

**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
5 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
5 de maio de 2016

c2014

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

Este exemplar é de propriedade do Instituto Militar de Engenharia, que poderá incluí-lo em base de dados, armazenar em computador, microfilmar ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade dos autores e do orientador.

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.

20f. : 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, 5 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
5 de maio de 2016

Resumo

Resumo em pt

Palavras-chave: botnets, clustering.

Abstract

Abstract in English

Keywords: botnets, clustering.

Lista de ilustrações

Figura 1 – Ciclo de Vida das Botnets	13
Figura 2 – Arquitetura Centralizada	14
Figura 3 – Arquitetura Híbrida	15
Figura 4 – Cronograma	18

Lista de tabelas

Lista de abreviaturas e siglas

C&C	Comando e Controle
IRC	<i>Internet Relay Chat</i>
HTTP	<i>HyperText Transfer Protocol</i>
P2P	<i>Peer-to-peer</i>
IDS	<i>Intrusion Detection System</i>

Sumário

1	INTRODUÇÃO	10
1.1	Contextualização	10
1.2	Objetivo	10
1.3	Motivação	10
1.4	Justificativa	10
1.5	Metodologia	10
1.6	Estrutura	10
2	BOTNETS	11
2.1	Elementos das Botnets	11
2.2	Ameaças e Formas de Defesa	11
2.3	Ciclo de Vida das Botnets	12
2.4	Arquitetura das Botnets	13
2.5	Deteccção de Botnets	14
2.6	Sistema Integrado de Defesa Cibernética para Deteccção de Botnets	16
3	CRONOGRAMA	17
4	CONCLUSÃO	19
	Referências	20

1 Introdução

1.1 Contextualização

1.2 Objetivo

O objetivo deste trabalho é desenvolver e analisar uma ferramenta para detectar possíveis hospedeiros de bots em uma botnet, utilizando algoritmos de agrupamento que utilizarão dados de consultas de DNS realizadas. Para esta análise, serão utilizados logs de consulta com os hospedeiros previamente mapeados.

1.3 Motivação

1.4 Justificativa

A justificativa desse trabalho é reduzir o trabalho manual de identificar possíveis botnets em uma rede, funcionando como um filtro que já informará com maior velocidade as máquinas suspeitas. Por isso, os algoritmos de agrupamento se destacam para auxiliar nesse processo, além de poder servir no futuro como um módulo primário de um sistema mais complexo de detecção.

1.5 Metodologia

1.6 Estrutura

2 Botnets

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

Com o crescimento e diversificação do uso da Internet, o meio cibernético se tornou mais relevante e mais atraente para a realização de ataques maliciosos. Isso motivou o crescimento do número de botnets existentes e aumentou o potencial de contaminação das mesmas, além disso, para evitar os mecanismos de detecção existentes, elas se tornaram cada vez mais sofisticadas.

Para que o detector se torne mais robusto, e leve em conta as configurações existentes e até detecte possíveis novas configurações das botnets, é preciso compreender o funcionamento das botnets e seus objetivos, para que possamos identificar características constantes na botnet, mesmo quando o botmaster está tentando evitar os mecanismos de detecção.

2.1 Elementos das Botnets

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). Geralmente, o malware é inicializado quando o hospedeiro inicializa a máquina, porém isso pode ser configurado pelo botmaster para dificultar a detecção da atividade maliciosa.

O canal de C&C é o meio que o botmaster tem para se comunicar com a sua botnet, e é a parte chave do funcionamento, pois é necessário para o envio dos comandos necessários para a atividade maliciosa aos hospedeiros. Dessa forma, grande parte das características da botnet, como robustez, facilidade de detecção/desativação, estabilidade, etc., são definidas pela forma que a infraestrutura de C&C está organizada.

2.2 Ameaças e Formas de Defesa

O crescimento do número de máquinas conectadas constantemente à enlaces de alta velocidade e rodando sistemas com vulnerabilidades consideráveis, criou um ambiente favorável à formação de botnets. Esse crescimento, aliado à alta efetividade e potencial de causar danos, fez com que as botnets se tornassem um dos maiores desafios de pesquisa em segurança no espaço cibernético atual. [2]

Existem características que tornam o 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 despercebido e a contribuir com maior capacidade de banda ao botmaster, facilitando ataques como os de negação de serviço.

