivos. Ambos, tipo e endereço, são usados na entrada da camada de abstração para indicar a localização e a forma como lidar com o arquivo, já que os diretórios são processados de maneira diferente. Nesse caso a saída da camada de *inode* não é a única entrada para a próxima camada porque todo o sistema de arquivos é necessário para localizar o endereço do bloco referente ao arquivo em questão.

Os sistemas computacionais são sujeitos a alterações nas diversas camadas que o constituem [14]. Assim, os aplicativos, o sistema operacional e até os dispositivos do hardware podem apresentar resultados não correspondentes ao real processamento. Portanto, cada camada de abstração pode introduzir erros que compõem a saída de cada camada. Na busca da eliminação desses erros é necessário conhecer as entradas, conjunto de regras e saídas de cada camada e verificar a tradução.

Para ilustrar a questão, que denominam ilusão em camadas, Farmer & Venema [14] fazem referência a René Magritte, um pintor que produzia imagens ilusionistas feitas em camadas e a capa do livro deles, que mostra a figura de um cachimbo e a frase "Isto não é um cachimbo". Assim, apresentam a idéia de que não é possível, apenas olhando para a imagem, dizer se ela corresponde a uma transposição de um objeto real ou a uma figura da imaginação do autor. É necessário dissecar os objetos de análise nas investigações para concluir a respeito do seu conteúdo.

Uma das conseqüências da estrutura em camadas é a relação inversa entre precisão e significância dos dados [14]. A medida em que são verificadas camadas mais próximas do hardware mais precisos são os resultados pois menos processamento é envolvido. Por outro lado, é mais difícil extrair significado útil das estruturas físicas.

A abstração em camadas tem também implicações para a destruição e recuperação de informações apagadas dos sistemas digitais. Farmer & Venema [14] afirmam:

> *"Tem sido provado que destruir informação é surpreendentemente difícil. Chips de memória podem ser lidos mesmo após uma máquina ter sido desligada. Embora projetada apenas para ler uns e zeros, chips de memória tem modos de diagnóstico não documentados que possibilitam o acesso a pequeninos rastros de fragmentos de bits. Dados em um disco magnético pode ser recuperados mesmo após terem sido várias vezes sobrescritos. Embora os drives de disco sejam projetados para apenas ler os uns e zeros que foram escritos por último, traços de padrões magnéticos antigos permacem no meio físico."*[1]

O desafio de um investigador inclui a recuperação de informações que foram parcialmente destruídas, ou seja, dar sentido ao lixo digital [14]. Contudo, sem o auxílio da aplicação que criou o arquivo é difícil entender o conteúdo dele. Ainda, reconstruir arquivos a partir dos blocos de disco agrupados pelos sistemas de arquivos, sem o auxílio deles, é como montar um quebra-cabeças gigante. Em resumo, quanto mais camadas forem afetadas quando da destruição de dados, mais difícil é de compreender o que se vê.

Por exemplo, apagar um arquivo a partir do sistema de arquivos é relativamente simples, mas não é suficiente para destruir o conteúdo ou atributos. Informações dos arquivos apagados permanecem nos blocos de disco

---

1 "*Destroying information turns out to be surprisingly difficult [Gutmann, 1996] and [Gutmann, 2001]. Memory chips can be read even after a machine is turned off. Although designed to only read ones and zeroes, memory chips have undocumented diagnostic modes that allow access to tiny left-over fragments of bits. Data on a magnetic disk can be recovered even after it is overwritten multiple times. Although disk drives are designed to only read the ones and zeroes that were written last, traces of older magnetic patterns still exist on the physical media [Veeco, 2004]."*

que foram alocados para ele. O conteúdo de um arquivo apagado é mantido até que dados de um arquivo novo sejam gravados por cima deles. Em sistemas de arquivos com boas propriedades de agrupamento o conteúdo dos arquivos apagados permanece por anos. Analogamente, arquivos apagados são como fósseis: um osso ou outro pode ser perdido mas o fóssil remanesce até que seja completamente destruído [14].

Esse fenômeno de eliminação e remanescência também pode acontecer em outros níveis de abstração. Por exemplo, os sinais analógicos captados pelas cabeças dos discos magnéticos podem indicar dados remanescentes, bem como padrões nos domínios magnéticos das trilhas de um disco rígido podem ser percebidos com uso de técnicas de microscopia.

Em cada camada da hierarquia de abstração que constitui os sistemas computacionais, informações ficam congeladas quando são apagadas. Apesar das informações ficarem cada vez mais ambíguas a medida que níveis mais baixos da hierarquia são alcançados, maior também é o grau de persistência. Portanto, a abstração, que faz dos computadores algo útil, produz a volatilidade.

3.2 Investigação de sistemas em uso e desligados

Para preservar os dados analisados em seu estado original, da mesma forma que a cena de um crime é isolada, são tomadas providências para evitar alterações posteriores ao momento em que o sistema sob investigação foi comprometido. Assim, são feitas cópias dos dados e mantidos os originais a salvo, enquanto a investigação é conduzida verificando-se as cópias feitas. Isso

confere credibilidade às conclusões da investigação e permite que elas sejam atestadas quando necessário.

De forma ideal, é desejável uma cópia completa de todo o sistema porém existem elementos que impossibilitam a realização disso. A medida que os dados estão sendo coletados, usuários e programas, intencionalmente ou não, podem disparar alterações no estado do sistema destruindo evidências valiosas.

É por conta desse tipo de problema que a análise forense tradicional é realizada em sistema que não está em atividade, chamada de análise pós-morte. A doutrina recomenda que o sistema comprometido seja desligado e logo em seguida seja feita uma cópia dos dados que restaram, a exemplo de *logs* dos programas, horários de acesso, o conteúdo dos arquivos, etc. Essa abordagem facilita a captura dos dados e garante uma cadeia lógica irrefutável.

Contudo, a abordagem de Farmer & Venema [14] não é essa. Eles preferem uma compreensão abrangente dos fatos a uma quase certeza do ocorrido, sabendo que em uma ação judicial suas conclusões seriam mais facilmente colocadas sob suspeita. No entanto, para eles parece haver um paradoxo, especialmente nos métodos de coleta de dados, já que compreender melhor o sistema sob análise propicia maior certeza dos resultados. Porém, alertam que o método requer mecanismos consistenstes para obtenção dos dados e bom entendimento dos efeitos colaterais do processo. Eles acreditam que mecanismos automatizados para coleta de dados são essenciais para obtenção de resultados consistentes.

Certamente deve-se ter cuidado e um plano de ação quando um sistema em funcionamento é analisado. Isolar o computador da rede e dos usuários é o primeiro passo. Uma vez que certos dados são mais sujeitos a alterações do que outros, a captura das informações deve seguir uma ordem que leve em conta a expectativa de vida dos dados. A Tabela 1 relaciona o tempo de vida dos dados e o tipo do dado, sugerindo uma ordem para colher os dados.

Tabela 1 - Expectativa de vida dos dados – extraída de [14]

Registradores, memória de dispositivos periféricos, caches, etc.	nano-segundos
Memória principal	nano-segundos
Estados de rede	mili-segundos
Processos em execução	segundos
Discos	minutos
Disquetes, mídia de *backup*, etc.	anos
CD-ROMs, impressões, etc.	décadas

No entanto, pode não haver justificativa para pegar os dados residentes na memória de um equipamento se o evento a ser investigado ocorreu a muito tempo atrás. Fora casos como esse, seguir a ordem de volatilidade possibilita maior chance de resultados mais claros.

A impossibilidade de pegar todos os dados de uma só vez reside no fato de que quando os dados de um tipo são obtidos, dados de outro tipo são alterados. Farmer & Venema [14] fazem analogia desse efeito como princípio da física quântica descrito por Werner Heisenberg que informa não ser possível conhecer com precisão os dados sobre posição e movimento de uma partícula

atômica ao mesmo tempo. Portanto, é impossível (a não ser em máquinas virtuais devidamente configuradas [14]) obter todos os dados existentes em um sistema computacional.

