Hamilton Fonte II

Coletando dados de memória de uma máquina em nuvem para análise forense

São Paulo, Brasil

2016, v-0.1

Hamilton Fonte II

Coletando dados de memória de uma máquina em nuvem para análise forense

Projeto de pesquisa para a disciplina Metodolodia de Pesquisa Científica em Engenharia de Computação.

Universidade de São Paulo – USP

Escola Politécnica - Engenharia de Computação

Programa de Pós Graduação em Engenharia Elétrica - Mestrado

Orientador: Marcos Antonio Simplício Jr

São Paulo, Brasil

2016, v-0.1

Sumário

1 Introdução

Aqui vai a introdução

2 Justificativa

Aqui vai a justificativa

3 Objetivos

Aqui vão os objetivos

4 Métodos

Aqui vão os métodos

5 Revisão Bibliográfica

- **Digital forensics framework for a cloud environment (GEORGE; VENTER; THOMAS, 2012)** : Framework para coleta de dados de máquinas virtuais. Tem duas formas de acionamento, a manual e outra automaticamente, integrada com algum sistema de detecção de ameaça. Quando acionado, escuta a rede, determina qual a máquina é objeto de investigação e coleta informações de log e tráfego de rede e associa ao usuário das respectivas máquinas. Propõe o armazenamento das evidências em local fora da núvem para escapar de problemas de jurisdição e multi-inquilino mas tem inteligência para usar a própria nuvem como armazenamento caso o espaço fora da nuvem acabe.

  A proposta dá a entender que é aplicável apenas a um sistema virtual estático, onde o número e organização das máquinas é constante. De informação volátil, coleta apenas tráfego de rede, não coleta memória. Com a forma de acionamento descrito ele não consegue descrever, com as evidências, como era o sistema antes do ataque. Apesar de armazenar a evidëncia fora da nuvem, não dá detalhes de cadeia de custódia, garantia de integridade e confidencialidade. Esta abordagem tem um viés mais técnico, leva poucas questões jurídicas em consideração.

  De melhor, eu proponho autilização de container para associar a evidência a sua origem tornando o processo independente de máquina. Como eu implemento uma janela de x dias de coleta antes da detecção do ataque eu consigo descrever, através de evidência como era o sistema antes do mesmo. Com isso consigo evidências em um cenário de infra-estrutura dinâmica. Eu tomo precauções para garantir a integridade e confidencialidade dos dados transportando via TLS os dados para um local fora da nuvem e estabelecendo um controle de acesso.

- **Evidence and cloud computing the virtual machine introspection approach (POISEL; MALZER; TJOA, 2013)** : Descreve um método 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 automaticamente por demanda integrado a um sistema de detecção de ameaça mas também suporta acionamento manual.

  A técnica descrita cobre apenas o processo de coleta de informações, não explica onde ou como elas serão armazenadas. 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. Segundo a comunidade, tal estratégia é imune a

técnicas anti- forenses empregadas por usuários maliciosos pois esta localizada fora da máquina virtual. Como a abordagem não tem conhecimento do que esta rodando dentro da máquina virtual precisade uma copia 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.

De melhor, eu proponho um arcabouço para coleta e armazenamento de evidências, não apenas a coleta. Minha abordagem usa uma estratégia diferente pois coleto a memória diretamente dentro da máquina virtual. Nessa estratégia eu evito o problema do gap semântico típico das soluções por introspecção, não preciso realizar tradução de endereços de memória para viabilizar a análise forense mas, de acordo com a comunidade fico mais sucetível a técnicas anti-forenses. Minha abordagem teria também a vantagem de conhecer o que esta rodando dentro da máquina e assim ter mais eficiência na coleta.