Existem duas formas para o combate das 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 tráfego 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.

O mecanismo que estamos desenvolvendo é da forma reativa, já que o algoritmo encontrará padrões em botnets que já estão atuando. Porém, uma característica desejável para um detector reativo é a detecção em tempo real, com o objetivo de minimizar os danos causados e o tempo de reação do botmaster. Porém, essa característica é um desafio, devido ao grande número de dados que devem ser tratados e analisados. Dessa forma, nosso objetivo neste projeto será se aproximar disso, utilizando a detecção dos dados coletados ao longo de um dia para detectar bots que atuaram nas últimas 24 horas.

2.3 Ciclo de Vida das Botnets

Na maioria dos casos, existe um ciclo com fases bem definidas de como uma botnet é criada e mantida, a Figura 1 mostra essas fases para cada novo hospedeiro que é contaminado.

Na primeira fase, chamada de injeção inicial, o atacante procura vulnerabilidades na máquina do futuro hospedeiro para explorá-las e infectá-lo com o malware, tornando-se um bot em potencial, isso pode ocorrer, por exemplo, através de um download indesejado ou através de um anexo em um e-mail. Após a infecção ser bem sucedida, ocorre a injeção secundária: o host infectado, através do malware inicial instalado, busca em uma rede os reais binários do malware do bot, os quais após baixados e executados concluirão a infecção e tornam o host em um bot real.[3].

Durante a fase de conexão, o bot estabelece conexão com o canal de C&C, isso se repete sempre que o host é reiniciado, podendo ser considerada uma fase vulnerável já que segue um padrão. Após a efetivação da conexão, o bot se torna ativo na botnet, e passa a realizar os comandos enviados pelo botmaster através do canal de C&C, efetivando as

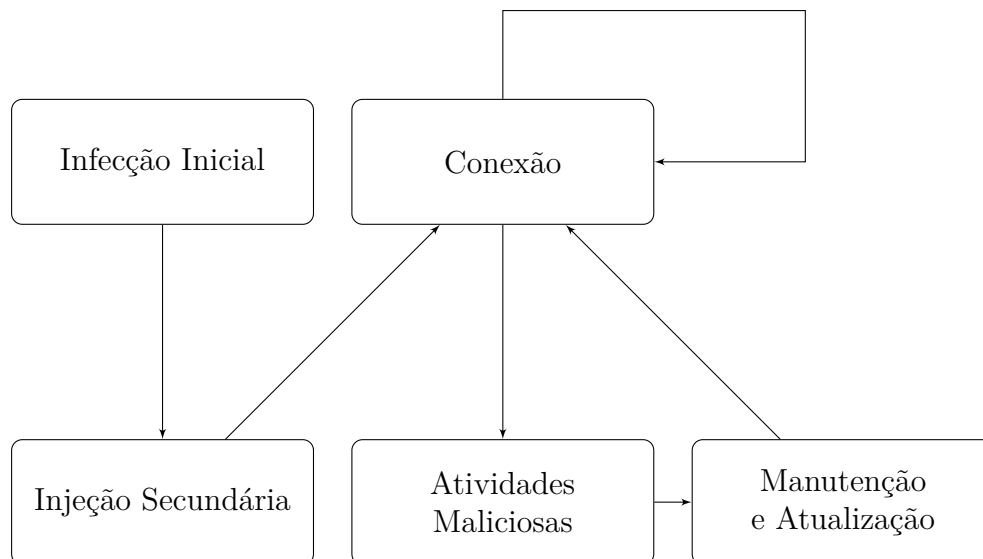


Figura 1 – Ciclo de Vida das Botnets

atividades maliciosas solicitadas. A última fase é a de manutenção e atualização, e tem por objetivo manter a botnet ativa e atualizada, já que se o botmaster deseja que os bots possam evitar novas técnicas de detecção, adicionar novas funcionalidades ou até mesmo alterar o servidor de C&C, os binários do programa bot devem ser modificados.

2.4 Arquitetura das Botnets

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

Na arquitetura centralizada, mostrada na Figura 2 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. Ela é muito utilizada pelo protocolo IRC (*Internet Relay Chat*), porém pelo fato de tráfego desse protocolo ser incomum e raramente utilizado, ele costuma ser bloqueado, inutilizando a botnet. Por isso, o uso do protocolo HTTP (*HyperText Transfer Protocol*) se popularizou já que ele é muito utilizado, disfarçando as comunicações das botnets.