Além disso, o princípio de Heisenberg não é o principal fator a restringir a obtenção de todos os dados em um computador. Isso por que os sistemas computacionais não são conhecidos por seu estado em um dado momento, mas pela sequência contínua de eventos que executa. E, já que a memória, processos e arquivos são alterados muito rapidamente, registrar a atividade de uma maneira precisa e periódica causaria distúrbios dramáticos à operação do sistema. Para ilustrar, Farmer & Venema [14] apontam que ao rodar um programa simples como o date e acompanhar a execução com o strace constata-se a chamada de centenas de funções do sistema em uma fração de segundo, correspondentes a milhares de instruções de máquina a serem consideradas.

Pior, mesmo acompanhando todos os programas não é possível conhecer toda a história da máquina. As placas de vídeo, controladores de disco e outros periféricos também contam sua própria história com os respectivos processadores, memória e formas de armazenamento.

Portanto, em investigações de sistemas em atividade e, de maneira mais pronunciada, naquelas realizadas a partir de cópias de sistemas desligados nunca será possível recuperar o passado, mas isso não é necessário para concluir sobre o que aconteceu.

3.3 Validade das evidências

Mesmo não sendo possível desvendar toda atividade passada em um sistema computacional, devem ser dadas respostas a algumas perguntas sobre o ocorrido. Dessa forma, quanto mais precisos e completos forem os dados coletados melhor será a compreensão que se pode construir do sistema e a consistência das conclusões do investigador. A questão é como contornar a possibilidade de os dados obtidos dos sistemas comprometidos não serem verdadeiros.

Para abordar o assunto, Farmer & Venema [14] fazem referência a um teste de Alan Turing que intentava estabelecer o grau de "inteligência" dos equipamentos. No teste um entrevistador fazia perguntas digitando-as em um terminal de teletipo e, recebia a resposta, dada ou pelo computador ou por um ser humano. O entrevistador devia discernir se a resposta fora dada pelo computador ou não.

Assim como o teste de Turing, a descoberta forense consiste em observar um sistema computacional e concluir sobre o que ocorreu com ele. A questão é se estão sendo observados rastros do invasor ou marcas deixadas por ele para despistar quem tenta achá-lo. Para não cair em armadilhas deixadas por um invasor, uma abordagem válida é verificar os detalhes dos sistemas e conciliar os dados de pontos diversos para ter maior certeza do evento.

A listagem da Figura 4 apresenta dados gerados por três fontes diferentes [14], o *log* do TCP Wrapper (inetd), o comando `last` e o comando

`lastcomm`. Em conjunto fornecem informação sobre a conecção de um usuário a partir de uma máquina a outra, a duração da sessão e os processos executados durante a sessão, elementos suficientes para concluir sobre a atividade de alguém usando uma conta do sistema.

```
May 25 10:12:46 spike telnetd[13626]: connect from hades
|
| wietse       ttyp1    hades        Thu May 25 10:12 - 10:13  (00:00)
| |
| | hostname     wietse       ttyp1      0.00 secs Thu May 25 10:12
| | sed          wietse       ttyp1      0.00 secs Thu May 25 10:12
| | stty         wietse       ttyp1      0.00 secs Thu May 25 10:12
| | mesg         wietse       ttyp1      0.00 secs Thu May 25 10:12
. . .            .            .          .    .    .    .  .    .
| | ls           wietse       ttyp1      0.00 secs Thu May 25 10:13
| | w            wietse       ttyp1      0.00 secs Thu May 25 10:13
| | csh          wietse       ttyp1      0.03 secs Thu May 25 10:12
| | telnetd      root         __         0.00 secs Thu May 25 10:12
| |
| wietse       ttyp1    hades        Thu May 25 10:12 - 10:13  (00:00)
```

Figura 4 - Três fontes de informação sobre uma sessão de login - extraído de [14]

Em sistemas reais pode haver mais informação a respeito de uma sessão do sistema, algumas dessas informações presentes no próprio equipamento acessado, a exemplo dos horários de acesso dos arquivos acessados pelos processos executados, e outras em roteadores, sistemas de prevenção de intrusos, no equipamento de onde partiu o acesso, dentre outros. Todas essas informações podem ser correlacionadas e fornecem o poder necessário para a descoberta forense.

4 PROCEDIMENTOS APLICÁVEIS A INVESTIGAÇÕES DE SISTEMAS DIGITAIS

O objetivo desse capítulo é apresentar as pesquisas sobre o processo de investigação de sistemas digitais que exploram as fases existentes. Após essa visão geral, o foco é direcionado para os procedimentos de análise dos dados coletados.

O processo de investigação inclui várias etapas para garantir que as conclusões extraídas pelo investigador sejam consistentes e baseadas em evidências sob rigoroso controle. Em cada uma dessas etapas o perito vale-se de técnicas e ferramentas para executar o trabalho. Assim, o conhecimento de todo o processo de investigação permite a identificação do ponto em que o trabalho está e discernir como proceder para alcançar o objetivo final passando pelo caminho necessário.

Em muitos crimes digitais os procedimentos para desempenhar o trabalho forense não são consistentes nem padronizados [13]. O maior desafio é que *"procedimentos analíticos e protocolos não são padronizados nem uma terminologia padrão é usada pelos pesquisadores e profissionais da área*"[1] [apud 17]. Nos anos recentes, surgiram algumas tentativas para criar linhas gerais, mas essas foram escritas com foco em uma tecnologia específica e sem considerar um processo genérico.

1 *analytical procedures and protocols are not standardized nor do practitioners and researchers use standard terminology*

Farmer & Venema [13], por exemplo, propuseram linhas gerais que incluem *"proteger e isolar, gravar a cena, conduzir uma busca sistemática por evidências, coletar e empacotar as evidências e manter a cadéia de custódia"*[1]. A definição deles sobre o processo forense, bem como as idéias sobre métodos específicos de como desempenhar cada passo poderiam ter sido abstraídas para tornarem-se aplicáveis a sistemas computacionais em geral.

Outra tentativa de descrever o processo forense foi feita por Mandia e Prosise dentro de uma metodologia de resposta a incidentes. Essa metodologia consiste dos seguintes passos: *"preparação pré-incidente, detecção de incidente, resposta inicial, formulação da estratégia de resposta, duplicação, investigação, implementação de medidas de segurança, monitoramento da rede, recuperação, relatório e conferência"*[2] [apud 17]. A metodologia se propõem a fornecer fôlego e profundidade a análise forense computacional, e é abstrata o suficiente para poder ser aplicada a sistemas computacionais em geral.

O Departamento de Justiça dos Estados Unidos da America (Department of Justice – DOJ[3]) também tentou descrever o processo da forense computacional, levando em conta os benefícios da abstração do processo em relação a tecnologias específicas. Essa abstração inclui as fases de coleta,

1 *secure and isolate, record the scene, conduct a systematic search for evidence, collect and package evidence, and maintain chain of custody*

2 *"pre-incident preparation, detection of incidents, initial response, response strategy formulation, duplication, investigation, security measure implementation, network monitoring, recovery, reporting, and follow-up"*

3 http://www.usdoj.gov

exame, análise e relatório. Eles fizeram progressos significativos na identificação dos aspectos centrais do processo forense e então definiram passos para sustentá-lo, sem atrelar-se aos detalhes de uma tecnologia ou metodologia particular.

A ciência da forense digital foi definida no Digital Forensics Research Workshop I [apud 16] como:

> *"O uso de métodos derivados e validados cientificamente voltados para preservação, coleta, validação, identificação, análise, interpretação, documentação e apresentação de evidências derivadas de fontes digitais com o propósito de facilitar ou proporcionar a reconstrução de eventos considerados criminosos, ou de ajudar a antecipar ações não autorizadas aparentemente perturbadoras das operações planejadas."*[1]

O Digital Forensics Research Workshop (DFRW) é formado por um consórcio de grande abrangência liderados pela academia e não pelos, assim chamados, agentes da lei. É um representante da comunidade científica para enfrentar os desafios do desenvolvimento do processo forense digital.

