

# PROGETTO DATA SCIENCE

## ANALISI DI TRAFFICO DATI

### TRAMITE SPLUNK

Professore:

Domenico Ursino

Dipartimento di Ingegneria dell'Informazione

Corso: Data Science

Università Politecnica delle Marche, Ancona

**splunk >**

Di:

Claudio Menotta

Mattia Ippoliti

Italo Cervigni

# Indice

<b>1</b>	<b>Introduzione</b>	<b>1</b>
<b>2</b>	<b>Dataset</b>	<b>3</b>
<b>3</b>	<b>Analisi sugli indirizzi IP</b>	<b>8</b>
3.1	Indirizzi IP dai quali sono partiti più flussi dati . . . . .	8
3.2	Distribuzione dei flussi generati dai diversi IP . . . . .	9
3.3	Indirizzi IP nel quale sono arrivati più flussi dati . . . . .	9
3.4	Distribuzione dei flussi arrivati ai diversi IP . . . . .	10
<b>4</b>	<b>Caratteristiche generali del traffico dati</b>	<b>11</b>
4.1	Pacchetti inviati . . . . .	11
4.2	Uso dei flags . . . . .	12
4.3	Protocolli utilizzati . . . . .	12
4.4	Numero di flussi dati per classe . . . . .	13
<b>5</b>	<b>Analisi sul numero di Packets inviati per classe</b>	<b>15</b>
5.1	Pacchetti scambiati per classe . . . . .	15
5.2	Packets inviati per classe e indirizzo IP . . . . .	16
<b>6</b>	<b>Analisi sui protocolli</b>	<b>18</b>
6.1	Traffico dati su Source IP Address per protocollo TCP . . . . .	18
6.2	Traffico dati su Destination IP Addres per protocollo TCP . . . . .	19

## INDICE

6.3 Classificazione del traffico dati per protocollo . . . . .	21
<b>7 Analisi sui pacchetti</b>	<b>22</b>
7.1 KPI sul numero di pacchetti trasmessi giorno per giorno . . . . .	22
7.2 Andamenti temporali dei pacchetti trasmessi per source IP address, per source port, per destination IP address e per destination port . . . . .	23
<b>8 Analisi sui bytes</b>	<b>25</b>
8.1 KPI sul numero di bytes trasmessi giorno per giorno . . . . .	25
8.2 Andamenti temporali dei bytes trasmessi per source IP address, per source port, per destination IP address e per destination port . . . . .	26
<b>9 Analisi temporale sull'utilizzo dei protocolli</b>	<b>28</b>
9.1 Analisi sul numero di bytes trasmessi per giorno e ora, per ogni protocollo	28
9.2 Analisi sul numero di pacchetti trasmessi per giorno e ora, per ogni protocollo . . . . .	29
9.3 Analisi sulla durata delle trasmissioni di dati per giorno e ora per ogni protocollo . . . . .	30
<b>10 Analisi su Bytes e Pacchetti inviati e ricevuti</b>	<b>31</b>
10.1 KPI . . . . .	32
10.2 Bubble Chart con i maggiori indirizzi IP di partenza . . . . .	32
10.3 Bubble Chart con i maggiori indirizzi IP di destinazione . . . . .	33
10.4 Maggiori venti indirizzi IP di destinazione per somma di Bytes e Pacchetti	33
<b>11 Analisi sugli attacchi avvenuti nel 08/04/2017</b>	<b>35</b>
11.1 Numero totale di pacchetti inviati durante gli attacchi del 08/04/2017 . . .	36
11.2 Numero totale di pacchetti inviati durante la risposta agli attacchi del 08/04/2017 . . . . .	36
11.3 Numero totale di Bytes inviati durante gli attacchi del 08/04/2017 . . .	37
11.4 Numero totale di bytes inviati durante la risposta agli attacchi del 08/04/2017	37

## INDICE

11.5 Durata totale degli attacchi del 08/04/2017 . . . . .	38
11.6 Durata totale della risposta agli attacchi del 08/04/2017 . . . . .	38
11.7 Numero totale di Bytes usati durante gli attacchi del 08/04/2017 divisi per porte di partenza . . . . .	39
11.8 Numero totale di Bytes usati durante gli attacchi del 08/04/2017 divisi per porte di destinazione . . . . .	39
11.9 Numero totale di Bytes usati durante gli attacchi del 08/04/2017 divisi per protocollo . . . . .	40
11.10 Numero totale di Bytes usati durante la risposta agli attacchi del 08/04/2017 divisi per protocollo . . . . .	41
11.11 Numero totale di Bytes usati durante gli attacchi del 08/04/2017 divisi per "Attack Type" . . . . .	41
11.12 Tecniche di attacco preferite dagli attaccanti . . . . .	42
<b>12 Analisi sugli attacchi avvenuti il 10/04/2017</b>	<b>43</b>
12.1 Numero totale di bytes trasmessi per "Src Pt" e "Dst Pt" . . . . .	44
12.2 Numero totale di bytes trasmessi per Protocollo e "Attack Type" . . . . .	45
<b>13 Analisi Port Scan</b>	<b>46</b>
13.1 KPI . . . . .	47
13.2 Bytes inviati nei due tipi di attacco . . . . .	47
13.3 Analisi in base al numero totale di bytes, pacchetti e durata per "Attack Type" . . . . .	47
13.4 Numero di attacchi di tipo Port Scan avvenuti . . . . .	49
13.5 Flag utilizzati durante gli attacchi . . . . .	49
13.6 Connessioni tentate in ciascun attacco . . . . .	50
<b>14 Conclusioni</b>	<b>52</b>
<b>Bibliografia</b>	<b>53</b>

# 1. Introduzione

Splunk è stato nominato nel Magic Quadrant 2018 di Gartner come uno dei principali leader del mercato per quanto riguarda i sistemi SIEM (Security Information and Event Management). Migliaia di organizzazioni in tutto il mondo utilizzano Splunk come SIEM per il monitoraggio della sicurezza, la difesa dalle minacce, le indagini sugli incidenti, la risposta agli incidenti e molto altro. Uno dei prodotti principali di Splunk è Splunk Enterprise, un software che permette di raccogliere una grande quantità di dati generati dai vari componenti dell'infrastruttura IT, da Web Server, database, sensori e altro. I dati generati dalle macchine sono una delle aree dei big data con la crescita più rapida e, allo stesso tempo, con maggiore complessità. È anche una delle aree più preziose, che contiene una registrazione definitiva di tutte le transazioni degli utenti, del comportamento dei clienti, del comportamento delle macchine, delle minacce alla sicurezza, delle attività fraudolente e altro ancora. Splunk trasforma i dati delle macchine in informazioni preziose, qualsiasi sia il settore industriale in cui ci troviamo, e ciò viene definito come Operational Intelligence. Con funzionalità di analisi intuitive, machine learning, applicazioni predefinite e API aperte, Splunk Enterprise è una piattaforma flessibile che può operare da casi d'uso molto specifici fino alla gestione della dorsale di analisi di tutta l'azienda. Splunk Enterprise:

- Raccoglie e indicizza i dati di log e di macchina proveniente da qualsiasi sorgente.
- Fornisce potenti funzionalità di ricerca, analisi e visualizzazione che consentono di potenziare le capacità di ricerca, analisi e visualizzazione da tutta l'organizzazione.
- Possiede un esteso ecosistema di applicazioni basate su di esso che fornisce soluzioni per la sicurezza, per valutare e incrementare l'operatività dell'azienda, per effettuare analisi di business e molto altro ancora.

## CAPITOLO 1: INTRODUZIONE

- È disponibile sia come software on-premise che come servizio cloud.

Nelle sezioni successive verrà per prima cosa descritto il dataset scelto per poi presentare le analisi effettuate con Splunk Enterprise.

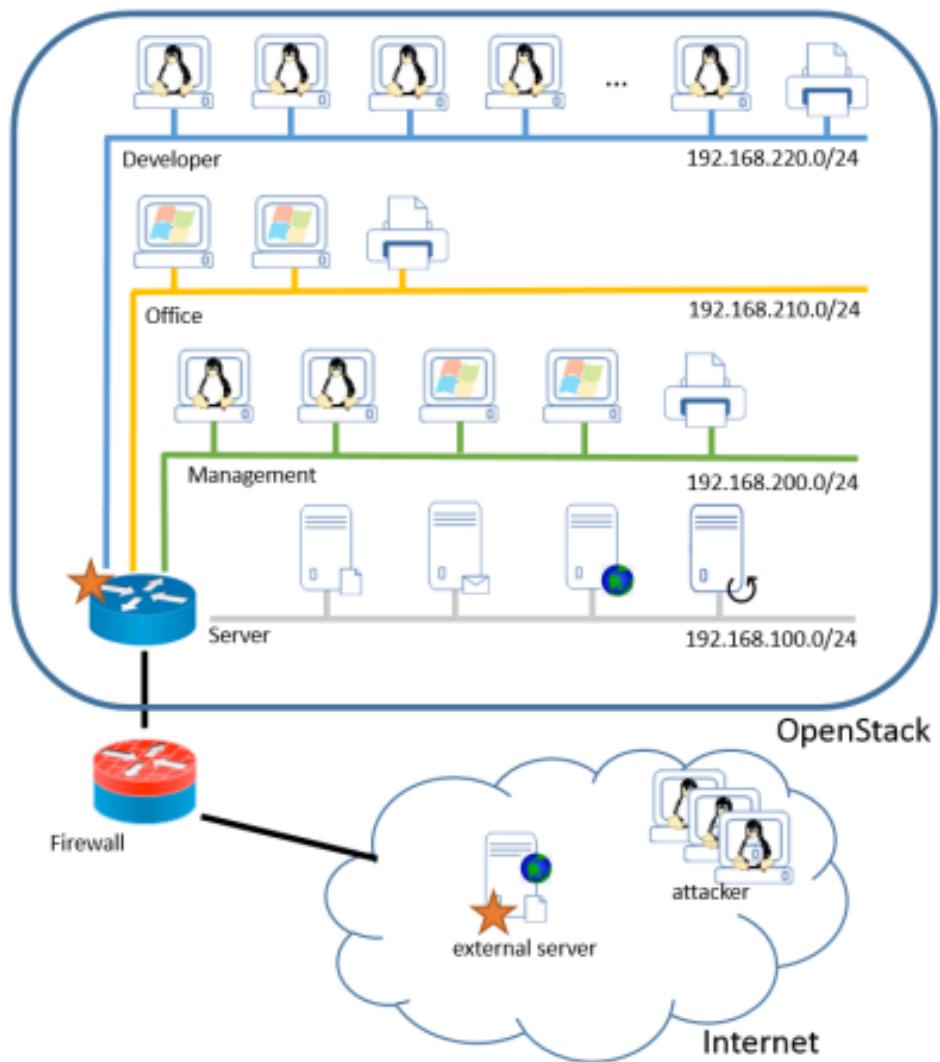
## 2. Dataset

CIDDS (Coburg Intrusion Detection Data Sets) è un dataset per la valutazione per sistemi di rilevamento delle intrusioni basati sulla rete e anomalie, sviluppato da un gruppo di ricerca dell'università di Coburg in Germania. Poiché il settore IT è in continua evoluzione, gli aggressori sono costretti ad adattarsi e trovare nuovi modi per penetrare nel loro obiettivo di interesse. Quindi, lo sviluppo di sistemi di rilevamento delle intrusioni può essere visto come un processo in continua evoluzione tra i tentativi degli attaccanti e gli adattamenti innescati dei difensori. Da questo punto di vista, non è opportuno testare gli attuali sistemi di rilevamento delle intrusioni con vecchi set di dati. Pertanto, l'obiettivo principale di CIDDS è la generazione di set di dati personalizzabili e aggiornati. Per affrontare questo obiettivo, l'idea di base alla base di CIDDS è creare set di dati basati sul flusso etichettati in un ambiente virtuale utilizzando OpenStack (la più famosa struttura software open source legata al cloud computing).

Di seguito, si vuole spiegare l'idea generale di CIDDS con il set di dati CIDDS-001 generato in modo esemplare nell'articolo [1]. Per creare il set di dati CIDDS-001, è stato emulato un ambiente di piccole imprese. Questo ambiente include diversi client e server tipici come un server di posta elettronica o server Web. Per emulare il comportamento benigno degli utenti come navigare sul Web, inviare e ricevere e-mail e scambiare file, vengono utilizzati degli script Python. Per garantire un comportamento dell'utente il più realistico possibile, i client svolgono le loro attività rispettando un programma di lavoro individuale che considera l'orario di lavoro e le pause pranzo. E' stato anche tenuto conto del fatto che dipendenti diversi tendono ad avere compiti lavorativi diversi. Un manager, ad esempio, parteciperebbe a più riunioni, generando così meno traffico di rete rispetto a un ricercatore che naviga costantemente. Quindi, le caratteristiche di ogni utente vengono impostate utilizzando un file di configurazione.

## CAPITOLO 2: DATASET

Per l'inclusione del traffico di rete effettivo, che ha origine al di fuori dell'ambiente OpenStack, è stato distribuito un server esterno. Il server esterno viene distribuito direttamente su Internet. Offre due servizi: una homepage pubblica raggiungibile sia per i clienti interni che per gli interessati esterni e fornisce un servizio di sincronizzazione dei file (Seafire) per i client interni. Questo servizio viene utilizzato da tutti i client della sottorete di Management e Developer. Da sottolineare il fatto che i client interni dell'ambiente OpenStack comunicano con lo stesso indirizzo IP pubblico al server esterno. Oltre al server esterno, si ha il controllo su altri tre server su Internet. Da questi server, sono stati eseguiti diversi attacchi al server esterno. L'intera struttura della rete è visibile nella seguente figura.



Il traffico registrato sul server esterno viene salvato su 4 dataset, creati monitorando il traffico per 4 settimane consecutive. Per l'analisi in Splunk viene scelto il traffico

## CAPITOLO 2: DATASET

riguardante la terza settimana, che è stata quella nella quale il server esterno ha subito più attacchi. Il dataset usato è quindi *CDDIS-external-week3*.

Gli attributi all'interno del dataset sono riportati all'interno di questa tabella.

External Server	
Caratteristiche	Descrizione
Data first seen	Ora di inizio flusso dei dati
Duration	Durata del flusso
Proto	Transport Protocol (esempio: ICMP, TCP, or UDP)
Src IP Addr	Indirizzo IP della sorgente del flusso
Src Pt	Source Port
Dst IP Addr	Indirizzo IP della destinazione del flusso
Dst Pt	Destination Port
Packets	Numero di pacchetti trasmessi
Bytes	Numero di Bytes trasmessi
Flags	Concatenazione di tutti i TCP flags
Class	Nome della classe attribuita al flusso (normal, attacker, victim, suspicious or unknown)
AttackType	Tipo di attacco (portScan, dos, bruteForce)
AttackID	ID dell'attacco univoco, ovvero tutti i flussi che appartengono allo stesso attacco hanno lo stesso ID
AttackDescription	Fornisce informazioni aggiuntive sui parametri di attacco impostati (ad esempio il numero di tentativi di password indovinate per attacchi SSH-Brute-Force o le impostazioni dei parametri per portScans )

## CAPITOLO 2: DATASET

Ci sono particolari considerazioni da fare negli attributi. La prima è che gli ultimi tre attributi forniscono informazioni aggiuntive sugli attacchi eseguiti. Questi attributi vengono utilizzati solo se il flusso appartiene alla classe *attacker* o *victim*. Se il flusso appartiene alla classe *normal*, *suspicious* o *unknown*, il valore di questi tre attributi viene impostato su un valore predefinito "- - -". Inoltre il secondo attributo è chiamato *attackID*. Un ID univoco viene assegnato a ciascun attacco eseguito, di conseguenza, tutti i flussi che appartengono allo stesso attacco condividono lo stesso valore in questo attributo.

La seconda considerazione da fare è che non è stato attaccato il server esterno dall'ambiente OpenStack. Pertanto, si etichettano tutti i flussi del server esterno che hanno la loro origine o destinazione nell'ambiente OpenStack come *normal*. Inoltre, si ha il controllo su tre server che vengono distribuiti direttamente su Internet (*attacker1*, *attacker2*, *attacker3*), e questi server effettuano solo gli attacchi al server esterno. Poiché si conoscono le origini, l'obiettivo e i timestamp degli attacchi eseguiti da questi server, si è in grado di etichettare tutti i flussi corrispondenti con le etichette di classe *attacker* o *victim*. In particolare il flusso sarà *attacker* se va da uno dei degli attaccanti al server esterno, mentre sarà *victim* se sarà la risposta del server esterno ad uno degli attaccanti. Il server esterno fornisce una homepage per le persone interessate. Pertanto, tutto il traffico verso le porte 80 e 443 potrebbe essere traffico normale o tentativi di intrusione. Di conseguenza, il traffico verso le porte 80 e 443 sul server esterno viene etichettato come *unknown*. Il traffico di rete rimanente viene etichettato come *suspicious*, poiché non vengono offerte ulteriori informazioni per gli utenti pubblici.