A fragilidade da arquitetura centralizada, motivou o desenvolvimento da arquitetura descentralizada, na qual uma variedade de protocolos P2P (*Peer-to-peer*) é utilizada. A flexibilidade e robustez dessa arquitetura, permite que mesmo que muitos bots sejam desativados a botnet possa continuar funcionando, já que não existem pontos centralizados de C&C.

As arquitetura híbridas apresentam características de ambas as arquiteturas centralizadas e descentralizadas, como mostrado na Figura 3, na qual os bots são classificados

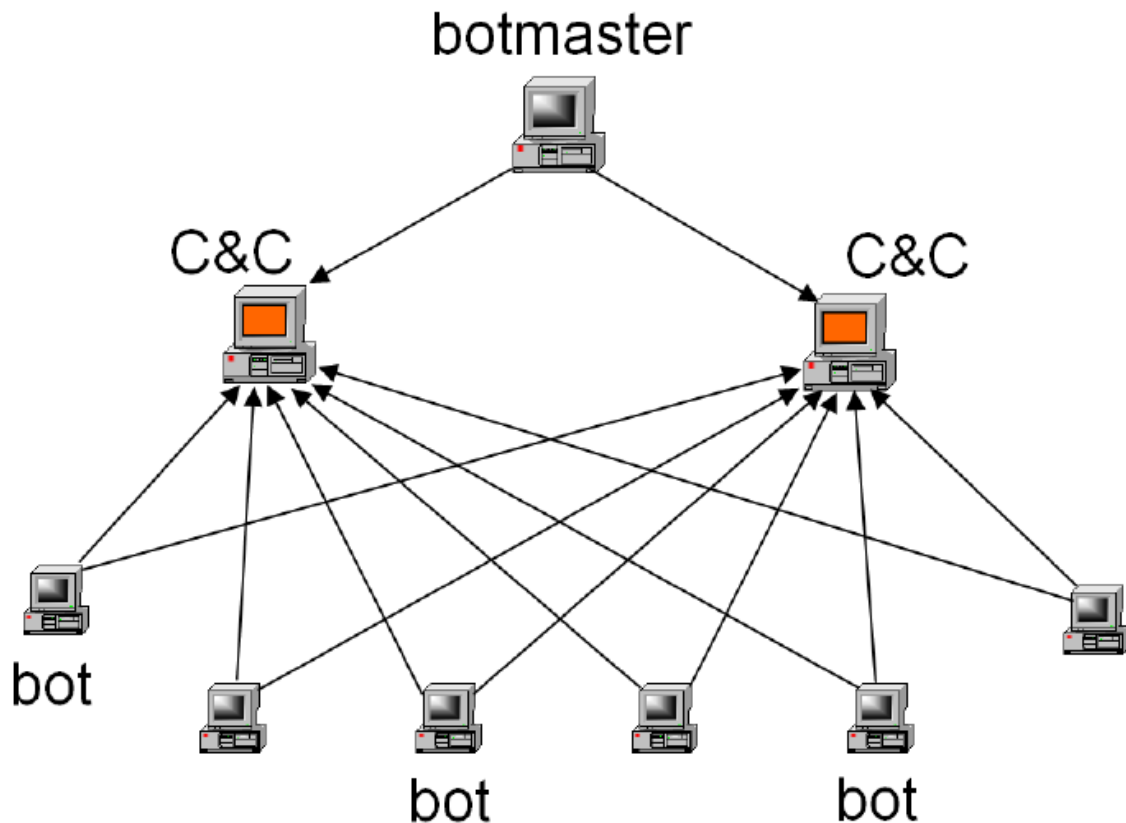


Figura 2 – Arquitetura Centralizada[4]

em dois grupos: clientes e servos. Os servos exercem os papéis tanto de clientes quanto servidores, possuindo endereço de IP estático e público para serem acessíveis globalmente, sendo utilizados para repassar os comandos enviados pelo botmaster. Os demais bots, são denominados clientes pois não aceitam comunicações de entrada, deitam forma e podem apresentar IP dinâmico, privado ou protegidos por *firewall* para não serem roteados facilmente. Por fim, a arquitetura aleatória é um modelo até agora 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.

