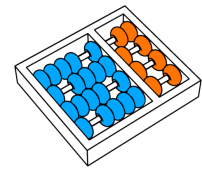


Alexandre Or Cansian Baruque

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

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 APRESEN-
TADA À BANCA EXAMINADORA POR
ALEXANDRE OR CANSIAN BARUQUE,
SOB ORIENTAÇÃO DE PROF. DR.
PAULO LICIO DE GEUS.

Assinatura do Orientador(a)

CAMPINAS
2015

Visualização de dados de honeypots

Alexandre Or Cansian Baruque¹

Banca Examinadora:

- Prof. Dr. Paulo Licio de Geus (Supervisor/*Orientador*)
-
-
-

¹Financial support: CNPq scholarship (process XYZ) 2010–2012

Abstract

The abstract must contain at most 500 words...

Observation: the text in this page *has to* fit it in. The Unicamp's printing service would reject this work if more than one page were used.

Resumo

O presente projeto tem como objetivo a implantacao de um sistema para o tratamento e visualizacao de dados coletados por meio de sensores participantes de uma infraestrutura 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 posicao geografica da origem do potencial ataque com os dados dos ataques e data e hora da coleta.

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

Sumário

Abstract	vii
Resumo	ix
1 Introdução	1
2 Related work	3
3 Atividades desenvolvidas	5
4 Metodologia	7
5 Resultados	9
6 Discussão / Considerações finais	11
7 Conhecimentos adquiridos	13
7.1 Aplicabilidade	13
7.2 Incorporacao de novas tecnicas	13
7.3 Geracao de produtos e processos	13
7.4 Contribuicao da participacao no projeto para sua formacao	14
8 Outros resultados	15

Lista de Tabelas

Lista de Figuras

Capítulo 1

Introdução

Capítulo 2

Related work

Capítulo 3

Atividades desenvolvidas

Para alcançar os objetivos propostos no projeto, as seguintes atividades foram desenvolvidas: - Estudo de conceitos sobre visualização de dados através do uso de plataformas de desenvolvimento Web, como Django e Javascript. Em específico, foi estudado extensivamente o uso de bibliotecas para D3 (data driven documents) e Jvectormap. - Definição da arquitetura utilizada para a implantação da infra-estrutura básica utilizada pelo servidor Web utilizado para a visualização dos dados e em outros sistemas necessários para a administração dos serviços. - Instalação das ferramentas e outros programas (e suas dependências) necessários para a implementação de um prototipo em uma máquina desktop provida pelo orientador. Ao longo de todas as atividades desenvolvidas foi também feita a documentação do projeto, assim como outros relatórios e atividades requeridas (demonstração, 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 requisitos especificos 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

Discussão / Considerações finais

Conforme apresentado na seção anterior, foram obtidos resultados concretos no projeto durante o período da bolsa. Entretanto, existem atividades importantes a serem feitas para os futuros trimestres, como previsto no plano de trabalho, as quais irão permitir a obtenção de informações de mais qualidade do ponto de vista de segurança. As principais propostas dessas atividades consistem do desenvolvimento de procedimentos para o tratamento dos dados obtidos, agregação de outros tipos de dados (provenientes de outros sensores) e a correlação de dados de diferentes endereços de origem e fontes visando a descoberta de ataques em associação. Além disso, há espaço para melhorias no que se refere a usabilidade da ferramenta, entre as quais incluem-se: uma melhor escolha de cores para a representação dos dados no mapa, melhor disposição do gráfico de protocolos, apresentação de mais informações sobre os potenciais ataques e uma interface mais avançada para a manipulação da dimensão temporal dos dados, permitindo ao analista definir intervalos de tempo para apresentar os dados com o objetivo de identificar padrões com maior facilidade. Por último, pretende-se buscar uma melhoria na performance do sistema em relação ao tempo de resposta entre a busca dos dados no banco de dados e a renderização dos pontos no mapa. Isto é possível por meio do uso de cache e outras técnicas de otimização disponíveis na plataforma Django, mas tais funcionalidades estavam além do escopo do protótipo desenvolvido até 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 confiáveis 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.

Capítulo 8

Outros resultados

Exemplo: Relatorios tecnicos, apostila/descricoes de metodos, normas, produto tecnologico, servico tecnologico, processos, tecnicas, sistemas, modelos, metodos, desenho tecnico e diagrama, documento de software e sistema da qualidade, especificacao de macroprocesso, especificacao de produto e de software, estudo prospectivo, instrucao para execucao de procedimento, manual de instrucoes - tecnico e usuario, folha de dados de produto, pareceres sobre artigo cientifico (revisor de periodico), parecer tecnico e laudo, relatorio de afericao, relatorio de atividades de participacao em eventos tecnicos cientificos, textos didaticos. - Tutorial para instalacao e configuracao do sistema de visualizacao usando mapas. - Relatorio tecnico das atividades desenvolvidas durante o projeto.