Infine, l'ultima considerazione da fare riguarda gli indirizzi IP, al quale si applica un processo di anonimizzazione per questioni di privacy. I seguenti indirizzi IP vengono gestiti specificatamente durante il processo di anonimizzazione:

- Tutti i server e i client dell'ambiente OpenStack comunicano con stesso indirizzo IP pubblico al server esterno. E' stato sostituito questo indirizzo IP pubblico con *OPENSTACK\_NET*.
- Tutte le macchine virtuali all'interno dell'ambiente OpenStack utilizzano lo stesso server DNS. Si rinomina l'indirizzo IP del server DNS come *DNS*.
- L'indirizzo IP del server esterno viene sostituito con *EXT\_SERVER*.

## CAPITOLO 2: DATASET

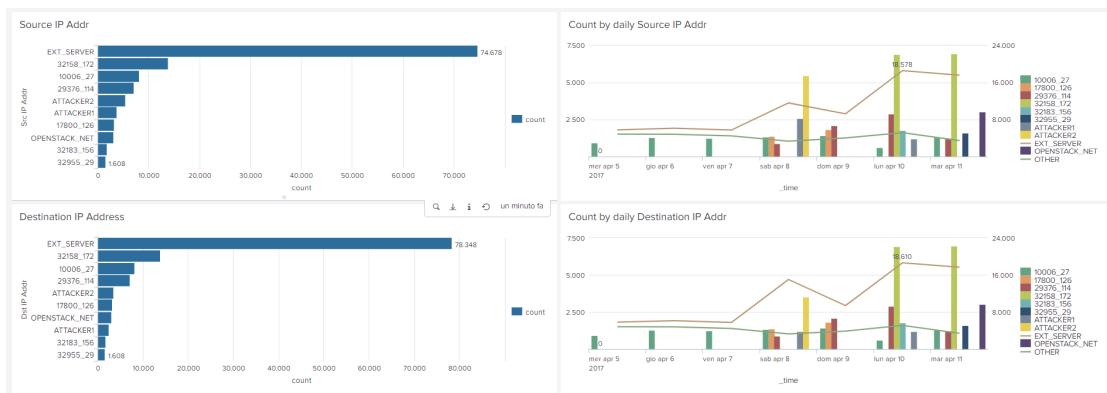
- Gli indirizzi IP degli aggressori esterni vengono sostituiti con *ATTACKER1*, *ATTACKER2* e *ATTACKER3*.

Per i restanti indirizzi IP pubblici è stato utilizzato il seguente processo di anonimizzazione: i primi tre byte di ogni indirizzo IP vengono sostituiti con un numero generato in modo casuale, mentre il quarto byte dell'indirizzo IP viene mantenuto. Ciò consente di conservare le informazioni sulle strutture di rete, poiché tutti gli indirizzi IP della stessa sottorete vengono sostituiti con lo stesso numero generato in modo casuale. La tabella mostra alcuni esempi di questo processo di anonimizzazione.

#	IP Address	Anonymized IP Address
1	8.8.8	4711_8
2	8.8.8.9	4711_9
3	8.8.8.18	4711_18
4	9.9.9.9	13107_9
5	9.9.9.173	13107_173
6	8.8.8.9	4711_9
7	7.7.7.7	2311_7

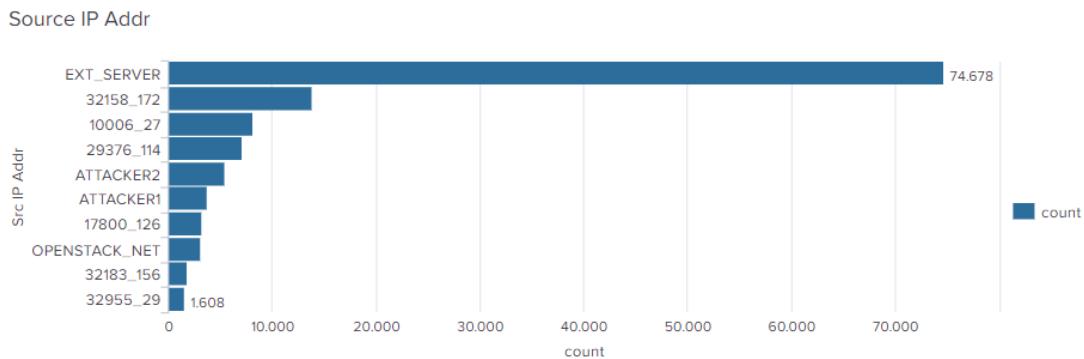
### 3. Analisi sugli indirizzi IP

Si inizia l'analisi del traffico dati sul server esterno andando a studiare quali sono stati gli indirizzi IP che sono stati più utilizzati.



#### 3.1 Indirizzi IP dai quali sono partiti più flussi dati

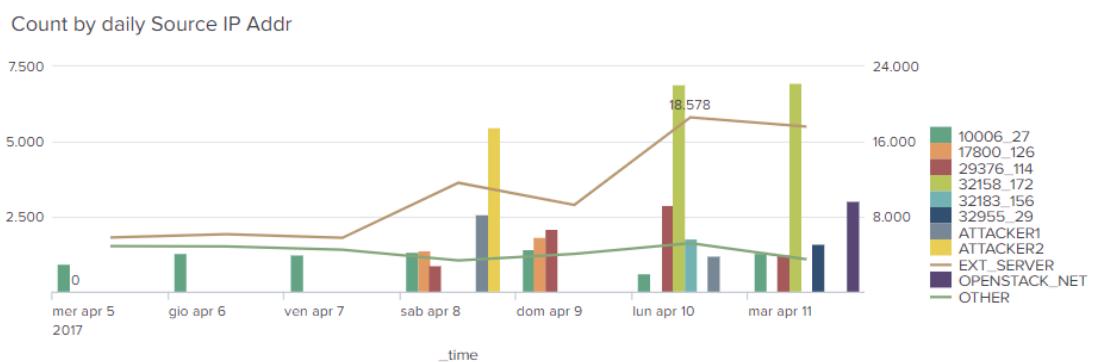
Si parte dal primo istogramma in alto a sinistra, nel quale viene rappresentato il conteggio di ciascun IP per l'attributo *Src IP Addr*, visualizzando quindi il numero di flussi dati partiti da ciascun indirizzo IP.



Si nota immediatamente come l'IP che identifica il server esterno sia quello dalla quale partono più flussi di dati, proprio perché tutti gli IP, sia interni all'OpenStack che quelli pubblici di utenti nel Web, comunicano con il server esterno, che quindi dovrà rispondere alle loro richieste creando un flusso dati che parte proprio da qui.

### 3.2 Distribuzione dei flussi generati dai diversi IP

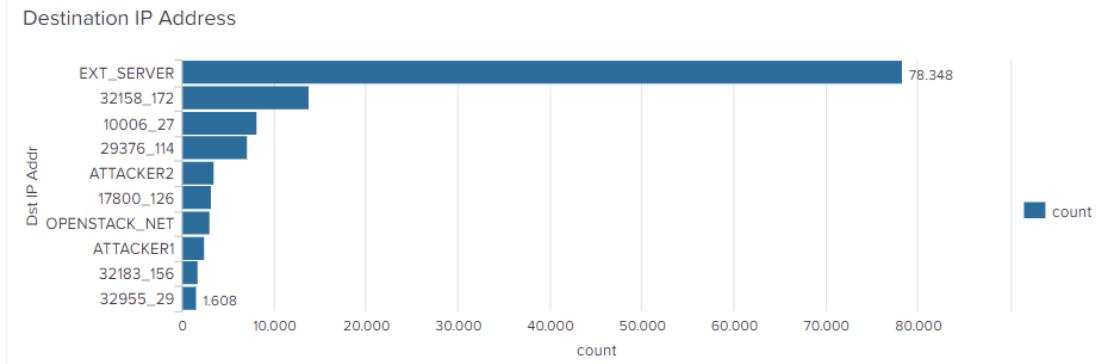
Interessante è anche vedere quanto questi IP siano stati utilizzati, e in particolare quante volte hanno iniziato un traffico di dati, durante la settimana presa in esame. La settimana che si considera va da mercoledì 5 Aprile 2017 a martedì 11 Aprile 2017.



Si nota come nei vari giorni ci siano stati picchi di accessi al server da vari indirizzi IP, che sono rappresentati nelle barre verticali, mentre la somma di tutti gli accessi fatti dagli altri IP con valori trascurabili è riportata nella linea OTHER. Tra gli IP più interessanti ci sono *ATTACKER1* e *ATTACKER2*, indirizzo IP di chi ha eseguito l'attacco al server, che presentano picchi di traffico dati nei giorni 8 e 10 Aprile. Infine si vede come per tutti i giorni il numero di flussi di dati sia stato sempre elevato nel server esterno.

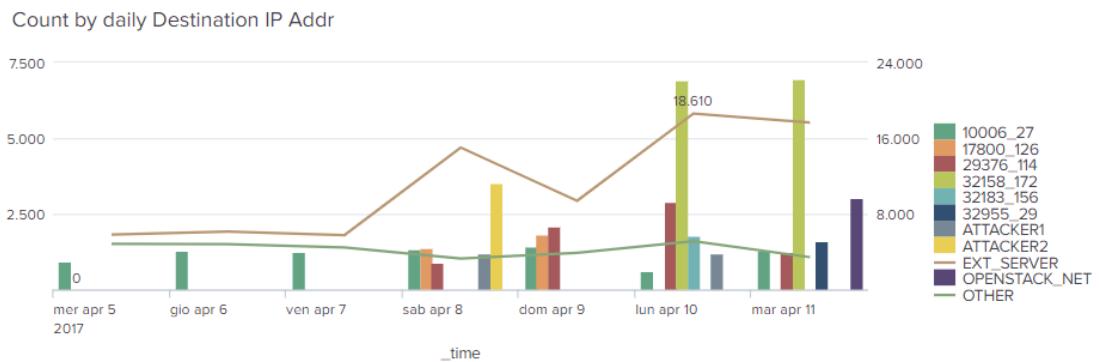
### 3.3 Indirizzi IP nel quale sono arrivati più flussi dati

In maniera analoga è stata fatta un analisi sugli IP di destinazione del flusso dati. Come ci si poteva aspettare anche in questo caso l'indirizzo IP di destinazione più usato è quello del server esterno, dato che tutti gli indirizzi IP del Web e all'interno dell' OpenStack comunicano con esso.



### 3.4 Distribuzione dei flussi arrivati ai diversi IP

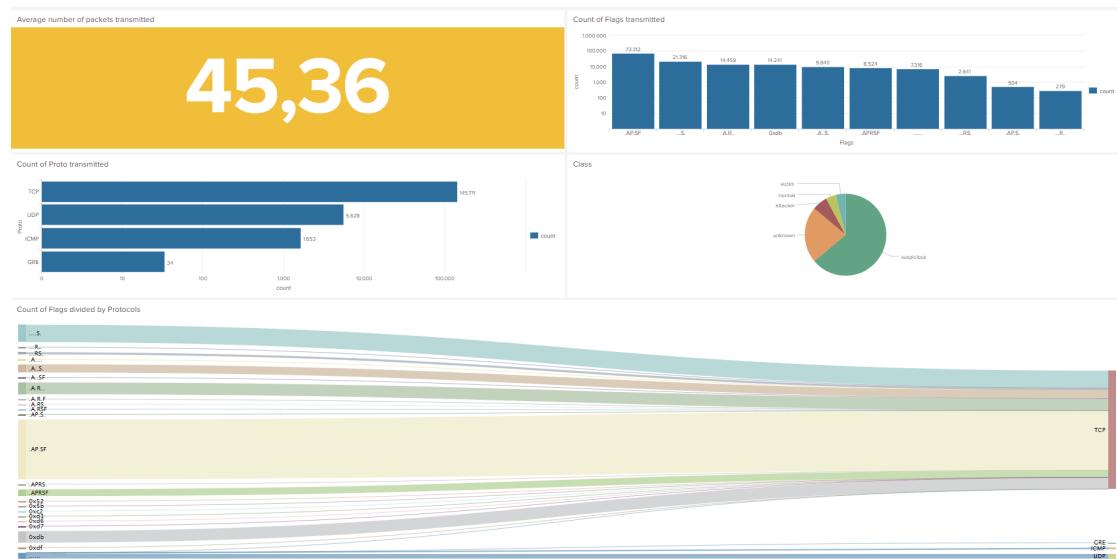
Infine anche per i *Dst IP Addr* viene graficata la distribuzione temporale all'interno della settimana dell'uso degli IP.



Anche in questo caso l'andamento è molto simile a quello della *Src IP Addr*, dato che per quasi tutti i flussi che arrivano al server esterno, c'è una sua risposta che torna indietro all'IP di partenza tramite un nuovo flusso di dati.

## 4. Caratteristiche generali del traffico dati

Nella seconda dashboard si vanno a studiare le caratteristiche generali dei flussi di dati trasmessi durante la settimana.



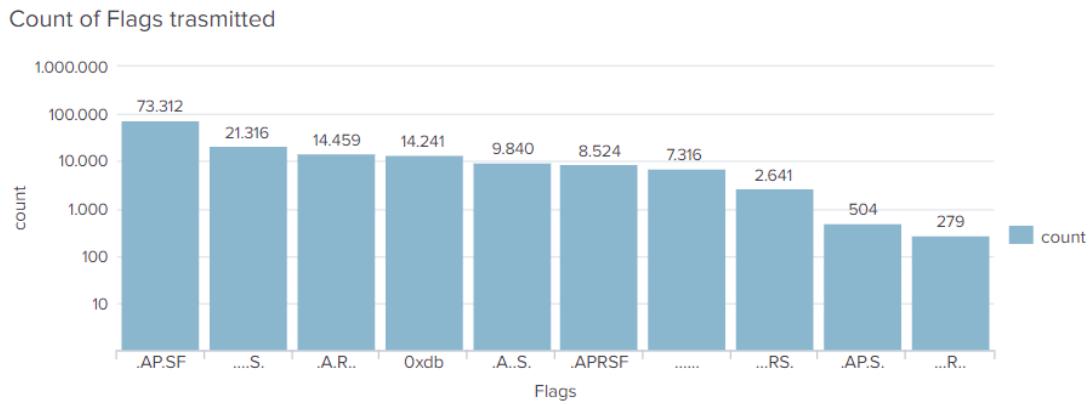
## 4.1 Pacchetti inviati

In alto a destra si può immediatamente vedere il numero medio di pacchetti inviati per flusso dati.



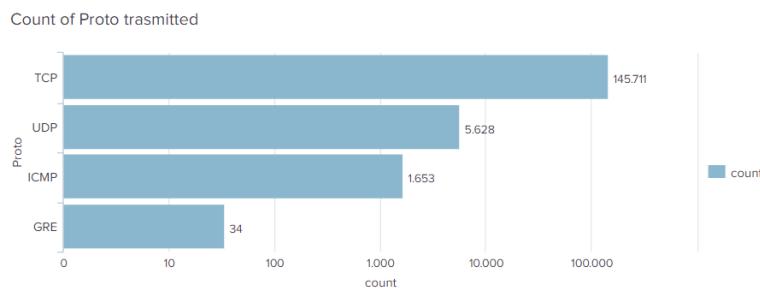
## 4.2 Uso dei flags

Un'altra cosa molto interessante da osservare sono i flags trasmessi durante il traffico dei dati. Poiché la maggior parte del traffico è avvenuto attraverso il protocollo TCP, si ha che la maggior parte dei flag utilizzati sono proprio di questo protocollo.



## 4.3 Protocolli utilizzati

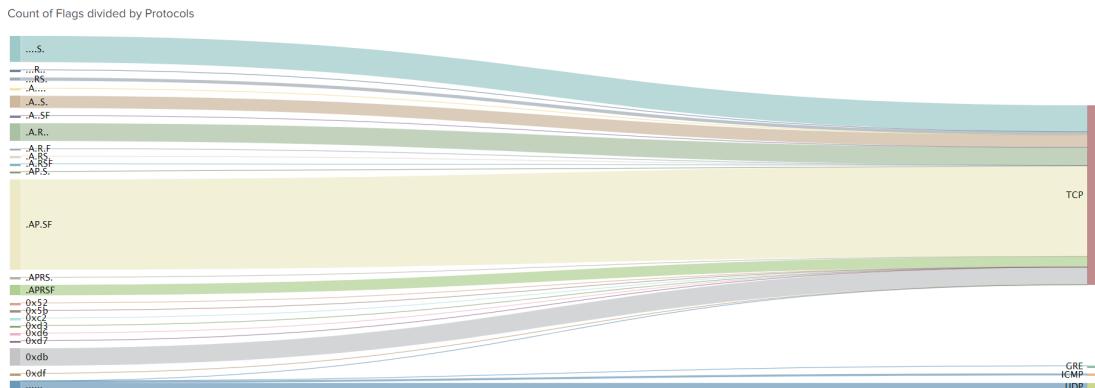
Si vanno poi a studiare i protocolli più usati per la trasmissione dei dati.



