

Hamilton Fonte II

**Modelo de plano de pesquisa - PCS5012 -
2016/1**

São Paulo

2016

Resumo

Esta pesquisa vai demonstrar que as abordagens para realização de análise forense de memória volátil disponíveis hoje não lidam com as características elásticas, multi-inquilino e multi-jurisdição das soluções baseadas em nuvem de forma satisfatória. A pesquisa tem como resultado esperado comprovar, através da implementação de um arcabouço de coleta de informações de memória, que a solução está em uma mudança na abordagem da coleta e armazenamento da evidência. Tendo êxito daremos um passo para a solução dos dois maiores problemas hoje na forense em nuvem, volume de dados e conformidade com requisitos jurídicos.

1 Descrição do problema de pesquisa

Forense digital é um conjunto de técnicas de coleta e análise de evidências geradas por computadores que tem por objetivos, entender a sequência de eventos que permitiu que um ataque fosse perpetrado, impedir que a vulnerabilidade que tornou possível um ataque seja explorada novamente e apoiar um processo legal (SANG, 2013). A forense digital cresceu a partir de técnicas usadas na forense tradicional. Começou de forma ad-hoc, passou por uma fase de adaptação aos requisitos legais até a era atual de ferramental avançado de coleta e análise (CHARTERS; SMITH; MCKEE, 2008).

A utilização crescente de virtualização, ferramentas online e hospedagem em nuvem (AMAZON, 2016), esta inviabilizando algumas práticas forenses (SHARMA; SABHARWAL, 2012). Especificamente a funcionalidade da elasticidade de carga ofertada pelos provedores de nuvem por meio da qual infraestrutura pode ser alocada e desalocada dinamicamente, trouxe o problema da volatilidade dos dados nas máquinas virtuais. Com algumas ameaças que não deixam evidências em disco (RAFIQUE; KHAN, 2013), a memória de uma máquina despejada de um pool e seus recursos reciclados seria para sempre perdida e com ela evidências importantes. O simples armazenamento do conteúdo da memória não satisfaz o princípio de Daubert. A abordagem de armazenar constantemente todas as alterações na memória não contribui para a solução do crescente backlog de dados que os investigadores tem para analisar (QUICK; CHOO, 2014). O ferramental forense disponível hoje esta pouco adaptado a desafios trazidos pela nuvem (DYKSTRA; SHERMAN, 2012), focam em completude, resposta a incidente e raros geram evidências aceitaveis em um processo legal (REICHERT; RICHARDS; YOSHIGOE, 2015).

A solução deste problema passa pela confirmação das seguintes hipóteses:

1. É possível conseguir o mesmo resultado da coleta mesmo se a máquina e o processo de origem não existirem mais.
2. Não é necessário todo o histórico de alterações da memória nem a cópia bit a bit da mesma para a análise do incidente.

Essa pesquisa se justifica pela necessidade da forense em se adaptar a computação em nuvem. Não ser capaz de levantar evidências impede investigações, prisão de criminosos e correção de vulnerabilidades. Tal cenário é um convite a criminalidade e garantia de impunidade.

2 Objetivos

O princípio de Daubert é uma norma da lei Norte-Americana que versa sobre a admissibilidade de evidências oriundas do trabalho de peritos em um processo legal. Pertinentes a esta pesquisa estão a obrigatoriedade em garantir que as provas não foram comprometidas ou alteradas durante o processo de coleta, a reprodutibilidade do processo e por fim conhecer sua taxa de erro. Sendo assim, o objetivo principal deste projeto é provar que é possível coletar dados de memória de processos em máquinas na nuvem de modo que estes se submetam com sucesso a parte cabível do Princípio de Daubert. Para alcançar este objetivo, os seguintes subobjetivos precisam ser alcançados:

1. Conseguir reproduzir o processo de coleta mesmo se a máquina de onde se originaram os dados tiver sido despejada e seus recursos desalocados.
2. Conseguir realizar análise de um incidente com no máximo 10% da informação de memória resultantes de processos de coleta existentes hoje.
3. Conseguir realizar a coleta sem violar jurisdição e privacidade de usuários não relacionados a investigação.
4. Conhecer a taxa de erro relacionada ao processo.