- **Design and implementation of FROST: FoRensic tools for Open STack (DYKSTRA; SHERMAN, 2013)** : Framework para coleta de dados de máquinas virtuais através da API do hypervisor. Isola a máquina virtual afetada do pool original para realização da coleta. Precisa ser acionado quando uma ameaça é detectada. É o mais bem acabado arcabouço de todas as propostas encontradas até agora mas ainda assim é uma proposta técnica, faz breve menção a aspectos jurídicos. Por estar integrado ao Open Stack o arcabouço depende de cooperação do provedor de serviços de núvem onde ele esta rodando, isso é considerado problemático pela comunidade pois a prioridade do mesmo é manter o serviço funcionando e não coletar evidencias forenses. Como roda no hypervisor não conhece o que esta rodando dentro da máquina. Depende da existência da máquina virtual para realização da coleta.

  De melhor, eu proponho a utilização de container para associar a evidência de memória a sua origem tornando o processo independente de máquina e a janela de x dias antes da detecção do ataque para conseguir descrever o sistema antes do mesmo. Minha proposta não depende de cooperação do provedor do serviço de nuvem. Minha abordagem teria também a vantagem de conhecer o que esta rodando dentro da máquina e assim ter mais eficiência na coleta.

- **Automated Forensic Data Acquisition in the Cloud (REICHERT; RICHARDS; YOSHIGOE, 2015)** : Propõe um modelo que tira snapshots de máquinas virtuais atrelado a algum mecanismo de detecção de ameaça baseado no hypervisor. Usa o Google Rapid Response para salvar as informações coletadas fora da núvem de forma a driblar os problemas de multi-jurisdição e multi-inquilino. Descreve satisfatóriamente a cadeia de custódia da evidência.

O modelo proposto só começa a coletar evidência após a detecção da ameaça, e toma um snapshot da máquina toda o que já foi julgado pela comunidade como um processo custoso em termos de espaco em disco pois piora o problema do volume de dados a ser analisado. Pessoalmente acho ariscado depender de snapshots pois caso precise, repetir o processo de coleta pode não ser possível. Um exemplo é editar um HD virtual que esta atrelado a uma máquina virtual da qual se gerou snapshots, isso pode gerar perda de dados.

De melhor eu proponho apenas a coleta de informações de memória e uso a janela de coleta de x dias antes do ataque para manter sob controle a quantidade de informação que precisa ser analisada. Eu proponho autilização de container para associar a evidência a sua origem tornando o processo independente de máquina.

- **A log based approach to make digital forensics easier on cloud computing (SANG, 2013)** : Método que sugere salvar a informação coletada fora da núvem 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 e não descreve como controla o acesso a evidência armazenada. Segundo o próprio autor, o método não funciona em IaaS. Precisa da cooperação do provedor de nuvem pois depende das informações que esta último decidiu adicionar ao log. O método não é aplicável a coleta de informações de memória.

  A proposta não coleta dados de memória por decisão do autor, esta proposta entrou na lista pela abordagem baseada em log. Neste quesito, minha proposta é melhor no que garante todos os pontos da cadeia de custódia, integridade e confidencialidade. No âmbito da informação coletada eu não dependo das decisões do que guardar no log do provedor de serviço de nuvem para conseguir a evidência.

- **Volatile memory acquisition using backup for forensic investigation (DEZFOULI et al., 2012)** : Técnica desenvolvida para dispositivos móveis que sugere a utilização do próprio como repositório das evidências coletadas da memória. Para manter a utilização de espaço ao mínimo sugere manter apenas o último estado conhecido da memória.

  É uma técnica interessante do ponto de vista de estratégia de armazenamento quando guarda apenas o último estado da memória. Essa abordagem porém perde a informação do momento do ataque e não consegue descrever o sistema antes do mesmo. Do resto da proposta não é aplicável para este projeto pois, armazenando a evidência na máquina a mesma seria perdida quando a máquina fosse despejada do pool e seus recursos liberados. Outro aspecto é a cadeia de custódia que não é abordada na

proposta.