Dalla figura si osserva che il protocollo più largamente utilizzato è il TCP, un protocollo di rete a pacchetto di livello di trasporto, appartenente alla suite di protocolli Internet, che si occupa di controllo della trasmissione ovvero rendere affidabile la comunicazione

dati in rete tra mittente e destinatario. Da notare infatti che l'asse del conteggio è di tipo logaritmico, per permettere una buona visualizzazione del numero di utilizzi di ciascun protocollo. In seconda posizione troviamo il protocollo UDP, con un utilizzo molto inferiore. Infatti l'utilizzo del protocollo TCP rispetto a UDP è, in generale, preferito quando è necessario avere garanzie sulla consegna dei dati o sull'ordine di arrivo dei vari segmenti (come per esempio nel caso di trasferimenti di file). Al contrario UDP viene principalmente usato quando l'interazione tra i due host è idempotente (in informatica, in matematica, e in particolare in algebra, l'idempotenza è una proprietà delle funzioni per la quale applicando molteplici volte una funzione data, il risultato ottenuto è uguale a quello derivante dall'applicazione della funzione un'unica volta) o nel caso si abbiano forti vincoli sulla velocità e l'economia di risorse della rete (es. streaming in tempo reale, videogiochi multiplayer).

Altra cosa da osservare nei protocolli, già parzialmente detta, è il Sankey Diagram relativo ai flags e protocolli utilizzati. Dalla figura si vede che tutti i flags inviati sono



caratteristici del protocollo TCP, mentre per gli altri tre protocolli per la quasi totalità dei casi non vi sono stati flags inviati, infatti il valore tipico è "....." che sta ad indicare appunto l'assenza di flag .

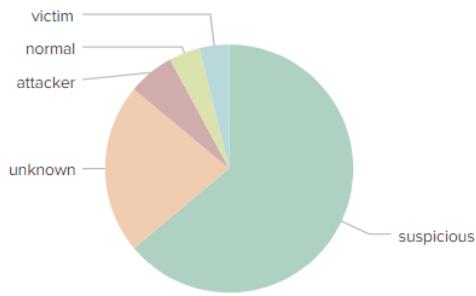
#### 4.4 Numero di flussi dati per classe

L'ultimo grafico analizzato in questa dashboard riguarda l'entità attribuita a ciascun flusso di dati avvenuto durante la settimana, attraverso l'attributo *Class*.

Si può subito notare come la maggior parte del traffico avvenuto sia caratterizzato come *suspicious*, ovvero traffico di utenti del web con indirizzo IP pubblico di cui non si hanno

## CAPITOLO 4: CARATTERISTICHE GENERALI DEL TRAFFICO DATI

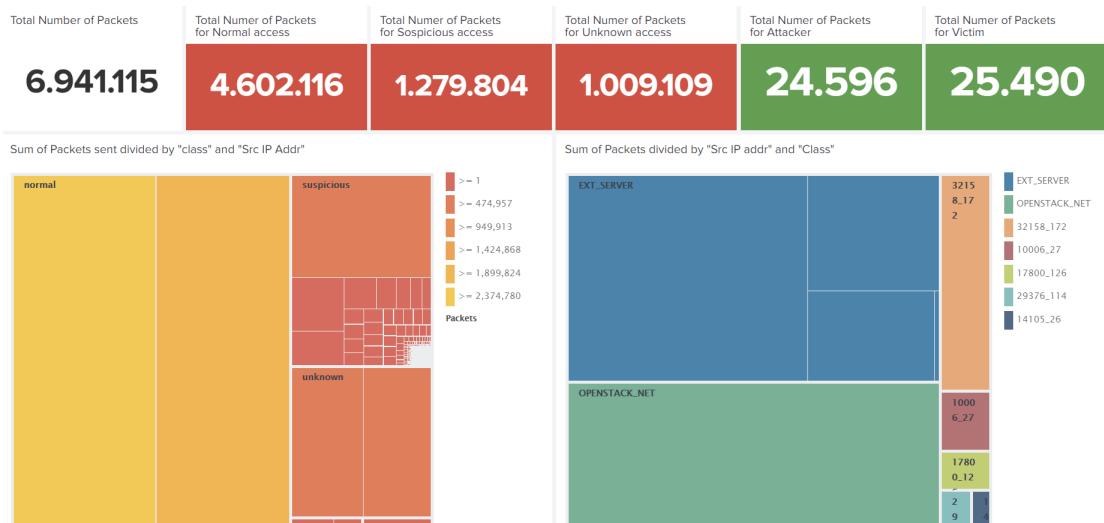
Class



informazioni. Invece i flussi di dati verso l'interno dell'azienda, quindi verso l'ambiente OpenStack, classificati con *normal*, sono relativamente bassi. Si osserva inoltre che il traffico nelle porte 80 e 443 è significativo (denominato con *unknown*). Infine si può notare il fatto che durante questa settimana una buona percentuale del traffico dati è dovuta ai diversi attacchi eseguiti dai tre attaccanti presenti in internet.

# 5. Analisi sul numero di Packets inviati per classe

In questa terza dashboard si valuta il numero totale di pacchetti scambiati per le varie classi e per indirizzo IP.



## 5.1 Pacchetti scambiati per classe

Si inizia l'analisi partendo dalle 6 KPI poste nella parte superiore della dashboard.

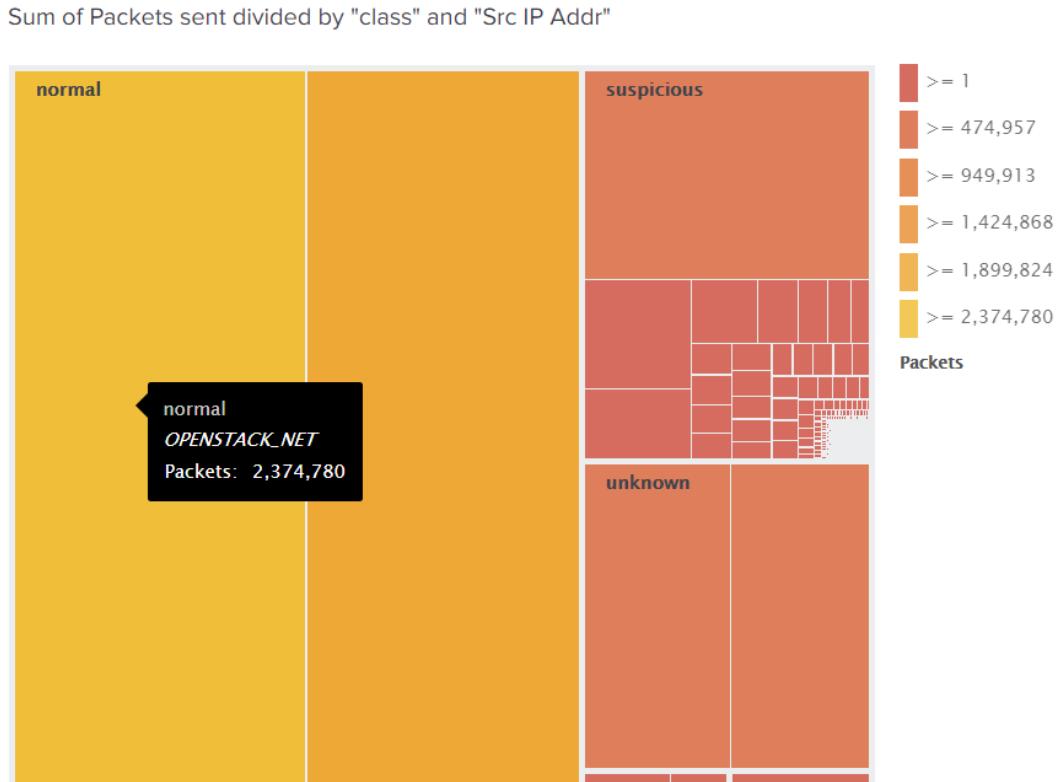


Nella prima KPI si legge il numero di pacchetti dati inviati nel corso della settimana. Nelle altre KPI si può vedere come la somma dei pacchetti inviati sia distribuita nelle

cinque classi etichettate. Si nota immediatamente che circa l' 80% dei pacchetti sia stato inviato nel flusso di dati etichettato come normale, nonostante il fatto che il numero di flussi *normal* sia molto minore di quello *suspicious* e *unknown* come visto sulla dashboard precedente. Questo significa che per ciascun flusso di dati di tipo *normal*, ovvero quelli che partono o finiscono all'interno dell'azienda, si ha un gran numero di pacchetti scambiati rispetto alle altre classi, e ciò probabilmente perché si ha con il server esterno un servizio di sincronizzazione dei file. Si vede infine come il numero di pacchetti scambiati negli attacchi sia invece molto minore rispetto alle altre classi.

## 5.2 Packets inviati per classe e indirizzo IP

Il secondo grafico da prendere in considerazione è la heatmap in basso a sinistra, nella quale viene fatta una distribuzione dei pacchetti inviati per classe nei vari indirizzi IP.

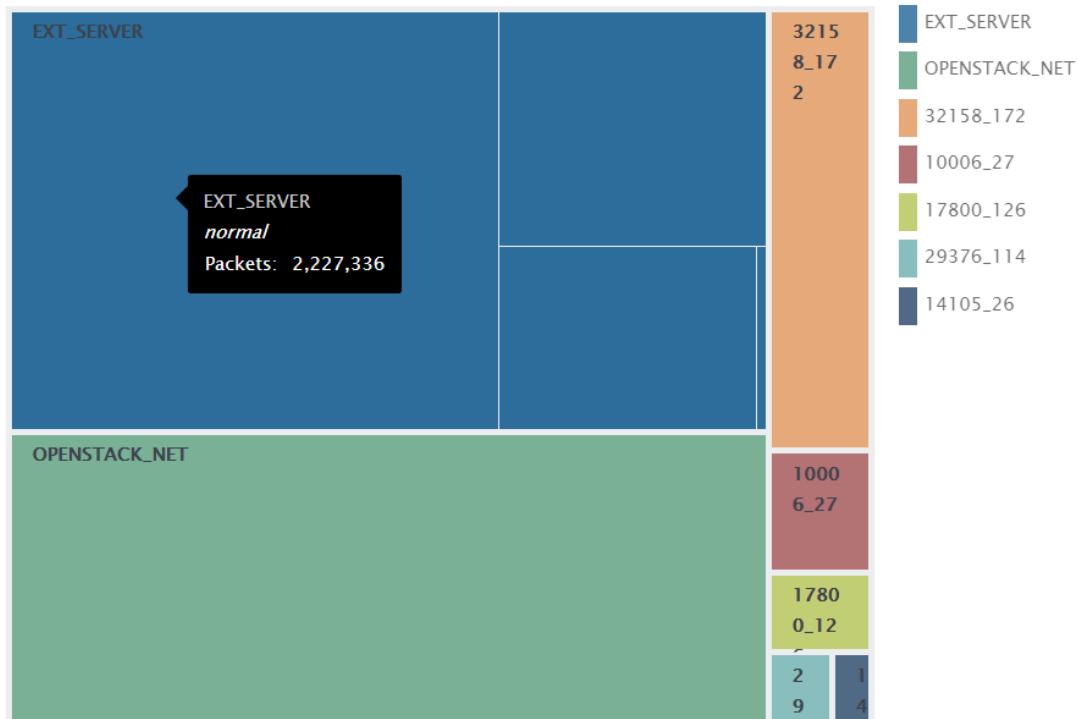


Come già detto la classe *normal* può avere solo due tipi di indirizzi IP dal quale parte il flusso di dati, che sono *OPENSTACK\_NET* e *EXTERNAL\_SERVER*. Ovviamente per tutte le classi circa la metà dei pacchetti sono inviati dal server esterno poiché ad ogni domanda di un client vi è la risposta del server. Per la classe *suspicious* avremo

la distribuzione dei pacchetti inviati sui vari IP pubblici della rete, e si può osservare che quello che ha inviato più dati è 1006\_7, seguito poi da altri IP. Mentre per la classe *unknown* i pacchetti vengono inviati da un solo IP, ovvero 32158\_172, che sarà l'unico ad utilizzare quindi o la porta 80 o 443. Infine per le classi *attacker* si ha che che i pacchetti sono inviati solo dagli IP *ATTACKER1* e *ATTACKER2*, mentre nella classe *victim* solo dal server esterno, essendo classificati in questa classe i flussi dati che partono dal server in risposta ad uno degli attaccanti.

Infine nell'ultima heatmap in basso a destra vi è una analisi simile a quella fatta precedentemente, ma questa volta si valutano per gli IP che hanno avuto un maggior numero di invio di pacchetti, come essi siano distribuiti nelle varie classi.

Sum of Packets divided by "Src IP addr" and "Class"



Si può vedere facilmente che gli indirizzi IP da cui sono partiti più pacchetti di dati sono il server esterno e dalla rete interna aziendale. Il primo, come già detto, contiene flussi di dati che sono *normal*, *suspicious* e *unknown*, mentre per il secondo sono solo di tipo *normal*. Infine tra i vari IP che hanno inviato più pacchetti dati vi sono diversi IP pubblici di client esterni all'azienda, che ovviamente sono classificati con *suspicious*.

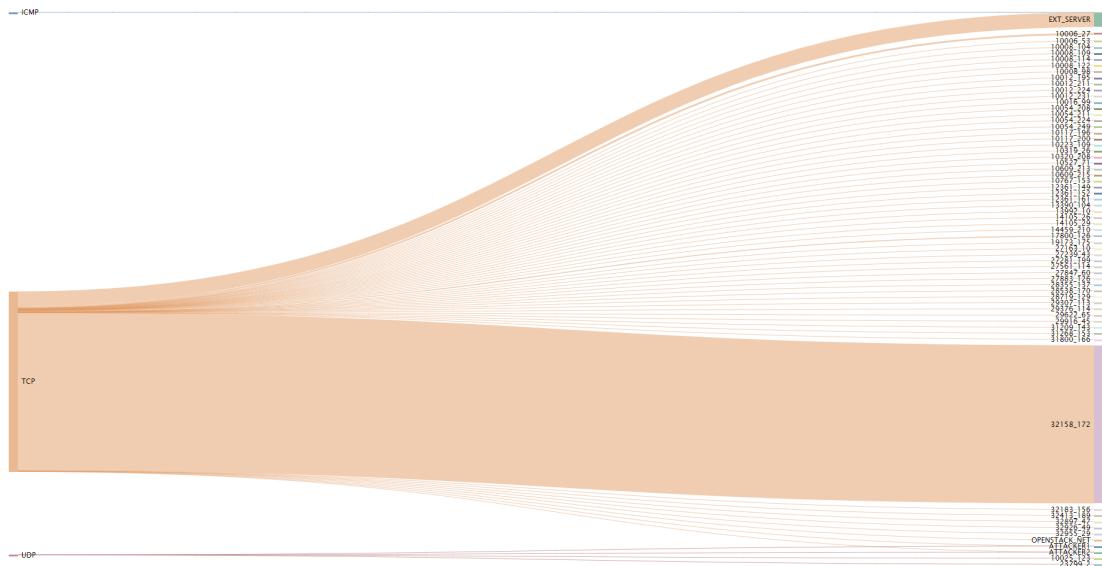
# 6. Analisi sui protocoli



In questa dashboard sono stati analizzati i traffici di dati concentrandosi sui protocolli utilizzati. In particolare, come sottolineato nelle dasboard precedenti il TCP costituisce il protocollo più utilizzato e rappresenta la base per la maggior parte delle reti pubbliche e locali e per i servizi di rete. I tre grafici presenti in questa dashboard sono delle Sankey chart ognuna con uno scopo differente.

## 6.1 Traffico dati su Source IP Address per protocollo TCP

Nella prima ci si concentra su un protocollo, cioè quello prevalente nel nostro dataset, il TCP, in quanto si è deciso di non visualizzare la totalità dei dati a causa della caoticità del grafico che ne risulterebbe (come si può vedere nella Sankey chart [6.1] che corrisponde a questo grafico considerando però la totalità dei dati senza restrizioni).

