Clasificación de tráfico en Internet utilizando métodos estadísticos

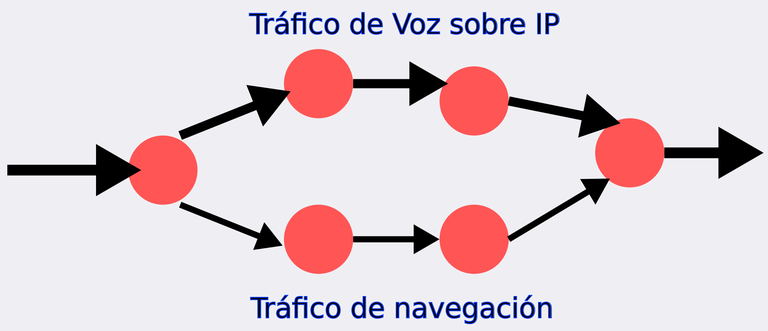
Autor: Gabriel Gómez Sena
Tutores: Pablo Belzarena y Paola Bermolen

Motivación

¿Por qué interesa clasificar el tráfico en Internet?

En una red de mediano o gran porte, es importante conocer la composición del tráfico que circula para la planificación del crecimiento

Conociendo el tipo de tráfico se pueden aplicar políticas de tratamiento diferenciado, por ejemplo para ofrecer calidad de servicio a las diferentes aplicaciones o para controlar aplicaciones que estén haciendo un uso abusivo de los recursos de la red



¿Por qué interesa clasificar el tráfico en linea?

Si se quieren tomar acciones sobre el tráfico circulante (cambio de rutas, aplanamiento, descarte, asignación de clase de servicio) es necesario identificar el tráfico a su ingreso a la red para poder aplicar las acciones necesarias antes que el flujo de tráfico finalice

Métodos clásicos

Clasificación por número de puerto de TCP/UDP

Las aplicaciones actuales negocian puertos dinámicamente y otras se "esconden" utilizando puertos predeterminados de otras aplicaciones (por ejemplo muchas usan el puerto 80 del protocolo HTTP pero no pueden considerarse navegación web) y por lo tanto este método ya no sirve

Clasificación en base a búsqueda de patrones característicos en los datos del usuario

Es un método de buena precisión pero de alto costo computacional y con objeciones en cuanto a la privacidad de los datos del usuario
No funciona cuando el tráfico está encriptado

Análisis estadístico de características de los flujos TCP/UDP

Técnica utilizada en esta tesis que tiene la ventaja que no necesita acceder a los datos del usuario y analiza el comportamiento de los flujos

Propuesta y Resultados

Análisis estadístico del tamaño de los segmentos del flujo TCP/UDP

El tamaño de los segmentos es una propiedad de los flujos que permite clasificar el tipo de tráfico

Los tamaños se obtienen facilmente, no se afecta la privacidad de los datos del usuario y puede usarse para clasificar tráfico encriptado

Solamente en base a los primeros segmentos intercambiados

Con menos de **5** segmentos es posible identificar el tráfico con muy buena precisión

Análisis estadístico con Support Vector Machines

Esta técnica de clasificación utilizada mejora la precisión de la clasificación en comparación con técnicas más tradicionales y tiene alto desempeño, lo que permitiría usarla para clasificación en línea

Ejemplos