# Estudio del tráfico de fondo de Internet mediante un telescopio de red situado en España

Rodolfo García-Peñas Rafael A. Rodríguez-Gómez Gabriel Maciá-Fernández
Universidad de Granada - *Network Engineering & Security Group* (NESG)
rodgar@correo.ugr.es, {rodgom, gmacia}@ugr.es

***Resumen*—El tráfico de fondo de Internet está formado por aquellos paquetes de red que son recibidos de forma no solicitada. Es un tráfico generado usualmente en las fases preliminares de ataques por equipos que realizan enumeraciones de objetivos y servicios disponibles, enviado como respuestas a ataques de denegación de servicio, o enviado por error en configuraciones y comandos incorrectos. Su adquisición y análisis permite observar lo que está ocurriendo en Internet y es una importante herramienta para la identificación de nuevos tipos de ataques y de atacantes. La adquisición de este tráfico se realiza mediante telescopios de red, nodos que anuncian bloques de direcciones IP sin uso y almacenan el tráfico enviado hacia estas direcciones.**

**El presente artículo realiza un estudio del tráfico recibido por un telescopio de red situado en España durante el año 2023, con más de 4.700 millones de paquetes. Las características del tráfico son comparadas con las de estudios previos, resaltando los cambios de comportamiento y los ataques más comunes.**

***Index Terms*—IBR (*Internet Background Radiation*), IBN (*Internet Background Noise*), *network telescope*, *backscatter*.**

**Tipo de contribución:** *Investigación original*

## I. Introducción

La adquisición y análisis de tráfico de red se ha utilizado de forma habitual para la identificación de comportamientos anómalos que pudieran indicar la existencia de código malicioso, ataques a los sistemas informáticos e incluso de atacantes. Dentro de los diferentes tráficos de red, el tráfico de fondo de Internet, IBR (*Internet Background Radiation*), es el tráfico de red que transita por Internet pero que no ha sido solicitado por ningún sistema [1]. Desde hace años, la adquisición y análisis de tráfico IBR ofrece una información relevante sobre lo que está ocurriendo en Internet, permitiendo identificar desde donde se envía el tráfico, hacia qué destinos, tipos de ataques que están apareciendo y cuáles han dejado de ser interesantes para los atacantes, etc.

Los sistemas de adquisición de tráfico de fondo de Internet más habituales son los telescopios de red [2]. Principalmente consisten en redes que anuncian un conjunto de direcciones IP que no están en uso, por lo que no se origina tráfico desde ellas y, por lo tanto, todo el tráfico recibido es tráfico no solicitado o tráfico IBR [1].

Dos características importantes de los telescopios de red son el número de direcciones IP que anuncian en Internet y la dispersión geográfica de estas [3]. Un mayor número de direcciones IP permite una superficie de adquisición de tráfico más amplia y con ello una mayor probabilidad de encontrar más casos de tráfico IBR, más orígenes y más tipos de ataque, o realizar esta adquisición en menos tiempo [3]. Sin embargo, una consecuencia indeseada de este aumento en la superficie de adquisición es la obtención de un mayor volumen de tráfico repetido, que no aportará información adicional. Por otro lado, la dispersión geográfica implica la utilización de diferentes bloques de direcciones, anunciadas mediante distintas rutas BGP (*Border Gateway Protocol*). De esta forma, sería posible que un telescopio de red distribuido [3] pudiera conocer si el tráfico IBR tiene un carácter más global o más localizado. La consecuencia indeseada de esta dispersión es la necesidad de disponer de infraestructura en un mayor número de emplazamientos, con diferentes conexiones a Internet y mayores costes.

Los estudios relativos a la adquisición de tráfico IBR se dividen entre los que describen, analizan y operan los sistemas de adquisición (telescopios de red), y los que centran sus esfuerzos en el análisis de la información recolectada [2]. La puesta en marcha de un telescopio de red requiere de infraestructura y direccionamiento IP, por lo que la mayoría de los análisis utilizan telescopios ya creados, basándose en información de un mismo conjunto de direcciones IP de destino. Debido a esto, algunos telescopios de red tienen direcciones conocidas o que se han filtrado, y para evitarlos, los atacantes pueden omitir el envío de paquetes hacia ellos.

El objetivo del presente artículo es la identificación, clasificación, búsqueda de patrones y comportamientos de tráfico mediante la descripción y análisis del tráfico IBR almacenado por un telescopio de red español, así como la comparación de los resultados obtenidos con otros estudios realizados.

Para la realización de este estudio se ha utilizado la información de un telescopio de red con 1.024 direcciones, divididas en cuatro redes diferentes y situado en España. Si bien existen distintos artículos sobre el estudio del tráfico IBR, no se han encontrado ninguno centrado en este país ni utilizando este direccionamiento.

La estructura de este artículo es la siguiente. En la Sección II se muestran los antecedentes y el estado del arte del tráfico IBR, analizando los diferentes tipos de tráfico IBR y la forma de detectarlos y clasificarlos. En la Sección III se describen los datos obtenidos por el telescopio de red, así como la metodología de análisis utilizada. En la Sección IV se presenta un análisis estadístico de la información recogida durante el periodo de un año por el telescopio de red. En la Sección V se exploran de algunos de los ataques detectados. Finalmente, en la Sección VI se discuten las conclusiones del estudio y se proponen unas posibles líneas de trabajo futuro.

## II. Antecedentes y estado del arte

La mayor parte del tráfico IBR tiene como origen varias causas, tales como la generación del mismo por parte de *malware*, la respuesta a los ataques de denegación de servicio distribuido (DDoS, Distributed Denial of Service) [2][4], tráfico de reflexión [5], la generación de tráfico por parte

de *botnets* [6], la enumeración de nodos y puertos [1] y un largo etcétera. Existen varias clasificaciones de este tipo de tráfico [2][7][8] que de forma resumida permiten organizar el tráfico IBR en tres categorías principales: *i*) tráfico de actividad directa, *ii*) tráfico reflejado o *backscatter* y *iii*) errores de configuración.

El tráfico de actividad directa es aquel en que el sistema que origina el tráfico intenta obtener información o impactar de alguna forma en el nodo destino. Estos ataques pueden ser de tipo lógico, explotando vulnerabilidades de los protocolos y servicios, o ataques de inundación (*flood*). Ejemplos de este tipo de tráfico serían la enumeración de nodos y servicios, la propagación de diferentes tipos de *malware*, etc.

El tráfico *backscatter* [9] es el que se produce como consecuencia de los ataques DDoS, donde el emisor del ataque realiza *spoofing* de su dirección IP, provocando que las respuestas de este tráfico lleguen a destinos que no generaron las solicitudes.

En los errores de configuración, el tráfico generado puede ser la causa de una configuración incorrecta de servicios, direcciones IP mal codificadas en software, etc.

