Arquitetura de um sistema integrado de defesa cibernética para detecção de botnets

Autoria do artigo: Sérgio dos Santos Cardoso Silva e Ronaldo Moreira Salles

Apresentação: Bruno Follmann

Apresentação

- Apresentar uma proposta de arquitetura para um sistema integrado de defesa cibernética;

- Propor uma técnica de detecção de bots baseada em análise de logs DNS, empregando a arquitetura proposta.

Botnets

- Empregadas em ações de ataque cibernético com grande frequência;
- São redes formados por hosts escravos, os bots;
- Controladas por um mais atacantes, denominados botmasters;
- Uma das maiores ameaças ao funcionamento da rede, devido a sua flexibilidade;
- Permitem roubo de informações, envio de spam, ataques DDoS, manipulação de jogos e serviços online, etc.

Botnets

- A detecção de botnets é dificultosa;
- Elas empregam técnicas que visam ocultar o tráfego gerado;
- Um exemplo é a simples minimização do número de troca de mensagens;
- Trabalhos atuais apresentam soluções isoladas para a detecção ou desativação de botnets;
- Soluções comerciais existentes são baseadas em assinatura, tendo assim problemas com detecção de novos bots.

Ciclo de vida dos bots

Fase 1: Injeção inicial

- O invasor varre uma subrede buscando vulnerabilidades;
- Ele infecta a vítima por meio de diversos possíveis métodos, como exploits e anexos de mensagens;

Fase 2: Injeção secundária

- A máquina infectada executa um script, buscando o código de malware em um repositório;

Fase 3: Conexão ou rally

- Toda a vez que a máquina for reiniciada, contatará o servidor de comando e controle;
- Assim, pode confirmar sua participação e receber comandos.

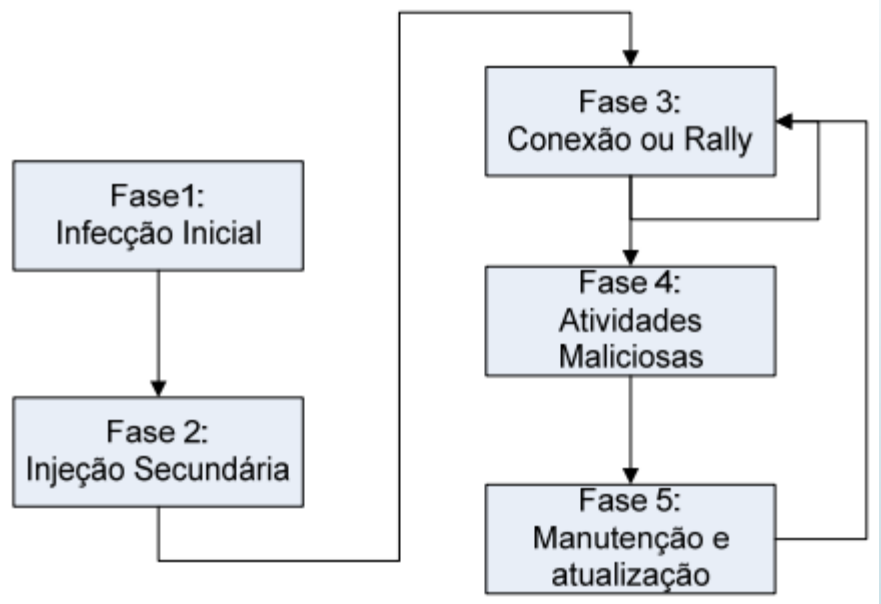
Ciclo de vida dos bots

Fase 4: Atividades maliciosas

- O bot está pronto para receber comandos e efetuar ataques;

Fase 5: Manutenção e atualização

- Inclui evasão de técnicas de detecção, adição de novas funconalidades, e até mesmo migração de servidor.