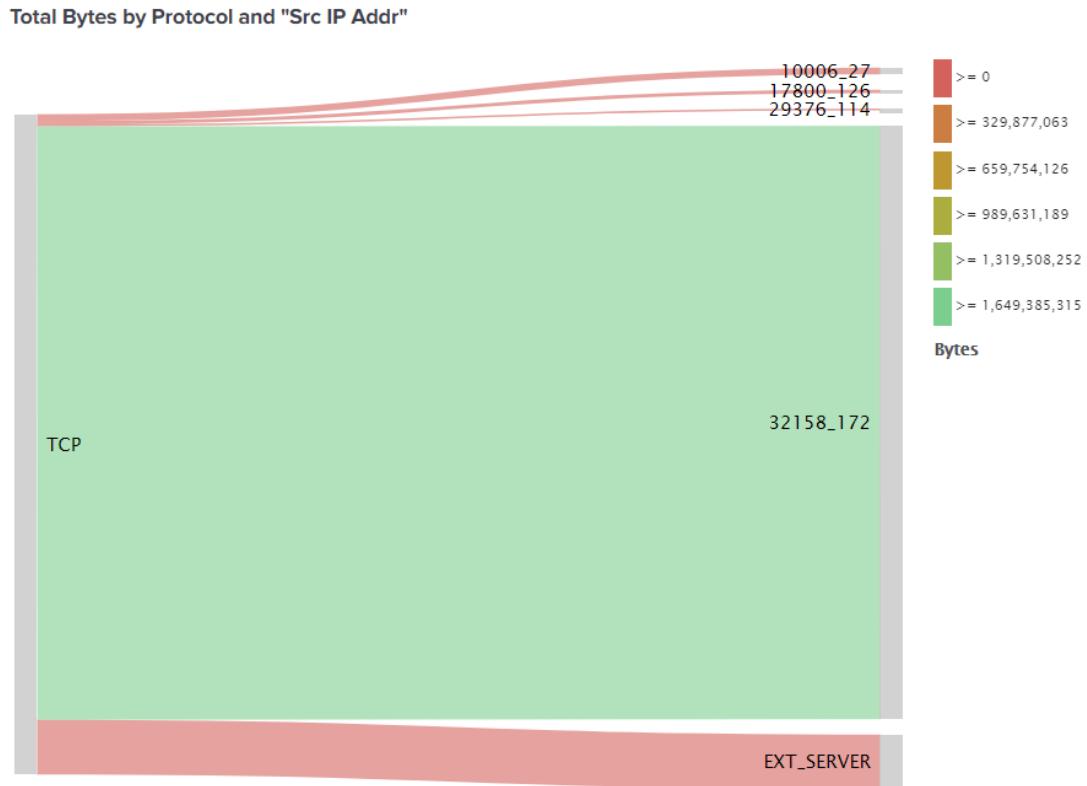


**Figura 6.1:** Grafico Completo del Traffico dati su Source IP Address per protocollo TCP

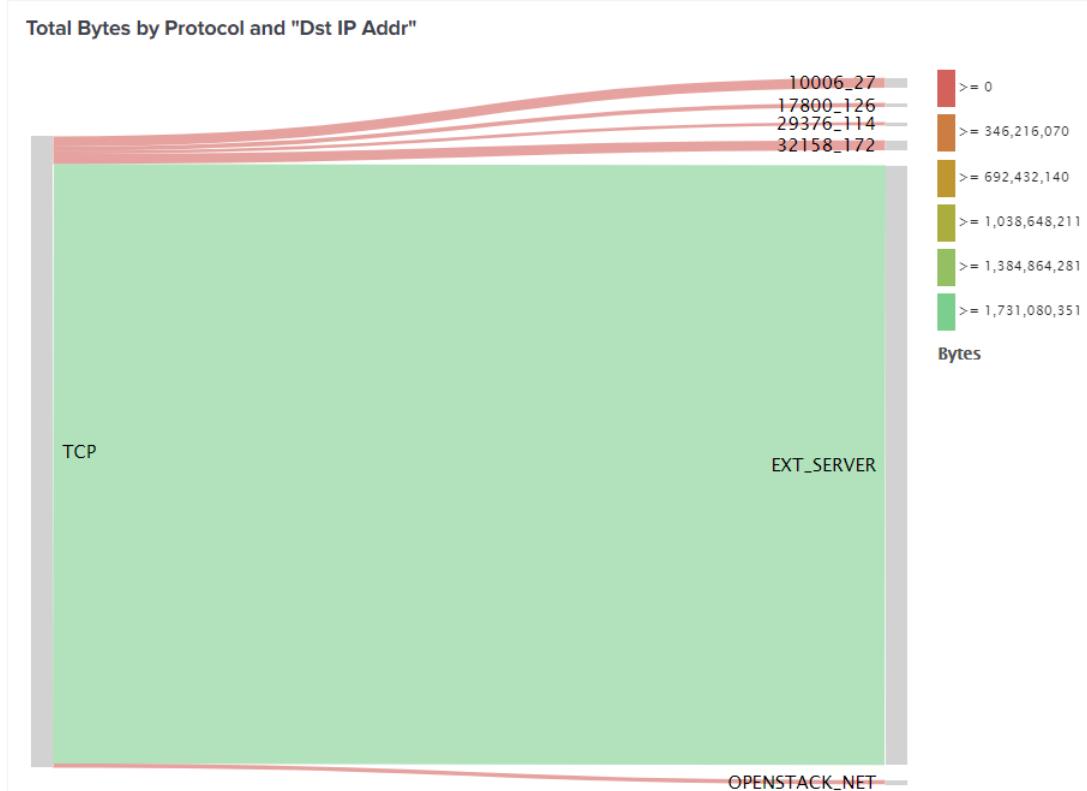
Si vuole mostrare tra i traffici effettuati tramite TCP, quali sono stati i Source IP address dai quali sono stati trasmessi un numero di bytes maggiori nella figura [6.2]. È associato un colore secondo una scala in base al numero medio di bytes scambiati durante la trasmissione dati a quell'IP address. Si può notare che la maggior parte degli accessi con protocollo TCP sono stati fatti dall'indirizzo 32158 – 172 con la trasmissione di un gran numero di bytes. Molti bytes sono stati trasferiti anche dal server esterno, ma le risposte che esso ha fornito hanno una dimensione molto più piccola dei dati che gli sono stati inviati.

## 6.2 Traffico dati su Destination IP Address per protocollo TCP

Nel secondo grafico [6.3] invece viene fatta un'analisi simile, sempre concentrandosi sul protocollo TCP, ma in tal caso vedendo quali sono stati i Destination IP address nei quali c'è stata una trasmissione di bytes maggiori. In questo caso gran parte dei dati trasmessi sono arrivati al server esterno, che come osservato precedentemente ha fornito delle risposte di dimensione notevolmente inferiore.



**Figura 6.2:** Traffico dati su Source IP Address per protocollo TCP



**Figura 6.3:** Traffico dati su Destination IP Address per protocollo TCP

### 6.3 Classificazione del traffico dati per protocollo

Nel terzo ed ultimo grafico della dashboard corrente si mostra tra i traffici di dati effettuati tramite tutti i vari protocolli, come sono stati classificati (sospetti, normali o sconosciuti). Inoltre, è associato un colore secondo una scala in base al numero di bytes scambiati durante quel tipo di trasmissione. Tra gli accessi TCP si osserva che la maggior parte delle trasmissioni di dati è stata classificata come normale, anche se una parte consistente è stata classificata come sospetta, e un'altra come sconosciuta. Si deduce anche che tra le trasmissioni con protocollo TCP quelle sconosciute sono state caratterizzate in media da un elevato numero di bytes trasmessi. Per quanto riguarda gli altri protocolli si ha che sono stati utilizzati molto meno e tra essi troviamo il GRE con cui sono stati effettuati soprattutto traffici di dati sospetti, UDP E ICMP con cui sono stati effettuati sia dei traffici sospetti che normali, ma questi due protocolli sono stati utilizzati anche per la simulazione di attacchi al sistema, denominati “attacker” e per la risposta a tali attacchi, denominate come “victim”.



**Figura 6.4:** Classificazione del traffico dati per protocollo

# 7. Analisi sui pacchetti



In questa dashboard ci si concentra principalmente sul numero di pacchetti che sono stati trasmessi durante gli accessi nei giorni del periodo considerato dal dataset.

## 7.1 KPI sul numero di pacchetti trasmessi giorno per giorno

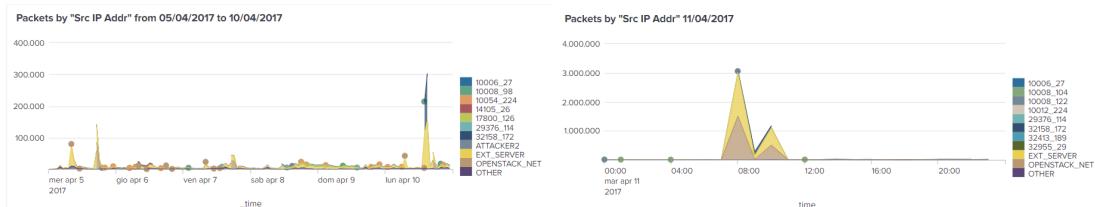
Nella parte alta della dashboard si hanno sette KPI che mostrano giorno per giorno il numero totale dei pacchetti trasmessi. Il giorno in cui si è avuto il maggior numero di pacchetti trasmessi è stato l'11 aprile (con 4.695.742 pacchetti). Invece il giorno in cui si è avuto il minor numero di pacchetti trasmessi è stato il 6 aprile (con 213.593 pacchetti). Accanto ai KPI è mostrato il trend che si è avuto rispetto al giorno precedente in percentuale e si nota che si è avuto un calo nei pacchetti trasmessi solo il 6 e l'8 aprile, mentre nei restanti giorni si è avuto un trend positivo. In particolare la maggiore crescita

percentuale si è avuta l’11 aprile con una crescita del 461% rispetto al giorno precedente, ciò indica che quel giorno c’è stato un grande traffico di dati nel sistema.

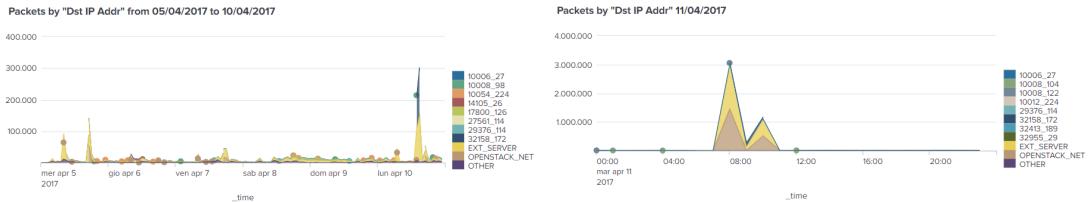


## 7.2 Andamenti temporali dei pacchetti trasmessi per source IP address, per source port, per destination IP address e per destination port

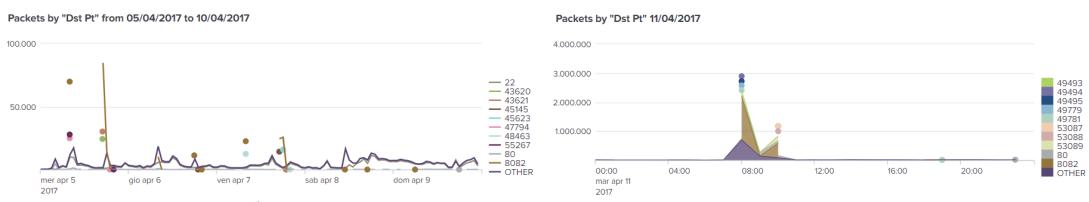
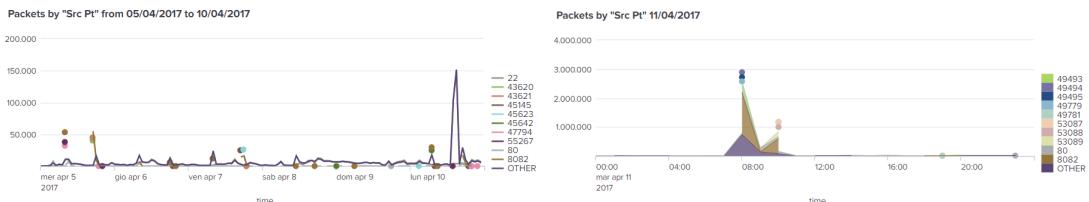
Nella parte bassa della dashboard sono invece riportati dei grafici in cui sono rappresentati gli andamenti temporali dei pacchetti trasmessi per source IP address, per source port, per destination IP address e per destination port. Come già notato l’11 aprile ha visto un traffico di dati molto maggiore rispetto ai giorni prima, per questo si è deciso di dividere ogni grafico in due intervalli temporali (l’intervallo dei giorni prima dell’11 aprile e quello dell’11 aprile stesso), altrimenti non si sarebbero potuti apprezzare gli andamenti dei giorni prima dell’11 aprile a causa dell’ordine di grandezza minore dei valori che si sono avuti in quei giorni. Infatti, confrontando i vari grafici affiancati si nota che si ha una differenza di un ordine di grandezza nelle ordinate. Come si nota tra i source IP address quelli che hanno mandato più pacchetti, a partire dal picco dell’11 aprile, e in generale durante quasi tutta la durata del periodo analizzato è la voce corrispondente al server esterno, seguita dalla voce “OPENSTACK-NET”, seguiti poi da altri vari indirizzi IP. Una cosa analoga si nota per quanto riguarda i destination IP address.



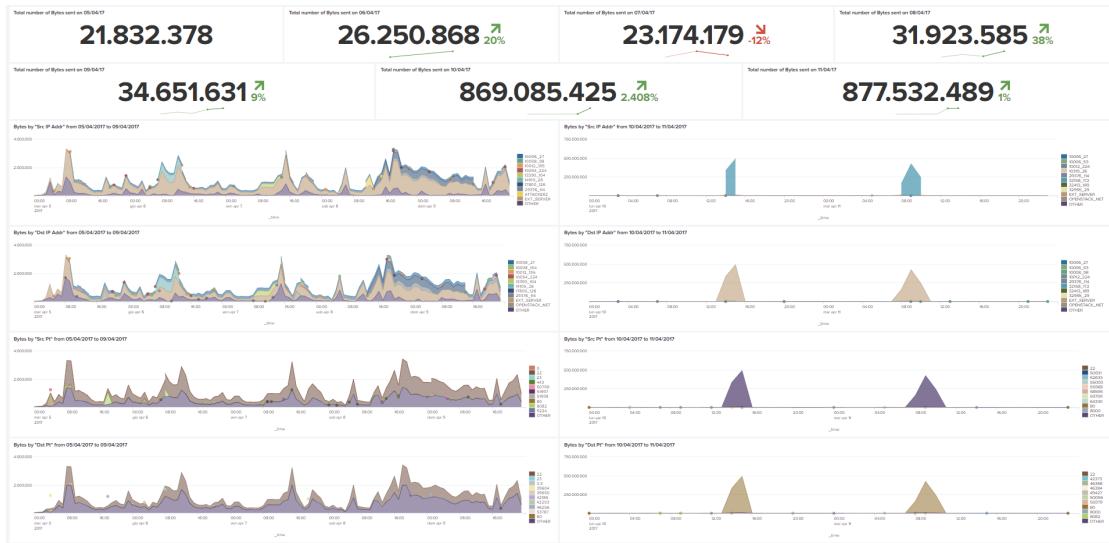
## CAPITOLO 7: ANALISI SUI PACCHETTI



Si osserva che a picchi di trasmissioni di dati verso un indirizzo corrispondono altrettanti picchi di trasmissione di dati in uscita da quello stesso indirizzo, ciò perché quasi sempre dopo l'accesso ad un indirizzo da esso vi è una risposta e quindi una trasmissione di dati inversa. Per quanto riguarda le destination port che hanno visto più flussi dati in termini di pacchetti si può notare che prima del picco dell'11 aprile quelle più frequenti erano la 8082 e la voce "other", che racchiude la somma di tutti gli altri IP, mentre durante il picco dell'11 aprile si sono avuti picchi di flussi dati su porte differenti dal periodo precedente, cioè la 49493, la 8082 e la 49494. Situazione analoga con le stesse porte si è avuta per le source port.



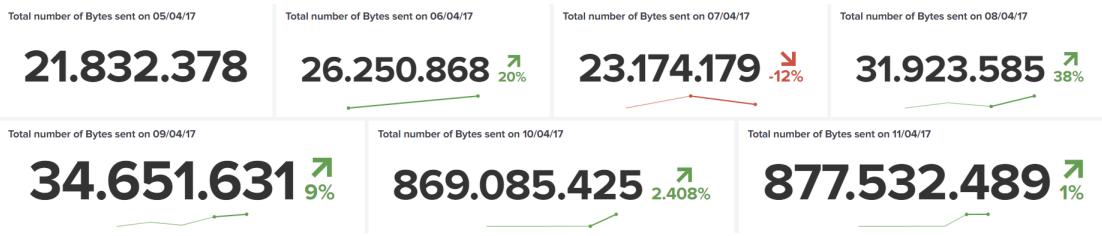
# 8. Analisi sui bytes



In questa dashboard invece ci si concentra principalmente sul numero di bytes che sono stati trasmessi durante i traffici di dati nei giorni del periodo considerato dal dataset.