Categorizar el IBR dentro de alguno de estos tres tipos puede ser una tarea compleja. Por ejemplo, no hay distinción entre el tráfico generado por una enumeración de nodos mediante el comando `ping`, o la codificación errónea en software de una dirección de destino. Ataques de reflexión por amplificación podrían ser erróneamente identificados como tráfico de actividad directa al nodo destino de las direcciones IP. El análisis del tráfico en conjunto, donde se puedan observar múltiples orígenes o destinos, puede ayudar a clasificar los tipos de ataque. Por ejemplo, si varias direcciones IP de un telescopio de red muestran solicitudes de `ping`, indicaría una enumeración de equipos y no un error de codificación en software.

La adquisición de tráfico IBR se ha realizado desde hace mucho tiempo, por ejemplo, en 2007 Allman [10] documentó el análisis de más de 12 años de registros de red, mostrando que ya se adquiría tráfico IBR en 1995. Los sistemas de adquisición reciben diferentes nombres, como *darknets*, telescopios de red, sumideros de red, etc. si bien el funcionamiento de todos ellos es el mismo: Anunciar rangos de direccionamiento que no están en uso y adquirir el tráfico que tiene como destino estas direcciones.

Desde 1995 la adquisición de tráfico IBR ha ido evolucionando. Para la realización de ciertos análisis se han utilizado redes con nodos activos y direcciones no usadas, en lo que se conoce como *greynets* [11][12][13], que hacen más complicado a los generadores de tráfico IBR identificar que la red no tiene uso y de esta forma poder evitarla. También se han creado sistemas que responden a ciertos tipos de paquetes [14], emulando la existencia de nodos y servicios y permitiendo que el origen envíe tráfico adicional, con lo que es posible capturar mayor información y diversidad de tipos de ataque.

## III. Metodología

En este apartado se describen los conjuntos de datos utilizados, así como los procedimientos seguidos para el análisis de los mismos.

### *III-A. Descripción de los datos*

El conjunto de datos sobre el que se ha realizado el análisis se ha obtenido de un telescopio de red formado por 1.024 direcciones IPv4. De cara a no difundir el direccionamiento utilizado en el telescopio, de forma que continúe siendo anónimo y pueda seguir realizando su función, durante este estudio se agruparán en cuatro redes /24 diferentes (*Red_1*,...*Red_4*). *Red_1* y *Red_2* se encuentran en una clase A (red /8) y *Red_3* y *Red_4* en otra red de clase A diferente.

Las redes están registradas en la base de datos de RIPE NCC [15] mediante un objeto *inetnum* (representa un bloque de direcciones) como en estado *Allocated*, lo que indica que han sido entregadas por RIPE NCC para su uso a un LIR (*Local Internet Register*, una empresa, universidad, etc.). No existe un objeto de tipo *inetnum* con estado *Assigned* (asignado), por lo que las redes pertenecen al LIR pero no están registradas como en uso. El registro indica que el país (*country*) es España (*ES*) para todas las direcciones.

Las redes se anuncian mediante el protocolo BGP. Se ha verificado la visibilidad de las rutas durante todo el periodo de tiempo de captura mediante la utilidad de estadísticas de red de RIPE NCC [16], siendo visible durante todo el periodo y la mayor parte del tiempo por los 372 *peers* existentes.

Los datos han sido recogidos en ficheros PCAP (*Packet CAPture*), formato de fichero estándar utilizado en la captura de datos por los telescopios de red, por lo que se dispone de toda la información (cabeceras y datos) de los paquetes. El periodo de tiempo de captura es desde el día 2 de enero hasta el 31 de diciembre del año 2023.

### *III-B. Metodología de análisis*

Debido al elevado volumen de datos a tratar, los análisis se han realizado a diferentes niveles de profundidad:

1. Análisis estadístico de todo el periodo. Se ha desarrollado un análisis estadístico de la información de todo el año 2023, considerando estadísticos como es el número de paquetes diarios recolectados, el tamaño medio del paquete, el volumen de datos recogido, etc. Para analizar esta información se han utilizado herramientas como `capinfos` o `tshark`.
2. Análisis detallado de un mes. Se ha realizado un análisis en mayor profundidad, utilizando para ello un volumen de datos inferior (datos recogidos durante un mes) que permite un análisis en mayor detalle. En este caso, se ha elegido el mes de octubre de 2023, debido a que es un mes relativamente reciente, con un número de paquetes superior a la media y sin grandes picos de tráfico. Para este análisis se han utilizado herramientas como `tshark`, `tcpdump` y `wireshark`, con las que se analiza la distribución de protocolos, puertos, etc. Este análisis también se ha desarrollado en momentos específicos, como días con un elevado número de paquetes o días con tamaños de paquete superiores a la media.
3. Análisis específicos y de carga útil. Se ha analizado en mayor detalle el contenido de paquetes en conjuntos de datos acotados. Para este análisis, además de las herramientas anteriores, se ha utilizado `Python` y la librería `Scapy`.

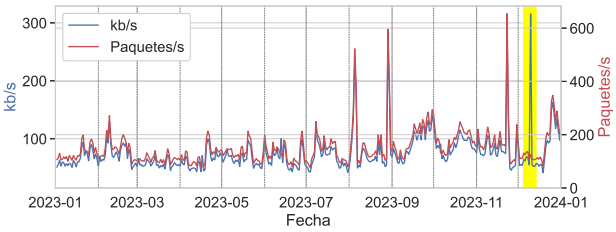


Figura 1. Número de paquetes y kilobits recibidos por segundo.

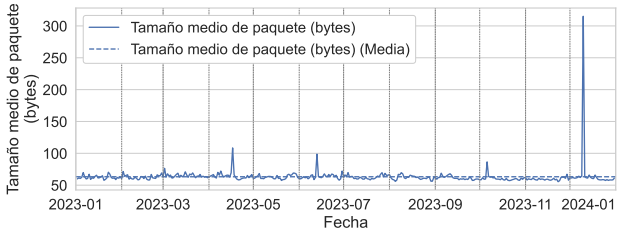


Figura 2. Tamaño medio del paquete recibido.

## IV. Análisis del tráfico IBR

Durante el periodo completo de la captura se han adquirido 4.751.072.356 paquetes IPv4 en las cuatro redes, generando un total de 362,39 gigabytes de datos, con una media diaria cercana al gigabyte.

