Resumo do Traballo de Fin de Grao

El correo electrónico es una de las formas de comunicación más utilizadas en la actualidad. El hecho de que sea gratuito y su simplicidad lo convierten en una opción muy atractiva. Y precisamente por ello, también ha dado lugar a que empresas u organizaciones lo utilicen de una forma no tan beneficiosa, habitualmente como publicidad, lo que se conoce como spam. Este hace referencia a mensajes no solicitados, no deseados o con remitente desconocido que son enviados con fines propagandísticos y que pueden perjudicar al receptor. De hecho, estudios recientes demuestran que estas comunicaciones componen más de la mitad de los emails que se envían.

Es por esto que han aparecido diferentes metodologías para combatir el spam, unas mejores que otras, pero sin dar con la solución perfecta. Entre estas se diferencian medidas preventivas, reactivas y proactivas. El objetivo de este trabajo de fin de grado es el desarrollo de una solución proactiva al problema del spam. La razón es que estas medidas son mucho más interesantes, dado que implican que la transacción SMTP entre el servidor origen y el destino no finaliza, por tanto, los mensajes no llegan a entrar en el dominio del destinatario y el remitente es notificado de que la entrega no se ha producido. En las aproximaciones reactivas, por el contrario, el spam sólo es "escondido" a la vista de los destinatarios finales, ya que el remitente desconoce que su mensaje ha sido descartado.

Hay que tener en cuenta que, para el envío de cualquier correo se hace uso del protocolo SMTP, el cual contiene un campo DATA donde se encuentra el contenido del mismo. Más concretamente, el estándar RFC 5322 determina que, en dicho campo, tras otras cabeceras importantes y un espacio en blanco, se encuentra lo que propiamente sería la información que se quiere transmitir. Por tanto, la solución antes mencionada se encargaría de analizar dicha información haciendo uso de eBPF, una herramienta para filtrado de paquetes, que permitirá establecer reglas para la búsqueda de un patrón conocido.

Así pues, en el contexto de este TFG, se va a desarrollar un programa que pueda construir filtros eBPF automáticamente, en función del mensaje concreto que se quiera descartar. Para ello, se procederá a comparar el tamaño del mismo con el de cada paquete que atraviese el filtro, así como un número a determinar de caracteres aleatorios, ya que, si estos coinciden, la probabilidad de que sean el mismo mensaje es muy alta.

Para ello se hará uso de la metodología scrum, un modelo de desarrollo ágil y flexible que se divide en sprints. En cada ciclo se trabaja sobre una tarea o requisito concreto, realizándose así entregas parciales y regulares del producto final que permiten alcanzar mejores resultados.

El desarrollo se hará utilizando un equipo con sistema operativo Ubuntu Desktop 19.04 que cuenta con: procesador Intel Core i7 7500U, 12Gb de RAM y HDD de 2Tb.