3 Método

1. Desenvolver uma aplicação de coleta de informações de memória em nuvem de modo que se consiga relacionar a mesma a sua origem e que o processo seja repetível. Este estudo será realizado em máquinas virtuais rodando sistemas operacionais windows e linux. A verificação do funcionamento correto será feita da seguinte forma:
 - a) Realizar a primeira coleta.
 - b) Despejar a máquina virtual do pool e liberar seus recursos.
 - c) Recriar a máquina e executar o processo de coleta novamente.
 - d) Mensurar a diferença entre as coletas através de comparação simples e guardar o percentual de diferença.
 - e) Repetir os passos (b), (c) e (d) até <limite>.
2. Desenvolver e documentar um arcabouço de armazenamento dos dados que garanta a cadeia de custódia da evidência guardando dados o suficiente para conseguir distinguir o momento anterior ao ataque.
3. Entrevistar analistas forenses e juizes. Apresenta-los o arcabouço de armazenamento e coletar opiniões sobre a aceitabilidade do mesmo.
4. Entrevistar analistas forenses e luizes. Apresenta-los a estratégia de armazenamento mínimo e coletar opiniões sobre a aceitabilidade do mesmo.
5. Publicar resultados.
6. Escrita e defesa de tese.

4 Revisão bibliográfica

- **Automated Forensic Data Acquisition in the Cloud (REICHERT; RICHARDS; YOSHIGOE, 2015)** : Propõe um modelo de coleta de dados de máquinas virtuais acionado por um sistema de detecção de ameaça. Usa o Google Rapid Response para salvar as informações coletadas fora da nuvem de forma a driblar os problemas de multi-jurisdição e multi-inquilino.

O modelo proposto não cobre o cenário em que uma máquina é despejada do pool e seus recursos liberados após a detecção da ameaça e não explica como vai manter a cadeia de custódia da evidência.
- **Digital forensics framework for a cloud environment (GEORGE; VENTER; THOMAS, 2012)** : Framework para coleta de dados de máquinas virtuais e associa-las aos usuários da nuvem. Usa a estratégia de armazenar tudo para evitar perda de informação decorrente do despejo de uma máquina do pool e desalocação de seus recursos. Armazena os dados coletados na própria nuvem.

O framework proposto, apesar de cobrir o cenário em que uma máquina é despejada do pool e os recursos liberados, não dá detalhes o suficiente para demonstrar que consegue reproduzir o processo e obter o mesmo resultado. É focado no usuário e não no hardware e leva algum tempo para identificar os usuários envolvidos. Usando a própria nuvem para armazenar os dados coletados ele depende de cooperação do provedor de nuvem para obter tais dados.
- **Evidence and cloud computing the virtual machine introspection approach (POISEL; MALZER; TJOA, 2013)** : Descreve o processo de coleta de informações de máquinas em nuvem através da técnica de introspecção em máquina virtual, onde se acessa os dados das máquinas virtuais através do hypervisor. Propõe que o processo seja disparado por demanda atrelado a um sistema de detecção de ameaça ou intervenção humana.

A técnica descrita cobre apenas o processo de coleta de informações. No que tange as informações de memória, como os endereços de memória são os do host, estes precisam ser traduzidos para que a análise forense seja feita. Como a abordagem não tem conhecimento do que está rodando dentro da máquina virtual precisa de uma cópia bit a bit da evidência. Embora pareça possível, não descreve como lida com o cenário onde uma máquina é despejada do pool e os recursos liberados.
- **Design and implementation of FROST: FoRensic tools for Open STrack (DYKSTRA; SHERMAN, 2013)** : Framework para coleta de dados de máquinas virtuais através da API do hypervisor, usa técnicas reconhecidas atualmente para

coleta e armazenamento dos dados. Isola a máquina virtual afetada do pool original para realização da coleta. Precisa ser acionado quando uma ameaça é detectada. O framework proposto não cobre o cenário em que uma máquina é despejada do pool e os recursos liberados após a detecção da ameaça. Não conhece o que esta rodando na máquina virtual, tem acesso apenas as informações que são disponibilizadas pela API do hypervisor e não explica como vai manter a cadeia de custódia.

- **A log based approach to make digital forensics easier on cloud computing (SANG, 2013)** : Método que sugere salvar a informação coletada fora da nuvem de modo a driblar os problemas de multi-inquilinato e multi-jurisdição, usa um mecanismo de hash para garantir a autenticidade e integridade da informação mas não dá detalhes da implementação. Dá exemplos de como funcionaria em PaaS e SaaS.

Segundo o próprio autor, o método não funciona em IaaS. Precisa da cooperação do provedor de nuvem pois depende que este último decida quais informações serão adicionadas ao log. O método não é aplicável a coleta de informações de memória.