La variación en el volumen de datos adquirido diariamente, y con ello el tamaño de los ficheros de captura, es debida a dos factores, el tamaño medio de los paquetes y el número de paquetes recibidos, o ambos factores de forma simultánea. En la Figura 1, para cada uno de los días del conjunto de datos, se puede observar en rojo el número de paquetes recibidos por segundo y en azul el número de kilobits recibidos por segundo. Ambas medidas mantienen una relación directa durante prácticamente todo el periodo de tiempo, mostrándose con una evolución paralela en la figura. Esto indicaría que, de forma general, cuando el tamaño del fichero es mayor se debe a que se recibe un mayor número de paquetes. En el único momento donde existe una gran diferencia es el día 10 de diciembre (resaltado en amarillo), donde el volumen de kilobits aumenta, lo que indica que el tamaño del paquete es significativamente mayor que la media del resto de días. Este caso específico se analizará posteriormente en el Apartado V. En la Figura 1 también se puede observar que durante los meses de septiembre y diciembre el volumen medio diario fue superior a otros meses, con aproximadamente 1,5 gigabytes de media al día. Existen varios días puntuales en los que el tráfico supera los tres gigabytes, son el 5 de agosto (3,24 gigabytes), 29 de agosto (3,71 gigabytes), 23 de noviembre (4,03 gigabytes) y 10 de diciembre (3,57 gigabytes).

La Figura 2 muestra el tamaño medio diario de los paquetes recibidos e indica también, mediante una línea punteada, la media total del tamaño del paquete, que se sitúa en los 63,2 bytes. La gráfica mantiene un valor prácticamente constante durante todo el periodo, a excepción de varios días puntuales. El más significativo es el día 10 de diciembre (como se observaba en la Figura 1), donde el tamaño medio del paquete aumenta hasta 315,13 bytes. Otros días con valores que resaltan especialmente en la gráfica son el 6 de octubre (86,55 bytes), el 13 de junio (98,74 bytes) y el 17 de abril (108,48 bytes).

En la Figura 3 se puede observar que la distribución de tráfico entre las diferentes redes del telescopio es homogénea, con un volumen aproximado del 25 % cada una de ellas (recuérdese que el número de direcciones IP de cada red es el mismo). Si observamos la Figura 1, vemos que hay periodos donde el tráfico es menor, como por ejemplo el mes de marzo, mientras que en otros meses el tráfico es mayor (por ejemplo, septiembre). Sin embargo estos crecimientos de tráfico no se ven reflejados en los porcentajes mostrados en la Figura 3, por lo tanto, de forma habitual, cuando hay un aumento de tráfico, éste ocurre en todas las redes por igual.

Con el objetivo de observar el tráfico de cada red de forma independiente y a la vez poder comparar los tráficos entre todas ellas, se ha decidido agrupar gráficamente la información de las cuatro redes de la siguiente manera. Se han situado las redes 1 y 3 en la parte inferior de la figura y las 2 y 4 en la parte superior. Esto permite observar cómo la suma de las redes 1 y 3 (lo mismo ocurre con la 2 y 4), representa de forma muy fiel el 50 % del volumen durante todo el periodo de captura. Así, se puede concluir que el tráfico recibido por el grupo de redes 1 y 2, que se encuentran en la misma clase A, es similar entre ellas, ocurriendo lo mismo con las redes 3 y 4 de la otra clase A. Además, se muestra cómo las variaciones de volúmenes de tráfico son simétricas entre las redes de las mismas clases A, representándose como pequeños picos en el 25 % y 75 % y ocurriendo de forma habitual en todo el gráfico.

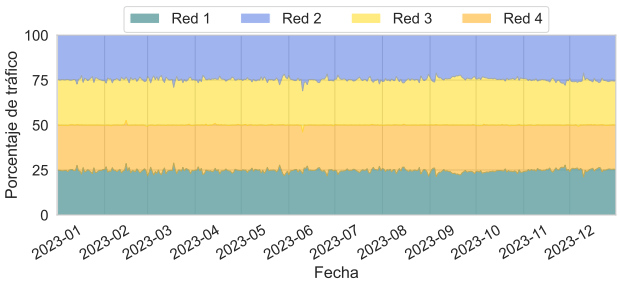


Figura 3. Porcentaje diario de paquetes recogidos por cada red.

*Analisis detallado del mes de octubre*

Teniendo en cuenta que la distribución entre las diferentes redes del telescopio de red es homogénea durante todo el periodo de la muestra anual, así como que el tamaño del paquete se mantiene en un valor próximo a la media, debido a limitaciones computacionales, se ha decidido realizar un análisis más profundo utilizando un periodo de tiempo menor, el mes de octubre del año 2023.

La información en formato PCAP del mes de octubre tiene un tamaño superior a 46,41GB de datos, con un total de

493,33 millones de paquetes. De media, cada dirección IP del telescopio de red recibió 10,79 paquetes IP por minuto. Esta tasa de paquetes está alineada con la obtenida en otros estudios previos [8] [17][21], si bien es ligeramente superior.

En la Tabla I se detalla la distribución de protocolos especificados en la cabecera IP de los paquetes capturados por el telescopio de red durante el mes de octubre. La gran mayoría es de tipo TCP (95,69 %), recibiéndose con frecuencia paquetes de tipo UDP (3,74 %), ICMP (0,51 %) y, en mucha menos cantidad, GRE (0,06 %). El resto de protocolos recogidos en la Tabla I son prácticamente residuales.

Tabla I
DISTRIBUCIÓN DE LOS PROTOCOLOS ESPECIFICADOS EN LA CABECERA IP CAPTURADOS DURANTE EL MES DE OCTUBRE DE 2023.

| Protocolo (Número) | Paquetes | Porcentaje |
|---|---|---|
| TCP (6) | 472.057.445 | 95,69 % |
| UDP (17) | 18.474.138 | 3,74 % |
| ICMP (1) | 2.501.577 | 0,51 % |
| GRE (47) | 297.557 | 0,06 % |
| IPv6 (41) | 411 | 0,00 % |
| IGMP (2) | 55 | 0,00 % |
| IP Encapsulado (4) | 44 | 0,00 % |
| IPv6 *Hop by Hop* (0) | 2 | 0,00 % |
| *Core Bases Trees* (CBT) (7) | 1 | 0,00 % |

Nótese que solamente se observan tres paquetes correspondientes a los protocolos CBT *Core Bases Trees*, descrito por la RFC2189, e IPv6 (encapsulado en IPv4). Se ha observado que el *payload* de datos de estos tres paquetes es parecido, comenzando por los bytes `01 bb`, y seguidos de seis bytes de datos, cuatro bytes a cero y dos bytes con valor `50 04`. Finaliza con bytes a cero hasta llegar a una longitud de veintiséis bytes en total. Coincidiría con una cabecera TCP mal formada, donde los bytes `01 bb` indicarían el puerto 443 y los bytes `50 04` la longitud de la cabecera (5 bytes) y el *flag* de RST activo. Se considera por tanto que estos son ejemplos de paquetes mal formados y en realidad no se habría recibido ningún paquete válido de estos protocolos de transporte.