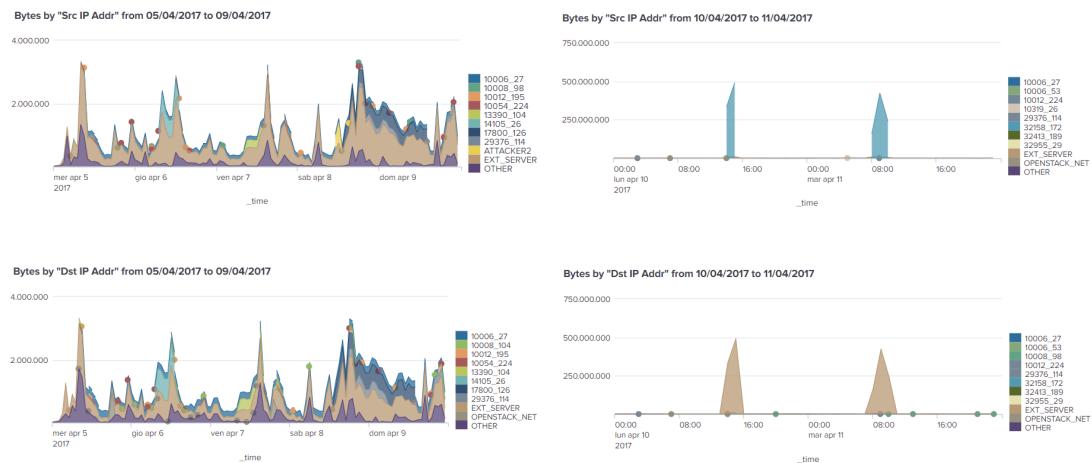
## 8.1 KPI sul numero di bytes trasmessi giorno per giorno

Nella parte alta della dashboard si hanno sette KPI che mostrano giorno per giorno il numero totale dei bytes trasmessi. Il giorno in cui si è avuto il maggior numero di bytes trasmessi è stato l'11 aprile (con oltre 870 milioni di bytes). Invece il giorno in cui si è avuto il minor numero di bytes trasmessi è stato il 5 aprile. Accanto ai KPI è mostrato il trend che si è avuto rispetto al giorno precedente in percentuale e si nota che si è avuto un calo nei bytes trasmessi solo il 7 aprile, mentre nei restanti giorni si è avuto un trend positivo, soprattutto il 10 aprile con un incremento di bytes trasmessi rispetto al giorno precedente del 2.408%



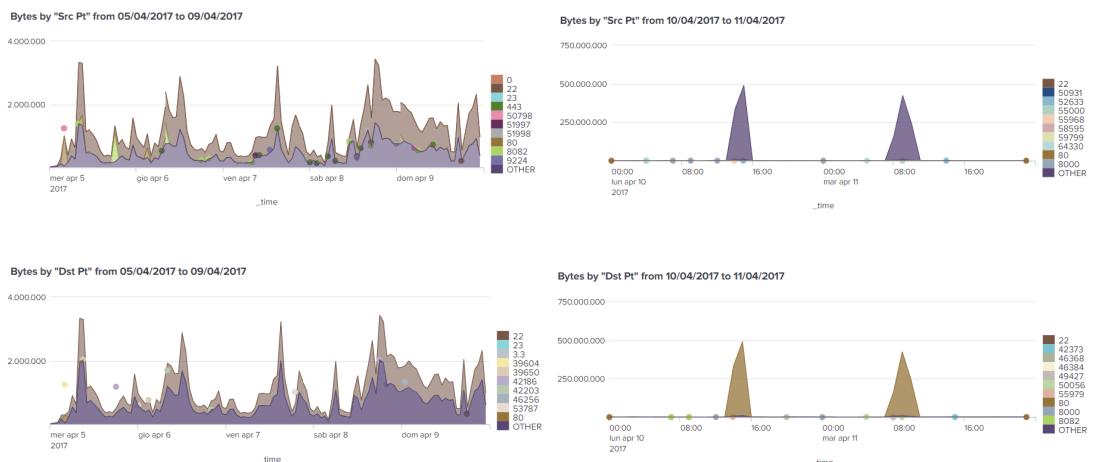
## 8.2 Andamenti temporali dei bytes trasmessi per source IP address, per source port, per destination IP address e per destination port

Nella parte bassa della dashboard sono invece riportati dei grafici in cui sono rappresentati gli andamenti temporali dei bytes nelle varie trasmissioni dati per source IP address, per source port, per destination IP address e per destination port. Come si nota i giorni in cui si sono registrati i maggiori picchi nei bytes trasmessi (in numero assoluto) sono stati il 10 e l'11 aprile. Come nella dashboard precedente si è deciso di dividere ogni grafico in due intervalli temporali (l'intervallo dei giorni prima dell'10 aprile e quello del 10 e 11 aprile), altrimenti non si sarebbero potuti apprezzare bene gli andamenti dei giorni prima del 10 aprile a causa dell'ordine di grandezza minore dei valori che si sono avuti in quei giorni. Infatti, confrontando i vari grafici affiancati si nota che si ha una differenza di due ordini di grandezza nelle ordinate. Come si nota tra i source IP address quelli che hanno mandato più bytes, soprattutto durante i picchi del 10 e 11 aprile, ma in generale durante gran parte della durata del periodo analizzato, sono stati il server esterno e l'indirizzo 32158-172. Una cosa analoga si nota per quanto riguarda i destination IP address.



## CAPITOLO 8: ANALISI SUI BYTES

Per quanto riguarda le destination port che hanno visto più flussi dati in termini di bytes si può notare che prima dei picchi del 10 e 11 aprile quelle più frequenti erano la 22 e la voce "other", che racchiude la somma di tutte le altre, mentre durante i picchi del 10 e 11 aprile si sono avute grandi quantità di flussi dati su una porta differente da quelle del periodo precedente, cioè la 80. Invece per le source port che hanno visto più flussi dati in termini di bytes si può notare che prima dei picchi del 10 e 11 aprile quelle più frequenti erano la 22 e la 9224, mentre durante i picchi del 10 e 11 aprile si sono avute grandi quantità di flussi dati su altre porte ("other").



# 9. Analisi temporale sull'utilizzo dei protocolli

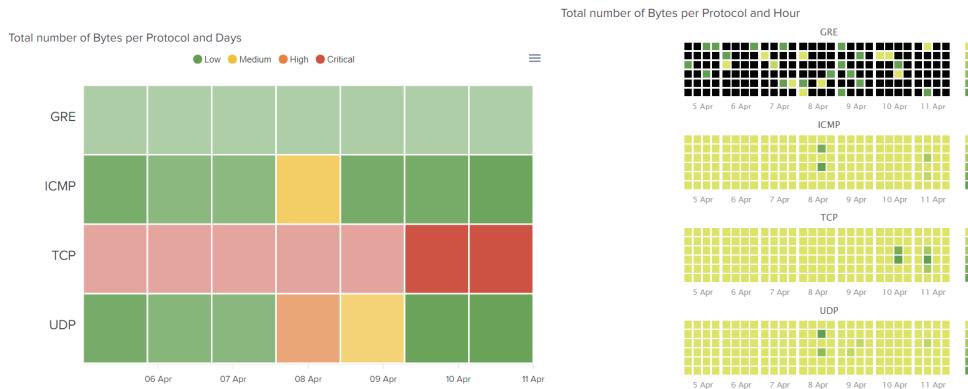


In questa dashboard viene fatta un'analisi su come i vari protocolli sono stati utilizzati durante i giorni considerati.

## 9.1 Analisi sul numero di bytes trasmessi per giorno e ora, per ogni protocollo

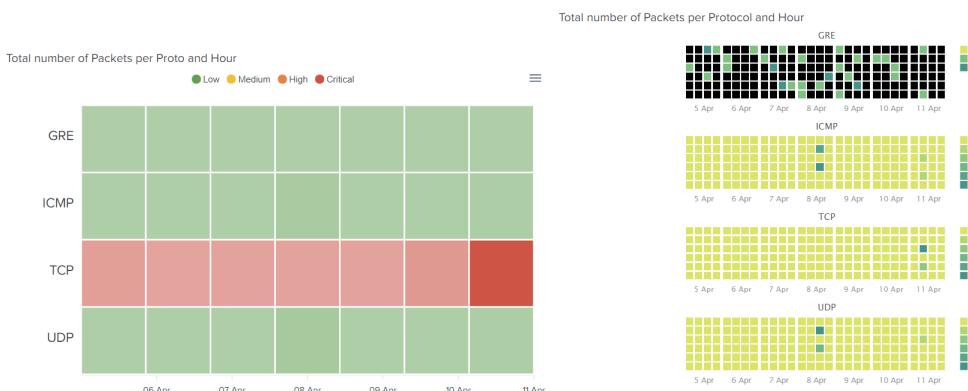
Per prima cosa si è fatta un'analisi in termini di numero di bytes trasmessi attraverso delle heatmap. Nella prima heatmap a sinistra della dashboard si hanno le righe che corrispondono ai quattro protocolli utilizzati e le colonne che corrispondono ai vari giorni. Sono stati stabiliti dei colori correlati al numero di bytes trasmessi su ogni combinazione protocollo-giorno. Come già ripetuto più volte il TCP risulta essere il protocollo più

utilizzato. Si conferma poi che i giorni che hanno registrato più bytes trasmessi sono stati il 10 e 11 aprile. Poi per studiare meglio gli accessi giornalieri considerando anche le ore in cui essi sono avvenuti, sono state realizzate altre quattro heatmap, una per protocollo. Anche in tal caso è attribuito un colore in base al numero di bytes trasmessi in una data ora di un giorno. Con questo maggior dettaglio si deduce che il picco di accessi avuti il 10 e 11 aprile sono stati effettuati in una fascia oraria abbastanza ristretta di due ore il 10 aprile e di tre ore l'11 aprile.



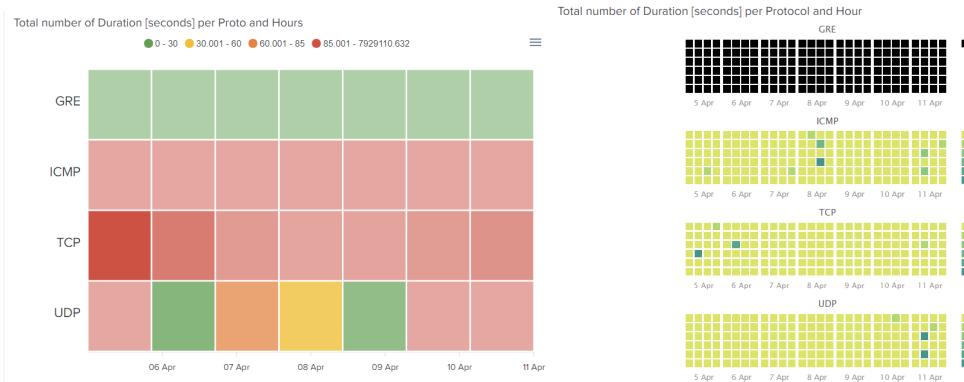
## 9.2 Analisi sul numero di pacchetti trasmessi per giorno e ora, per ogni protocollo

Una ulteriore analisi è fatta con heatmap analoghe sul numero di pacchetti trasmessi per protocollo e per giorno. Si osserva che a livello di pacchetti trasmessi con protocollo TCP, l'11 aprile è stato il giorno che ha registrato i valori più alti in assoluto, in particolare in una fascia di tre ore (la stessa per cui si è avuto anche un picco di bytes trasmessi).



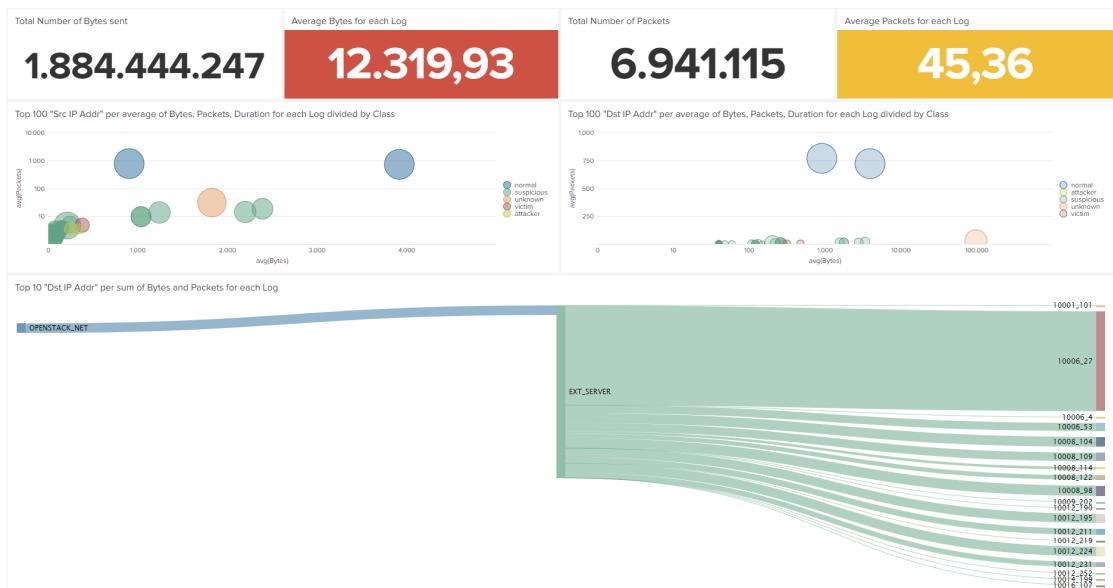
### 9.3 Analisi sulla durata delle trasmissioni di dati per giorno e ora per ogni protocollo

Infine una ulteriore analisi è fatta con heatmap analoghe sulla durata delle trasmissioni di dati per protocollo e per giorno. Si osserva che considerando il protocollo TCP a livello di durata delle trasmissioni il 6 aprile è stato il giorno che ha registrato i valori più alti in assoluto. Non si registrano invece durate particolarmente alte delle trasmissioni il 10 e 11 aprile in cui si sono avuti invece i picchi di bytes e pacchetti trasmessi, questo perché la durata non è strettamente correlata a queste variabili.



# 10. Analisi su Bytes e Pacchetti inviati e ricevuti

Un ulteriore studio interessante riguarda il numero di bytes e di pacchetti inviati divisi per l'indirizzo IP da dove sono partiti e quello di destinazione.



Le reti di calcolatori all'interno della rete operano secondo il modello detto packet switching. In una rete a commutazione di pacchetto l'informazione è trasmessa in pacchetti formati da una intestazione (header) ed un payload: l'header contiene informazioni di controllo, tra le quali un indirizzo destinazione che serve ad identificare il terminale a cui il pacchetto deve essere consegnato.

## 10.1 KPI

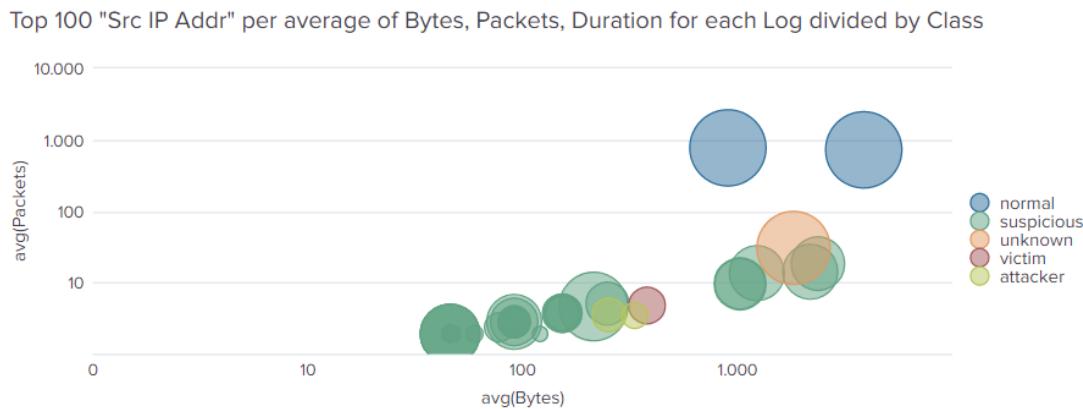
Si è voluto inanzitutto fare una panoramica iniziale attraverso le quattro KPI mostrate in figura.



In particolare la prima mostra il numero totale di bytes inviati mentre la seconda una media di questi. Stesso discorso per la somma e la media di pacchetti inviati e ricevuti.