Arquitetura do sistema

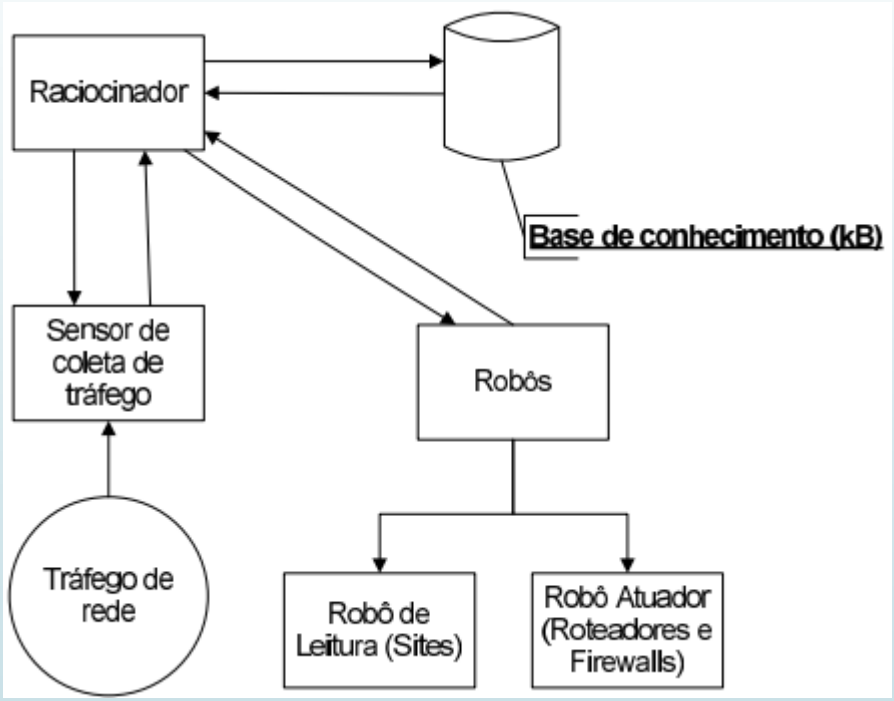
É desenvolvida de forma modular e permite tanto a análise do tráfego da rede quanto a análise semântica dos dados das máquinas atacantes e atacadas. Formada por:

- Sensor de coleta de tráfego: Monitora passivamente a rede;
- Raciocinador: Analisa as informações, considerando aspectos estatísticos e semânticos;
- Base de conhecimento (KB): Armazena as informações utilizadas pelo raciocinador;
- Robôs de atuação e leitura: Responsáveis pelas funções de proteção.

Arquitetura do sistema

Sensores adquirem dados das formas:

- Bruta, como registros de tráfego (traces);
- Com certo nível de agregação (fluxos);
- Análises de registros (logs).



Arquitetura do sistema

- O racicionador empregará algoritmos específicos para análise dos dados dos sensores;
- Conforme os resultados, serão acionados robôs de leitura ou atuação;
- O disparo dos robôs é baseado em valores pré definidos na KB;
- O robô de leitura tem como objetivo obter informações adicionais para a tomada de decisão;
- Confirmado que algum domínio é um bot ou C&C, o robô de atuação será ativado;
- Robôs de atuação podem ser especializados, como para firewalls, servidores e roteadores. Eles tomarão alguma medida para mitigar o problema.