Ciardhuáin [18] apresenta uma proposta de modelo generalista criado a partir da combinação de outros modelos e de acréscimos antes inexistentes. O modelo abrange todo o campo de uma investigação não apenas o processamento

1 *"The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations."*

das evidências. A Tabela 2 apresenta um quadro comparativo do modelo proposto Ciardhuáin e outros modelos de referência existentes.

Tabela 2 – Comparativo de terminologia de modelos - extraída de [18]

Activity in new model	***Models***			
	Lee et al	***Casey***	***DFRW***	***Reith et al***
Awareness				*Identification*
Authorisation				
Planning				*Preparation*
Notification				
Search/Identification	*Recognition, identification*	*Recognition*	*Identification*	
Collection	*Collection and preservation*	*Preservation, collection, documentation*	*Preservation, collection*	*Preservation, collection*
Transport				
Storage				
Examination	*Individualization*	*Classification, comparation, Individualization*	*Examination*	*Examination*
Hypothesis	*Reconstruction*	*Reconstruction*	*Analysis*	*Analysis*
Presentation	*Reporting and Presentation*		*Presentation*	*Presentation*
Proof/Defence			*Decision*	
Dissemination				

Não existe um único processo de investigação. Muito depende do estilo do investigador. Desde que as perguntas possam ser corretamente respondidas e não sejam violadas normas legais, os métodos usados são válidos. A diferença pode ser resumida à eficiência do método.

A abordagem de Carrier [15] é baseada nos procedimentos usados em investigações de crimes tradicionais, o ambiente da cena do crime digital é constituído por elementos de *hardware* e *software*. O processo é composto de três grandes fases: preservação do sistema, pesquisa das evidências e reconstrução de eventos. Essas fases não tem um sequência estabelecida. O fluxo geral pode ser visualizado na Figura 5.

Segundo Carrier [15], o uso da abordagem é possível em investigações de sistemas no próprio equipamento comprometido, chamada de "análise em vida", ou em equipamento seguro, chamada de "análise pós-morte". No primeiro caso existe o risco das informações obtidas terem sido manipuladas e serem incorretas. Já o segundo caminho nem sempre pode ser executado por exigir recursos as vezes indisponíveis.

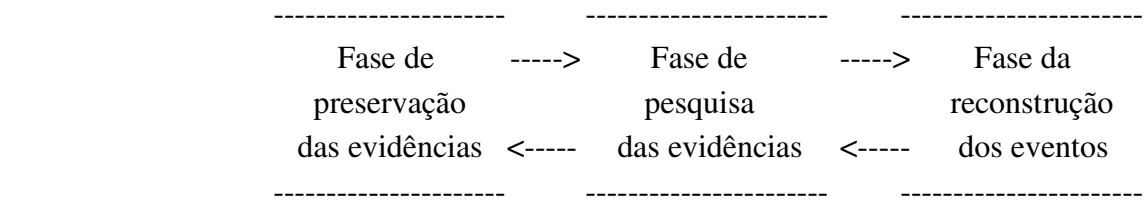

Figura 5 – Fluxo do processo de investigação segundo Carrier [15]

4.1 Preservação das evidências

Na fase de preservação do sistema o objetivo é cuidar para que as possíveis evidências existentes no sistema comprometido sejam preservadas. Os procedimentos para tal dependerão dos objetivos da investigação digital. Se o resultado das investigações fará parte de um processo judicial pode ser necessário desligar o cabo de alimentação do equipamento e realizar uma cópia

completa do disco rígido. Por outro lado, caso o sistema seja um *honeypot* ou tenha sido comprometido apenas por *spyware* nenhuma tarefa será necessária para preservar os dados. Entre os dois extremos estão a maioria dos casos que ocorrem nas organizações. A necessidade de preservar as evidências se mantém após a coleta de dados, uma vez que análises futuras serão necessárias.

O objetivo de evitar que as evidências sejam sobrescritas pode ser alcançado evitando-se que os processos escrevam nos dispositivos de armazenamento.

Em uma análise "pós-morte" isso é conseguido desligando o sistema e copiando os dados para outro sistema posteriormente. Nas "análises em vida" os processos podem ser suspensos ou mortos e o cabo de rede desconectado (conectando-o a uma porta sem conecção externa para evitar mensagens de quebra de link nos arquivos de *log*) ou podem ser aplicados filtros de pacotes que impeçam qualquer interferência no sistema. Além disso, dados importantes devem ser salvos para evitar perda por eventual alteração, a exemplo da hora de último acesso aos arquivos, que são alteradas quando o arquivo é lido.

Em qualquer dos casos devem ser gerados códigos resumo (HASH) dos dados analisados para garantir que eles não foram alterados após a análise. Os algoritmos que geram os códigos de resumo, a exemplo do MD5, do SHA-1 e do SHA-256, produzem um código relativamente pequeno para uma entrada de dados qualquer e que não é repetido para entradas diferentes.

4.2 Fase de pesquisa das evidências

Após os cuidados com a preservação dos dados deve-se procurar evidências que confirmem ou refutem as hipóteses sobre o ocorreu para o comprometimento do sistema. O processo é iniciado com a pesquisa dos lugares conhecidos para registro da atividade típica. Por exemplo, se a investigação é sobre os hábitos de uso da Internet procura-se nos arquivos de cache, no histórico e bookmarks por dados que confirmem ou rechassem a suspeita.

A teoria por trás do processo de pesquisa é simples, consiste em estabelecer o que se deseja e, em seguida, procurar dados com as características procuradas, Ou seja, deve-se definir o que esperamos encontrar e onde é provável achar.

A maioria das pesquisas são feitas nos arquivos ou nos sistemas de arquivos. Técnicas comuns incluem a pesquisa de arquivos de acordo com um padrão de nome especificado ou uma palavra-chave do conteúdo do arquivo, além de pesquisas baseadas nos horário de acesso ou alteração dos arquivos.

Pode-se pesquisar os valores de *hash* de arquivos conhecidos numa base de dados como a do National Software Reference Library[1] e compará-los com os resultados das somas MD5 ou SHA-1. Bases de dados de *hash* podem ser usadas tanto para verificar arquivos úteis do sistema como para encontrar arquivos conhecidos pelos estragos que produz. Também podem ser pesquisadas assinaturas contidas nos arquivos de modo a encontrar arquivos mesmo que o nome deles tenha sido alterado.

1 http://www.nslr.nist.gov

Quando análise é nos pacotes de rede, pode-se pesquisar pacotes de um determinada origem, com destino a determinada porta, ou ainda, que contenham uma palavra-chave.

4.3 Fase de reconstrução de eventos

Após colecionar dados existentes no sistema sob análise é preciso correlacioná-los para tentar reconstituir o que possa ter havido. Através de exemplos Carrier [15] esclarece o objetivo dessa fase:

> *"Durante a Fase de Pesquisa das Evidências, nós poderíamos encontrar vários arquivos que violam leis ou uma política da organização, mas isso não responde questões sobre os eventos. Um dos arquivos poderia ter sido baixado em um evento, mas nós ainda deveríamos tentar determinar qual aplicação o baixou. Existe evidência de que o navegador Web o baixou, ou poderia ter sido um software malicioso?"*[1]

Para realizar a reconstituição de eventos é necessário conhecer o sistema e os aplicativos instalados, pois sistemas operacionais tem respostas diversas para eventos específicos. Até mesmo versões diferentes de um mesmo software, como um *browser*, produz efeitos diferentes para determinado conjunto de eventos.

---

1 *"During the Evidence Searching Fase, we might have found several files that violate a corporate police or law, but that does not answer questions about events. One of the files may have been the effect of an event that downloaded it, but we should also try to determine which application downloaded it. Is there evidence that Web browser downloaded then, or could it be from malware?"*

5 FERRAMENTAS DE CÓDIGO ABERTO PARA USO EM INVESTIGAÇÕES DIGITAIS

Conforme visto no *Capítulo 4,* o processo de investigação forense inclui as etapas de coleta e análise dos dados sobre o evento investigado, além da extração, preservação e apresentação das evidências. No mesmo sentido, as ferramentas usadas no processo de investigação de sistemas digitais são especificadas de acordo com a etapa que apoiam e tem características próprias de cada atividade.

