Youtube.com Análisis

Carlos Carrillo

Pablo Cid

2019-01-13

Vamos a realizar el análisis del dominio https://www.youtube.com para ver cuantos enlaces debería seguir un usuario de este dominio para alcanzar un dominio potencialmente peligroso con contenido de malware o spam.

Obtener información del dominio

La primero que necesitamos realizar es una exploración del dominio para obtener todos los posibles links que podemos alcanzar. Para ello vamos a descargar si seguir todos los links que hemos encontrado. Para evitar que el proceso no acabe nunca, hemos limitado la busqueda de links con los siguientes parametros:

- Máximo número de links a seguir dentro de un mismo dominio 5
- Máximo número de links a seguir en profundidad 3

A continuación podemos ver un grafo de los dominios con mayor probabilidad de alcanzar, hemos limitado el grafo a 10 dominios para poder visualizarlo correctamente.

```
GroupAssignmentPackage::build_domain_graph(domain_info$links,max_domains_to_show)
```

Obtener información de dominios peligrosos

Hemos utilizado varias fuentes de información para poder decidir si un dominio contiene malware o spam.

- Malware Domain List
- Ultimate Hosts Blacklist

Ambas fuentes nos acaban proporcionando un listado de dominios poco fiables.

Actualmente tenemos un listado con 1360706 dominios con malware.

Análisis

Con la información de los links accesible desde el dominio analizado y con el listado de dominios con malware, ya podemos cruzar los datos y comprobar si hay links accesibles potencialmente peligrosos o no, y en caso afirmativo cuantos enlaces debería seguir el usuario.

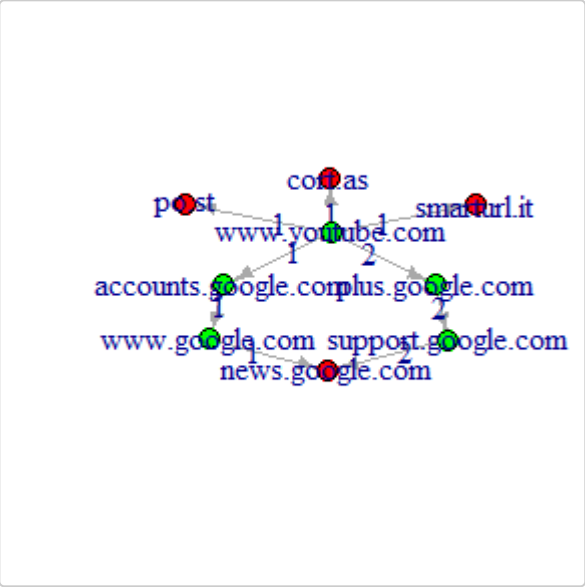
En este caso podemos concluir que el dominio si tiene malware accesible

A continuación podemos ver el listado de malware y desde que dominio es accesible

```
analysis %>% filter(contains_malware == T) %>% select("link", "originDomain")
#>                                    link       originDomain
#> 1                  http://po.st/SbMDYT    www.youtube.com
#> 2                 http://cort.as/-Deqz    www.youtube.com
#> 3                  http://cort.as/YGC9    www.youtube.com
#> 4  http://smarturl.it/NocheDeFantasia    www.youtube.com
#> 5 https://news.google.com/nwshp?hl=es     www.google.com
#> 6 https://news.google.com/nwshp?hl=en support.google.com
#> 7 https://news.google.com/nwshp?hl=en support.google.com
```

De forma más visual, podemos ver el grafo de conexiones para alcanzar todos los dominios con malware:

```
if (contains_malware)
{
  GroupAssignmentPackage::build_malware_graph(analysis)
}
```



Conclusiones

Como podemos observar es muy fácil que un usuario acabe en un dominio potencialmente peligroso. Por otro lado, con diferentes análisis que hemos realizado podemos observar que las fuentes de información de malware/spam quizá se tendría que realizar un filtrado, con una fuente de dominios de confianza ya que producen muchos falsos positivos.