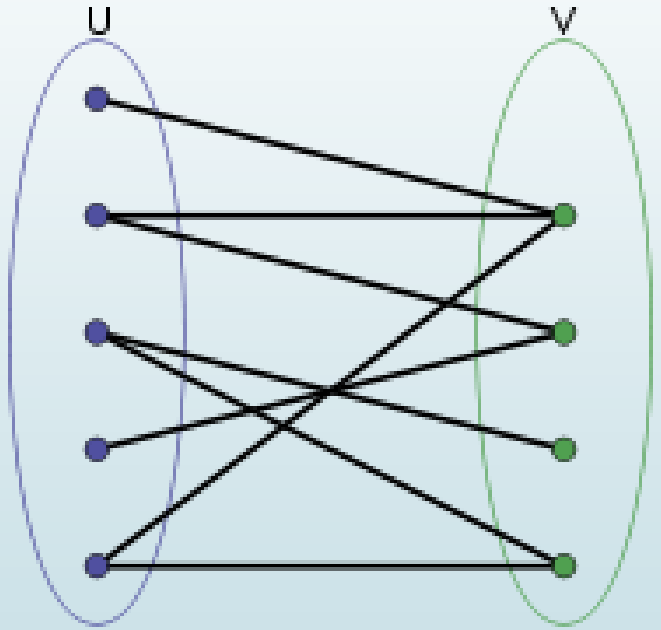
Estudo de caso - DNS

- Primeiro módulo implementado foi para a análise de logs DNS;
- Optou-se por esse estudo pois na fase de conexão (Rally) o bot precisa encontrar o C&C, traduzindo o domínio especificado;
- Os registros obtidos pelo sensor são mapeados em grafos de contatos.

Grafo de contatos

Grafo direcionado bipartido e ponderado Gdns = (U, V, E, P):

- Nós são formados pelos endereços IP (U) e domínios (V);
- Existirá aresta entre u (u ϵ U) e v (v ϵ V) se u solicitar consulta DNS para domínio v;
- Quantidade de consultas determina o peso P;
- α: limiar de desvio padrão;
- β: limiar de peso das arestas.



Premissas

O objetivo é determinar subgrafos que formam comunidades suspeitas. Isso se dá por meio de algumas premissas que descrevem comportamento geral de bots.

I. **Interesse comum:** Bots normalmente infectam mais de uma máquina em uma mesma rede;

II. **Comportamento robótico:** Sendo um software pré programado, a comunicação de hosts infectados com a botnet terá comportamento semelhante;

III. **Stealthiness:** Para evadir detecção, bots usam técnicas como fazer poucas consultas e a muitos domínios distintos.

Raciocinador

1. Filtragem inicial (elimina domínios que com certeza não são suspeitos);
2. Separação em subgrafos desconexos para cada domínio;
3. Busca por domínios com arestas incidentes que apresentam comportamento atípico;
4. Busca por arestas com baixo peso.

Robôs

- Os subgrafos resultantes da filtragem indicam endereços IP e domínios considerados suspeitos;
- O robô de leitura é acionado para verificar as suspeitas, por meio de análise semântica do site ou verificação em uma blacklist;
- Caso a suspeita seja confirmada, o robô de atuação é ativado, podendo realizar ações como criar regras no firewall, remover o domínio, comunicar o administrador da rede, entre outras.

Metodologia

- Análise de 3 dias do tráfego em uma instituição de ensino;
- Etapa 1 eliminou 97% dos IPs e domínios;
- Para os valores de a e β, foram considerados suspeitos todos os domínios que pertecem a ccTLD .cn (95% de infecção), e ao .ws;
- Calculou-se a mediana e desvio padrão para encontrar os valores de corte.

Etapa	Dia 12/03/12		Dia 13/03/12		Dia 14/03/12	
	# IP	# Dom	# IP	# Dom	# IP	# Dom
Gdns	20.529	238.852	30.027	383.977	37.355	499.777
1	540	6.962	784	11.428	949	15.040
2	540	6.962	784	11.428	949	15.040
3	290	1.359	260	2.244	122	2.918
4	245	1.149	221	1.899	104	2.492

Resultados preliminares

- A análise resultante revelou que 5 endereços IP e 253 domínios perteciam à botnet Conficker;
- Vantagem: excelete taxa de acerto, pois todas as botnets foram detectadas;
- Desvantagem: alta taxa de falsos positivos, por volta de 90%;
- Conclui-se que é preciso aprimorar a criação da whitelist;
- Também pode ser preciso refinar o cálculo dos valores de corte;
- Por fim, é preciso também aprimorar o robô de leitura.