Coleta e análise de características de fluxo para classificação de tráfego em Redes Definidas por Software

Autor: Rodolfo Vebber Bisol (rvbisol@inf.ufrgs.br)
Orientador: Alberto Egon Schaeffer Filho (alberto@inf.ufrgs.br)

Classificação de Tráfego

- Resiliência em redes de computadores é a capacidade de manter um nível aceitável de operação mesmo na presença de anomalias, como ataques maliciosos, sobrecarga operacional ou problemas de configuração.
- Classificação de tráfego é realizada com base nas características dos fluxos de tráfego existentes na rede, tais como: a quantidade de pacotes, bytes por segundo, protocolo, entre outras. Esta é uma estratégia frequentemente utilizada para prover resiliência através da detecção de fluxos de tráfego maliciosos.
- A seleção destas características é uma etapa muito importante da classificação de tráfego pois características irrelevantes podem gerar desperdício de recursos computacionais e ruídos que dificultam a classificação.

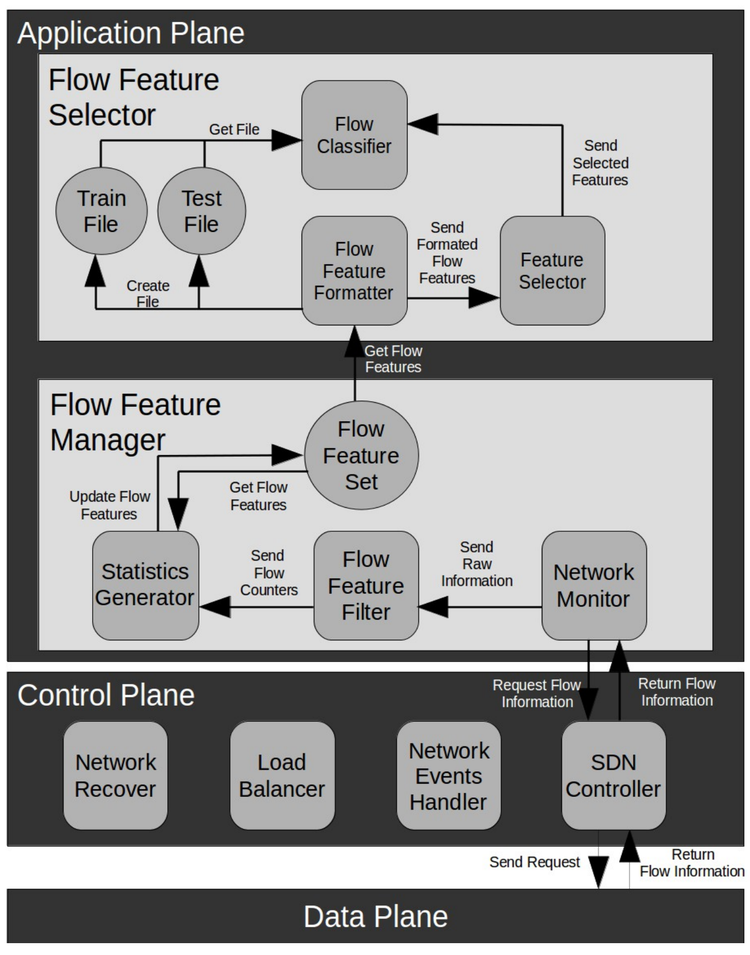
Redes Definidas por Software

- Redes Definidas por Software (SDN) possuem o controle dos dispositivos de rede, como roteadores e switches, centralizados em um único controlador, diferentemente das redes de computadores tradicionais que possuem o controle distribuído em cada dispositivo.
- Esta nova organização de redes oferece uma arquitetura modular favorável para o desenvolvimento de novas tecnologias na área de redes de computadores.

Coleta de Características

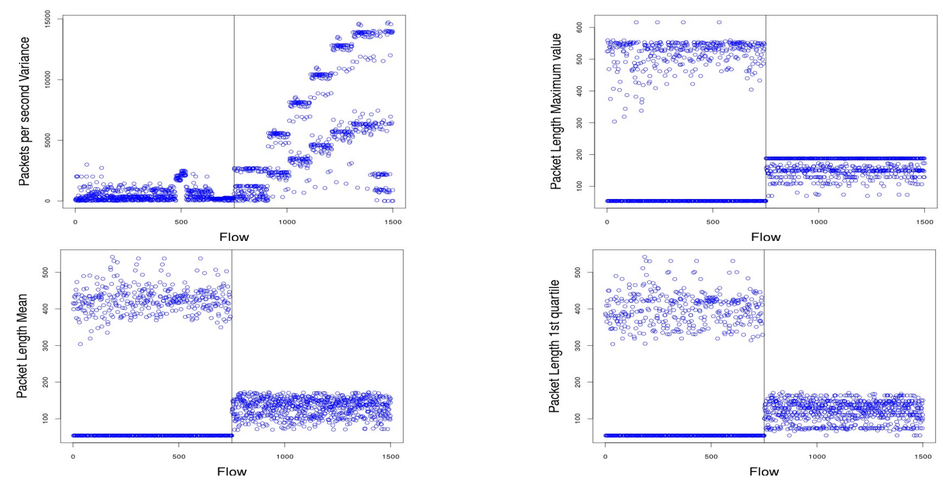
- Este projeto de pesquisa tira proveito da arquitetura SDN para monitorar e coletar características de tráfego em um ambiente controlado.
- Utiliza-se o protocolo OpenFlow para acessar dois contadores: *byte count* e *packet count*.
- Baseando-se nesses dois contadores, são derivadas 33 características sobre cada fluxo para serem analisadas, incluindo: tamanho médio dos pacotes, duração dos fluxos e variação do tempo entre a chegada de dois pacotes consecutivos.

