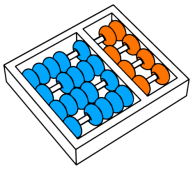
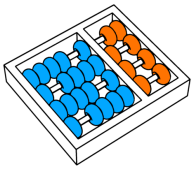


Alexandre Or Cansian Baruque

CAMPINAS
2015



Universidade Estadual de Campinas
Instituto de Computação

Alexandre Or Cansian Baruque

Visualização de dados de honeypots

Orientador(a): **Prof. Dr. Paulo Licio de Geus**

Co-Orientador(a): **Prof. Dr. André Abed Grégio**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Computação da Universidade Estadual de Campinas para obtenção do título de em Ciência da Computação.

Este exemplar corresponde à versão da Dissertação apresentada à banca examinadora por Alexandre Or Cansian Baruque, sob orientação de Prof. Dr. Paulo Licio de Geus.

Assinatura do Orientador(a)

CAMPINAS
2015

Instituto de Computação
Universidade Estadual de Campinas

Visualização de dados de honeypots

Alexandre Or Cansian Baruque

Banca Examinadora:

- Prof. Dr. Paulo Licio de Geus ()
-
-
-

Resumo

O projeto tem como objetivo a implantacao de um sistema para o tratamento e visualização de dados coletados por meio de sensores participantes de uma infra-estrutura de coleta de programas maliciosos.

Esse sistema foi desenvolvido fazendo o uso de tecnologias Web para o desenvolvimento de um servico de visualizacao dos dados que correlaciona a posição geográfica da origem do potencial ataque com os dados dos ataques e data e hora da coleta.

Isto torna possível que um analista que faz uso das funcionalidades integradas no sistema consiga identificar padrões e tendências das atividades maliciosas registradas, tirando assim conclusoes acerca dos dados observados.

Sumário

Capítulo 1

Introdução

Programas maliciosos (malware) representam uma das maiores ameaças aos usuários e sistemas conectados a Internet. A facilidade da geração de novos exemplares a partir de malware já existente, as chamadas variantes, faz com que a atuação dos mecanismos de defesa seja dificultada, seja na criação de vacinas, seja no processamento das centenas de milhares de variantes em operação. Por exemplo, no caso de antivirus com detecção baseada em assinaturas, e necessário em geral a atividade de um analista humano que descubra um modo de detecção para um dado malware e produza a assinatura em questão. Para tanto, o analista precisa ter em mãos a variante ainda não detectada. A fim de obter programas maliciosos e suas variantes, necessita-se de uma infraestrutura de coleta distribuida. A motivacão principal para tal infraestrutura e a de obter e armazenar exemplares de malware em circulação no Brasil para análise posterior. Com isso, pretende-se obter informacões acerca dos tipos de malware que atacam as redes que possuem um sensor de coleta (honeypot) a fim de agregar dados que permitam mostrar as tendências de ataques de uma maneira local (em cada rede) e global (conjunto das redes com sensores instalados). As informações obtidas desses honeypots de coleta incluem trafego de rede, registros de auditoria (logs) das vulnerabilidades exploradas e tipos de ataques utilizados no comprometimento dos serviços/sistemas, e do binário que representa o programa malicioso. De posse de tais informações, e possível representa-las por meio de técnicas de visualização e disponibiliza-las por meio de uma arquitetura Web, permitindo a observação de tendências e geração de estatisticas.

Capítulo 2

Trabalhos relacionados

ipviking honeynet

Capítulo 3

Atividades desenvolvidas

Para alcancar os objetivos propostos no projeto, as seguintes atividades foram desenvolvidas:

1. Estudo de conceitos sobre visualizacao de dados atraves do uso de plataformas de desenvolvimento Web, como Django e Javascript. Em especifico, foi estudado extensivamente o uso de bibliotecas para D3 (data driven documents) e Jvectormap.

2. Definicao da arquitetura utilizada para a implantacao da infra-estrutura basica utilizada pelo servidor Web utilizado para a visualizacao dos dados e em outros sistemas necessarios para a administracao dos servicos.

3. Instalacao das ferramentas e outros programas (e suas dependencias) necessarios para a implementacao de um prototipo em uma maquina desktop provida pelo orientador.

Ao longo de todas as atividades desenvolvidas foi tambem feita a documentacao do projeto, assim como outros relatorios e atividades requeridas (demonstracao, levantamento de dados, etc).