### *IV-A. Protocolo TCP*

Analizando el protocolo TCP, que es el mayoritario en el tráfico, se pueden observar en la Tabla II los *flags* activos en los paquetes recibidos a lo largo de todo el mes. Más del 99 % de los paquetes tienen únicamente el *flag* "S" (SYN) activo, que muestra la existencia de nodos origen que buscan establecer una conexión nueva. El envío de estos paquetes SYN es una práctica habitual en la fase de enumeración de servicios. En estudios previos [2][18] el *flag* SYN se muestra en una proporción similar, si bien estos estudios no hacen referencia a los *flags* ECE y CWR que en este caso han sido diferenciados.

En el proceso de establecimiento de una sesión TCP, denominado *three-way handshake*, el origen de la comunicación envía un paquete con el *flag* SYN activo. Si el nodo destino está dispuesto a establecer la comunicación, entonces responderá con un paquete con los *flags* SYN+ACK y el origen responderá con un paquete con el *flag* ACK, pasando la comunicación al estado de establecida. Sin embargo, si el destino no desea establecer la comunicación responderá con el *flag* RST y si no puede hacerlo, entonces responderá

Tabla II
DISTRIBUCIÓN DE LOS *FLAGS* DE LOS PAQUETES TCP CAPTURADOS DURANTE EL MES DE OCTUBRE DE 2023.

| *Flags* | Paquetes | Porcentaje |
|---|---|---|
| S | 467.373.974 | 99,01 % |
| SEC | 500.680 | 0,10 % |
| SC | 35.539 | 0,01 % |
| SA | 2.582.881 | 0,55 % |
| R | 796.095 | 0,16 % |
| A | 358.930 | 0,08 % |
| FSRPAU | 167.563 | 0,04 % |
| *Sin flags* | 91.917 | 0,02 % |
| PA | 78.466 | 0,02 % |
| RA | 54.676 | 0,01 % |
| *Otras combinaciones* | 16.724 | 0,00 % |

***Flags*:** S: SYN, A: ACK, F: FIN, R: RST, U: URG, P: PUSH, E: ECE, C: CWR

con los *flags* ACK+RST. De esta forma, en estudios sobre el tráfico *backscatter* se identifican los paquetes con *flags* SYN+ACK, RST, ACK y RST+ACK como respuestas a tráfico de DDoS [18]. Estos casos (SA, R, A y RA) aparecen entre los tráficos identificados de la Tabla II, complementando prácticamente el 100 % de los tipos de tráfico. Adicionalmente, hay dos configuraciones especiales de *flags* que agrupan un 0,06 % de los paquetes: Todos activados (*Xmas Tree Scan*) y todos desactivados (*Null Scan*) [19]. Estas configuraciones se utilizan habitualmente en escaneos de puertos avanzados, sin embargo, el comportamiento de los telescopios de red de no responder al tráfico enviado, puede indicar que el puerto esta abierto o filtrado. En el caso de todos los *flags* activos, hay 192 nodos origen que han realizado este envío de paquetes, y de ellos únicamente 2 (un paquete cada dirección IP) han utilizado ICMP para comprobar si el nodo destino estaba activo. Estas estaciones tampoco han realizado ningún otro envío de tráfico TCP con otros *flags*.

En relación con los orígenes del tráfico, una de las cuestiones es si todas las direcciones destino reciben el mismo tráfico desde los mismos orígenes, lo que ocasionaría grandes volúmenes de información redundante. Analizando los pares "IP origen" – "IP destino", se ha observado que durante el periodo del mes de octubre hubo 472.057.445 paquetes TCP recibidos en 26.417.262 envíos de tráfico entre pares diferentes. Observando el número de orígenes, existen 1.443.904 direcciones IP de origen distintas, por lo que de media, cada IP origen ha enviado tráfico al menos a 18 direcciones distintas. Para analizar esta relación entre direcciones IP origen y destinos contactados, en la Figura 4 se enfrentan en el eje Y las direcciones IP origen (el eje muestra el primer octeto de la dirección) y en el eje X las direcciones de las cuatro redes del telescopio de red, representadas por el identificador de la red y el *offset* de la dirección IP. Por ejemplo, suponiendo que las redes del telescopio de red fueran desde "192.168.0.0/24" hasta "192.168.3.0/24", la dirección IP de destino "192.168.1.128" se representaría en el eje X como "2 – 128", correspondiendo el "2" a la segunda de las redes (comienzan en la red "0") y el 128 a la dirección IP de la red.

Los puntos muestran el número de paquetes recibidos en ese par "IP origen" – "IP destino" durante el mes.

Mediante diferentes colores la figura muestra el número

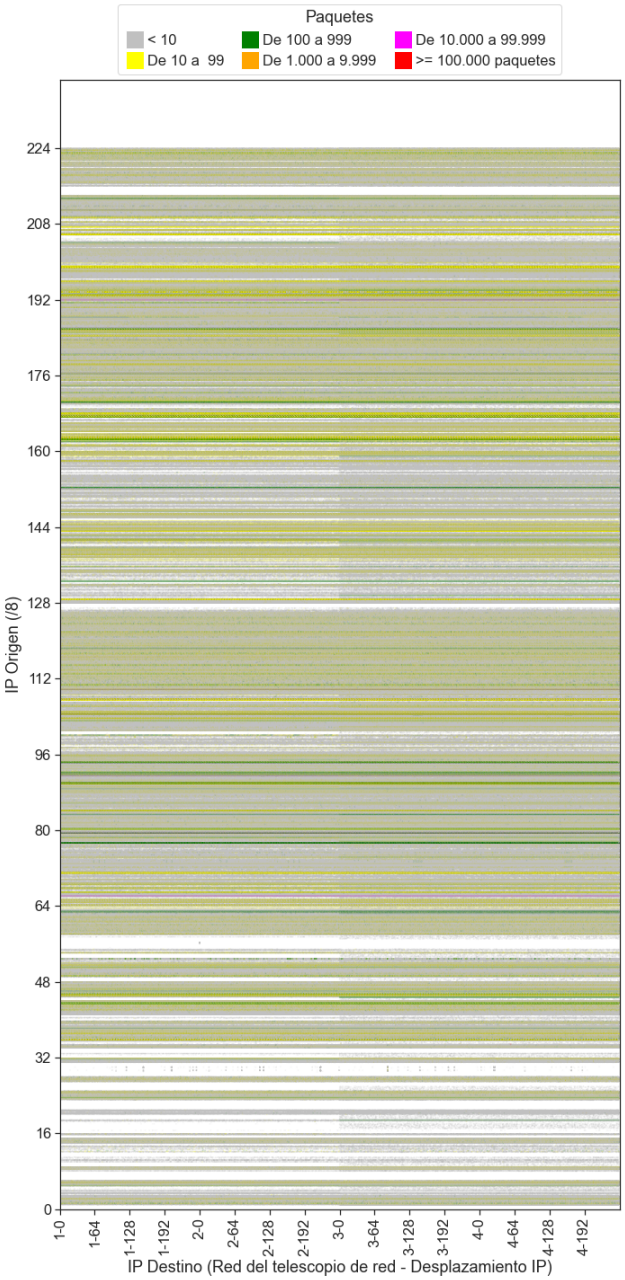


Figura 4. Paquetes desde IP origen a las IP de las redes del telescopio.

de paquetes enviados por un origen a un destino. Se puede observar como la mayoría de los puntos representan el envío de menos de 10 paquetes (color gris) o menos de 100 paquetes (color amarillo). En algunos casos puntuales se pueden observar líneas que muestran el envío de gran cantidad de paquetes.

