Universidade Presbiteriana Mackenzie

Pós-Graduação

Computação Forense

Robustez da Prova Digital
A Importância do Hash no Processo Judicial

Trabalho de Fundamentos da Investigação Criminal

Professora Dra. Juliana Abrusio

Alexandre Teixeira
T.I.A. Nº 70961093
alexandre.abreu@gmail.com

Nicholas Istenes Eses
T.I.A. Nº 70957878
nistenes@gmail.com

Carlos Cabral
T.I.A. Nº 70950555
kbralx@gmail.com

Robson da Silva Ramos
T.I.A. Nº 70970580
robsramos@ig.com.br

Resumo

Este trabalho tem por objetivo apresentar um histórico sobre as funções de hash, suas limitações e suas aplicações práticas dando ênfase às aplicações ligadas à computação forense e à investigação criminal.

1 Introdução

Ao passo que a utilização do computador evoluía para preencher os mais impensados espaços da vida cotidiana da humanidade esta evolução arrastava consigo um problema inerente à própria da utilização de informações criadas digitalmente ou digitalizadas a partir de documentos físicos: como garantir a integridade destes dados, visto que à sua "natureza" digital os torna altamente voláteis?

A volatilidade das informações digitais não é relacionada somente à alteração não autorizada dos dados, mas possui uma condição mais sutil, geralmente presente no comportamento dos próprios sistemas (operacionais ou não) que podem executar alterações nas informações sem a ação intencional humana. Desta maneira, como garantir que um memorando escrito por Alice com destino a Bob não teve o seu conteúdo alterado por uma pessoa (ou sistema) em seu computador, no computador de Bob ou no tráfego de um para outro?

Quando analisamos o problema sob a perspectiva da investigação criminal a situação fica ainda mais complexa: imaginemos que Alice e Bob são acusados de realizar fraudes de milhões de reais através de seus computadores e que é deferido pelo juiz um pedido de busca e apreensão para estas máquinas. O que – sob a perspectiva tecnológica - deveria ser feito para garantir que os dados de seus computadores não foram alterados no processo de coleta das informações?

A solução destes problemas surgiu em decorrência da solução de outro problema tão grande quanto. De acordo com Singh[1], em abril de 1977 os pesquisadores Ronald Rivest, Adi Shamir e Leonard Adleman do Laboratório de Ciência de Computação do MIT elaboraram a C*riptografia Assimétrica* aperfeiçoando o conceito de *Criptografia de Chave Pública* originalmente criado por Whitfield Diffie, Martin Hellman e Ralph Merkle.

Embora existam controvérsias[2] acerca da autoria da Criptografia de Chave Pública, o advento da Criptografia Assimétrica mudou definitivamente a forma pela qual criptografamos informações em computadores e trouxe para o centro do debate a figura do algoritmo, seu principal componente. É nos algoritmos que estão contidos os métodos pelos quais, a partir da utilização de chaves uma informação será criptografada e descriptografada.

Como substrato da elaboração dos algoritmos de criptografia, foram desenvolvidos outros tipos de algoritmos com a função de garantir que uma determinada informação não foi modificada

1 SINGH, 2002, p. 298

2 De acordo com o governo britânico, o conceito de *Criptografia de Chave Pública* foi originalmente criado pelo Quartel General de Comunicações do Governo (GCHQ) como relata SINGH (1999). Historicamente as pesquisas em criptografia sempre tiveram forte apoio militar e conseqüentemente foram cercadas de sigilo, deixando até de fazer a justiça em reconhecer o trabalho de muitos atores que contribuíram para o progresso da ciência de forma anônima. No entanto, a história da ciência possui diversos eventos nos quais cientistas chegaram a um mesmo resultado através de pesquisas legítimas.

tornando este valor uma espécie de “selo de garantia” da autenticidade da informação. Caso a informação seja alterada e submetida novamente ao algoritmo o valor de saída, chamado de message digest, não será mais o mesmo denunciando a quebra da integridade da informação. Esta forma de gerar valores únicos para informações a través de algorítimos, foi batizada de hash.

2 O Hash

As funções de Hash geralmente se traduzem na produção de uma seqüência de bits de tamanho fixo a partir do mapeamento dos bits da informação digital a qual será submetida ao algoritmo. Essa informação pode ser qualquer tipo de arquivo digital, por exemplo uma imagem JPG, um arquivo Word, um arquivo do sistema operacional, etc.

A sequência de bits produzida como resultado da aplicação do hash em uma informação pode também ser chamada de message digest ou código hash e o seu tamanho depende do algoritmo utilizado. Dentre os algoritmos utilizados para gerar o código hash os mais populares são os algoritmos das famílias MD5 (Message-Digest algorithm 5) desenvolvido por Ronald Rivest e SHA (Secure Hash Algorithm) desenvolvido pela Agência de Segurança Nacional Americana – NSA. A utilização da família SHA para o tratamento das informações oficiais é o padrão adotado em vários governos, dentre eles o americano (através da FIPS PUB[3] 140-2) e o brasileiro (através da resolução Nº65 de junho de 2009 da ICP-Brasil).

