MINISTÉRIO DA DEFESA
EXÉRCITO BRASILEIRO
DEPARTAMENTO DE CIÊNCIA E TECNOLOGIA
INSTITUTO MILITAR DE ENGENHARIA
ENGENHARIA DE COMPUTAÇÃO - SE/8

Jonas Rocha Lima Amaro, Yago Guimarães Coimbra

Ferramenta para detecção de padrões de botnet baseado em algoritmos de agrupamento de aprendizagem de máquina

Rio de Janeiro
2 de maio de 2016

Jonas Rocha Lima Amaro, Yago Guimarães Coimbra

Ferramenta para detecção de padrões de botnet baseado em algoritmos de agrupamento de aprendizagem de máquina

Trabalho Apresentado ao Curso de Engenharia de Computação do Instituto Militar de Engenharia como Verificação Especial do Projeto de Fim de Curso.

Instituto Militar de Engenharia

Orientador: Sergio dos Santos Cardoso Silva

Rio de Janeiro
2 de maio de 2016

Instituto Militar de Engenharia
Praça General Tibúrcio, 80 - Praia Vermelha
Rio de Janeiro - RJ CEP: 22290-270

Amaro, Jonas e Guimarães, Yago
S586d Ferramenta para detecção de padrões de botnet baseado em algoritmos de agrupamento de aprendizagem de máquina / Jonas Rocha Lima Amaro, Yago Guimarães Coimbra. - Rio de Janeiro: Instituto Militar de Engenharia, 2016.

18f. : il., graf., tab. : -cm.

Projeto de Fim de Curso - Instituto Militar de Engenharia
Orientador: Sergio dos Santos Cardoso Silva.

1 - Botnets 2 - Clustering

CDU 631.317.35

Jonas Rocha Lima Amaro, Yago Guimarães Coimbra

Ferramenta para detecção de padrões de botnet baseado em algoritmos de agrupamento de aprendizagem de máquina

Trabalho Apresentado ao Curso de Engenharia de Computação do Instituto Militar de Engenharia como Verificação Especial do Projeto de Fim de Curso.

Trabalho aprovado. Rio de Janeiro, 2 de maio de 2016:

Prof. Sergio dos Santos Cardoso Silva
Orientador, D. Sc., do IME

Profa. Raquel Coelho Gomes Pinto
Convidada, D. Sc., do IME

Prof. Julio Cesar Duarte
Convidado, D. Sc., do IME

Rio de Janeiro
2 de maio de 2016

Resumo

Resumo em pt

Palavras-chave: botnets, clustering.

Abstract

Abstract in English

Keywords: botnets, clustering.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

Fig.	Figura

Lista de símbolos

Γ	Letra grega Gamas

Sumário

1 Introdução

1.1 Objetivo

1.2 Motivação

1.3 Justificativa

1.4 Metodologia

1.5 Estrutura

2 Botnets

Botnets são redes formadas por máquinas infectadas com malware, permitindo que o botmaster (o atacante) realize diversas atividades criminais remotamente, como roubo de informações, ataques de negação de serviço, envio de SPAM, etc.[1]

Devido a sua alta efetividade e grande potencial de causar danos, as botnets são consideradas uma das maiores ameaças de segurança no espaço cibernético.

Existem duas formas de agir contra as botnets, reativamente ou preventivamente. A forma reativa é a mais comum e envolve detectar a existência da botnet e reagir ao ataque tentando reduzir o tráfico malicioso para níveis aceitáveis, uma desvantagem é que o ataque já vai ter sido inicializado quando for detectado, ou seja, já vai haver causado danos antes de ser solucionado. A forma preventiva busca evitar que a botnet possa realizar alguma atividade maliciosa, porém essa atividade não é simples, já que o atacante pode aprimorar seus bots, tornando os mais sofisticados, exigindo grandes investimentos para manter os recursos de segurança atualizados.

Estruturalmente, as botnets são formadas pelos bots, que são malwares instalados nos computadores das vítimas que podem realizar as ações maliciosas que o botmaster envia através do canal de comando e controle (C&C).

Existem características que tornam os host mais interessantes ao botmaster como: altas taxas de transmissão, baixos níveis de segurança e monitoração, alta disponibilidade e localização distante (dificultando que as agências reguladoras detectem as atividades, já que os bots estarão espalhados por diversas nações). Esses fatores ajudam o bot a passar desapercibido e a contribuir com maior capacidade de banda ao botmaster, facilitando ataques como os de negação de serviço.

