Hamilton Fonte II

Coletando dados de memória de máquinas virtuais em nuvem para análise forense - PCS5012 - 2016/1

São Paulo

2016

Resumo

Esta pesquisa tenta demonstrar que as abordagens para a coleta de informações de memória volátil de máquinas em nuvem com o propósito de análise forense disponíveis hoje não lidam com as características elásticas, multi-inquilino, multi-jurisdição da nuvem e os requisitos jurídicos para as evidências de forma satisfatória. Espero comprovar através da implementação de um arcabouço de coleta destas informações, que a solução para estes problemas 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 crescente e a pouca 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 ocorresse, impedir que a vulnerabilidade que tornou possível o ataque seja explorada novamente e apoiar os processos jurídicos que visam submeter os culpados as punições previstas na lei (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), está criando dificuldades para a coleta de informações, análise e utilização em processos legais (SHARMA; SABHARWAL, 2012). A funcionalidade de 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 virtual despejada de um pool e seus recursos liberados 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 da 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 está pouco adaptado a desafios trazidos pela nuvem (DYKSTRA; SHERMAN, 2012), focam em completude e poucos geram evidências aceitáveis em um processo jurídico (REICHERT; RICHARDS; YOSHIGOE, 2015). A cadeia de custódia, um processo de coleta e armazenamento de evidências que visa garantir que a evidência não foi alterada, destruída ou manipulada por pessoas não autorizadas, é pouco abordada nas soluções existentes hoje.

A solução destes problemas 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 foi despejada do pool e teve seus recursos liberados. Como a maioria das soluções de coleta hoje se baseiam em isolamento da máquina virtual afetada ou no armazenamento de informação de log, comprovar esta hipótese

nos dá a oportunidade de fechar uma forma de se conseguir impunidade em crimes cometidos na nuvem.

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 ao 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'.

Esta pesquisa tenta resolver os 2 maiores problemas da forense em nuvem hoje sugerindo uma mudança no paradigma da coleta de informações de memória das máquinas. Os desafios apresentados neste capítulo mostram que as abordagens atuais estão se esgotando e acompanhar a evolução da nuvem requer uma mudança nos procedimentos de coleta.

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 de gerar as mesmas evidências quando executado novamente por outra pessoa 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 temos:

Objetivo geral: Provar que é possível coletar dados de memória de processos em máquinas na nuvem de modo que estes atendam aos 3 requisitos legais mencionados anteriormente.

Objetivos específicos:

1. Conseguir reproduzir o processo de coleta mesmo se a máquina de onde se originaram os dados tiver sido despejada e seus recursos liberados.
2. Descobrir o mínimo de informação necessária antes do incidente de modo a viabilizar a análise.
3. Conseguir realizar a coleta sem violar jurisdição e privacidade de usuários não relacionados a investigação mantendo a cadeia de custódia.
4. Conhecer a taxa de erro do processo.

3 Método

Para atingir os objetivos:

- **Conseguir reproduzir o processo de coleta mesmo se a máquina de onde se originaram os dados tiver sido despejada e seus recursos liberados**
- **Conhecer a taxa de erro do processo**

é preciso construir os seguintes artefatos:

1. Construir uma máquina virtual e desenvolver uma aplicação que serão o alvo da coleta. A máquina virtual será construída usando programas de gerenciamento de configuração como Puppet (WIKIPEDIA, 2015b) ou Vagrant(WIKIPEDIA, 2016b) e a aplicação será instalada na máquina virtual usando um programa automatizador de instalação baseado em conteiner como o Docker (WIKIPEDIA, 2016a). A aplicação será uma página web dinâmica hospedada em um servidor Nginx (WIKIPEDIA, 2015a) na máquina alvo que responde um HTML contendo a palavra 'ACK' para qualquer requisição.
2. Desenvolver uma ameaça simples que altera a memória da aplicação alvo da coleta.
3. 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 resultado seja o mesmo.

   Com os artefatos construídos acima conduzir o experimento:

   a) Criar a máquina virtual.
   b) Executar a aplicação de coleta de evidência.
   c) Aplicar a ameaça.
   d) Despejar a máquina virtual e liberar seus recursos.
   e) Identificar e guardar a evidência.
   f) Repetir os passos anteriores.

   A medição: Estamos interessados em avaliar o quão idênticas são as evidências coletadas no processo. É esperado que haja alguma diferença pela natureza dinâmica da memória.

   Formalmente não há um limite máximo ou mínimo de erro aceitável no processo de coleta. No procedimento jurídico, a evidência é anexada junto com a descrição do

método de coleta e o erro associado onde então, caso uma das partes entenda ser necessário, pode ser contestada. Assim sendo a medição do erro da coleta será feita comparando a evidência coletada com as coletadas anteriormente. Esta comparação será feita bit a bit e o resultado será o cálculo da % de bits diferentes entre as duas coletas. Cada nova evidência gerada será comparada com as coletas anteriores e a média das comparações será a **taxa de erro do processo de coleta**.

Para atingir os objetivos:

- **Descobrir o mínimo de informação necessária antes do incidente de modo a viabilizar a análise**
- **Conseguir realizar a coleta sem violar jurisdição e privacidade de usuários não relacionados a investigação mantendo a cadeia de custódia**

é preciso construir o seguinte artefato:

1. O diagrama de um arcabouço para transportar a evidência da máquina virtual alvo construida para o experimento anterior até um local de armazenamento fora da nuvem. Este transporte será realizado via TLS e o ambiente de destino da evidência terá o acesso controlado por chave privada.