Atualmente a utilização de hashes é popular e muitos usuários fazem uso deste mecanismo para assinar documentos e mensagens digitalmente, autenticar usuários, garantir a integridade de documentos e de provas digitais na cadeia e custódia.

3 A Utilização do Hash

3.1 Assinatura Digital

A assinatura digital é um mecanismo que visa garantir a integridade e autenticidade dos arquivos eletrônicos e mensagens. Ela é utilizada para provar que a informação em questão não foi alterada e que esta foi assinada pela entidade que possui a chave privada e o certificado digital utilizados na assinatura.

No Brasil a aplicação do conceito ainda é pequena e atualmente é essencialmente utilizada em processos do governo e no relacionamento de áreas financeiras com a Receita Federal. De modo geral o processo de assinatura digital em uma mensagem se dá da seguinte forma:

3 Pela abreviação FIPS PUB entende-se *Federal Information Processing Standards Publication.* As FIPS PUBs são documentos editados pelo NIST – *National Institute of Standards and Technology*, orgão do governo americano responsável por determinar os padrões de tecnologia e seus usos nos Estados Unidos.

Processo de envio:

1 - O signatário do documento aplica a função de hash de forma a gerar o message digest do arquivo a ser enviado.

2 - O message digest gerado é criptografado com a chave privada do signatário.

3 - Por último o certificado digital é agregado a este pacote (com o arquivo e o message digest criptografado) e é enviado ao destinatário

Processo de recepção

1 - O certificado digital enviado possui a chave pública que será utilizado para descriptografar o message digest.

2 - Aplica-se então novamente a função de hash no arquivo recebido e se compara o resultado com o message digest recebido.

Todas as operações acima são feitas de maneira automática por softwares de assinatura digital. Se todas as operações ocorrem com sucesso estão provadas a integridade do arquivo e a sua autoria.

3.2 *Autenticação de usuários*

Os mecanismos de hash são largamente usados em serviços de autenticação como o serviço de diretório Active Directory da Microsoft que ao invés de armazenar as senhas dos usuários, armazena o message digest destas senhas. No momento do logon do usuário ao invés de o serviço de diretório comparar a senha digitada pelo usuário com uma senha armazenada no banco, este compara o hash da senha gerado após a digitação com o hash que tem armazenado.

3.3 Integridade de Arquivos

A utilização de hash para checar integridade de arquivos possui várias aplicações, desde a função de fornecer informações arquivos específicos no tráfego por redes Peer-to-Peer (P2P) ou calcular message digests de arquivos armazenados em Hard Disks no processo de cadeia de custódia.

3.3.1 P2P

As redes P2P são utilizadas para a troca dos mais variados tipos de arquivos (obtidos de maneira legal ou não) e boa parte dos softwares utilizados nestas redes utilizam de funções de hash para a ajudar o usuário a determinar se o arquivo o qual esta prestes realizar o download é o que está realmente procurando ou se trata de um código malicioso disfarçado.

Profissionais de segurança e policiais se aproveitaram deste mecanismo para identificar o tráfego de pirataria e pornografia infantil nestas redes. Como exemplos destes novos softwares podemos citar o Híspalis desenvolvido em 2005 pelo espanhol Albert Gabás que, a partir do message digest de arquivos (foto e vídeo) apreendidos pela polícia espanhola, consegue fornecer informações para a localização dos criminosos e o EspiaMule[4], criado recentemente com o mesmo propósito pelos peritos criminais Guilherme Martini Dalpian e Carlos Augustus Armelin Benites da Polícia Federal Brasileira.

3.3.2 Cadeia de Custódia

Na cadeia de custódia de provas digitais, a utilização do hash ocupa uma posição importante, pois em seu processo de calculo e em sua devida documentação na ata notarial que se garante:

- Que a imagem usada na análise continua íntegra;
- Que o perito que realizou a cópia não introduziu mudanças nas informações coletadas;
- Que o processo de captura foi realizado corretamente .

No entanto, o fato de se calcular o hash para cada uma das informações de um disco não o dispensa dos cuidados de um processo de cadeia de custódia. Tais como:

- Calcular o hash do disco original antes de copiá-lo;
- Após a cópia do disco, comparar o hash do disco original com o do disco de análise;
- É necessário que, além da cópia para o disco de análise, seja feita outra cópia que será lacrada para o uso em caso qualquer questionamento técnico das partes;
- Esta cópia deve ser lacrada com a documentação contendo o message digest de cada arquivo do disco;
- A cada novo contato com as cópias, deve-se registrar o responsável pelo acesso, a data e o horário.

3.3.2.1 Redução do tempo de análise da prova

Todos os computadores possuem quantidades enormes de arquivos de sistemas operacionais ou de outros softwares que não são alterados pelos usuários ou que são alterados por eles com o objetivo de esconder alguma informação em um arquivo de software.