Capítulo 4

Metodologia

Houve uma fase preliminar na qual foi adquirido conhecimento para tomar decisoes acerca da arquitetura inicial do projeto, tais como escolha de ferramentas e plataformas de desenvolvimento a serem utilizadas. Para tanto, observou-se as ferramentas e sistemas para visualizacao de dados de seguranca ja existentes, entre as quais as disponiveis pelos Projetos Honeynet, Viking e DioneaFR, visto que os objetivos e funcionalidades providas por essas ferramentas possuem componentes similares aos propostos neste projeto. Entretanto, tais ferramentas/sistemas sao limitados em certos aspectos, por exemplo, tipo e abrangencia de dados, necessidade de compartilhamento de informacoes remotamente, entre outros, o que gera a necessidade de desenvolvimento de um sistema interno, de acesso controlado e extensivel para os requisitosespecificos dos dados coletados. Apos essa analise inicial e tendo definido as ferramentas a serem utilizadas, o enfoque do estudo foi em adquirir proficiencia nas plataformas de desenvolvimento e ferramentas escolhidas para uso no projeto: Django, Jquery, Jvectormap, D3js, GeoIP, iNotify. Para o controle de versao dos componentes de codigo produzidos ao longo do projeto foi utilizado o sistema "git". Tais codigos sao armazenados em um repositorio privado criado no GitHub atraves de uma conta de estudante habilitada pelo bolsista. O repositorio consiste de dois ramos: o ramo mestre, no qual se mantem a versao mais atual do projeto com suas funcionalidades mais recentes, e o ramo demonstracao, onde fica a ultima versao estavel com o objetivo de ser utilizada para demonstracoes funcionais do prototipo do sistema. O framework Django em linguagem Python foi escolhido para constituir da infra-estrutura principal do projeto, tendo como funcao prover servicos de administracao no back-end, assim como servir o codigo em JavaScript a ser executado pelo cliente em um navegador Web. O Django tambem permite um facil gerenciamento da base de dados (sqlite3) atraves da criacao de procedimentos para administracao. Foram desenvolvidas rotinas em shell script para consolidar os dados coletados na base de dados por meio dos procedimentos mencionados. Uma arquitetura REST (Representational State Transfer) foi implantada fazendo uso de funcionalidades do Django, implicando que o acoplamento entre o cliente e o servidor nao seja tao rigido, permitindo assim

que modificacoes sejam feitas em ambos os lados sem muito impacto ao restante do sistema. O sistema implementado sobre tal arquitetura realiza consultas na base de dados e serializa os dados para serem encaminhados ao cliente executando codigo JavaScript no formato JSON. Para solucionar o problema de relacionar o endereco IP de origem do ataque com sua posicao geografica no mundo real, utilizou-se a ferramenta GeoIP, que permite baixar uma base relacional de IPs e latitude/longitude. Com isso, e possivel traduzir a maioria dos enderecos para localizacoes no mapa, obtendo tambem o pais e a cidade de origem. A visualizacao dos dados pelo cliente foi feita pelo uso de bibliotecas de codigo aberto baseados em JavaScript, sendo elas: jVectorMap, que permite a facil criacao e modificacao de mapas interativos; D3js para a geracao de graficos contendo estatisticas gerais sobre os dados coletados; jQuery, que permite a criacao de elementos de interface para prover interacao com o usuario. Um prototipo do sistema foi instalado em uma maquina provida pelo orientador, a qual ja continha dados de seguranca coletados em um projeto envolvendo outro bolsista (infraestrutura distribuida para coleta e analise de dados de honeypots). Nesta maquina, os procedimentos previamente instalados fazem o polling dos dados coletados por sensores distribuidos e os agrega em uma estrutura no sistema de arquivos. Com todos os sistemas em funcionamento, o servico Web que consome e processa tais dados atualiza sua informacoes de forma autonoma, provendo a visualizacao desejada.

Capítulo 5

Resultados

O resultado final foi um prototipo de uma sistema cuja ferramenta principal permite a visualizacao de dados de potenciais ataques a protocolos de rede e aplicacoes de sistema, bem como de servidores comprometidos realizando o provimento de codigos maliciosos para sensores distribuidos. A visualizacao desses dados e feita em um mapa, no qual sao correlacionadas a origem do ataque (coordenadas geograficas do endereco IP obtido) com a frequencia das ocorrencias do ataque em questao (raio do circulo impresso em sua coordenada especifica). O mapa permite observar os dados sobre os ataques em uma dimensao temporal, atualizando as informacoes de acordo com o horario no qual tais dados foram coletados pelos sensores. Dessa forma, pode-se observar as tendencias dos tipos de ataques no decorrer do dia, analisando-se os protocolos, servicos e locais de origem. Desenvolveu-se uma imagem em maquina virtual para VirtualBox. A configuracao desta maquina virtual ja esta corretamente aplicada com todos os requisitos para o funcionamento adequado do servidor Web utilizado no provimento da visualizacao de dados formatados de acordo com os requisitos do projeto. O desenvolvimento da maquina virtual citada tem como objetivo armazenar a ferramenta de uma forma portatil para ser utilizada em demonstracoes. A Figura abaixo representa um mapa em um dado instante de tempo com dados de potenciais ataques apresentados de acordo com a localizacao geografica do endereco IP obtido, bem como outras informacoes (vide quadro escuro na America do Norte), tais como, a cidade estimada por meio das coordenadas obtidas do IP, o tipo do ataque (servico alvo) e a porta de rede vulneravel que recebeu o ataque. Logo abaixo do mapa, pode-se notar uma barra de rolagem, a qual movimenta-se de forma automatica para mostrar as tendencias de ataque ao longo do dia e, mais abaixo, um grafico que representa os protocolos monitorados e a frequencia das conexoes recebidas por estes protocolos no instante de tempo em questao.