- **Volatile memory acquisition using backup for forensic investigation (DEZ-FOULI et al., 2012)** : Técnica que sugere a utilização da própria máquina como repositório de informações de memória volátil, para manter a utilização de espaço ao mínimo sugere manter apenas o último estado conhecido de cada processo. A técnica não cobre o cenário em que uma máquina é despejada do pool e seus recursos são reutilizados, de fato, esta técnica provocaria a perda de todas as evidências caso a liberação de recursos da máquina ocorra.

5 Resultados esperados

O primeiro resultado esperado é um arcabouço que consiga coletar informações de memória de processos de máquinas na nuvem e armazena-las em um local físico conhecido. Este armazenamento deve garantir a cadeia de custódia e o volume de informação armazenada será o suficiente para diferenciar o sistema antes e depois que a ameaça foi detectada. A utilização deste arcabouço sob as mesmas condições deve gerar um conjunto idêntico de evidências. Com isso, será possível fechar uma das formas de impunidade para os atos ilícitos em nuvem e contribuir para a diminuição do backlog de investigação.

Referências

- AMAZON. *Amazon Media Room Press Release*. [S.l.], 2016. 2 p.
- CHARTERS, I.; SMITH, M.; MCKEE, G. The Evolution of Digital Forensics. In: *Techno Forensics 2008 Conference*. [S.l.: s.n.], 2008. p. 1–39.
- DEZFOULI, F. N. et al. Volatile memory acquisition using backup for forensic investigation. *Proceedings 2012 International Conference on Cyber Security, Cyber Warfare and Digital Forensic, CyberSec 2012*, p. 186–189, 2012.
- DYKSTRA, J.; SHERMAN, A. T. Acquiring forensic evidence from infrastructure-as-a-service cloud computing: Exploring and evaluating tools, trust, and techniques. *Digital Investigation*, Elsevier Ltd, v. 9, n. SUPPL., p. S90–S98, 2012. ISSN 17422876. Disponível em: <<http://dx.doi.org/10.1016/j.diin.2012.05.001>>.
- DYKSTRA, J.; SHERMAN, A. T. Design and implementation of FROST: Digital forensic tools for the OpenStack cloud computing platform. *Digital Investigation*, Elsevier Ltd, v. 10, n. SUPPL., p. S87–S95, 2013. ISSN 17422876. Disponível em: <<http://dx.doi.org/10.1016/j.diin.2013.06.010>>.
- GEORGE, S.; VENTER, H.; THOMAS, F. Digital Forensic Framework for a Cloud Environment. p. 1–8, 2012.
- POISEL, R.; MALZER, E.; TJOA, S. Evidence and cloud computing: The virtual machine introspection approach. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, v. 4, n. 1, p. 135–152, 2013. ISSN 20935374 (ISSN). Disponível em: <<http://www.scopus.com/inward/record.url?eid=2-s2.0-84885399460&partnerID=40&md5=0e332690d4cb1f01934b540b535fd771>>.
- QUICK, D.; CHOO, K. K. R. Impacts of increasing volume of digital forensic data: A survey and future research challenges. *Digital Investigation*, Elsevier Ltd, v. 11, n. 4, p. 273–294, 2014. ISSN 17422876. Disponível em: <<http://dx.doi.org/10.1016/j.diin.2014.09.002>>.
- RAFIQUE, M.; KHAN, M. N. A. Exploring Static and Live Digital Forensics: Methods, Practices and Tools. *International Journal of Scientific & Engineering Research*, v. 4, n. 10, p. 1048–1056, 2013. Disponível em: <<http://www.ijser.org/researchpaper/%5CExploring-Static-and-Live-Digital-Forensic-Methods-Practices-and-Tools.pdf>>.
- REICHERT, Z.; RICHARDS, K.; YOSHIGOE, K. Automated forensic data acquisition in the cloud. *Proceedings - 11th IEEE International Conference on Mobile Ad Hoc and Sensor Systems, MASS 2014*, p. 725–730, 2015.
- SANG, T. A log-based approach to make digital forensics easier on cloud computing. *Proceedings of the 2013 3rd International Conference on Intelligent System Design and Engineering Applications, ISDEA 2013*, p. 91–94, 2013.
- SHARMA, H.; SABHARWAL, N. Investigating the Implications of Virtual Forensics. *Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on*, p. 617–620, 2012.