Já que a maioria das ferramentas disponíveis atualmente são para coleta e análise dos dados, optou-se nesse capítulo pelo estudo das ferramentas voltadas para análise dos dados obtidos em sistemas digitais. São apresentados os tipos e as características dessas ferramentas e indicado o diferencial das ferramentas de código aberto para essa atividade. Por fim, são apresentadas algumas ferramentas focadas no tema. Convém alertar que esse capítulo não é um tutorial.

5.1 Características e tipos de ferramentas para análise

Tanto as ferramentas para investigações digitais quanto os utilitários dos sistemas operacionais permitem visualizar evidências contidas em arquivos residentes no espaço alocado para o sistema de arquivos. Porém, somente com ferramentas específicas podem ser investigados os espaços não alocados e somente com elas algumas condições são satisfeitas. Por exemplo, ao listar o conteúdo de um diretório com um utilitário do sistema operacional a marca de tempo de acesso desse diretório é alterada. Isso não ocorreria com o uso de uma

ferramenta especialista. O raciocínio pode ser estendido às ferramentas destinadas a outras tarefas no processo de investigação. Portanto, faz-se necessário definir as características das ferramentas usadas em investigações digitais.

Para tanto, Carrier [16] considera dois tipos de problemas associados ao desafio da investigação digital. O primeiro diz respeito a quantidade de dados existente nos sistemas computacionais que precisam de ser analisados. Identificar e analisar todos os tipos de dados é uma tarefa gigantesca devido a crescente quantidade de dados processados pelos sistemas digitais. O segundo tem relação com a complexidade dos dados existentes em cada camada de abstração, característica inerente a complexidade dos sistemas digitais.

Dado o volume de dados presentes nos sistemas digitais, não é eficiente analisar todo e qualquer pedaço de informação colhida. Ao invés disso, técnicas são utilizadas para reduzir a quantidade de dados a serem analisados. Essas técnicas são um tipo de camada de abstração, dentre as quais estão a identificação dos pacotes de rede conhecidos usando sistemas de detecção de intrusos (IDS), a identificação das entradas desconhecidas nos arquivos de *log*, a identificação dos arquivos conhecidos no bancos de dados de HASH, etc.

O problema da complexidade na forense computacional é que os dados coletados para análises tem estruturas muito específicas e muito diversas. Requerer dos investigadores o conhecimento necessário para analisar tais dados seria limitar o desenvolvimento das investigações digitais.

Para resolver o problema da complexidade, ferramentas são usadas para traduzir os dados das camadas de abstração de forma a possibilitar a compreensão deles. Por exemplo, para exibir o conteúdo de um diretório presente em um sistema de arquivos, ou uma imagem dele, as ferramentas processam a estrutura do sistema de arquivos para mostrar os dados correspondentes. Os dados que representam os arquivos em um diretório estão em um nível muito baixo para serem facilmente compreendidos. Os diretórios constituem uma camada de abstração no sistema de arquivos. Outros exemplos incluem a codificação ASCII, os arquivos HTML, os pacotes de rede e os códigos fonte de aplicativos (Figura 6).

<table>
<tr><th colspan="3">Physical Media</th><th>Media Management</th><th colspan="3">File System</th><th>Application</th></tr>
<tr><td>Hed</td><td>Cyl</td><td>Etc</td><td></td><td colspan="3" rowspan="2"></td><td rowspan="4"></td></tr>
<tr><td colspan="3">Sectors</td><td>Partition table</td></tr>
<tr><td colspan="3" rowspan="2"></td><td rowspan="2">Partition</td><td>Boot sector</td><td>FAT</td><td>Data area</td></tr>
<tr><td colspan="3">....</td></tr>
<tr><td colspan="3"></td><td colspan="3">file</td><td>ASCII</td></tr>
<tr><td colspan="3"></td><td colspan="3"></td><td>HTML</td></tr>
</table>

Figura 6 – Exemplo de quatro níveis de camadas de abstração – extraída de [16]

O propósito das ferramentas de análise nas investigações digitais é apresentar precisamente os dados de uma camada de abstração num formato que torne possível ao investigador identificar as evidências [16]. A depender da habilidade do investigador e das necessidades de investigação, pode ser

suficiente a visualização do conteúdo dos blocos de dados ou processar os blocos pelas estruturas do sistema de arquivos.

Conforme visto no *Capítulo 3,* qualquer sistema digital e, portanto, todas as ferramentas usadas para analisar os dados que transitam nesses sistemas são baseados no modelo de camadas. Partindo da classificação das camadas de abstração, são definidas as categorias genéricas das ferramentas de análise. Essa forma de categorizar as ferramentas é diferente de outras já propostas que concentravam-se mais nas habilidades necessárias ao investigador [16].

Análise do meio físico de armazenamento: Ferramentas que traduzem o conteúdo e formato específicos (discos rígidos, *chips* e cartões de memória) para uma interface padrão, IDE ou SCSI por exemplo. O propósito dessa ferramenta inclui o processamento do formato específico e a recuperação de dados apagados, após terem sido sobrescritos.

Análise do gerenciamento da mídia: Ferramentas para análise do gerenciamento da mídia, vale dizer organização do meio de armazenamento, a exemplo da divisão de um disco rígido em partições, da organização de vários discos em um volume e da integração de vários *chips* de memória em um espaço de memória. Esse tipo de ferramenta não tem utilidade em alguns casos. Um banco de dados, por exemplo, acessa o disco rígido diretamente.

Análise do sistema de arquivos: Ferramentas que traduzem os *bytes* e setores das partições em arquivos e diretórios. O propósito desse tipo de ferramenta inclui ver arquivos e diretórios e recuperar arquivos apagados.

Análise de aplicações: Ferramentas que traduzem dados, tipicamente advindos do sistema de arquivos, em formados específicos necessários às aplicações. A análise nesse nível inclui a visualização de arquivos de *log*, arquivos de configuração, imagens, documentos e engenharia reversa de executáveis. Os dados de entrada são normalmente originados no sistema de arquivos.

Análise de rede: Ferramentas que traduzem os dados de mais baixo nível de uma rede física ou sem fio (pacotes de rede e alertas de IDS) para dados usados nas aplicações. Análises de *logs* de serviços de rede, servidor web ou *firewall*, fazem parte da categoria de análise de aplicações.

Análise de memória: Ferramentas que traduzem os *bytes* da memória física para os processos e dados do sistema. A análise dessa área inclui identificar o código que um processo rodava e extrair dados não disponíveis em outro lugar.

Os outros elementos indicados por Carrier [16] para caracterização de qualquer tipo de ferramenta de análise forense são os tipos e as propriedades dos erros existentes naquelas ferramentas. Segundo ele as ferramentas utilizadas

nas investigações digitais podem introduzir dois tipos de erro: erro de implementação da ferramenta de análise e erro de abstração. Erros de implementação são introduzidos devido a falhas na especificação e codificação das ferramentas de análise forense. Exemplos desse tipo incluem falhas de programação, uso de especificação incorreta ou uso de aplicação diversa da usada pela aplicação sob investigação. Esse tipo de erro é mais difícil de calcular por que exige extensa revisão do código e realização massiva de testes. Os esforços do Computer Forensics Tool Testing Group – CFTT do NIST[1] são de grande valia para o propósito de mensurar esse tipo de erro. Em geral, poder-se-ia assumir que erros de implementação encontrados em uma versão de ferramenta de análise seriam corrigidos em próximas distribuições. Assim, os investigadores podem manter o erro em valores mínimos atualizando a versão da ferramenta de análise.

A partir da apuração dos erros de implentação das ferramentas de análise é possível identificar os riscos associados a falhas não conhecidas dessas ferramentas. Os valores seriam baseados no número de falhas registradas num período recente e na severidade dessas falhas. O cálculo desse risco para ferramentas de código fechado pode ser falho, na medida do interesse dos fornecedores de não publicar erros e concertá-los silenciosamente, devido a vontade de manter os erros em níveis os mais baixos possíveis.