## 10.2 Bubble Chart con i maggiori indirizzi IP di partenza

Nel seguente grafico si è andati a studiare i maggiori 100 indirizzi IP di partenza divisi per media di bytes, pacchetti inviati e durata in secondi della richiesta del flusso.



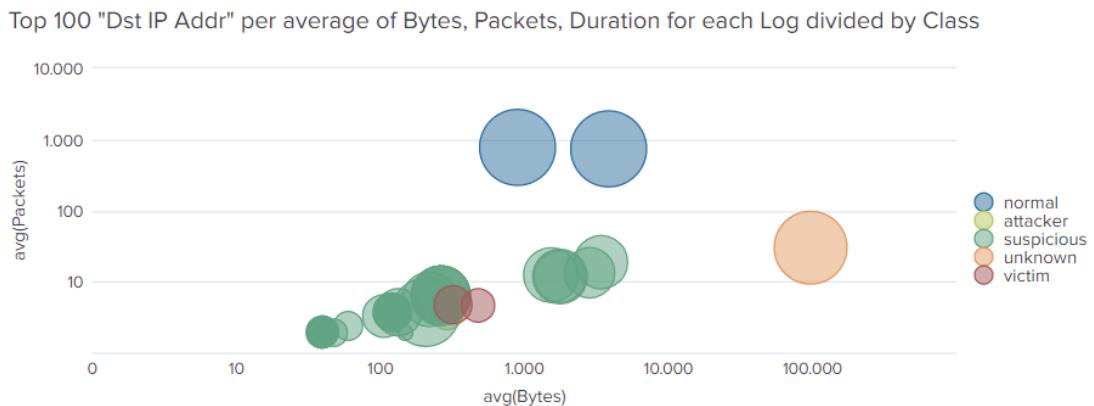
Come si può vedere si è andato ad utilizzare uno scatter plot (grafico a dispersione) che è un tipo di grafico in cui due variabili di un set di dati sono riportate su uno spazio cartesiano. I dati sono visualizzati tramite una collezione di punti ciascuno con una posizione sull'asse orizzontale determinato da una variabile e sull'asse verticale determinato dall'altra. In particolare ogni punto rappresenta un indirizzo IP di partenza e più è alta la media di pacchetti inviati (ordinate) e la media dei bytes inviati (ascisse), più tale indirizzo si posizionerà sulla parte in alto a destra del grafico. Per quanto riguarda la dimensione dei punti la variabile discriminante è in questo caso la durata media del flusso dell'operazione, mentre la gradazione di colori è divisa in base alla classe

della etichetta, e scorrendo con il puntatore sopra i vari punti viene visualizzato il nome dell'indirizzo IP.

Utilizzando una scala logaritmica si riesce a vedere più chiaramente la distinzione che c'è tra tali indirizzi IP, dove in particolare quelli che inviano in media più pacchetti e più bytes sono la *OpenStack Net* e la *External Server* e sono di classe "*normal*". Questo non ci stupisce dato che si etichettano tutti i flussi del server esterno che hanno come origine o target l'ambiente OpenStack Environment come "*normal*".

### 10.3 Bubble Chart con i maggiori indirizzi IP di destinazione

Stessa cosa si è fatta per lo studio dei maggiori 100 indirizzi IP di destinazione divisi per media di bytes, pacchetti inviati e durata in secondi della richiesta del flusso.

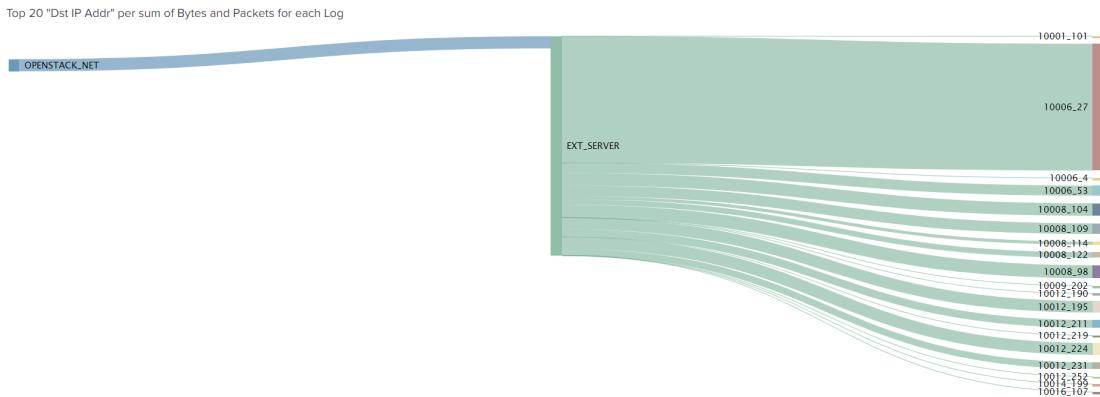


Si nota come anche in questo caso i maggiori due per media di pacchetti ricevuti sono la *OpenStack Net* e la *External Server*, di classe "*normal*". Al contrario il maggiore per numero di bytes ricevuti di media è l'*External Server*, questa volta però con classe della label "*unknown*".

### 10.4 Maggiori venti indirizzi IP di destinazione per somma di Bytes e Pacchetti

Per tale analisi si è scelto di utilizzare il diagramma di Sankey o diagramma di flusso in cui l'ampiezza delle frecce è disegnata in maniera proporzionale alla media di bytes

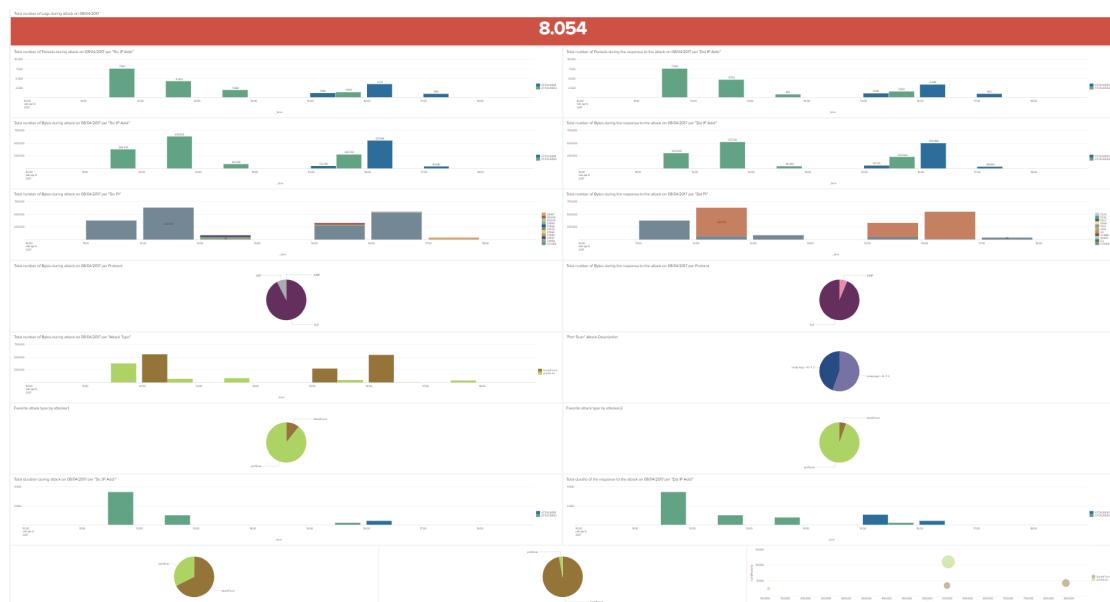
e pacchetti inviati. In generale i diagrammi di Sankey accentuano visivamente i grandi trasferimenti o flussi all'interno di un sistema: sono perciò utili in questo caso per individuare i contributi dominanti in un flusso complessivo.



In particolare tale diagramma si dirama due volte: da *OpenStack Net* arriva a *External Server* e successivamente si dirama di nuovo nei maggiori venti indirizzi IP di destinazione. Ciò coincide con quello già detto e cioè che tutti i client interni dell'ambiente OpenStack comunicano solo con il server esterno, che poi a sua volta indirizzerà un flusso dati verso un nuovo IP o di nuovo verso l'ambiente OpenStack.

# 11. Analisi sugli attacchi avvenuti nel 08/04/2017

Nel seguente paragrafo ci si è voluti incentrare sull’analisi degli attacchi provenienti dalla rete esterna durante il normale traffico network.

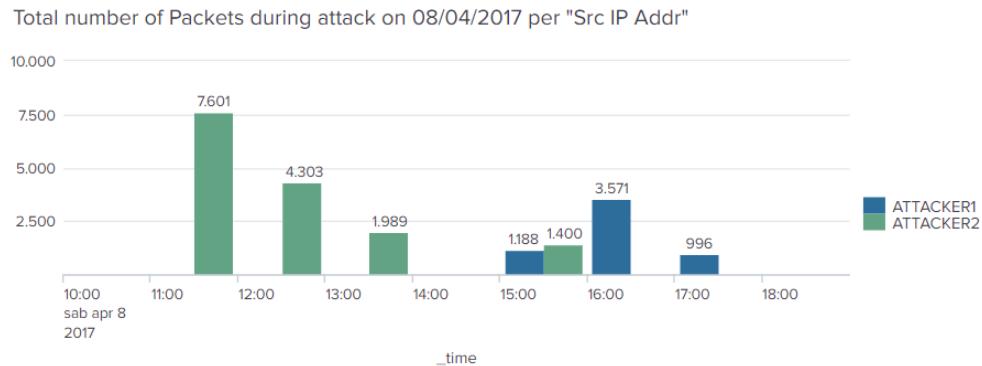


Dalla KPI in alto si nota subito come il numero totale dei flussi dati che riguardano gli attacchi avvenuti in quel giorno sono in gran numero: 8054.



## 11.1 Numero totale di pacchetti inviati durante gli attacchi del 08/04/2017

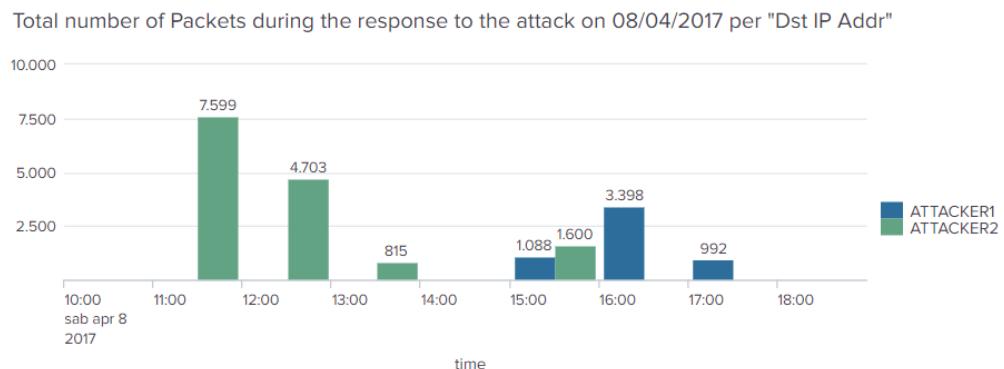
Un'informazione interessante è il numero totale di pacchetti che sono stati inviati dall'indirizzo IP di partenza durante quella giornata, discriminando da quale attaccante sono appunto partiti.



Dalla figura si nota come gli unici attaccanti sono due e che hanno operato nelle seguenti fasce orarie della giornata, con una maggiore concentrazione nella mattina da parte di "Attacker2" e del pomeriggio da "Attacker1".

## 11.2 Numero totale di pacchetti inviati durante la risposta agli attacchi del 08/04/2017

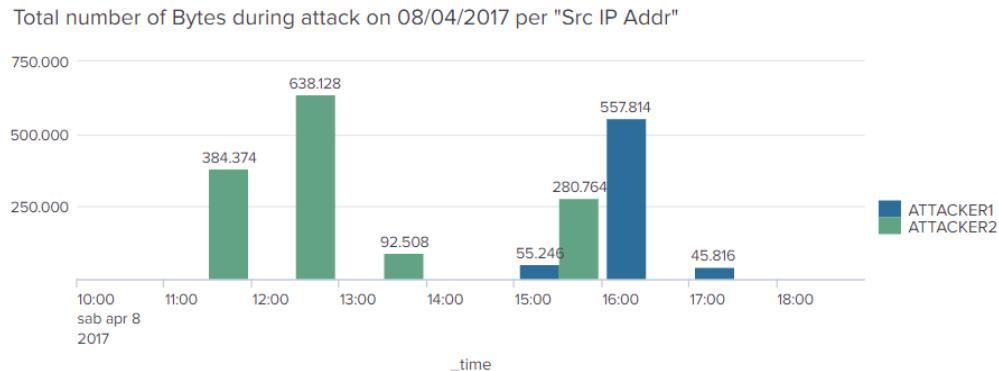
Allo stesso modo è interessante analizzare il numero totale di pacchetti inviati dal server esterno, che sono arrivati come risposta all'indirizzo IP di destinazione.



In questo caso il numero di pacchetti dati in risposta sono all'incirca lo stesso numero di pacchetti che erano stati inviati dagli attaccanti.

### 11.3 Numero totale di Bytes inviati durante gli attacchi del 08/04/2017

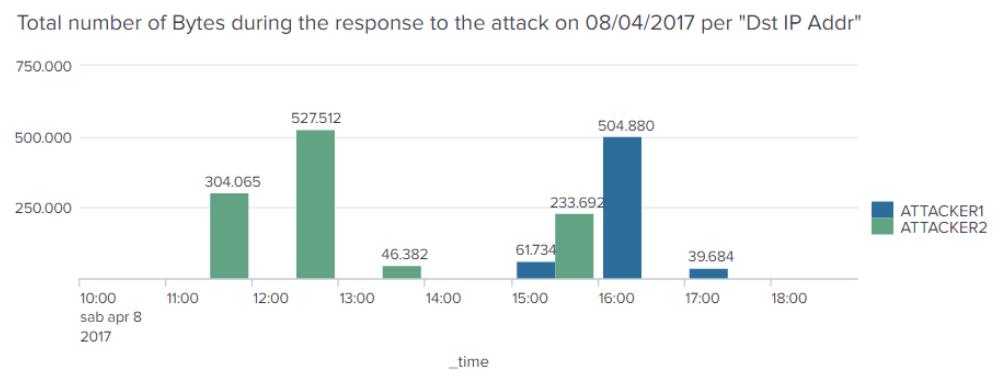
Stessa analisi si può attuare sul numero totale di bytes che sono stati inviati dall'indirizzo IP di partenza durante l'08/04/2017, discriminando da quale attaccante sono appunto partiti.



In questo caso c'è stata una maggiore quantità di bytes inviata nella fascia oraria delle 12:00 alle 13:00 mentre nel caso precedente il picco di aveva dalle 11:00 alle 12:00 da parte dell'"Attacker2".

### 11.4 Numero totale di bytes inviati durante la risposta agli attacchi del 08/04/2017

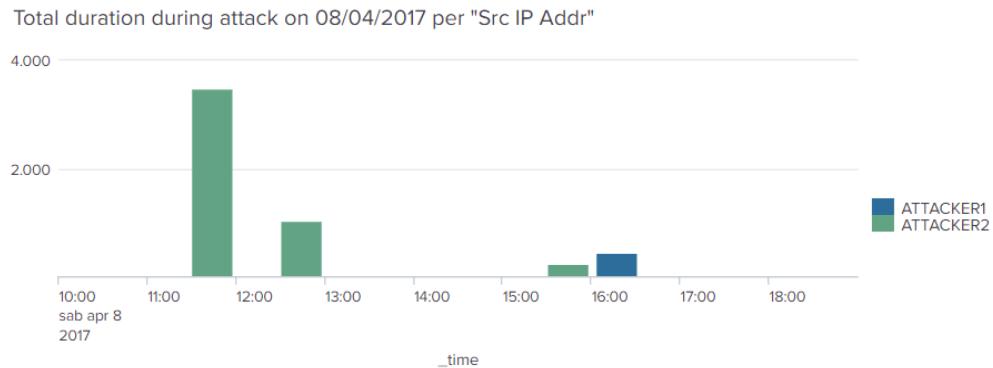
Il numero totale di bytes che sono arrivati come risposta dall'indirizzo IP di destinazione durante quella giornata sono raffigurati nella seguente figura.



In questo caso il numero di bytes inviati dal server esterno in risposta ai due attaccanti è leggermente minore di quello inviato durante l'attacco.