2.5 Detecção de Botnets

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, por isso elas são consideradas mais efetivas para compreender as características

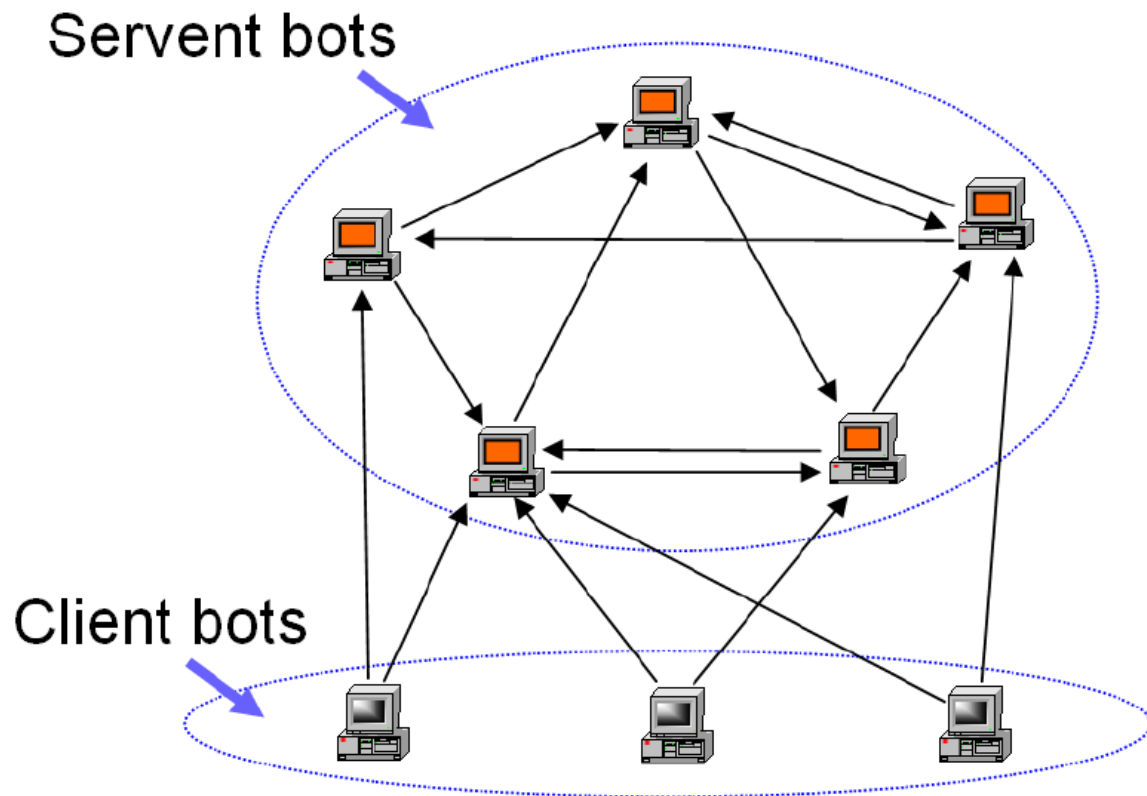


Figura 3 – Arquitetura Híbrida[4]

de uma botnet do que a detecção propriamente dita.

A detecção por 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 e envolve a posse de um banco de dados enorme com o maior número de informações existentes sobre as botnets previamente detectadas. 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 esta última a forma de detecção mais utilizada e pesquisada atualmente.

A monitoração passiva de uma rede consiste em analisar 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.