En esta figura se pueden distinguir fundamentalmente las siguientes situaciones:

1. Es habitual que desde una dirección IP origen se envíe tráfico a todas las direcciones IP de las diferentes redes (línea horizontal completa). Esta situación se repite de forma constante para una gran proporción de las direcciones origen, lo que indicaría tráfico de escaneo y enumeración de servicios y que coincide con el tipo de paquete TCP enviado (con el *flag* SYN).
2. Las redes 3 y 4 reciben un mayor número de paquetes. Puede verse cómo la mitad derecha de la figura presenta un mayor sombreado en gris. Esto coincide con la información mostrada en la Figura 3, donde se observa ligeramente cómo los ejes del 25 % y 75 % son superados por estas redes en el mes de octubre, mostrando la diferencia de paquetes.
3. Si bien hay tráfico desde prácticamente todas las redes /8, existen varias de ellas que no muestran, o muestran muy poco tráfico. Estas redes, mostradas en blanco al no existir envíos de tráfico, coinciden con las redes de tipo *LEGACY* en las asignaciones de direccionamiento realizadas por IANA [20]. Este tipo de redes son las primeras que se asignaron, antes de la creación de los RIR y cada una de ellas pertenece normalmente a una única entidad. Desde las redes reservadas (0/8, 127/8, etc.) tampoco se recibe tráfico.

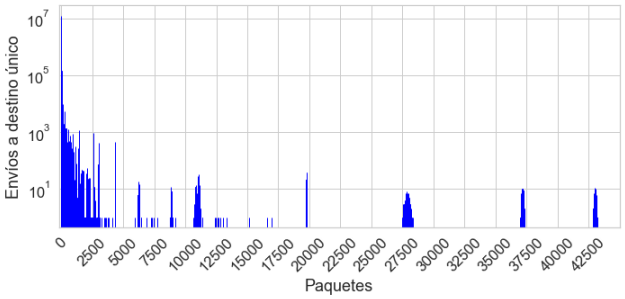


Figura 5. Distribución de paquetes enviados a un destino (Hasta 45.000).

*Mayores generadores de tráfico TCP*

Con el objetivo de analizar los orígenes que generan una mayor cantidad de tráfico en la Figura 5 se representa la distribución del número de paquetes enviados con un mismo par "IP origen" - "IP destino" durante el mes de octubre. Para mejorar la visualización, la Figura 5 presenta solamente aquellas muestras que no exceden de 45.000 paquetes, excluyendo la generación de 131.070 y 597.418 paquetes por parte de dos direcciones IP diferentes. La primera de ellas corresponde a tráfico de enumeración (paquetes con el *flag* SYN y sin *payload*), mientras que la muestra con mayor número de paquetes corresponde a tráfico *backscatter* con puerto origen 443 y *flag* RST. Nótese en la figura que la mayoría de los envíos de paquetes tienen un valor inferior a 100. Es interesante analizar algunos orígenes que envían grandes volúmenes de paquetes TCP a un conjunto de destinos. Se han seleccionando los 20 orígenes con más tráfico y se han observado dos casos:

1. Orígenes que envían un elevado número de paquetes a un único destino (4 casos) o a unos pocos (menos de 10 destinos, 7 casos). Todos los casos son de tipo enumeración, con el *flag* SYN activado.
2. Orígenes que envían un elevado número de paquetes a todas las direcciones de todas las redes del telescopio de red (8 casos) o a alguna de las redes (1 caso, a las redes 3 y 4). Se asocian a escaneos completos a todos

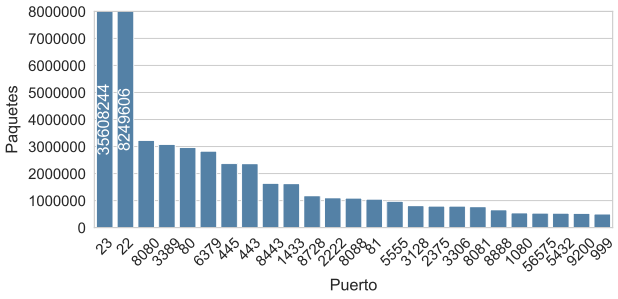


Figura 6. Distribución de puertos TCP con más paquetes recibidos.

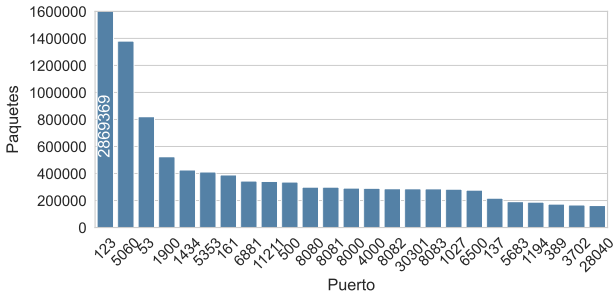


Figura 7. Distribución de puertos UDP con más paquetes recibidos.

los nodos y todos sus puertos, y presentan también el *flag* SYN activado.

Por lo tanto, el mayor número de paquetes corresponde a tráfico asociado con una fase de enumeración. Los tráficos de *backscatter*, con la excepción del envío descrito anteriormente (que representa el 0,15 % del tráfico TCP del mes de octubre), no son los que están generando los tráficos más elevados.

*Análisis de puertos*

En la Figura 6 se representan los 25 puertos TCP que más paquetes han recibido. Al igual que en otros estudios previos [2][21], los puertos que han recibido más paquetes son los relacionado con accesos remotos, como son *Secure SHell* (SSH, puerto 22 y alternativamente 2222), Telnet (puerto 23) y acceso mediante RDP (*Remote Desktop Protocol*, puerto 3389). Los accesos web (puertos 80, 81, 8080, 8081, 443, 8443) y los puertos relativos a servidores *proxy*, como Squid (puerto 3128) o SOCKS (puerto 1080), también están entre los que más paquetes reciben. En relación con bases de datos se pueden encontrar el servicio Redis (puerto 6379), ElasticSearch API (puerto 9200), Microsoft SQL Server (puerto 1433), PostgreSQL (puerto 5432) y MySQL (puerto 3306). El servicio de compartición de ficheros mediante SMB (puerto 445) también está entre los que reciben mayor número de paquetes. Es de especial relevancia tener en cuenta que algunos de los puertos indicados previamente, relacionados con accesos remotos y también algunos relacionados con accesos web, incluyendo los puertos 5555 (originalmente para ADB y TR069) y 56575, están relacionados con troyanos y diversas *botnets* como Mirai [22].