O segundo tipo de erro é aquele introduzido pelas reduções inseridas nos sistemas computacionais para resolver o problema da quantidade maciça de

1 http://www.cftt.nist.gov

dados, a exemplo dos sistemas IDS que reduzem ataques diversos a um só ataque. Uma vez que os IDS não tem 100% de certeza que os pacotes de rede são parte de um ataque podem cometer erros em cada ataque que tentam identificar. Com o desenvolvimento de pesquisas e produção de melhores técnicas de abstração, esses erros devem ser minorados. Vale observar que esse tipo de erro só ocorre quando há simplificações introduzidas no projeto da camada de abstração. Por exemplo, os sistemas de arquivos possuem diversas camadas de abstração em seu projeto, mas nenhum erro é produzido por elas por não haver esse tipo de simplificação.

Para resolver esse problema deve-se calcular a margem de erro de cada camada envolvida numa investigação e considerar essa margem quando da análise dos resultados obtidos [16]. Na busca da eliminação desses erros é necessário conhecer as entradas, conjunto de regras e saídas de cada camada e verificar a tradução.

A partir das definições dos erros existentes nas ferramentas usadas em investigações digitais e das diferentes formas de organização das camadas de abstração em um sistema digital, Carrier [16] propõem alguns requisitos das ferramentas de análise usadas em investigações digitais. São eles:

Usabilidade: A ferramenta deve apresentar os dados em um formato claro e preciso para que o investigador não interprete incorretamente o resultado;

Completeza: A ferramenta deve fornecer ao investigador acesso a todos os dados de saída de uma dada camada de abstração;

Precisão: As ferramentas devem garantir que os dados são precisos e uma margem de erro é calculada para que os resultados sejam interpretados adequadamente.

Exatidão: A ferramenta deve sempre produzir a mesma saída para uma determinada entrada e um conjunto de regras.

Auditabilidade: A ferramenta deve propiciar a verificação dos resultados que produz, sendo necessário garantir o acesso tanto às entradas quanto às saídas para que essas últimas sejam verificadas.

Além desses atributos, as funcionalidades de somente leitura e de verificação de integridade são recomendadas. Garantir que a ferramenta de análise não irá alterar dados e que irá informar quando uma saída produzida não é válida facilita o trabalho dos investigadores.

Percebe-se que as ferramentas de código aberto baseadas no GNU/Linux podem atender perfeitamente aos requisitos elencados, já que tanto elas quanto o próprio sistema base contam com a transparência necessária para assegurar os requisitos de completeza, exatidão, precisão e auditabilidade. O mesmo não pode ser tido das ferramentas de código fechado ou das baseadas em sistemas de código fechado, pois os testes de verificação dessas ferramentas são do tipo “caixa preta”, em que são conhecidos as entradas e saídas apenas.

Dado a variedade de possibilidades a serem testadas não seria possível atestar com a mesma certeza o cumprimento das características requeridas.

Um elemento adicional para sustentar a assertiva acima foi trazido por Carrier [19]. Segundo ele, para que evidências sejam admitidas numa corte judicial nos EUA elas devem ser relevantes e confiáveis. A confabilidade de evidências científicas, tais como as saídas produzidas pelas ferramentas de investigação digital, é determinada pelo juiz em um processo preliminar denominado "Daubert Hearing". A responsabilidade do juiz no procedimento é determinar se a metodologia e as técnicas usadas para identificar as evidências são aceitáveis, e, portanto, se a evidência é confiável. Existem quatro pontos usados na avaliação:

- **Teste**: o procedimento pode ser e foi testado?
- **Taxa de erro:** existe uma taxa de erro conhecida para o procedimento?
- **Publicidade:** o procedimento foi publicado e revisado pelos profissionais da área?
- **Aceitação:** o procedimento é amplamente aceito na comunidade científica reconhecida?

Carrier [19] mostrou que as ferramentas de código aberto aderem de maneira mais clara e abrangente aos requisitos estabelecidos pelo teste de "*Daubert"* do que as ferramentas de código fechado.

As ferramentas usadas em investigações digitais são usadas para demitir empregados, imputar crimes as pessoas ou demonstrar a inocência delas. Todas questões sérias e o mercado de aplicações para investigações digitais não deveria ser tratado da mesma forma que outros mercados de software. O objetivo não deve ser o domínio de mercado pela manutenção do segredo das técnicas usadas nas ferramentas de investigação digital.

5.2 Ferramentas de código aberto

O projeto "*Open Source Digital Forensics*"[1] mantém um repositório de ferramentas de código aberto classificadas de acordo com o sistema operacional base da ferramenta, *Unix*® ou *Windows*®, e com relação a funcionalidade que desempenha em ambientes de inicialização, ferramentas de aquisição de dados, ferramentas de análise do gerenciamento da mídia, ferramentas de análise dos sistemas de arquivos, ferramentas de análise de aplicações e ferramentas de análise de rede.

Para que uma ferramenta seja incluída na lista deve atender os seguintes requisitos:

- deve ser útil para investigações de computadores ou resposta a incidentes;
- o código fonte deve ser facilmente acessível a partir de um sítio na Web ou instrusões claras devem ser fornecidas sobre como obtê-lo (mesmo que haja custo);

1 http://www.opensourceforensics.org/index.html

- caso a ferramenta seja distribuída no formato binário, uma cópia de todas as bibliotecas não usuais deve ser fornecida (OpenSSL por exemplo).

O objetivo dos requisitos é permitir que o investigador veja o código fonte das ferramentas que usa nas investigações e sustentar as próprias conclusões no caso delas serem colocadas em dúvida. Dada a relevância do objetivo, são relacionados junto com cada ferramenta o *link* para obtenção do código fonte da ferramenta.

Nas Tabelas 3, 4, 5 e 6 são apresentas algumas ferramentas de análise para GNU/Linux – três para cada categoria relacionada na página específica do projeto[1]. Nelas são incluídas além do nome, a descrição do *software*, o *link* para a página onde cada um pode ser baixado e informações adicionais consideradas de interesse.

Nas sessões 5.2.1 e 5.2.2 são apresentados o *The Coroner's Toolkit – TCT* e o *The SleuthKit* respectivamente. As duas ferramentas são destacadas por contemplarem utilitários usados em investigações digitais correspondentes a diversas categorias e serem comumente relacionados como ferramentas específicas para forense computacional, além de terem o código fonte disponível.

1 http://www.opensourceforensics.org/tools/unix.html

Tabela 3 – Ferramentas para análise do gerenciamento de mídias

Nome	Descrição, link e comentários
CDfs	Um sistema de arquivos para *Linux* que "exporta" todas as trilhas e imagens de boot de um CD como arquivos normais. Depois de exportados, os arquivos podem ser montados (e.g. imagens ISO), copiados e reproduzidos (trilhas de áudio e video). **Website:** http://www.elis.rug.ac.be/~ronsse/cdfs/ **Informações adicionais:** O sistema de arquivos permite o acesso a dados não visualizados com sistemas de arquivos padrão. São exemplos: a cópia inicial de um arquivo com o mesmo nome presente em duas seções de um CD multi-seção no formato ISO9660; e dados de discos com várias seções criadas com o mkisofs sem o parâmetro -C.
disktype	Detecta o formato do conteúdo do disco ou uma imagem do disco. Ele reconhece vários sistemas de arquivos, tabelas de partição, códigos de boot e outras extruturas comuns em sistemas computacionais. **Website:** http://disktype.sourceforge.net/ **Informações adicionais:** Escrita em C, não utiliza bibliotecas e cabeçalhos especiais.
TestDisk	Verifica e recupera partições apagadas. Suporta diversos tipos de partições. **Website:** http://www.cgsecurity.org/testdisk.html **Informações adicionais:** A ferramenta pesquisa as características dos discos rígidos (tamanho e geometria) verifica a estrutura dele e a compara com a tabela de particionamento existente. Pode consertar ou recriar partições permitindo também a nanálise de dados presentes nos discos.