No geral, existe um ciclo com fases bem definidas de como um host se torna um bot. A primeira fase, injeção inicial, ocorre quando o host é infectado pelo malware, tornando-se um bot em potencial, através de um download indesejado ou através de um anexo em um email, por exemplo. Após a infecção ser bem sucedida, ocorre a injeção secundária: o host infectado busca em uma rede os binários do malware, os quais após baixados e executados farão com o que o host comece a se comportar como um bot. A terceira fase, chamada de conexão ou rallying, envolve a conexão entre o bot e o C&C, e acontece sempre que o host é reiniciado, podendo ser considerada uma fase vulnerável já que segue um padrão. Após isso, o bot entra na quarta fase, na qual aguarda que o C&C envie os comandos para que ele comece a executar as atividades maliciosas solicitadas. A última fase é a de manutenção e atualização, sendo importante se o botmaster deseja que os bots possam evitar novas técnicas de detecção ou adicionar novas funcionalidades, por exemplo.

Existem 4 tipos de arquiteturas para as botnets: centralizada, descentralizada, híbrida

e aleatoria. Na arquitetura centralizada, todos os bots se comunicam com um número pequeno de servidores de C&C, embora ela ofereça vantagens ao botmaster, como baixa latência e facilidade de manutenção, ela também torna a botnet bastante vulnerável, permitindo que ela seja desligada após a identificação dos poucos pontos centrais de C&C. Isso motivou o desenvolvimento da arquitetura descentralizada, onde uma variedade de protocolos P2P é utilizada, permitindo que mesmo que muitos bots sejam desativados a botnet possa continuar funcionando, já que não existem pontos centralizados de C&C. A arquitetura híbrida apresenta características de ambas as arquiteturas centralizadas e descentralizadas, na qual os bots são classificados em dois grupos: clientes e servos, os servos exercem os papéis tanto de clientes quanto servidores, sendo utilizados para repassar os comandos enviados pelo botmaster. Por fim, a arquitetura aleatória é um modelo apenas teórico, no qual o bot não se comunica ativamente com o botmaster ou com outros bot, para realizar um ataque o botmaster vasculha a rede em busca de um bot para enviar o comando e realizar as atividades maliciosas.

Existem duas categorias de técnicas para detecção de botnets: honeynets e sistemas de detecção de intrusos (IDS). As honeynets consistem na criação de redes com a intenção de que elas sejam comprometidas, permitindo que as informações sobre a botnet sejam captadas.

A detecão via IDS, pode ser classificada entre duas técnicas: a baseada em assinaturas e a baseada em anomalias. A técnica baseada em assinaturas, consiste em extrair padrões da rede e comparar com um banco de dados onde se encontram os padrões que já foram vistos em botnets, ou seja, ela não permite que novas botnets sejam identificadas. Dessa forma, a técnica baseada em anomalias é a principal área de pesquisa para detecção de botnets, baseando-se em anomalias na rede, como alta latência, aumento no tráfego ou uso de portas incomuns para detectar a presença de bots na rede.

As técnicas baseadas em anomalias, podem ser baseadas no host, onde cada máquina possui uma ferramenta de monitoração instalada (o que não é muito escalável), e tem seu comportamento analisado para verificar a existência de atividades suspeitas. Além disso, a análise pode ser baseada na rede, ativa (que possuem a grande desvantagem de aumentar o tráfego da rede ao injetar pacotes com a finalidade de examinar se um cliente é humano ou um bot) ou passivamente, sendo a forma de detecção mais utilizada atualmente.

A monitoração passiva de uma rede consiste em analizar o tráfego da rede buscando por comunicações suspeitas que podem ter sido enviadas pelos bots ou canais de C&C. Essa monitoração é possível pois os bots de uma mesma botnet costumam apresentar padrões de comunicação, já que eles são pré programados pelo mesmo botmaster para entrar contato com o servidor de C&C.

Para que a análise do tráfego seja viabilizada, são empregadas diversas técnicas como métodos estatísticos, mineração de tráfego, teoria de grafos, clustering, modelos estocásticos, redes neurais, entre outras.

A detecção de botnets é uma tarefa bastante desafiadora porque os botmasters estão sempre aprimorando os bots, tornando os mais difíceis de serem detectados. Por exemplo, as primeiras detecções buscavam mensagens suspeitas nos conteúdos da mensagem, afim de evitar isso os botmasters passaram a utilizar criptografia tornando essa técnica de detecção obsoleta. Outra dificuldade para algoritmos de clustering é que podem ser evitados usando técnicas de randomização nas comunicações e atribuição de tarefas diferentes para os bots.

3 Cronograma

Embora o objetivo do trabalho seja o desenvolvimento de um projeto e não pesquisa, foi preciso começar por um intenso estudo do problema que queríamos resolver, ou seja das botnets. Isso é uma etapa importante de um projeto de aprendizagem de máquina, pois permite uma melhor identificação de quais features serão mais relevantes.

Em seguida, foi feito um estudo dos algoritmos de clustering existentes e os objetivos de cada técnica.