Se ha observado que la gran mayoría de los *payload* de los paquetes TCP están vacíos, dado que son paquetes de tipo SYN y no llevan carga de datos. Únicamente 5,18 millones de paquetes de los 467 millones de paquetes TCP con el *flag* SYN activado, que representa un 1,11 %, tienen algún tipo de *payload*. Es importante resaltar que la utilización de otro tipo de sistemas de recolección de tráfico, como aquellos que responden simulando que existe un nodo o un servicio activo [14][23], permite recoger un mayor volumen de tráfico TCP, incluyendo mayor número de *payloads*, con información más precisa y completa de diferentes tipos de ataque.

### *IV-B. Protocolo UDP*

Con relación al protocolo UDP, como se ha mostrado en la Tabla I, el número de paquetes es de 18,47 millones, representando un 3,74 % de los paquetes del mes de octubre. La principal diferencia con TCP es el tamaño medio del paquete, así en el caso de UDP es de 152,67 bytes, mientras que en TCP es de 41,52. En el caso de TCP un paquete ocupa como mínimo 40 bytes (20 de la cabecera IP y 20 de la cabecera TCP) lo que implica que la mayoría viajan sin *payload* (tamaño medio de 41,52 bytes). En el caso de UDP el tamaño mínimo es de 28 bytes (20 de la cabecera IP y 8 de la cabecera de UDP) y el tamaño medio de 152,67 bytes, así que, a diferencia de TCP, la mayoría incluye *payload*.

Un 30 % de los paquetes UDP (5,4 millones) presenta en la cabecera IP el identificador "54321" (el siguiente identificador más utilizado es el "0" con 201.896 paquetes). El identificador "54321" aparece escrito en el código fuente de ZMap [24] [25], una herramienta especialmente diseñada para realizar escaneos y enumeraciones masivas. En el caso de TCP, este identificador se ha utilizado en un 18,7 % del total de paquetes TCP (88,4 millones). Esta diferencia de porcentaje podría indicar el uso de herramientas diferentes para el escaneo en TCP y en UDP, o que los orígenes del envío de paquetes son diferentes para ambos protocolos, o ambas premisas. Para ello, se han listado las direcciones IP utilizadas en UDP y las utilizadas en TCP y se han comparado. En el caso de UDP se han encontrado 129.321 direcciones IP origen diferentes, mientras que en caso de TCP han sido 1.443.904, 11 veces más. Existen 14.023 direcciones IP origen que aparecen en ambos protocolos, un 9,2 % de las direcciones origen UDP. De esta forma, se puede derivar que la mayoría de los escaneos para TCP y UDP se realizan por orígenes diferentes.

*Análisis de puertos*

Analizando la información relativa a los puertos destino UDP con un mayor número paquetes (ver Figura 7), el puerto con mayor tráfico es el relativo al servicio NTP (puerto 123), con más de 2,8 millones de paquetes. Este puerto es habitualmente utilizado para realizar ataques de amplificación por reflexión. El puerto del servicio SIP (*Session Initiation Protocol*, puerto 5060) se utiliza para la realización de llamadas de voz sobre IP (VoIP) y es utilizado por los teléfonos VoIP y también por centralitas telefónicas, teniendo multitud de ataques diferentes y vulnerabilidades. El servicio de DNS (puerto 53) es utilizado también para la realización de ataques de amplificación por reflexión. Existen varios puertos utilizados para obtener información, siendo los principales NetBIOS (puerto 137), LDAP (puerto 389), memcaché (puer-

to 11211), ws-discovery (puerto 3702) y SNMP (puerto 161). Sin embargo, en UDP llama la atención el volumen de puertos utilizados por troyanos (similar a lo que ocurre en TCP), con puertos como el 4000, 8000, 8080, 8081, etc. Mediante UDP también se intentan realizar intrusiones, por ejemplo mediante acceso VPN (puerto 1194, OpenVPN).

En el Apartado V se describirán algunos de los *payloads* más utilizados en el tráfico capturado relativo al protocolo UDP.

### *IV-C. Protocolo ICMP*

El volumen de tráfico de paquetes ICMP es mucho menor que el destinado a los protocolos anteriormente descritos. La Tabla III muestra los diferentes tipos de paquetes y su volumen. Como se puede observar, la gran mayoría del tráfico (97,95 %) corresponde a solicitudes de *echo* (`ping`). Los paquetes recibidos de tipo 3 (Destino inalcanzable), corresponden habitualmente con tráfico *backscatter*, donde el destino de la información responde al origen con dirección falsificada (*spoofing*) y le informa que no es posible alcanzar el nodo o el puerto indicado. Esta misma situación ocurre con las respuestas de *echo* y de tiempo excedido (tipo 11).

Tabla III
DISTRIBUCIÓN DE LOS DE LOS PAQUETES ICMP CAPTURADOS DURANTE EL MES DE OCTUBRE DE 2023.

| Tipo, Código, Descripción | Paquetes | Porcentaje |
|---|---|---|
| Type:0 Echo Reply | | |
| Code:0 Echo Reply | 19.757 | 0,79 % |
| Code:>0 Código no válido | 198 | 0,01 % |
| Type:3 Destination Unreacheable | | |
| Code:0 Net Unreacheable | 187 | 0,01 % |
| Code:1 Host Unreacheable | 1.468 | 0,06 % |
| Code:2 Protocol Unreacheable | 3.363 | 0,13 % |
| Code:3 Port Unreacheable | 17.149 | 0,69 % |
| Code:4 Fragmentation Needed | 1.038 | 0,04 % |
| Code:10 Communication with Destination Host is Administratively Prohibited | 396 | 0,02 % |
| Code:13 Communication Administratively Prohibited | 1.601 | 0,06 % |
| Type:5 Code:1 Redirect | 3.642 | 0,15 % |
| **Type:8 Code:0 Echo Request** | **2.450.412** | **97,95 %** |
| Type:11 Time Exceeded | | |
| Code:0 Time to Live exceeded in Transit | 2.183 | 0,09 % |
| Code:1 Fragment Reassembly Time Exceeded | 17 | 0 % |
| Type:13 Code:0 Timestamp | 166 | 0,01 % |

## V. ANÁLISIS ESPECÍFICOS Y DE CARGA ÚTIL

En relación con la carga útil y situaciones anómalas, se han seleccionado algunos ejemplos significativos:

### *V-A. Ataques por reflexión NTP*

Los ataques por amplificación por reflexión tienen como objetivo enviar un volumen pequeño de tráfico que ocasionará una respuesta de tráfico con un volumen mucho mayor. Esta acción puede elevar la carga de proceso del nodo que recibe la consulta, así como el tráfico de red. La falsificación mediante *spoofing* de la IP origen permite que el tráfico devuelto sea utilizado para atacar a un tercer nodo, inundándole con el tráfico de la respuesta.