- **Narrowing the semantic gap in virtual machine introspection (DOLAN-GAVITT et al., 2011)** : Esta proposta é uma combinação da técnica de introspecção em máquina virtual e a integração com a API do hypervisor. O principal objetivo é diminuir o gap semântico para facilitar a análise da evidência. Para isso o autor implementa um API para transformar dados de baixo nível em informação de alto nível. Depende de cooperação do provedor de serviço de nuvem, não tem conhecimento da máquina hospedada e não vai além da coleta, não descreve como resolve a cadeia de custódia. Tem a vantagem de ser imune a técnicas anti-forenses.

  De melhor eu proponho um arcabouço de coleta e armazenamento de evidências não apenas um método de coleta. Como vou diretamente na máquina não tenho o problema do gap semântico e como conheço o contexto do que esta rodando dentro da máquina virtual sou mais eficiênte na coleta.

- **Digital Forensics as a Service: A game changer (BAAR; BEEK; EIJK, 2014)** : Esta proposta é focada em uma mudança no armazenamento e forma de trabalho dos peritos forenses. Propoe que a forense seja oferecida como um serviço e que as evidências sejam armazenadas em um local centralizado com o devido controle de acesso e garantia de integridade da evidência. Descreve a arquitetura de armazenamento, qual o perfil que deve ter acesso a evidência e como é este acesso. Esta proposta não é focada apenas em incidentes na nuvem mas em qualquer outro incidente.

  Embora seja uma ótima proposta de armazenamento de evidências e controle de acesso a elas, ele não descreve o processo de coleta nem de transporte. É uma proposta focada mais na solução de problemas relacionados a manipulação da informação após a coleta, transporte e armazenamento. Não toca no assunto de coleta qualquer que seja, em nuvem ou fisica.

- **Live Digital Forensics in a Virtual Machine (ZHANG; ZHANG; WANG, 2010)** : Proposta para coletar informações de memória de máquinas virtuais através de snapshots das mesmas. O metodo de coleta envolve tomar o snapshot da máquina, no diretório onde o snapshot foi armazenado pegar o arquivo referente a memória e abri-lo / analisá-lo usando um programa de leitura de memória de mercado. O autor não trás detalhes do transporte, armazenamento ou cadeia de custódia. Precisa que a máquina exista para conseguir coletar a informação e o processo é dependente de intervenção humana. Analisando com mais cuidado é possível repetir a coleta mesmo

sem a máquina existir uma vez que temos o snapshot mas o autor não dá detalhes do caso.

De melhor, eu limito a quantidade de informação necessária a investigação através da implementação da janela de x dias antes do incidente e o meu processo é automático, uma vez disparado não requer intervenção humana. Descrevo como garanto a cadeia de custódia da evidência e consigo reproduzir o processo mesmo se a máquina não existe mais pois a evidência esta atrelada ao container.

6 Análise dos Resultados

Aqui a análise dos resultados

Referências

BAAR, R. B. van; BEEK, H. M. A. van; EIJK, E. J. van. Digital Forensics as a Service: A game changer. *Digital Investigation*, Elsevier Ltd, v. 11, p. S54–S62, 2014. ISSN 17422876. Disponível em: <http://dx.doi.org/10.1016/j.diin.2014.03.007>. Citado na página 10.

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. Citado na página 9.

DOLAN-GAVITT, B. et al. Virtuoso: Narrowing the semantic gap in virtual machine introspection. *Proceedings - IEEE Symposium on Security and Privacy*, p. 297–312, 2011. ISSN 10816011. Citado na página 10.

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>. Citado na página 8.

GEORGE, S.; VENTER, H.; THOMAS, F. Digital Forensic Framework for a Cloud Environment. p. 1–8, 2012. Citado na página 7.

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>. Citado na página 7.

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. Citado na página 8.

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. Citado na página 9.

ZHANG, L.; ZHANG, D.; WANG, L. Live Digital Forensics in a Virtual Machine. In: *2010 Internation Conference on Computer Application and System Modelling (ICCASM 2010)*. [S.l.: s.n.], 2010. v. 6, p. 328–332. Citado na página 10.