Após esses estudos, começa a implementação do sistema detector de botnets. A primeira etapa é desenvolver um tratamento automatizado dos dados obtidos pela coleta dos logs DNS, já que o objetivo final é de que esse tratamento seja feito diariamente. Depois serão implementados algoritmos de agrupamento que usarão os dados tratados para identificar padrões de botnets no log DNS coletado. Por fim, os resultados dessas técnicas serão testados e analisados e servirão de motivação para possíveis refinamentos nos algoritmos.

Durante essas tarefas, desenvolveremos também os relatórios e apresentações para as seguintes avaliações:

- Verificação Especial em Maio,
- Verificação Corrente em Julho,
- Verificação Final em Setembro.

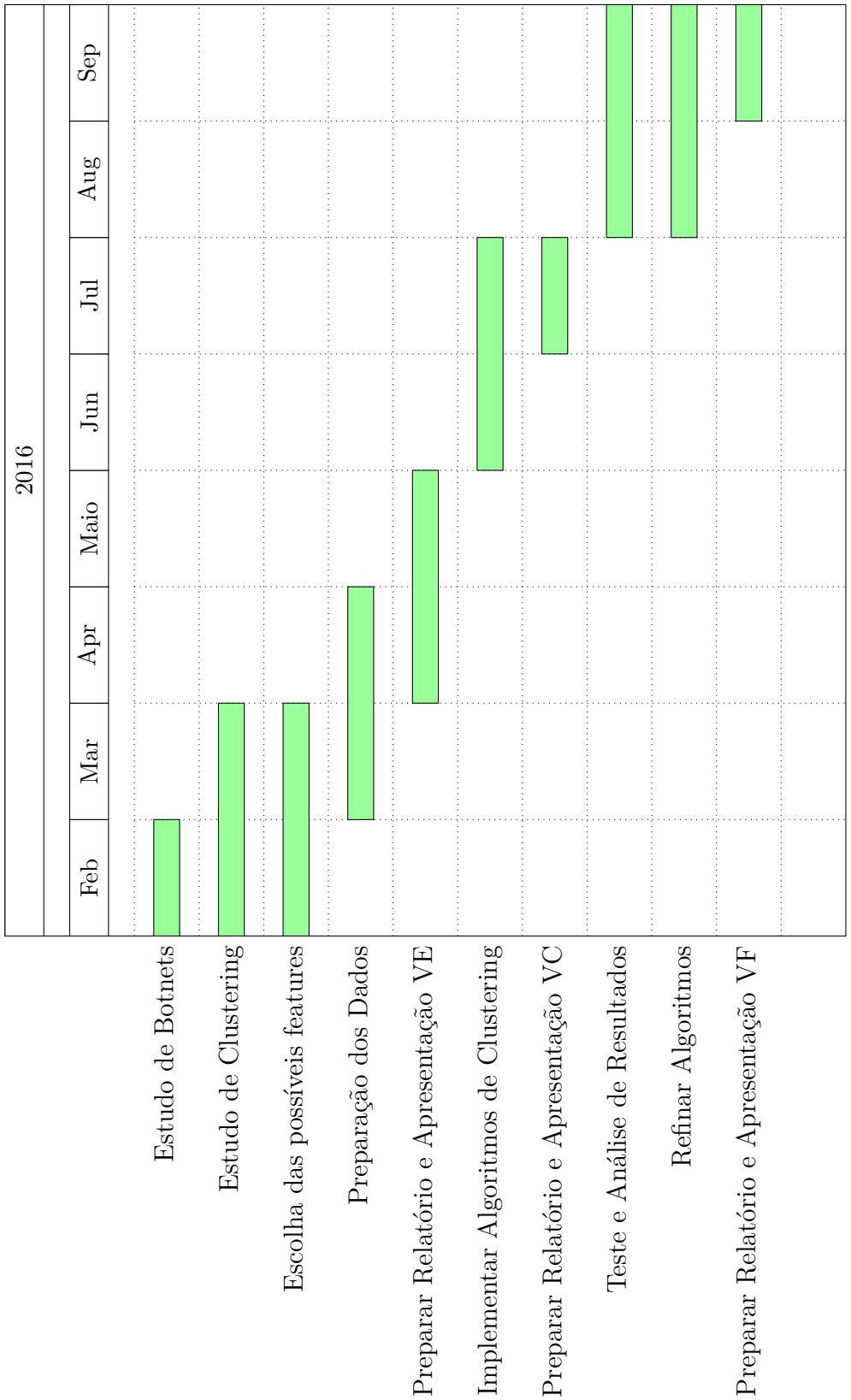

Figura 1 – Cronograma

4 Conclusão

Texto Conclusão

Referências

1 SILVA, S. S. et al. Botnets: A survey. *Computer Networks*, Elsevier, v. 57, n. 2, p. 378–403, 2013.