Tabela 4 – Ferramentas para análise do sistema de arquivos

Nome	Descrição, link e comentários
File System Investigator	Possibilita a visualização e extração de arquivos e estrutura de diretórios dos sistemas de arquivos ou imagens desses sem fazer uso dos mecanismos do sistema operacional. Não grava nada no sistema de origem preservando as marcas de tempo. Projetado para possibilitar o suporte a diferentes sistemas de arquivos, atualmente suporta a versão 3 do ReiserFS ale do EXT2/3. **Website:** http://www.rossi.com/fstools/intro.html **Informações adicionais:** Escrito em JAVA é independente de plataforma. Contudo, devido a limitações da linguagem, arquivos especiais como *device nodes, pipes, sockets e links* não podem ser extraídos.
pyflag	FLAG (Forensic and Log Analysis GUI) foi projetado para simplificar o processo de análise de *log* e investigações forenses. Permite a analise e correlação de grandes quantidade de dados fazendo uso de um banco de dados para gerernciar as operações. **Website:** http://pyflag.sourceforge.net/ **Informações adicionais:** Uma vez que é uma aplicação *web* é possível disponibilizar a ferramenta em um servidor central e disponibiliza-lo para vários usuários. Os dados de cada usuário são carregados em casos de forma a manter as informações separadas.
Enhanced Linux Loopback	O suporte a Loopback no kernel do Linux permite montar sistemas de arquivos apenas para leitura permitindo a análise forense dos dados alocados. O enhanced loopback modifica o driver nativo do Linux de forma a permitir a interpretação automática e mapeamento das partições contidas em uma imagem de disco rígido. **Website:** ftp://ftp.hq.nasa.gov/pub/ig/ccd/enhanced_loopback **Informações adicionais:** Também está disponível um conjunto de ferramentas específicas para o sistema de arquivos XFS.

Tabela 5 – Ferramentas para análise de aplicações

Nome	Descrição, link e comentários
GNU Binutils	Conjunto de ferramentas que podem ser usadas para análise binária, como o utilitário *strings* que lista sequências de caracteres imprimíveis, o *gprof* que mostra quais partes de um programa está consumindo mais tempo de execução e o *nm* que lista os símbolos dos arquivos objetos. **Website:** http://www.gnu.org/software/binutils/ **Informações adicionais:** Inclui também um montador multi-plataforma (*as)* e um linker (*ld)* que, ao contrário do padrão, tenta continuar a execução após a ocorrência de um erro possibilitando que outros erros sejam identificados e fornecendo um diagnóstico melhor.
FAUST	O *File AUdit Security Toolkit* – FAUST é um *script* escrito em *perl* que pode auxiliar a análise de arquivos presentes em equipamentos invadidos. O objetivo é extrair informações úteis para análise posterior. **Website:** http://www.security-labs.org/Security/FAUST/ **Informações adicionais:** suporta arquivos binários *elf* (executáveis e objetos) e *scripts bash*.
Forensic Hash Database	A combinação de várias fontes de somas de *hash* incluindo a da biblioteca do NIST (National Software Reference Library – NSRL)[1] disponíveis em uma base de metadados no formato RDBMS (*relational database management system*). **Website:** http://www.forinsect.de/forensics/ **Informações adicionais:** O arquivo com a base está disponível em http://www.forinsect.de/forensics/forensic_hash_database-1.02.tar.gz

1 http://www.nsrl.nist.gov

Tabela 6 – Ferramentas para análise de rede

Nome	Descrição, link e comentários
Ethereal	Apesar de ser um software em desenvolvimento, o Ethereal é adequado para uso rotineiro. Contempla todas as funcionalidades esperadas em um analisador de protocolos, além de outras incomuns. Inclui, por exemplo, a captura de dados a partir de uma conecção de rede ativa ou de um arquivo com tráfego previamente capturado por uma grande variedade softwares e dispositivos. **Website:** http://www.ethereal.com/ **Informações adicionais:** Reconhece 706 protocolos diferentes, salva ou imprime as saídas em texto ou PostScript® e permite a personalização da apresentação dos dados.
tcpflow	Captura dados transmitidos em conecções TCP e armazena-os em um formato adequado à análise de protocolo. O 'tcpflow' reconstrói o fluxo real de dados e armazena-o em arquivos separados por conecção para análise posterior. **Website:** http://www.circlemud.org/~jelson/software/tcpflow/ **Informações adicionais:** tcpflow é baseado na biblioteca *LBL Packet Capture Library* disponível em ftp://ftp.ee.lbl.gov/libpcap.tar.Z, entende sequenciamento de pacotes e reconstrói fluxo de dados mesmo no caso de retransmissões ou entrega fora de ordem. Entretanto, não consegue lidar com fragmentos IP.
tcpreplay	Conjunto de utilitários para sistemas UNIX para edição e reprodução de tráfego de rede que foi previamente capturado por ferramentas como o tcpdump e o etherreal. O objetivo do tcpreplay é fornecer condições de realizar testes confiáveis e reproduzíveis de dispositivos de rede diversos como *switches*, roteadors e *firewalls*. **Website:** http://tcpreplay.sourceforge.net/ **Informações adicionais:** possibilita a edição de pacotes das camadas 2 a 4 e a reprodução do tráfego em uma velocidade qualquer.

5.2.1 The Coroner's ToolKit – TCT

The Coroner Toolkit – TCT é uma coleção de utilitários para forense computacional desenvolvidos por Wietse Venema e Dan Farmer. A apresentação inicial ocorreu em 1999 no IBM T.J. Watson Research Center. A primeira distribuição geral ocorreu em 2000 nos sítios web dos autores[1]. O software foi ampliado em vários aspectos por Brian Carrier que disponibilizou a própria versão no SleuthKit[2] (uma distribuição linux que empacota diversos software para forense). A seguir são apresentadas os componentes do TCT.

O comando `grave-robber` obtem informações para forense e pode ser usado em uma máquina alvo em funcionamento ou numa imagem de disco de um sistema de arquivos sob investigação. Em análises de sistemas em funcionamento o programa procura respeitar a ordem de volatilidade buscando informações, a partir de vários utilitários do TCT, na seguinte ordem [14]:

- atributos de todos os comandos e arquivos que o TCT acessa para obter as informações. Isso é feito primeiro para preservar as marcas de tempo respectivas.
- informações sobre o estado dos processos e, opcionalmente, a memória dos processos em execução.
- arquivos apagados que ainda estão ativos.
- os arquivos executáveis de todos os processos.
- todos os atributos dos arquivos apagados.

---

1 http://www.porcupine.org/forensics/tct.html e http://www.fish.com

2 http://www.sleuthkit.org

- informação sobre o estado da rede.
- informação sobre o estado do *host,* por meio de comandos específicos que fornecem informações sobre a configuração do sistema.
- Atributos dos arquivos existentes; produzindo o corpo do arquivo (*body*) que é usado pela ferramenta `mactime` descrita mais abaixo.
- opcionalmente, informações sensíveis a segurança do sistema controladas pelos usuários, tais como arquivos que permitem acesso remoto a conta do usuário e relativos às tarefas automatizadas programadas pelos usuários.
- cópia dos arquivos de configuração e outros arquivos críticos.

Toda essa informação é armazenada em um "recipiente", uma estrutura de diretório protegida nomeada com o nome do host e o horário de início da obtenção dos dados. Para cada arquivo armazenado no recipiente o `grave-robber` calcula o *hash* MD5. No final, com o recipiente fechado, é calculado o *hash* MD5 de cada arquivo de com os *hashes* individuais.

O comando `mactime` gera um relatório cronológico de todos os acessos aos arquivos a partir das informações dos atributos desses arquivos presentes no arquivo '*body*' produzido pelo `grave-robber`. De forma alternativa o `mactime` pode produzir o arquivo '*body*' no momento da sua execução, enquanto varre o sistema de arquivos.