En el caso de NTP el atacante realiza una solicitud de tipo "`REQ_MON_GETLIST`" o de tipo

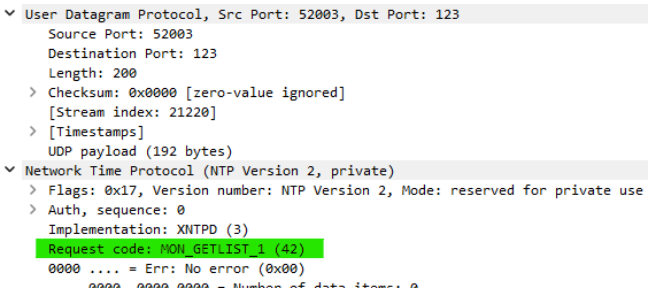


Figura 8. Ejemplo de ataque de reflexión NTP.

"`REQ_MON_GETLIST_1`" [26], solicitando las últimas 600 direcciones IP que han consultado el servidor de hora. La respuesta permite multiplicar por más de 200 el tráfico enviado. En la Figura 8 se puede observar un ejemplo de solicitud NTP de este tipo de ataque capturado en el telescopio.

### *V-B. Tráfico DDoS del día 10 de diciembre*

Tras analizar lo ocurrido el día 10 de diciembre se ha observado que a partir de las 10 horas y 48 minutos el comportamiento del tráfico IBR cambia y comienza a llegar un elevado volumen de tráfico de tipo DNS. Consiste en respuestas del protocolo de resolución de nombres a la petición de información de tipo "ANY" del TLD (*Top Level Domain*) de Sierra Leona (.sl) con el identificador de transacción "0x4567". Esta solicitud se puede realizar, falsificando la IP origen para que la respuesta del servidor DNS le llegue al objetivo del ataque DDoS, utilizando la herramienta `dig` mediante el comando siguiente:

```
dig -t ANY sl
```

Al solicitar al servidor DNS información del tipo "ANY", la respuesta incluye cualquier tipo de información existente (registros A, AAAA, NS, claves criptográficas de DNSSEC, etc.), haciendo la respuesta lo más voluminosa posible. Al ser una respuesta voluminosa (aproximadamente 4.000 bytes), esto produce la fragmentación del paquete, recibiéndose unos 3 o 4 fragmentos de tráfico.

Este es un claro ejemplo de un ataque de amplificación DNS, donde enviándose un volumen pequeño de datos, se pueden generar respuestas de un tamaño 100 veces mayor, inundando así al destino del ataque.

### *V-C. Tráfico de evasión de censura*

Dentro de los *payloads* TCP se ha encontrado que el más utilizado durante el mes de octubre (194.411 paquetes) es el siguiente (se ha modificado el nombre del campo *Host* para preservar la privacidad):

```
GET /?q=ultrasurf HTTP/1.1\r\n
Host: modificado.com\r\n\r\n
```

Son conexiones al puerto 80 (web) con el objetivo de encontrar proxies de entrada que permitan la creación de una VPN entre el usuario y un conjunto centralizado de servidores *proxy* de salida a Internet, que permitan evitar el filtrado, la censura y la monitorización realizada por algunos países [27]. *Ultrasurf* es el nombre del producto *freeware* que realiza la función de *proxy* de entrada.

## VI. Conclusiones

En el presente artículo se han analizado las características y los tipos de tráfico IBR capturados por un telescopio de red español durante el año 2023. En este periodo se han recibido más de 4,75 mil millones de paquetes IP, donde más del 95 % del tráfico corresponde a TCP. Se pueden extraer las siguientes conclusiones:

- Se ha realizado un análisis estadístico del tráfico de todo el año y del mes de octubre centrado en los protocolos ICMP, TCP y UDP. Se ha podido comprobar que el tráfico IBR ha aumentado ligeramente respecto a estudios previos, sin detectarse un mayor volumen de tráfico de redes con origen España.
- Se ha detectado principalmente tráfico de enumeración y de *backscatter*, reportados en estudios previos. Adicionalmente, se han analizado en mayor detalle algunos de los ataques detectados, como algún ataque de DDoS de amplificación y el de evasión de situaciones de censura, que no había sido reportado previamente.

Como posibles líneas de trabajo futuro se propone:

- Realizar una comparativa con otros telescopios de red y con telescopios distribuidos que respondan a ciertos patrones de tráfico (HoDiNT [23]).
- Utilizar herramientas de tipo *tcpflow* y *udpflow* para poder reensamblar los paquetes y realizar búsquedas de información.
- Clasificar automáticamente el tráfico capturado, priorizando aquellos que ofrezcan información relevante.

## Agradecimientos

Se agradece encarecidamente a la persona que nos ha facilitado el acceso a los datos del telescopio de red, sin los cuales no sería posible realizar este artículo.

Esta publicación es parte del proyecto NetSEA-GPT (C-ING-300-UGR23), cofinanciado/a por la Consejería de Universidad, Investigación e Innovación y por la Unión Europea con cargo al Programa FEDER Andalucía 2021-2027. Este artículo se ha publicado en el marco del proyecto de investigación SICRAC (PID2020-114495RB-I00), financiado por MCIN/ AEI /10.13039/501100011033.

## Referencias

[1] R. Pang, V. Yegneswaran, P. Barford, V. Paxson, y L. Peterson, ≪Characteristics of internet background radiation≫, en Proceedings of the 4th ACM SIGCOMM conference on Internet measurement - IMC '04, Taormina, Sicily, Italy, 2004, pp. 27-40. doi: 10.1145/1028788.1028794.

[2] E. Wustrow, M. Karir, M. Bailey, F. Jahanian, y G. Huston, ≪Internet background radiation revisited≫, en Proceedings of the 10th annual conference on Internet measurement - IMC '10, Melbourne, Australia, 2010, pp. 62-74. doi: 10.1145/1879141.1879149.

[3] D. Moore, C. Shannon, G. M. Voelker, y S. Savage, ≪Network Telescopes: Technical Report≫, UC San Diego: Department of Computer Science & Engineering, p. 14, 2004.

