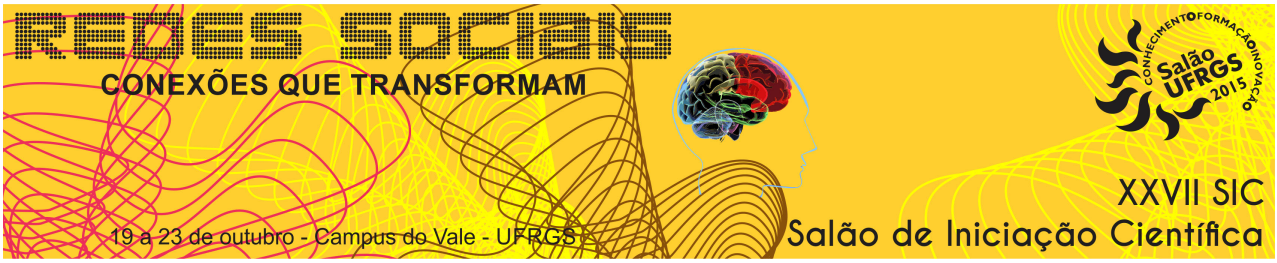

Evento	Salão UFRGS 2015: SIC - XXVII SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2015
Local	Porto Alegre - RS
Título	Coleta e análise de características de fluxo para classificação de tráfego em Redes Definidas por Software
Autor	RODOLFO VEBBER BISOL
Orientador	ALBERTO EGON SCHAEFFER FILHO

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)
Instituição: Universidade Federal do Rio Grande do Sul (UFRGS)

Classificação de fluxos de tráfego é um mecanismo importante para prover resiliência em redes de computadores, podendo ser aplicado com diversos objetivos, como detecção de fluxos de tráfego maliciosos. Estes mecanismos, apesar de já serem muito sofisticados, ainda precisam ser aprimorados pois certos ataques podem se camuflar entre o tráfego legítimo que normalmente existe na rede. Neste contexto, o objetivo desse projeto é investigar a utilização de Redes Definidas por Software (SDN) para aprimorar estes mecanismos de classificação através da identificação e seleção de características importantes na detecção de tráfego maliciosos. Essas são etapas necessárias, pois características irrelevantes podem adicionar ruído dificultando a classificação, além de causar desperdício de recursos computacionais.

SDN possui como característica o plano de controle dos dispositivos de encaminhamento centralizado em um controlador, diferentemente das redes tradicionais que possuem o controle distribuído entre os dispositivos. O controlador de uma SDN é responsável por tomar todas as decisões de encaminhamento de pacotes e enviá-las aos dispositivos utilizando um protocolo, como por exemplo, o OpenFlow. Esta importante característica oferece uma significante simplificação na coleta de informações da rede, que passa a ser feita através de uma única interface entre o controlador da rede e as aplicações que necessitam destas informações.

Neste projeto foi desenvolvido um mecanismo que tira proveito desta simplificação e é capaz de: coletar dados primitivos da rede e derivar um conjunto avançado de características que descrevem o perfil de cada fluxo de tráfego existente na rede; e analisar estas características de fluxo a fim de definir qual subconjunto de características oferece a melhor acurácia na classificação de fluxos de tráfego.

Para o desenvolvimento deste projeto foi criada uma rede onde foram inseridos diversos tipos de tráfego, por exemplo, ataques de negação de serviço (DoS) e *streaming* de vídeo. O objetivo de usar diversos tipos de tráfegos é avaliar as características de fluxo criadas em diversos cenários próximos da realidade.

O protocolo OpenFlow fornece uma série de informações sobre os fluxos existentes na rede que podem ser coletadas com o auxilio do controlador. Estas informações são constituídas pelas regras de encaminhamento enviadas aos *switches*. O mecanismo desenvolvido no projeto coleta somente as informações relevantes: para identificação de fluxos, que são endereços IP e portas TCP/UDP de origem/destino e protocolo de aplicação; e para criação de características de fluxo, que são os contadores nativos de *bytes* e pacotes. Com base nestes contadores foi criado um conjunto de características de fluxo avançadas composto por três classes de características: escalares, estatísticas e complexas.

Com o objetivo de selecionar as características mais importantes são empregados dois algoritmos de seleção de características: Principal Component Analysis (PCA) e Algoritmo Genético. Cada algoritmo utiliza princípios diferentes para a seleção de características, o que resulta em subconjuntos distintos. Cabe ao mecanismo selecionar qual deles resulta em uma melhor acurácia na classificação de fluxos de tráfego.

Os resultados obtidos apresentam uma melhora na acurácia da classificação utilizando os subconjuntos de características criados quando comparados com a acurácia da classificação realizada utilizando o conjunto completo. Além disso, os subconjuntos apresentam características distintas para cada cenário de teste, demonstrando que cada perfil de tráfego possui características específicas que melhor os descrevem.