## 11.5 Durata totale degli attacchi del 08/04/2017

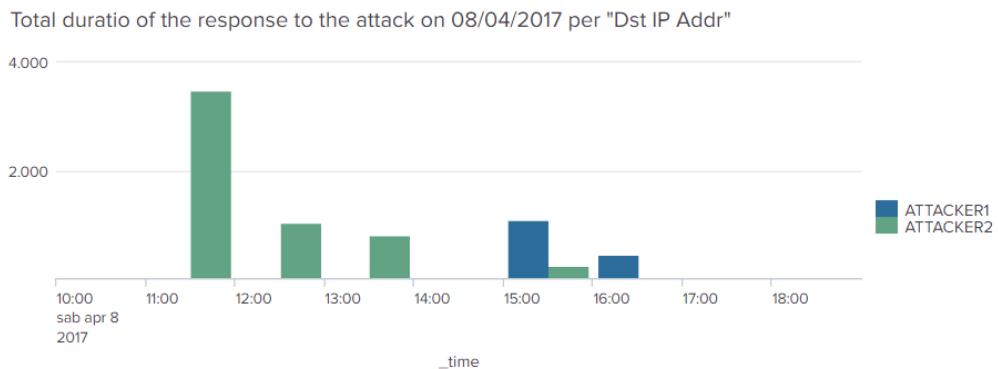
Ultima analisi sui flussi di dati trasmessi dagli indirizzi IP di partenza riguarda la loro durata, distribuita nelle ore della giornata del 08/04/2017, anche qui discriminando in base all'attaccante.



Si nota come l'attacco proveniente da "Attacker1" duri quasi 3496 secondi (poco meno di un'ora) ed è per questo che nel grafico 11.1 raggiunge il picco per pacchetti inviati. Interessante è inoltre notare come invece l'attacco presente nella fascia oraria che va dalle 12:00 alle 13:00 duri di meno: 1129 secondi (poco più di un'ora) ma abbia più bytes trasmessi. Infine si nota come l'"Attacker2" tenda a trasmettere dati per una durata inferiore di tempo, tuttavia con appena 445 secondi (circa 15 minuti) è riuscito a trasmettere quasi lo stesso numero di bytes dell'"Attacker1" in più di un'ora.

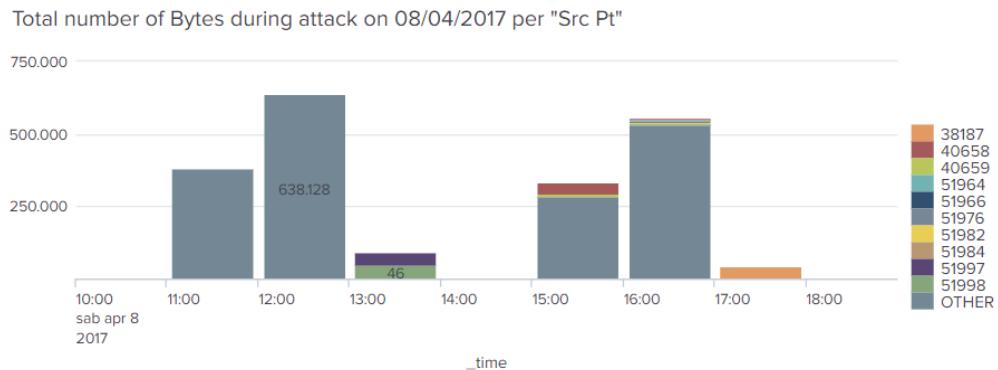
## 11.6 Durata totale della risposta agli attacchi del 08/04/2017

Le durate totali dei flussi di dati che sono arrivati come risposta dal server esterno ad uno dei due attaccanti durante quella giornata sono raffigurati nella seguente figura.



## 11.7 Numero totale di Bytes usati durante gli attacchi del 08/04/2017 divisi per porte di partenza

Ulteriore analisi la si può effettuare sul numero totale di bytes trasmessi divisi per "Source Port" durante l'08/04/2017. Nel grafico sono rappresentate le porte con maggiori bytes trasmessi e in grigio sulla voce "OTHER" le restanti porte con meno bytes trasmessi. Questo grafico mostra quindi come siano stati usati durante gli attacchi un numero piuttosto bilanciato di bytes per ogni Source Port e non un attacco massiccio su un'unica porta.



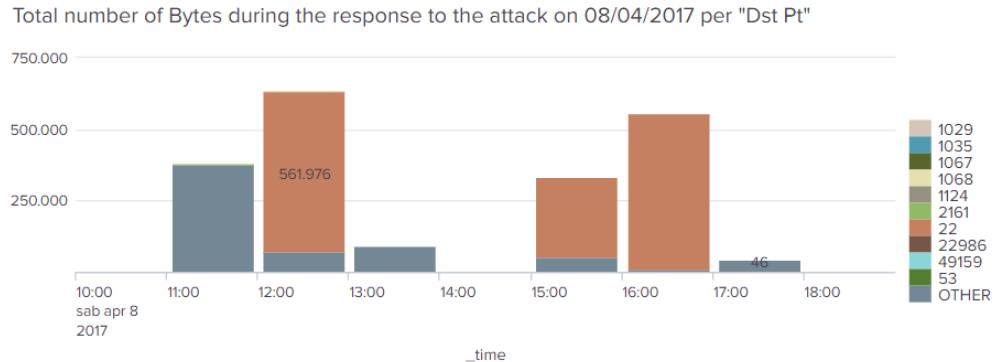
In questo caso le porta di partenza più utilizzate sono quelle raffigurate nella legenda, dove in particolare una divisione in base all'ora si può vedere dalla tabella sottostante:

_time	38187	40658	40659	51964	51966	51976	51982	51984	51997	51998	OTHER
2017-04-08 10:00											
2017-04-08 11:00					152	152	152	152	152	152	383310
2017-04-08 12:00											638128
2017-04-08 13:00									46231	46231	46
2017-04-08 14:00											
2017-04-08 15:00		40425	10957								284628
2017-04-08 16:00		6114	1656	2805	2805	2805	2805	2805	2701	533318	
2017-04-08 17:00		45770									46
2017-04-08 18:00											

## 11.8 Numero totale di Bytes usati durante gli attacchi del 08/04/2017 divisi per porte di destinazione

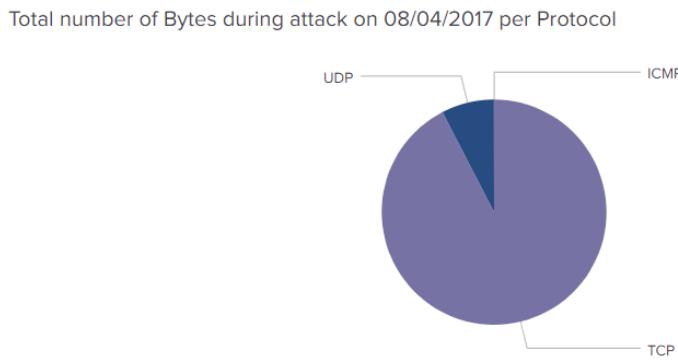
Parallelamente, l'analisi precedente la si può effettuare sul numero totale di bytes trasmessi divisi per "Destination Port" durante l'08/04/2017. Si può notare come la porta indubbiamente più utilizzata sia la **22**, che appartiene al protocollo **SSH**, dove l'SSH

(Secure SHell) è un protocollo che permette di stabilire una sessione remota cifrata tramite interfaccia a riga di comando con un altro host di una rete informatica.



## 11.9 Numero totale di Bytes usati durante gli attacchi del 08/04/2017 divisi per protocollo

Altro studio interessante sulla natura degli attacchi può esser fatto sempre sul numero totale di bytes trasmessi, ma in questo caso dividendoli per protocollo utilizzato:

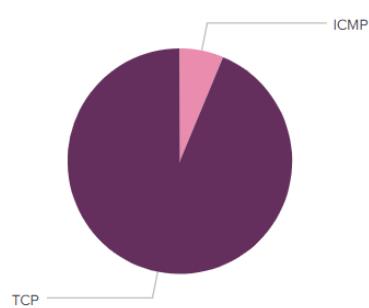


Si può notare come il protocollo più utilizzato è il TCP, che si occupa di controllo della trasmissione rendendo affidabile la comunicazione dati in rete tra mittente e destinatario. In minor parte è utilizzato l'UDP e in piccolissima parte l'ICMP.

## 11.10 Numero totale di Bytes usati durante la risposta agli attacchi del 08/04/2017 divisi per protocollo

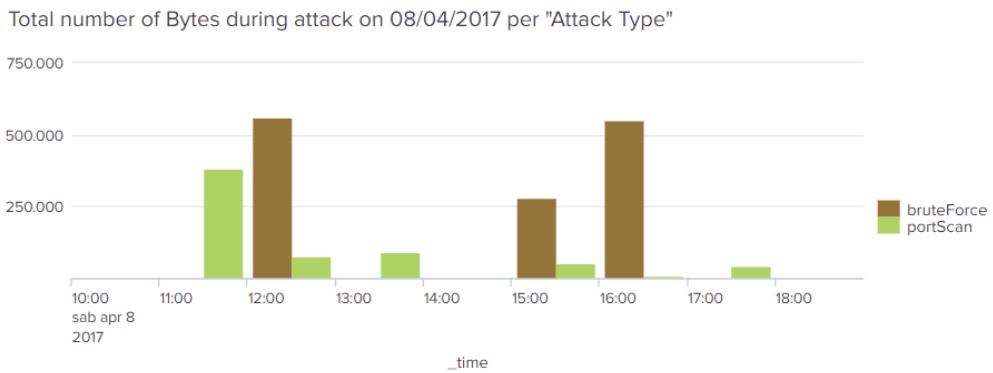
Per quanto riguarda il totale di bytes trasmessi come risposta dall'indirizzo Ip di destinazione dalle vittime, anche in questo caso il protocollo più usato è il TCP ed è all'incirca la stessa percentuale dei bytes trasmessi durante gli attacchi. Al contrario per quanto riguarda il secondo protocollo più usato in questo caso è l'ICMP (6.3%), mentre nella figura precedente era l'UDP (7.6%).

Total number of Bytes during the response to the attack on 08/04/2017 per Protocol



## 11.11 Numero totale di Bytes usati durante gli attacchi del 08/04/2017 divisi per "Attack Type"

Stessa analisi si può attuare sul numero totale di bytes che sono stati inviati durante l'08/04/2017 divisi per "Attack Type".

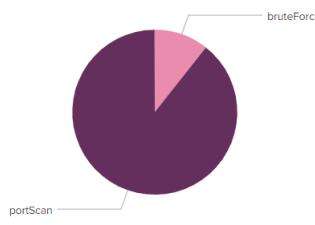


Grazie a quest'ultimo grafico si può vedere come la fascia oraria che andava dalle 11:00 fino alle 12:00 nella quale si aveva un picco dei pacchetti inviati [11.1](#) e una durata di più

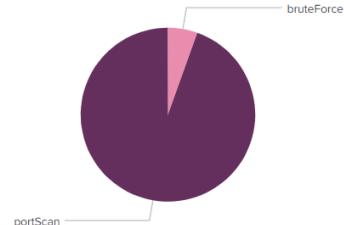
di un'ora [11.5](#), si hanno unicamente attacchi di tipo **Port Scan**. Inoltre gli attacchi nei quali si trasmettono maggior numero di bytes sono di tipo **Brute Force**.

## 11.12 Tecniche di attacco preferite dagli attaccanti

Un'altro studio portato avanti per capire la natura degli attacchi è stato quello di dividere il numero di attacchi per "Attack Type".



**Figura 11.1:** "Attacker 1"



**Figura 11.2:** "Attacker 2"

Si può notare che sia per l'"Attacker 1" nel 89% dei casi che per l'"Attacker 2" per il 94% degli attacchi, il tipo preferito è la **Port Scan**. Ricollegandosi al paragrafo precedente [11.11](#) si nota come gli attacchi di tipo **Brute Force** sono molto meno numerosi, ma in questo caso quando questi avvengono, inviano un numero molto più elevato di bytes rispetto alla **Port Scan**.

# 12. Analisi sugli attacchi avvenuti il 10/04/2017



Nel seguente paragrafo ci si è voluti incentrare sull'analisi degli attacchi avvenuti nella data del 10/04/2017 provenienti dalla rete esterna durante il normale traffico network.

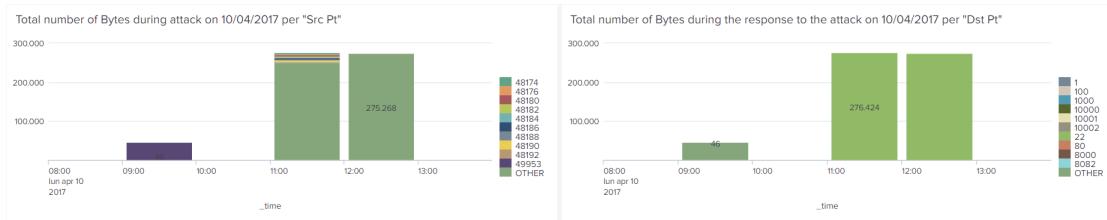
Si ha che in questo caso gli attacchi sono di meno (1201) e inoltre c'è solo un attaccante sulla rete in tale giornata. In particolare gli orari nei quali si hanno maggiormente pacchetti e bytes trasmessi sono dalle 11:00 alle 13:00, ma comunque si ha un buon traffico anche nella fascia oraria che va dalle 09:00 alle 10:00 .

_time	ATTACKERI
2017-04-10 08:00	0
2017-04-10 09:00	0.058
2017-04-10 10:00	0
2017-04-10 11:00	226.992
2017-04-10 12:00	219.504
2017-04-10 13:00	0

Questo dimostra la natura diversa dei due attacchi, quello della mattina con durata quasi istantanea a tanta trasmissione di informazione e il secondo nella tarda mattina

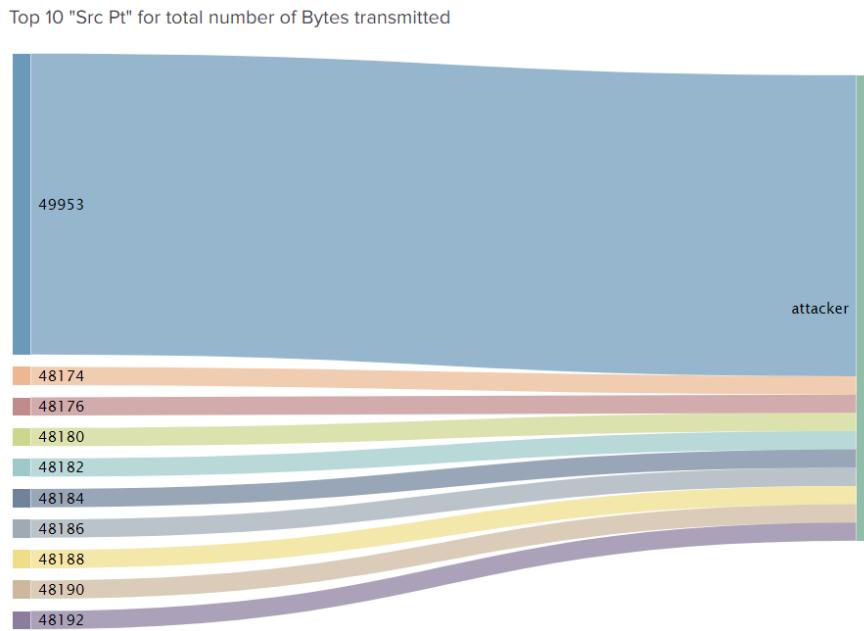
di media con molti meno bytes e pacchetti trasmessi ma per un tempo più elevato.

## 12.1 Numero totale di bytes trasmessi per "Src Pt" e "Dst Pt"



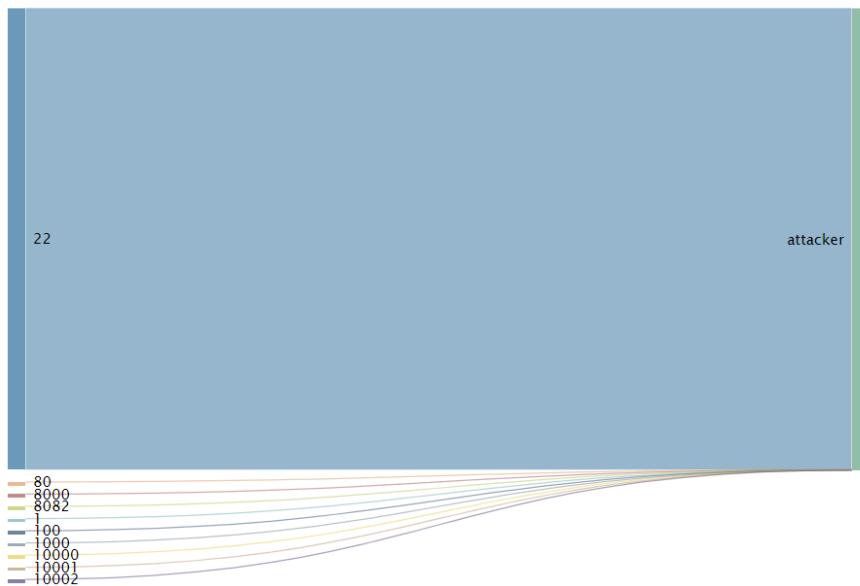
Come ci si poteva aspettare il maggior flusso di trasmissione di dati durante gli attacchi nella giornata del 10/04/2017 si ha nella tarda mattinata.

