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 em nuvem disponíveis hoje não lidam com as características elásticas, multi-inquilino e multi-jurisdição de forma satisfatória. Esperamos comprovar através da implementação de um arcabouço de coleta de informações de memória, que a solução esta 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 núvem, volume de dados e conformidade com requisitos juridicos.

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 o ataque seja explorada novamente e apoiar um processo legal para levar os culpados a prisão (SANG, 2013). A forense digital cresceu a partir de técnicas usadas na forense tradicional. Começou de forma artesanal, 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 criando dificuldades para a coleta de informações, análise e utilização em processos legais (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 requisito jurídico de se repetir o processo e conseguir os mesmos resultados. 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 analizar (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) e os que levam o fazem apenas até a cadeia de custódia. 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: Esta hipótese esta associada a necessidade jurídica de se repetir o processo de coleta e conseguir os mesmos resultados em um cenário onde as evidências de um ataque estavam na memória de uma máquina virtual que não esta mais no pool. Como as soluções de coleta hoje todas se baseiam em isolamento da máquina virtual afetada ou no armazenamento de informação de log, conseguir comprovar esta hipótese tráz um grande avanço para a pesquisa.

2. Não é necessário todo o histórico de alterações da memória nem a cópia bit a bit da mesma para a análise do incidente: Esta hipótese esta associada a realidade do

crescente volume de informações que precisam ser analisadas pelos investigadores forenses, em 2014 a média era de 6 a 12 meses de backlog (QUICK; CHOO, 2014). A confirmação desta hipótese envolve a aceitação de uma mudança no paradigma de coleta de informações forenses de 'completude' para 'a coleta do necessário'.

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

2 Objetivos

A aceitabilidade de evidências resultantes do trabalho de peritos é regulada por um conjunto de regras da lei Norte-Americana. Deste conjunto de leis derivou-se 3 requisitos que a evidência coletada de nuvem precisa se submeter são eles, a obrigatoriedade em garantir que as provas não foram comprometidas ou alteradas durante o processo de coleta, o processo de coleta ser capaz gerar as mesmas evidências quando executado novamente e por fim conhecer sua taxa de erro. A maioria das pesquisas relacionadas a coleta de informações tem viés técnico e poucos lidam com os aspectos legais relacionados a coleta de informações. Sendo assim, o objetivo principal deste projeto é provar que é possível coletar de dados de memória de processos em máquinas na nuvem de modo que estes atendam aos 3 requisitos legais mencionados anteriormente. 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 núvem 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 evidència 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

A revisão bibliográfica focou nos seguintes conceitos da coleta de informações forenses em nuvem:

- Acessar e coletar as informações em geral das máquinas em nuvem
- Conformidade com os seguintes requisitos legais:
  - reproduzir o processo com os mesmos resultados
  - não violar privacidade de outros usuários de nuvem
  - garantir a cadeia de custódia da evidência

Com isso constatou-se que, referente a coleta de informações a maioria dos trabalhos mais importantes na área (REICHERT; RICHARDS; YOSHIGOE, 2015), (POISEL; MALZER; TJOA, 2013), (DYKSTRA; SHERMAN, 2013), (GEORGE; VENTER; THOMAS, 2012) (SANG, 2013) foca em coleta reativa, isto é, ela acontece apenas após a ameaça ser detectada. Podem acontecer por acionamento manual ou integrada a um mecanismo de detecção de ameaça. Em casos de memória volátil, esta forma de coleta de informação não consegue descrever como era a memória antes do ataque, o que acredito é necessário para a estratégia de coletar o mínimo necessário para realizar a investigação e apoiar o processo jurídico. A única proposta encontrada que leva esta necessidade em consideração é (DEZFOULI et al., 2012) mas propoe que o dado seja armazenado no próprio dispositivo o que pode levar a perda das informações caso a máquina seja despejada do pool e seus recursos liberados.

Ainda na coleta de informações, (REICHERT; RICHARDS; YOSHIGOE, 2015), (GEORGE; VENTER; THOMAS, 2012) sugerem a abordagem de forense ao vivo onde os dados são coletados com o sistema rodando enquanto (POISEL; MALZER; TJOA, 2013), (DYKSTRA; SHERMAN, 2013), (SANG, 2013) adotam a estratégia de isolar e/ou pausar a máquina virtual e em seguida disparar o processo de coleta. Nos dois casos, nenhum dos autores cobre o cenário onde a máquina é despejada do pool e os recursos são liberados. Analisando cuidadosamente as propostas de (POISEL; MALZER; TJOA, 2013), parece que é possível cobrir o cenário mencionado anteriormente mas ele não dá detalhes da implementação o suficiente para termos certeza.

Referente aos requisitos legais, nenhuma das propostas consegue reproduzir os mesmos resultados ao repetir o processo no cenário em que uma máquina é despejada da nuvem e seus recursos liberados. Analisando cuidadosamente a proposta de (GEORGE; VENTER;

THOMAS, 2012) parece que é possível mas o autor não dá detalhes de implementação suficientes para termos certeza.

No armazenamento das informações coletadas, (REICHERT; RICHARDS; YOSHIGOE, 2015), (GEORGE; VENTER; THOMAS, 2012), (POISEL; MALZER; TJOA, 2013) e (DYKSTRA; SHERMAN, 2013) usam estratégia de armazenamento fora da nuvem para resolver o problema de violação de privacidade de outros usuários no processo da coleta. (SANG, 2013) e um caso específico de (GEORGE; VENTER; THOMAS, 2012) dependem de cooperação do provedor de nuvem para conseguir as informações necessárias à investigação.

Na garantia da cadeia de custódia apenas (SANG, 2013) faz alguma tentativa de resolver a questão porem cuida apenas da garantia que a evidência não foi destruida ou alterada a traves de cálculo de hashing mas não explica como impede o acesso não autorizado à evidência. As propostas dos outros autores estão focadas apenas no aspecto técnico da coleta.

5 Resultados esperados

O primeiro resultado esperado é um arcabouço que consiga coletar informações de memória de processos de máquinas na nuvem e armazená-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.