   A validação do arcabouço: O trabalho do perito é regido por uma norma que trata, entre outras coisas, dos procedimentos que devem ser executados para coleta. Esta norma no Brasil é o Código de Processo Penal que foi instituído pelo decreto lei Nº 3689 de 3 de outubro de 1941. O Código Penal Brasileiro em sua Parte Especial tipifica uma série de crimes digitais mas assim como o Código de Processo Penal possuem brechas (BARBARA, 2014) pois não evoluíram tão rápido quanto a tecnologia. Com isso, aferir se a solução atende aos objetivos mencionados depende de opinião de peritos assim, temos as seguintes ações:

- Entrevistar peritos forenses, apresentá-los o arcabouço de armazenamento e coletar opiniões sobre a sua aceitabilidade.
- Entrevistar peritos forenses, apresentá-los a taxa de erro do processo e coletar opiniões sobre a sua aceitabilidade.
- Entrevistar novamente peritos forenses, apresentá-los a estratégia de armazenamento mínimo e coletar opiniões sobre a sua aceitabilidade.

4 Revisão bibliográfica

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

- Acessar e coletar as informações, de memória ou não, das máquinas virtuais em nuvem.
- Conformidade com os seguintes requisitos jurídicos:
  - Reproduzir o processo com os mesmos resultados e conhecer a taxa de erro.
  - Não violar privacidade de outros usuários de nuvem ou jurisdição de outros países.
  - Garantir a cadeia de custódia da evidência.

Com a revisão 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) e (SANG, 2013) focam 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. Esta capacidade é necessária para viabilizar a abordagem 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 propõe que o dado seja armazenado no próprio dispositivo o que pode levar a perda de informações importantes caso a máquina virtual 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 parar a máquina virtual e em seguida executar o processo de coleta. Nos dois casos, nenhum dos autores cobre o cenário onde a máquina virtual 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 suficientes 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 virtual 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 jurisdição e 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 serviços de nuvem para conseguir as informações necessárias à investigação. Depender do provedor de serviços de nuvem é uma estratégia fraca pois o foco deles é garantir a continuidade do serviço não a coleta de evidências.

Na garantia da cadeia de custódia apenas (SANG, 2013) tenta resolver a questão mas toma cuidados apenas para garantir que a evidência não foi destruída ou alterada. Ele faz isso através 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 integridade e a confidencialidade da evidência. O volume de informações armazenadas 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.

6 Cronograma

O plano de trabalho é descrito pelas seguintes atividades e organizado de acordo com o apresentado na Figura 1. Este cronograma pode sofrer alterações ao longo da realização da pesquisa.

- Revisão bibliográfica
- Acompanhamento do estado da arte
- Estudo das aplicações baseadas em container
- Implementação da receita para a máquina virtual
- Implementação da aplicação de coleta
- Implementação da ameaça
- Execução do experimento de coleta
- Analise dos resultados e ajustes na aplicação de coleta
- Documentação do arcabouço
- Entrevista com peritos
- Ajuste no arcabouço
- Escrita e defesa de tese.

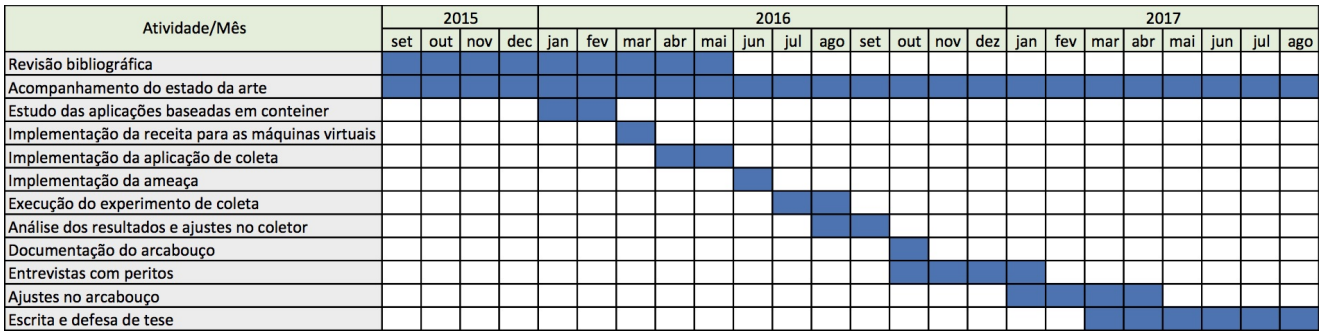

Figura 1 – Fonte: Autor

Referências

AMAZON. *Amazon Media Room Press Release*. [S.l.], 2016. 2 p.

BARBARA, D. *Desafios da perícia forense em um ambiente de computação nas nuvens*. [S.l.], 2014.

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.

WIKIPEDIA. *NginX*. 2015. Disponível em: <https://pt.wikipedia.org/wiki/Nginx>.

WIKIPEDIA. *Puppet*. 2015. Disponível em: <https://pt.wikipedia.org/wiki/Puppet>.

WIKIPEDIA. *Docker*. 2016. Disponível em: <https://en.wikipedia.org/wiki/Docker{_}{%}28software{%}29>.

WIKIPEDIA. *Vagrant*. 2016. Disponível em: <https://en.wikipedia.org/wiki/Vagrant{_}{%}28software{%}29>.