Si può vedere che per quanto riguarda le "Source Port" dell'attaccante, le migliori 10 che hanno trasmesso un numero maggiore di byte sono raffigurati nel seguente grafico:



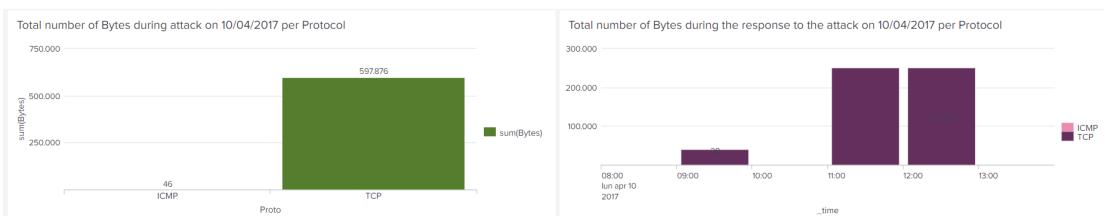
Al contrario per quanto riguarda le "Destination Port" come si poteva intuire la principale è la 22 che mi indica i protocolli SSH.

Top 10 "Dst Pt" for total number of Bytes transmitted

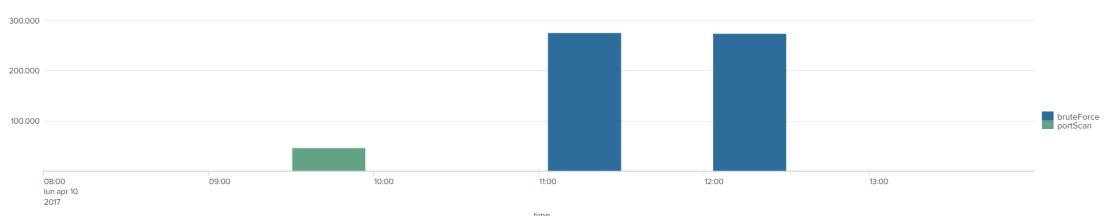


## 12.2 Numero totale di bytes trasmessi per Protocollo e "Attack Type"

Infine andando a studiare numero di bytes trasmessi divisi per protocollo durante gli attacchi del 10/04/2017 si può notare come la quasi totalità è di tipo TCP.



Inoltre si può notare come gli attacchi nel quale sono stati inviati un numero maggiore di bytes sono quelli di tipo **Brute Force**, avvenuti nella tarda mattinata.

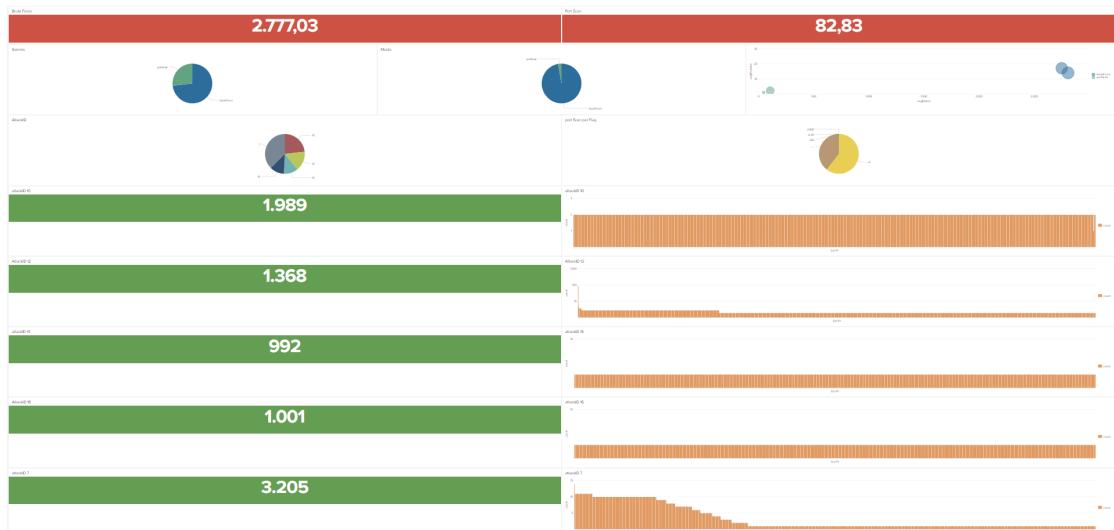


# 13. Analisi Port Scan

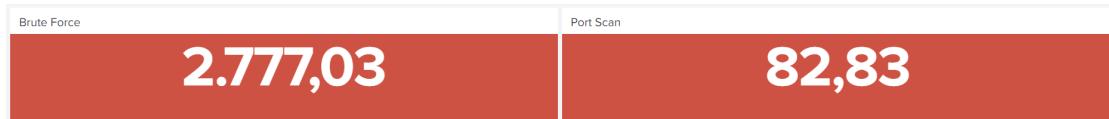
In questa dashboard si è voluto condurre un'analisi specifica sugli attacchi di tipo **Port Scan**, che probabilmente tra i due tipi eseguiti è il più interessante.

Con port scanning, in informatica, si indica una tecnica progettata per sondare un server o un host al fine di stabilire quali porte siano in ascolto sulla macchina. Questa tecnica è spesso utilizzato dagli amministratori per verificare le politiche di sicurezza delle loro reti, e dagli hacker per identificare i servizi in esecuzione su un host e sfruttarne le vulnerabilità.

Un **Port Scan** è un processo che invia le richieste dei client a un intervallo di indirizzi di porte su un host, con l'obiettivo di trovare una porta aperta. Elaborando le risposte è possibile stabilire (anche con precisione) quali servizi di rete siano attivi su quel computer: una porta si dice "in ascolto" ("listening") o "aperta" o attiva quando vi è un servizio, programma o processo che la usa.



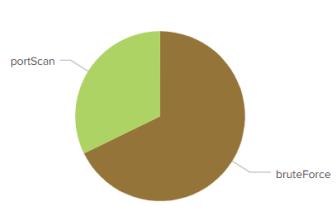
### 13.1 KPI



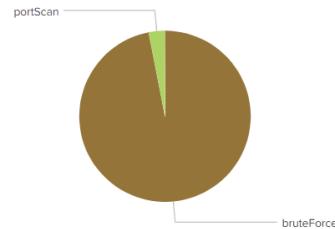
Nelle KPI sono indicati il numero di bytes medio che generalmente viene inviato in un flusso dati durante l'attacco di tipo **Brute Force** e **Port Scan**. Si vede immediatamente come nel **Port Scan** i bytes inviati su ciascuna porta siano molti meno che nell'altro tipo. Questo è dovuto al fatto che nel Port Scan si inviano pochi dati su ciascuna porta che si testa per capire semplicemente se essa sia aperta o meno.

### 13.2 Bytes inviati nei due tipi di attacco

La somma totale dei bytes inviati e la media di bytes che ciascun tipo di attacco invia, sono sintetizzati nella seguente figura:



**Figura 13.1:** Somma di Bytes



**Figura 13.2:** Media di Bytes

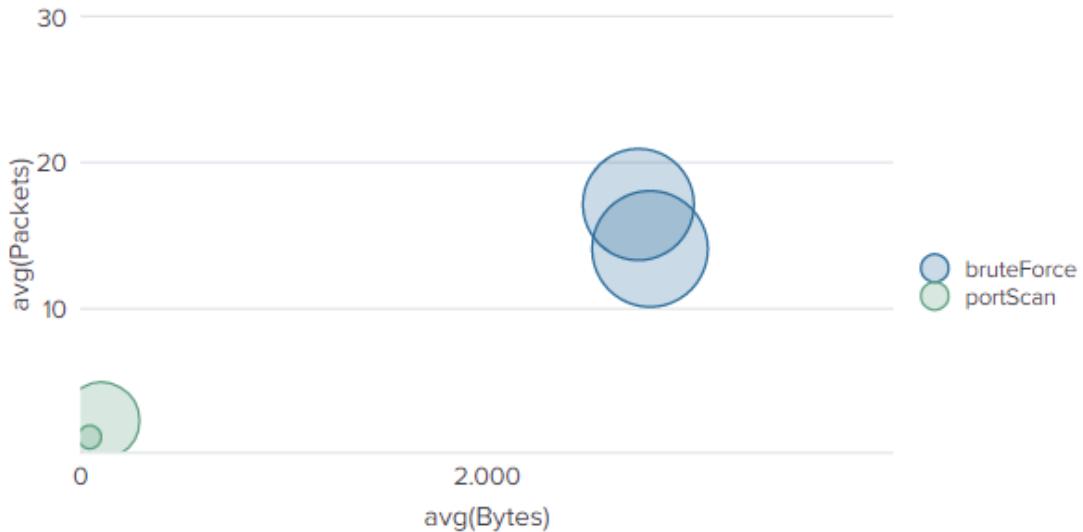
Si può osservare come mediamente i bytes trasmessi negli attacchi di tipo **Brute Force** sono molto maggiori di quelli di tipo **Port Scan**, come visto nella KPI precedente, e se si considera la somma totale si osserva che il 73% dei bytes è stato trasmesso durante attacchi di tipo **Brute Force**.

### 13.3 Analisi in base al numero totale di bytes, pacchetti e durata per "Attack Type"

Come ulteriore analisi è interessante notare che entrambi gli attaccanti, durante gli attacchi hanno inviato mediamente un numero di bytes e pacchetti simili, e in particolare

## CAPITOLO 13: ANALISI PORT SCAN

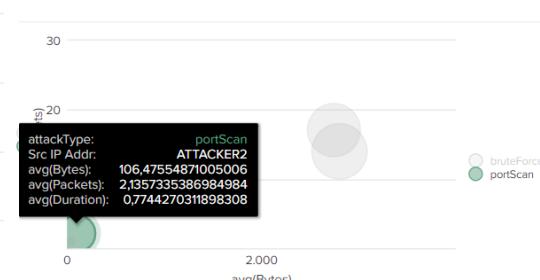
per gli attacchi di tipo **Port Scan** essi avevano valori molto ridotti rispetto all'altro tipo di attacco.



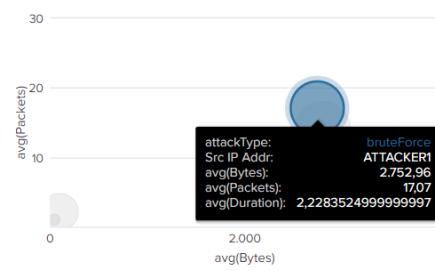
La grandezza delle bolle all'interno del grafico è data dalla durata complessiva degli attacchi, che sono di due tipi, da parte di uno dei due attaccanti. I valori sono riportati nella figura successiva:



**Figura 13.3: "Attacker 1" portScan**



**Figura 13.4: "Attacker 2" portScan**



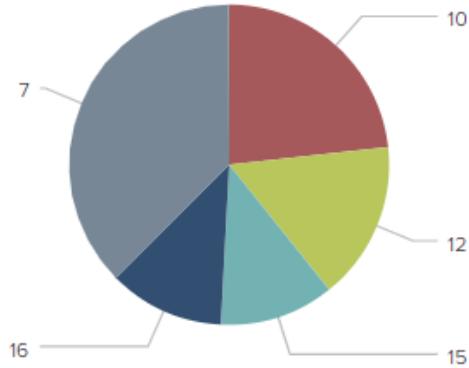
**Figura 13.5: "Attacker 1" Brute Force**



**Figura 13.6: "Attacker 2" Brute Force**

### 13.4 Numero di attacchi di tipo Port Scan avvenuti

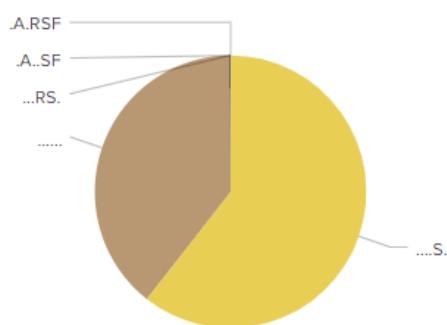
Come già detto, i flussi registrati nel dataset che appartengono allo stesso attacco, hanno nella voce *AttackID* uno stesso ID che identifica l'attacco.



Dalla torta si vede come gli attacchi effettuati siano effettivamente 5, poiché abbiamo altrettanti ID diversi. L'area associata a ciascun elemento nella torta è proporzionale al numero di flussi che hanno generato durante gli attacchi, e quindi al numero di porte a cui hanno provato ad accedere.

### 13.5 Flag utilizzati durante gli attacchi

Dalla torta si vede come per quasi la metà delle volte non sia stato utilizzato alcun flag, mentre il più utilizzato risulta essere il flag *S*. Nel protocollo TCP, un pacchetto con questo flag indica un tentativo di connessione.



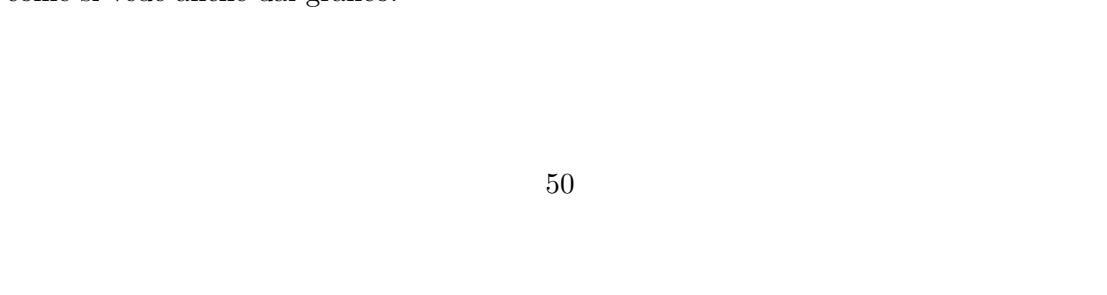
## 13.6 Connessioni tentate in ciascun attacco

Come ultima analisi si è andato a vedere a quali porte si è tentata la connessione in ciascun attacco, e quindi per ciascun ID. Per ogni attacco sono state tentate almeno un migliaio di connessioni a porte diverse, e spesso su ciascuna porta è stata tentata più di una connessione.

Iniziando dal primo attacco, quello con ID 10, dalla KPI si legge che il numero di connessioni tentate sono 1989, e in particolare sono state provate 995 porte diverse, con due tentativi di connessione a ciascuna, tranne per la porta 8.0 nella quale è avvenuta una sola connessione.



Nel secondo attacco, quello con ID 12, sono state tentate meno connessioni, ovvero 1368. In questo caso sono state provate 1001 porte diverse, con un picco di 89 tentativi per la porta 22986 (da notare sul grafico la scala logaritmica nell'asse Y per poter ottenere una visualizzazione migliore).



## CAPITOLO 13: ANALISI PORT SCAN



Nel quarto attacco, quello con ID 16, sono state tentate 1001 connessioni, e per ciascuna porta c'è stato un solo tentativo, come si vede anche dal grafico.



Infine nell'ultimo attacco, quello con ID 7, sono state tentate 3205 connessioni a 998 porte diverse. In questo caso per molte porte sono state tentate anche una decina di connessioni, e si vede dal grafico che la porta più utilizzata è stata la 53 con 14 tentativi. La porta 53 è un servizio DNS usato per *domain name resolution*, e spesso alcuni attacchi prendono di mira le vulnerabilità all'interno dei server DNS.



## 14. Conclusioni

In questo elaborato si è cercato di mostrare al meglio alcune delle potenzialità del software Splunk, dando una panoramica su quelle che sono le reali potenzialità dello stesso, con il quale si è riusciti anche ad ottenere un'analisi completa sul comportamento degli attaccanti al server. Ovviamente oltre alle analisi che sono state mostrate, Splunk è in grado di fare analisi in tempo reale, comportandosi come un Intrusion Detection System(IDS) e un Security Information and Event Management(SIEM). Si può affermare che Splunk è leggermente più complicato nell'uso rispetto ai tools di Business Intelligence che erano stati utilizzati precedentemente. Il linguaggio SPL (Search Processing Language) riesce ad estrapolare informazioni estremamente utili, mantenendo il suo avanzato modo di operare completamente trasparente nei confronti dell'utente, rendendolo quasi user-friendly dopo un po' di apprendimento nell'uso. Infatti nel campo dei SIEM è proprio Splunk il leader di mercato secondo il *Magic Quadrant Gartner*.

# Bibliografia

- [1] Markus Ring, Sarah Wunderlich, Dominik Gruedl, Dieter Landes, and Andreas Hotho. *Flowbased benchmark data sets for intrusion detection*. ACPI, 2017.