Capítulo 6

Considerações finais

Conforme apresentado na secao anterior, foram obtidos resultados concretos no projeto durante o periodo da bolsa. Entretanto, existem atividades importantes a serem feitas para os futuros trimestres, como previsto no plano de trabalho, as quais irao permitir a obtencao de informacoes de mais qualidade do ponto de vista de seguranca. As principais propostas dessas atividades consistem do desenvolvimento de procedimentos para o tratamento dos dados obtidos, agregacao de outros tipos de dados (provenientes de outros sensores) e a correlacao de dados de diferentes enderecos de origem e fontes visando a descoberta de ataques em associacao. Alem disso, ha espaco para melhorias no que se refere a usabilidade da ferramenta, entre as quais incluem- se: uma melhor escolha de cores para a representacao dos dados no mapa, melhor disposicao do grafico de protocolos, apresentacao de mais informacoes sobre os potenciais ataques e uma interface mais avancada para a manipulacao da dimensao temporal dos dados, permitindo ao analista definir intervalos de tempo para apresentar os dados com o objetivo de identificar padroes com maior facilidade. Por ultimo, pretende-se buscar uma melhoria na performance do sistema em relacao ao tempo de resposta entre a busca dos dados no banco de dados e a renderizacao dos pontos no mapa. Isto e possivel por meio do uso de cache e outras tecnicas de otimizacao disponiveis na plataforma Django, mas tais funcionalidades estavam alem do escopo do prototipo desenvolvido ate o momento.

Capítulo 7

Conhecimentos adquiridos

7.1 Aplicabilidade

Os conhecimentos adquiridos durante esse projeto possuem uma ampla area de aplicacao, sendo esta qualquer uma que necessite uma forma visual de interpretar dados. Alem disso, a experiencia que obtive com a criacao de servicos Web e utilizacao/programacao em plataformas e arquiteturas diferentes nao possui nenhuma restricao quanto a area de aplicacao, podendo ser util tanto em pesquisas futuras quanto em trabalhos de desenvolvimento.

7.2 Incorporacao de novas tecnicas

Neste projeto adquiri intimidade com o estilo de programacao de APIs no padrao REST, que permite o desenvolvimento de aplicativos Web fracamente acoplados, possibilitando assim a criacao de servicos Web escalaveis e capazes de se adaptar a mudanca de requisitos de forma flexivel. Alem disso, familiarizei-me com a plataforma Django, que tem sido muito utilizada em diversos meios, do academico ao industrial, para criacao de sistemas Web com diversas caracteristicas integradas e possibilidade de criacao de modulos em Python.

7.3 Geracao de produtos e processos

Ao longo do projeto desenvolvi um cliente em JavaScript que permite a visualizacao de dados nao especificos em um mapa de forma generica, podendo assim ser reutilizado para futuros projetos envolvendo representacao de dados nos quais sua posicao geografica possui importancia. Adquiri conhecimento para sobre o processo de geracao de imagens em Virtualbox, sendo util para encapsular sistemas em um ambiente virtual controlado, tornando-o assim portavel. Este processo de criacao de imagens executaveis (appliances) serve para a preparacao de demonstracoes de qualquer tipo,

tornando- as mais confiaveis ao custo de perda de performance, porem permitindo a disponibilizacao e utilizacao de maneira facilitada.

7.4 Contribuicao da participacao no projeto para sua formacao

Participar no projeto contribuiu para o meu conhecimento de ferramentas para rapida prototipagem de aplicacoes e implantacao de servicos Web. Tais conhecimentos nao sao restritos a area de pesquisa, mas sim tambem para o mercado de trabalho, onde no modelo atual de startups, o tempo para a execucao de um projeto e capaz de determinar o futuro da empresa. Aprendi tambem a estruturar relatorios tecnicos de maneira adequada de forma a expor ideias e mostrar resultados.