Arquitetura

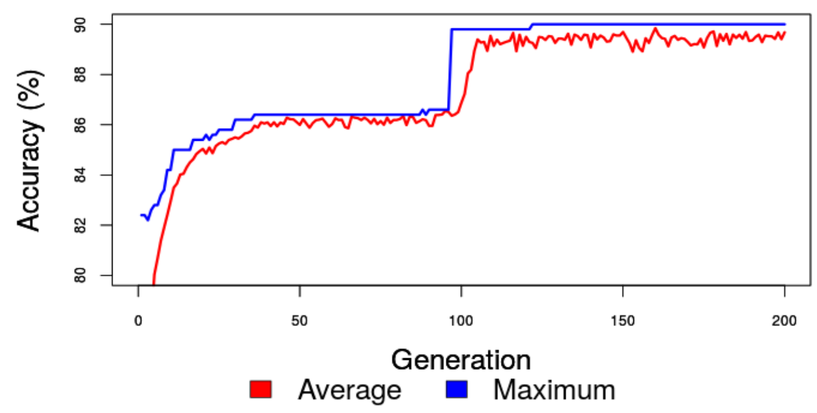


Análise

- Para analisar as características de fluxo criadas, são comparados 3 algoritmos de seleção de características: *Principal Component Analysis* (PCA), *Genetic Algorithm* (GA) e *Sequential Backward Selection* (SBS). Além disso, para a classificação dos fluxos de tráfego são utilizados 2 algoritmos: Support Vector Machine (SVM) e K-means.
- O PCA elimina características correlacionadas e seleciona as que apresentam maior variância. Os gráficos a seguir apresentam as 4 características mais importantes para classificação de ataques DDoS de acordo com o PCA.



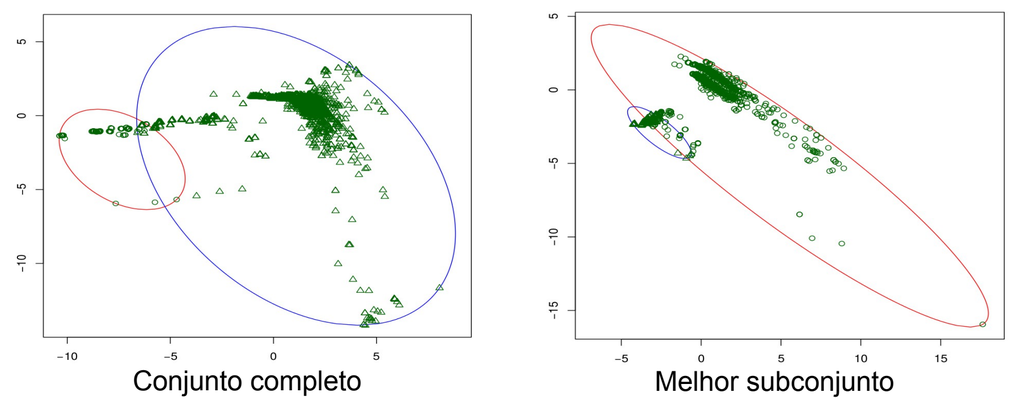
- O gráfico abaixo apresenta a evolução da qualidade dos subconjuntos de características selecionadas utilizando GA. A cada geração este algoritmo cria subconjuntos diferentes imitando o comportamento natural da evolução das espécies.



- A tabela a seguir apresenta os resultados da classificação dos fluxos de tráfego utilizando SVM.

	Conjunto completo	PCA	GA	SBS
Acurácia média (%)	86.2	91.26	95.93	95.13
Tamanho médio do conjunto	33	13	24	19

- Os gráficos abaixo apresentam a classificação dos fluxos utilizando o algoritmo K-means, onde as elipses representam um tipo de tráfego diferente.
- Pode-se observar que no gráfico da direita os dois conjuntos se encontram mais compactos, o que facilita a classificação realizada pelo K-means



Conjunto completo | Melhor subconjunto

- Identification and Selection of Flow Features for Accurate Traffic Classification in SDN. 14th IEEE International Symposium on Network Computing and Applications (IEEE NCA 2015), 28-30 September 2015, Cambridge, USA.

UFRGS UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
.inf INSTITUTO DE INFORMÁTICA UFRGS
Computer Networks UFRGS