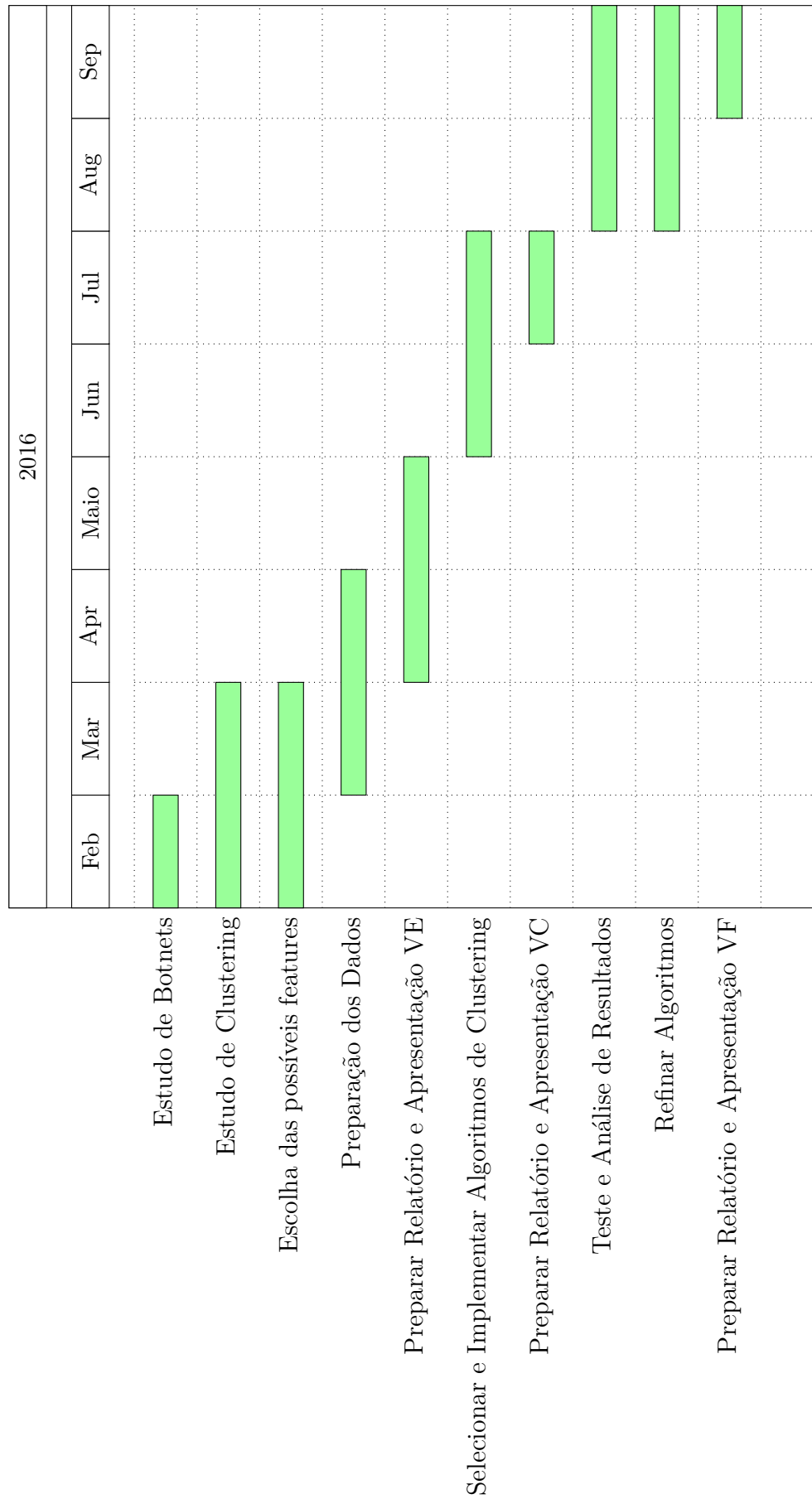


Figura 4 – Cronograma

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.
- 2 SOLTANI, S. et al. A survey on real world botnets and detection mechanisms. *International Journal of Information and Network Security*, IAES Institute of Advanced Engineering and Science, v. 3, n. 2, p. 116, 2014.
- 3 FEILY, M.; SHAHRESTANI, A.; RAMADASS, S. A survey of botnet and botnet detection. In: IEEE. *Emerging Security Information, Systems and Technologies, 2009. SECURWARE'09. Third International Conference on*. [S.l.], 2009. p. 268–273.
- 4 WANG, P.; SPARKS, S.; ZOU, C. C. An advanced hybrid peer-to-peer botnet. *IEEE Transactions on Dependable and Secure Computing*, v. 7, n. 2, p. 113, 2010.
- 5 SILVA, S. S.; SALLES, R. M. Arquitetura de um sistema integrado de defesa cibernética para detecção de botnets. *Programa de Engenharia de Defesa*, Instituto Militar de Engenharia, 2012.