Desta forma, a primeira fase do processo de análise do disco coletado, geralmente toma muito tempo do perito ou o assistente técnico pois eles precisam separar os arquivos não

4 O nome “EspiaMule” faz alusão ao eMule, um dos softwares mais populares para troca de arquivos em redes P2P e o próprio EspiaMule foi desenvolvido com base no código fonte do eMule.

manipulados daqueles que realmente sofreram interação do usuário.

Com o objetivo de colaborar com a rotina dos profissionais de Computação Forense, o NIST criou um grupo chamado de National Software Reference Library (NSRL) que constantemente alimenta uma base de message digests de arquivos de sistemas disponíveis no mercado. Desta maneira, ao início da análise o perito ou o assistente técnico podem comparar os message ddigests do disco de análise com a base do NSRL descartando rapidamente os arquivos de software e detectando um arquivo suspeito disfarçado de arquivo de software comercial.

4 Colisões

Os algoritmos de hash devem possuir dentre as suas propriedades a resistência à colisão. Por esta propriedade entende-se ser resistente o suficiente à possibilidade de se gerar um mesmo message digest para informações diferentes.

O tamanho das mensagens geradas possuem uma limitação e esta limitação proporciona que se gere dois códigos hash iguais para diferentes entradas. Já foi divulgado que o algoritmo MD5, hoje em descrédito, é passível de colisão em até 5 minutos.

Uma forma de contornar a situação apresentada acima, é a utilização de novos algoritmos ou então a utilização de mais de um algoritmo produzindo assim dois hashes diferentes para cada informação.

5 Considerações Finais

Neste trabalho buscamos descrever como foi criado e em quais atividades é utilizado o mecanismo de hash, sobretudo em sua aplicação na área de computação forense.

Entendemos que a busca constante por parte de pesquisadores da área de criptografia em gerar colisões para os algoritmos de hash é essencial para que estas tecnologias acompanhem a evolução da capacidade de processamento dos computadores é necessária para o progresso contínuo da ciência. No entanto, não pudemos encontrar em nossa pesquisa nenhuma ocorrência histórica na qual foi gerada uma colisão para invalidar ou alterar uma prova digital. Pelo contrário, todas descobertas de colisões foram geradas em ambiente controlado e em prova de conceito.

Em nosso grupo é unânime a concordância com a opinião de Lewis[5] e de Giuliano Giova[6] de que a utilização de hashes será , ainda por muito tempo, uma ferramenta muito útil na rotina dos profissionais de computação forense e que o que mais fortalece a manutenção da utilização de hashes é a segurança de uma eficiente cadeia de custódia.

5 LEWIS, 2009
6 Proferida em sala de aula no dia 23/09/2009

6 Referências Bibliográficas

SINGH, Simon. **O Livro dos Códigos:** A ciência do sigilo - do antigo Egito à criptografia quântica. 2. ed. São Paulo: Record, 2002.

LEWIS, Don L.. **The Hash Algorithm Dilemma:** Hash Value Collisions. Disponível em: <http://www.forensicmag.com/articles.asp?pid=238>. Acesso em: 26 set. 2009.

MUNDOPT. **Espanhol cria sistema que detecta pedófilos.** Disponível em: <http://www.mundopt.com/n-espanhol-cria-sistema-que-detecta-pedofilos-8126.html>. Acesso em: 26 set. 2009.

PROCEEDINGS OF THE FOURH INTERNATIONAL CONFERENCE OF FORENSIC COMPUTER SCIENCE, 4., 2009, Natal. **EspiaMule e Wyoming ToolKit:** Ferramentas de Repressão à Exploração Sexual Infanto-Juvenil em Redes Peer-to-Peer. Natal: Icofcs.org, 2009.

NIST. **Introduction to the NSRL.** Disponível em: <http://www.nsrl.nist.gov/new.html>. Acesso em: 26 set. 2009.

NIST. **FIPS PUB 140-2:** SECURITY REQUIREMENTS FOR CRYPTOGRAPHIC MODULES. Disponível em: <http://csrc.nist.gov/publications/fips/fips140-2/fips1402.pdf>. Acesso em: 26 set. 2009.

ICP BRASIL -. **RESOLUÇÃO No 65, DE 09 DE JUNHO DE 2009:** PADRÕES E ALGORITMOS CRIPTOGRÁFICOS DA ICP-BRASIL (DOC ICP-01.01).

CONSELHO DA JUSTIÇA FEDERAL. **O que é Assinatura Digital.** Disponível em: <http://www.jf.jus.br/cjf/tecnologia-da-informacao/identidade-digital/o-que-e-assinatura-digital>. Acesso em: 26 set. 2009.

RODRIGUES, Tony. **Nosso Amigo Hash:** Parte II. Disponível em: <http://forcomp.blogspot.com/2007/07/nosso-amigo-hash-parte-ii.html>. Acesso em: 26 set. 2009.