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

São Paulo, Brasil

2016, v-0.1

Sumário

1 Introdução

Os crescentes avanços tecnológicos que têm ajudado a vida do ser humano também trouxeram diversos novos tipos de crimes relacionados a tecnologia da informação. (citacao) Quando do evento de um crime, há a necessidade de preservar as evidências de modo que apoiem processos legais e investigativos mas também para que possamos compreender a sequência de eventos envolvidos no ataque de modo a evitar que as mesmas vulnerabilidades sejam exploradas novamente.

Forense Digital é uma ferramenta importante para a solução de crimes cometidos com um computador pela sua capacidade de reconstruir a cena de um crime a partir das evidências digitais deixadas por um ataque. A forense digital evoluiu nos últimos anos de uma fase ad hoc, passando por uma fase de adaptação aos requisitos legais até a era do ferramental avançado de coleta e análise (citação). Com o crescente uso de ferramentas online, virtualização e hospedagem em nuvem, o volume de dados gerado por esses sistemas que já era uma preocupação da comunidade forense em 2008 (citação), hoje começou a inviabilizar algumas de suas práticas (citação).

Com a funcionalidade da elasticidade de carga ofertada pelos provedores de nuvem por meio da qual infraestrutura pode ser alocada e desalocada dinamicamente, veio à tona o problema da volatilidade dos dados nas máquinas virtuais. Com algumas ameaças que não deixam evidências em disco(citação), a memória de uma máquina despejada de um pool seria para sempre perdida. Com a máquina mais poderosa hoje ofertada pela Amazon tendo 240Gb de memória (citação), o ferramental forense disponível hoje esta pouco adaptado a desafios trazidos pela nuvem (citação). A forense digital precisa evoluir.

2 Justificativa

A tecnologia hoje é parte de nossas vidas e também é parte da vida dos criminosos. O último crime registrado na America do Norte que não envolveu uma evidência digital data de 2011 (citação), no Brasil só em 2015 houve um aumento de 87% no registro de notas relacionadas a abusos ou crimes digitais (citação).

Do ponto de vista forense, praticantes e pesquisadores concordam que aspectos legais e o crescente volume de dados figuram entre as principais dificuldades trazidas pela computação em nuvem (citação). A maioria das técnicas forenses atuais são da época pré-nuvem e algumas características da nuvem estão inviabilizando tais práticas. O aspecto multi inquilino impede a remoção do hardware pois o mesmo é compartilhado com outros usários, removê-la seria uma violação da privacidade de usuários não relacionados a investigação. A característica distribuída pode alocar informação relevante à investigação em vários países dificultando a obtenção da mesma (citação). Por último, o crescente volume de dados envolvidos nas aplicações atuais deram aos investigadores 6 a 12 meses de backlog de dados para investigar (citação), download de terabytes de dados para análise leva horas para ser realizado e requer colaboração dos operadores de nuvem (citação). Mudar o paradigma de coleta de evidência tem sido proposto por pesquisadores e praticantes há alguns anos (citação), porém esta prática precisa submeter a legislação para que seja aceita em um processo legal.

Por fim, a criatividade dos criminosos para explorar vulnerabilidades trouxe a tona ataques que não deixam evidências em disco, toda a operação ocorre em memória (citação). Com o caráter elástico das soluções em nuvem, uma maquina despejada de um pool e cujos recursos foram reutilizados não pode mais prover as evidências necessárias para investigação criminal e dependendo da importâcia da evidência o criminoso fica impune. Coletar a memória de processos de máquinas em núvem de modo que sejam usadas com sucesso em um processo penal, sem sobrecarregar ainda mais os investigadores é a principal justificativa desta pesquisa.

A coleta de dados de memória não pode ser realizado diretamente de dentro da máquina virtual pois o próprio processo de leitura da memória a altera. A principal técnica de coleta de informações de memória é a Virtual Machine Introspection (citacao), nesta técnica a máquina é pausada e via hypervisor realiza-se a leitura da memória e em seguida a tradução dos endereços de memória virtual para o real. As desvantagens desta implementação estão na necessidade de tradução dos endereços de memória, para realização da análise e necessidade de credenciais pra acesso e integração com o hypervisor. (citacao). No que tange o volume de dados baixados temos as técnicas da Ïproof of retrieveabilityë

R̈proof of Data Possessionq̈ue se apoiam na transmissão e comparação de hashes gerados a partir do conteúdo que se quer coletar porém o uso destas requer quer a informação tenha baixa volatilidade o que não é o caso da informação contida em memória temporária. Na frente legal, as abordagens atuais de live forensics onde esta pesquisa se concentra carecem de credibilidade por não produzirem evidências aceitas em um processo legal.(citação).

3 Objetivos

- Objetivo Principal
  - O objetivo principal desta pesquisa é derivar um arcabouço que permita coletar a informação da memória de processos de máquinas em nuvem de modo que seja aceita em um processo legal, sem violar jurisdição internacional e a privacidade dos usuários de nuvem não relacionados a investigação. Com isso confirmando as hipóteses relacionadas a este trabalho.
    - Hipotese 1: É possível coletar dados de memória de processos em máquinas na nuvem de modo que se submetam com sucesso ao princípio de Daubert.
    - Hipotese 2: É possível realizar a coleta sem violar a privacidade de usuários da nuvem não envovidos no caso investigado.
    - Hipotese 3: É possível coletar tais dados sem envolver jurisdições de outros países.

- Sub Objetivos
  - Coletar apenas a memória das máquinas involvidas na investigação
  - Construir um método de coleta que produza erros menores que ...
  - Gerar evidências que se submetam com sucesso ao princípio de Daubert

4 Plano de Trabalho

- reescrever isso
  - implementação das imagens para realização de testes.
  - implementação dos containers a partir das imagens acima para realização dos testes.
  - verificar viabilidade do relacionamento imagem-container-máquina.
  - gerar primeiro lote de evidências.
  - buscar opiniões jurídias sobre a evidência gerada.
  - buscar % de erro na geração da evidência.
  - otimizar a geração da evidência para diminuir a taxa de erros

5 Material e Métodos

- reescrever isso
  - Implementação de virtualização de mercado.
  - Rede e laboratório para salvar informações coletadas.
  - Conta no dockerhub
  - laptop para [especificação aqui] desenvolvimento da solução

6 Análise dos Resultados

- reescrever isso
  - avaliando as opiniões sobre a validade da evidência por representantes legais brasileiros.
  - verificando se a evidência se encaixa no principio de Daubert.
  - repetir o processo de coleta e verificar o nível de erro.

Referências