Exemplificando os benefícios do `mactime` para os investigadores são apresentadas nas listagens das figuras 7 e 8 diferentes visões de uma sessão de *login* remota. A primeira mostra o que o usuário vê e a segunda apresenta o relatório do `mactime`. É possível identificar claramente o início do servidor *telnet*, do programa de login, o acesso aos arquivos de sistema quando o usuário entra e o início do processo do *shell* do usuário.

```
$ telnet sunos.fish.com
Trying 216.240.49.177...
Connected to sunos.fish.com.
Escape character is '^]'.

SunOS UNIX (sunos)

login: zen
Password:
Last login: Thu Dec 25 09:30:21 from flying.fish.com
Welcome to ancient history!
$
```

Figura 7 – visão do usuário de uma sessão de login remota – extraída de [14]

```
Time         Size MAC Permission Owner Group File name
19:47:04    49152 .a. -rwsr-xr-x root  staff /usr/bin/login
            32768 .a. -rwxr-xr-x root  staff /usr/etc/in.telnetd
19:47:08      272 .a. -rw-r--r-- root  staff /etc/group
              108 .a. -r--r--r-- root  staff /etc/motd
             8234 .a. -rw-r--r-- root  staff /etc/ttytab
             3636 m.c -rw-rw-rw- root  staff /etc/utmp
            28056 m.c -rw-r--r-- root  staff /var/adm/lastlog
          1250496 m.c -rw-r--r-- root  staff /var/adm/wtmp
19:47:09     1041 .a. -rw-r--r-- root  staff /etc/passwd
19:47:10   147456 .a. -rwxr-xr-x root  staff /bin/csh
```

Figura 8 – relatório do `mactime` relativo a listagem da figura 7. A coluna MAC indica o método de acesso ao arquivo (*Modify, read Access ou status Change*). Nomes de arquivos com a mesma marca de hora são apresentados em ordem alfabética. – extraída de [14]

O `lazarus` é um programa simples cujo objetivo é conferir aos dados sem estrutura uma forma para que possam ser visualizados e editados. Os sistemas de arquivos modernos minimizam o tempo de acesso aos arquivos

mantendo próximas informações semelhantes. Dentre outras coisas, isso reduz a fragmentação de arquivos individuais. O `lazarus` usa essa característica e outros princípios e heurísticas na tentativa de reconstruir a estrutura do conteúdo de arquivos apagados.

O TCT vem com utilitários que desconsideram a camada do sistema de arquivos. Isso possibilita ao *software* acessar arquivos existentes bem como informações de arquivos apagados. Em vez de nomes de arquivos esses programas usam a abstração dos números de *inode* e a representação de alocação de blocos ou mesmo a abstração mais baixa de números de blocos no disco.

O TCT suporta sistemas de arquivos populares no mundo UNIX como UFS (BSD e Solaris) e EXT2FS/EXT3FS (Linux). O Sleuth Kit adiciona suporte adicional que inclui os sistemas de arquivos NTFS, FAT16 e FAT32.

Os utilitários para sistema de arquivos do TCT original são:

`ils` – Acessa os atributos dos arquivos pelo número do *inode*. Por padrão, também são listados os atributos de arquivos não alocados.

`icat` – Acessa o conteúdo dos arquivos pelo número do *inode*. É o comando preferencial para pesquisar conteúdo de arquivos apagados.

`unrm` – Acessa blocos do disco pelo número do bloco do disco. Por padrão ão lidos todo o conteúdo de arquivo não alocado e produzida saída para uso dos programas como o `lazarus`.

As ferramentas de baixo-nível para memória são mais apropriadas para uso exploratório do que análises consistentes. A razão para isso é que a saída

que produzem contem pouca ou nenhuma informação sobre a estrutura, de forma que é adequada apenas para processamento por ferramentas que não fazem uso dessas informações.

`pcat` – descarrega a memória de um processo em execução.

`memdump` – descarrega a memória do sistema procurando evitar alterações na mesma. A saída deve ser enviada pela rede para evitar a interação com o *cache* do sistema de arquivos.

5.2.2 The Sleuth Kit – TSK

O TSK e o navegador forense Autopsy são ferramentas de código fonte abertos baseadas no UNIX desenvolvidas por Brian Carrier e disponibilizadas pela primeira vez no início de 2001 pelo autor. TSK é um conjunto de mais de 20 ferramentas de linha de comando que possibilitam a análise de discos e sistemas de arquivos a procura de evidências. O Autopsy é uma interface gráfica para as ferramentas do TSK que pode ser usada para facilitar a análise.

O TSK é composto por ferramentas de linha de comando organizadas em grupos, incluindo ferramentas de disco, de volume, de sistema de arquivos e de pesquisa.

Existem duas ferramentas de disco no TSK, `diskstat` e `disksreset`. Atualmente a `diskstat` roda somente no Linux e fornece estatísticas sobre o disco rígido. Ela pode ser usada por exemplo, para pesquisar a *Host Protetected Area* – HPA antes de copiar dados de um disco. A ferramenta mostra o número total de setores e quais os setores acessíveis pelos

usuários, permitindo concluir se existe a HPA. O comando `disksreset` remove temporariamente a HPA se ela existe. Depois que o disco é reiniciado a HPA retornará.

O conteúdo de um disco rígido é organizado em volumes e o TSK inclui uma ferramenta para listar a organização das partições dos volumes. O comando `mmls` suporta partições DOS (`dos`), APPLE (`mac`), BSD(`bsd`), SUN(`sun`) e GPT(`gpt`). O tipo da tabela de partições pode ser especificado na linha de comando usando o parâmetro `-t` seguido do tipo (conforme indicado entre os parênteses). A saída do `mmls` é ordenada pelo endereço inicial da partição, independente da posição dela na tabela. Também é mostrado quais os setores no volume não estão associados a uma partição.

Dentro da maioria dos volumes existe um sistema de arquivos. Grande parte do TSK destina-se ao sistema de arquivos. As ferramentas para o sistema de arquivos do TSK são baseadas nas ferramentas do *The Coroner's ToolKit* – TCT. As ferramentas atuais funcionam com partições reais ou imagens de disco. As ferramentas para o sistema de arquivos suportam os formatos ext2/3 (linux-ext2, linux-ext3), FAT (fat, fat12, fat16, fat32), NTFS (ntfs), UFS1/2(freebsd, openbsd, netbsd, solaris). Elas também permitem visualizar páginas individuais do conteúdo real do disco e de partições swap. O tipo do sistema de arquivos deve ser especificado com o parâmetro `-f` seguido de um dos tipos informados antes (entre parênteses).

As ferramentas para análise do sistema de arquivos são mostradas na Tabela 7. Para mais detalhes sobre as opções de *flags* disponíveis veja as páginas de manual ou o sítio *web*.

Tabela 7 – Ferramentas do The Sleuth Kit para sistema de arquivos– TSK

Nome	Descrição
`fsstat`	Mostra estatísticas do sistema de arquivos(e.g. estrutura, tamanho e rótulo).
`ffind`	Encontra nomes de arquivos alocados ou não que apontam para uma determinada estrutura de dados.
`fls`	Lista nomes de arquivos alocados e apagados de um diretório.
`icat`	Extrai unidades de dados de um arquivo, indicado pelo *inode* (no lugar do nome do arquivo).
`ifind`	Encontra a estrutura de meta dados cujo nome de arquivo aponta para ela ou a estrutura de meta dados que aponta para uma unidade de dados.
`ils`	Lista a estrutura e o conteúdo de meta dados.
`istat`	Mostra estatísticas e detalhes sobre uma estrutura de meta dados em um formato de fácil leitura.
`dcat`	Extrai o conteúdo de uma unidade de dados.
`dls`	Lista os detalhes sobre unidades de dados, podendo extrair o espaço não alocado ao sistema de arquivos.
`dstat`	Mostra estatísticas sobre uma unidade de dados num formato claro.
`dcalc`	Calcula a posição dos dados existentes numa imagem encontrados no espaço não alocado (obtidos com o dls). É útil quando uma evidência é encontrada no espaço não alocado.
`jcat`	Mostra o conteúdo de um bloco específico do *journal.*
`jls`	Lista as entradas na base de *journal* do sistema de arquivos.