[4] Pearson, D. T. ≪An exploration of the overlap between open source threat intelligence and active Internet Background Radiation≫, Rhodes University - Faculty of Science, Computer Science. 2020. Disponible en: http://vital.seals.ac.za:8080/vital/access/manager/Repository/vital:32299?site_name=GlobalView Visitada: 26/03/2024.

[5] R. R. Nuiaa, S. Manickam, y A. H. Alsaeedi, ≪Distributed reflection denial of service attack: A critical review≫, IJECE, vol. 11, n.º 6, p. 5327, dic. 2021, doi: 10.11591/ijece.v11i6.pp5327-5341.

[6] M. Antonakakis et al., ≪Understanding the mirai botnet≫, en Proceedings of the 26th USENIX Security Symposium, en Proceedings of the 26th USENIX Security Symposium. USENIX Association, 2017, pp. 1093-1110.

[7] D. Yates, ≪A System for Characterising Internet Background Radiation≫, 2014. Disponible en: http://www.cs.ru.ac.za/research/g11y1408/thesis.pdf. Visitada: 26/03/2024.

[8] M. Zolotykh, ≪Comprehensive Classification of Internet Background Noise≫, en 2020 Global Smart Industry Conference (GloSIC), IEEE, 2020, pp. 35-41. doi: 10.1109/GloSIC50886.2020.9267850.

[9] E. Balkanli y A. N. Zincir-Heywood, ≪On the analysis of backscatter traffic≫, en 39th Annual IEEE Conference on Local Computer Networks Workshops, Edmonton, AB, Canada: IEEE, sep. 2014, pp. 671-678. doi: 10.1109/LCNW.2014.6927719.

[10] M. Allman, V. Paxson, y J. Terrell, ≪A brief history of scanning≫, en Proceedings of the 7th ACM SIGCOMM conference on Internet measurement - IMC '07, San Diego, California, USA, 2007, p. 77-82. doi: 10.1145/1298306.1298316.

[11] W. Harrop y G. Armitage, ≪Defining and Evaluating Greynets (Sparse Darknets)≫, en The IEEE Conference on Local Computer Networks 30th Anniversary (LCN'05)l, Sydney, NSW, Australia: IEEE, 2005, pp. 344-350. doi: 10.1109/LCN.2005.46.

[12] L. Miao, W. Ding, y H. Zhu, ≪Extracting Internet Background Radiation from raw traffic using greynet≫, en 2012 18th IEEE International Conference on Networks (ICON), Singapore, Singapore: IEEE, dic. 2012, pp. 370-375. doi: 10.1109/ICON.2012.6506586.

[13] P. Richter y A. Berger, ≪Scanning the Scanners: Sensing the Internet from a Massively Distributed Network Telescope≫, en Proceedings of the Internet Measurement Conference, New York, NY, USA, oct. 2019, pp. 144-157. doi: 10.1145/3355369.3355595.

[14] R. Hiesgen, M. Nawrocki, A. King, A. Dainotti, T. C. Schmidt, y M. Wählisch, ≪Spoki: Unveiling a New Wave of Scanners through a Reactive Network Telescope≫, presentado en 31st USENIX Security Symposium (USENIX Security 22), 2022. Disponible en: https://www.usenix.org/conference/usenixsecurity22/presentation/hiesgen. Visitada: 26/03/2024.

[15] Réseaux IP Européens Network Coordination Centre. Disponible en: https://ripe.net/. Visitada: 26/03/2024.

[16] RIPEStats. Disponible en: https://stat.ripe.net/. Visitada: 26/03/2024.

[17] M. Bailey, E. Cooke, F. Jahanian, A. Myrick, y S. Sinha, ≪Practical Darknet Measurement≫, en 2006 40th Annual Conference on Information Sciences and Systems, Princeton, NJ, USA: IEEE, mar. 2006, pp. 1496-1501. doi: 10.1109/CISS.2006.286376.

[18] N. Furutani, T. Ban, J. Nakazato, J. Shimamura, J. Kitazono, y S. Ozawa, ≪Detection of DDoS Backscatter Based on Traffic Features of Darknet TCP Packets≫, en 2014 Ninth Asia Joint Conference on Information Security, Wuhan, China: IEEE, sep. 2014, pp. 39-43. doi: 10.1109/AsiaJCIS.2014.23.

[19] G. Lyon, ≪Nmap network scanning: official Nmap project guide to network discovery and security scanning≫, Zero-day Release: May 2008. Sunnyvale, CA: Insecure.Com LLC, 2010.

[20] ≪IANA IPv4 Address Space Registry≫ Disponible en: https://www.iana.org/assignments/ipv4-address-space/ipv4-address-space.xhtml. Última modificación 18-12-2023. Visitada: 26/03/2024.

[21] G. Huston, ≪Dark Traffic≫, The ISP Column, oct. 2019. https://www.potaroo.net/ispcol/2019-10/dark.html. Visitada: 26/03/2024.

[22] JPCERT Coordination Center, ≪JPCERT/CC Internet Threat Monitoring Report April 1, 2023 - June 30, 2023≫, JPCERT Coordination Center ago. 2023. Disponible en: https://www.jpcert.or.jp/english/doc/TSUBAMEReport2023Q1_en.pdf. Visitada: 26/03/2024.

[23] R. García-Peñas, R. A. Rodríguez-Gómez, y G. Maciá-Fernández, ≪HODINT: Arquitectura distribuida para la recolección y análisis del tráfico de fondo de Internet≫, en VIII Jornadas Nacionales de Investigación en Ciberseguridad JNIC2023, 2023.

[24] V. Ghiette, N. Blenn, y C. Doerr, ≪Remote Identification of Port Scan Toolchains≫, en 2016 8th IFIP International Conference on New Technologies, Mobility and Security (NTMS), nov. 2016, pp. 1-5. doi: 10.1109/NTMS.2016.7792471.

[25] Z. Durumeric, E. Wustrow, y J. A. Halderman, ≪ZMap: Fast Internet-Wide Scanning and its Security Applications≫, en 22nd USENIX Security Symposium (USENIX Security 13), Washington, D.C., ago. 2013, pp. 605–620. Disponible en: https://www.usenix.org/conference/usenixsecurity13/technical-sessions/paper/durumeric. Visitada: 26/03/2024.

[26] ≪National Vulnerability Database - CVE-2013-5211≫. Disponible en: https://nvd.nist.gov/vuln/detail/CVE-2013-5211. Visitada: 26/03/2024.

[27] ≪Radware. 2023 H1 Global Threat Analysis Report≫. Disponible en: https://www.radware.com/getattachment/75cca6a9-7008-49ed-a92b-5701bd033c1e/Radware-2023_H1_ThreatReport-08_2023-RW-396-FIN.pdf. Visitada: 26/03/2024.