As ferramentas para pesquisa incluem o `hfind` (pesquisa *hashes* nas bibliotecas NIST/NSRL, Hashkeeper[1] e numa base de dados personalizada criada com o `md5sum`[2]), o `mactime` (ferramenta mostrada na sessão 5.2.1), o `sorter` (ordena os arquivos de acordo com o tipo e realiza verificações e pesquisa em bases de *hash*) e o `sigfind` (pesquisa por um valor binário num local específico).

1 http://www.hashkeeper.org

2 Ferramenta para cálculo de *hash* parte do conjunto de ferramentas para manipulação de texto do projeto GNU Textutils. Disponível em http://www.gnu.org/software/textutils/textutils.html

6 CONCLUSÃO

A sociedade contemporânea depende das informações para o pleno funcionamento das estruturas que dão suporte as atividades cotidianas. Ao mesmo tempo, os sistemas digitais constituem a infra-estrutura por onde dados e informações são obtidos, processados, armazenados e transmitidos.

É necessário cuidar da segurança das informações para garantir que os serviços dos quais a sociedade depende sejam confiáveis. Já que os sistemas digitais são a base dessa organização, é preciso cuidar da segurança dos dispositivos que compõem esses sistemas.

A administração da segurança é um processo abrangente e contínuo de reavaliação de riscos e controles para minimizá-los. Esse processo inclui o tratamento dos incidentes que venham a ocorrer. As ocorrências dos incidentes são oportunidades para reavaliar os riscos a partir da identificação das causas que as possibilitaram.

As investigações digitais consistem nos procedimentos realizados com o uso de ferramentas e técnicas específicas para descobrir o que ocorreu em um dado sistema que tenha sido afetado por eventos quaisquer. Pode também subsidiar processos judiciais para imputar autoria ou responsabilidade pelos danos causados ao sistema.

As técnicas de investigação devem considerar a organização dos sistemas digitais para que produzam o efeito desejado. Em particular, a estrutura em camadas que caracteriza o projeto dos computadores e outros

dispositivos eletrônicos norteia o desenvolvimento do trabalho dos peritos. A análise dos dados de entrada e saída de cada camada permite a eles concluir sobre modificações feitas nos sistemas.

Os procedimentos para realização das investigações digitais vêm sendo estudados e aprimorados com o tempo. Uma característica almejada é a ampla aplicabilidade aos sistemas digitais que na prática são baseados em arquiteturas as mais diversas, tanto de hardware quanto de software.

Para viabilizar a aplicação de procedimentos abrangentes são necessárias ferramentas específicas para cada plataforma ou flexíveis o suficiente para adaptarem-se às diversas estruturas. Conforme visto no *Capítulo 5*, as ferramentas de código aberto baseadas no GNU/Linux surgem como alternativa preferencial para as investigações digitais. As razões para isso incluem:

- o sistema operacional do pinguim está disponível para diversas plataformas de hardware, possibilitando a investigação no próprio sistema comprometido quando existe essa necessidade;
- o rápido desenvolvimento possibilitado pelo envolvimento da comunidade adepta ao princípio do software livre;
- a garantia de não sepultamento, devido a possibilidade de qualquer um com conhecimentos suficientes continuar o projeto;
- a grande quantidade de ferramentas de código aberto voltadas para investigações digitais. Elas podem ser agrupadas em categorias

segundo a adequação a determinada tarefa no processo de investigação.

Ademais, o uso de ferramentas de código aberto nas investigações digitais é uma prática recomendável por que é preciso saber como as evidências foram obtidas e, assim, conferir a elas a credibilidade para sustentar ou refutar uma tese criminosa. A razão é que ferramentas de código fechado não podem ter sua idoneidade comprovada plenamente, ao contrário das de código aberto, pois os procedimentos para realizar as tarefas, incluídos no código dos programas, não podem ser verificados.

Existe uma ampla variedade de ferramentas voltadas para auxiliar o trabalho de investigação de sistemas digitais caracterizadas pela disponibilidade do código fonte. Isso auxilia a necessária capacidade do investigador de produzir suas próprias ferramentas para obtenção e análise de informações sobre um sistema sob avaliação. Portanto, o uso das ferramentas de código aberto baseadas no GNU/Lunix voltadas para investigações digitais é uma prática efetiva.

A avaliação das ferramentas de código aberto disponíveis na URL http://www.opensourceforensics.org é um desdobramento natural do trabalho. No entanto, a tarefa é deixada a trabalhos futuros nos quais possam ser escolhidas metodologias adequadas voltadas para a validação das ferramentas existentes.

7 REFERÊNCIAS BIBLIOGRÁFICAS

[1] Thomas H. Davenport. *Ecologia da Informação: Por que só a tecnologia não basta para o sucesso na era da informação*. São Paulo: Futura, 1998.

[2] Rosa Maria F. Martineli, *Tecnologia da Informação na construção do conhecimento: uma abordagem a partir do modelo de Nonaka e Takeuchi*. Florianópolis: UFSC, 2001. Disponível em: http://teses.eps.ufsc.br/defesa/pdf/8239.pdf Acesso em: maio de 2003.

[3] ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR ISO/IEC 17799 – Tecnologia da informação – Código de prática para a gestão da segurança da informação.* Rio de Janeiro, 2001.

[4] Marcos Sêmola. *Gestão da Segurança da Informação*. Rio de Janeiro. Campus, 2003.

[5] Edgard Costa Oliveira. *Controle Terminológico e Sistematização de Conceitos para a Produção Documentária Especializada: Casos de Variância no Repertório da Segurança da Informação*. Brasília: Universidade de Brasília, outubro de 2001. Disponível em: http://www.modulo.com.br/pdf/controleterminologico.zip. Acesso em: maio de 2003.

[6] Micki Krause et al. *Information Security Management Handbook*. 4. ed. [s.l.]: CRC Press. LLC, 1998.

[7] British Stantards Institute, *BS7799 – part 2: Specificaction for information security management systems.* England, 2002.

[8] International Standard Organization/International Elecctrotechnical Commission. *ISO/IEC 17799 - Information technology - Code of practice for information security management.* Switzerland, 2000.

[9] Terry Berstein et alli. *Segurança na Internet*. Rio de Janeiro: Campus,1997

[10] Fabricio Sergio de Paula. Uma arquitetura de segurança computacional inspirada no sistema imunológico. Tese de doutorado Instituto de Ciência da Computação – Unicamp, 2004. Disponível em: 20040713-PhD-Fabricio.Sergio.de.Paula-Uma.arquitetura.de.seguranca.computacional.inspirada.no.sistema.imunologico.pdf Acesso em: agosto de 2005.

[11] Brian Hatch & James Lee & George Kurtz. *Segurança contra hackers Linux*. 2.ed. São Paulo: Futura, 2003.

[12]Célio C. Guimarães, Flavio de S. Oliveira, Marcelo A dos Reis e Paulo L de Geus. *Forense computacional – Aspectos legais e padronização*. Instituto de Computação-UNICAMP, 2002. Disponível em http://www.unicamp.br/... Acesso em: agosto de 2005.

[13] Dan Farmer & Wietsie Venema. *Computer Forensics Analysis Class Handouts,* 1999. Disponível em: http://www.fish.com/forensics/class.html. Acesso em: junho de 2005

[14] Dan Farmer & Wietse Venema. *Forensic Discovery.* USA: Addison-Wesley, 2004

[15] Brian Carrier. *File System Forensics Analisys*. USA: Addison-Wesley, 2005

[16] Brian Carrier. *Defining Digital Forensic Examination and Analysis Tools Using Abstraction Layers*. International Journal of Digital Evidence, Winter 2003, Volume 1, Issue 4.

[17] Mark Reith & Clint Carr & Gregg Gunsch. *An Examination of Digital Forensic Models.* International Journal of Digital Evidence, Fall 2002, Volume 1, Issue 3.

[18] Séamus Ó Ciardhuáin. *An extended model of cybercrime investigations*. International Journal of Digital Evidence, Summer 2004, Volume 3, Issue 1.

[19] Brian Carrier. *Open Source Digital Forensics Tools: The Legal Argument* Disponível em http://www.digital-evidence.org/papers/opensrc_legal.pdf. Acesso em: agosto 2005