Università degli Studi di Padova

Dipartimento di Matematica "Tullio Levi-Civita"

Corso di Laurea in Informatica

Valutazione dell'Efficacia di Piracy Shield tramite l'Analisi dei Flussi Video e il suo Funzionamento

Tesi di laurea

Relatore
Prof. Alessandro Galeazzi

Correlatore
Prof. Mauro Conti

Laureando
Davide Maffei
Matricola 2042373

Anno Accademico 2023-2024

Dedicato a me stesso, a mio papà che mi ha trasmesso la passione per i computer e a mia mamma che mi ha sempre sostenuto.

Sommario

Il presente documento descrive il lavoro svolto durante il periodo di stage interno, della durata di circa trecentoventi ore, dal laureando Maffei Davide presso l'Università degli Studi di Padova. Il tirocinio è stato portato a termine sotto la supervisione del Prof. Alessandro Galeazzi, che ha ricoperto il ruolo di proponente, e con il supporto del Dott. Giacomo Quadrio e del Dott. Enrico Bassetti, membri del gruppo di ricerca SPRITZ del Dipartimento di Matematica dell'Università di Padova.

Il progetto di stage si concentra sull'analisi del traffico di rete che transita attraverso lo snodo VSIX[g] e sulla sicurezza informatica, con particolare attenzione a problematiche legate alla riproduzione illegale di contenuti in *streaming*. L'obiettivo è quello di stabilire delle tecniche pratiche e teoriche per l'analisi di dataset[g] complessi, al fine di classificare i diversi tipi di flussi di dati e identificare il traffico illegale. Il progetto nasce dall'esigenza di raccogliere informazioni sul funzionamento di Piracy Shield, nonché sulla valutazione della sua efficacia. Attraverso l'analisi dei flussi di rete, si vuole riuscire a distinguere i flussi video dagli altri tipi di traffico e a identificare le differenze tra il traffico ordinario e quello piratato.

Gli obiettivi del progetto includono l'acquisizione di competenze nella manipolazione di grandi dataset di traffico di rete, la progettazione e implementazione di un sistema di classificazione dei flussi, e l'identificazione di traffico legale e illegale utilizzando tecniche avanzate di analisi dei dati.

"Everybody want to know what I would do if I didn't win... I guess we'll never know."

— Kanye West

Ringraziamenti

Innanzitutto, vorrei esprimere la mia gratitudine al Prof. Alessandro Galeazzi, relatore della mia tesi, per l'aiuto e il sostegno fornitomi durante la stesura del lavoro.

Ringrazio in primis i miei genitori per il sostegno e per avermi dato la possibilità di fare questa esperienza, la quale molte volte ho dato per scontata, ma grazie ai loro sacrifici mi ha permesso di crescere e di imparare molto.
Ringrazio la mia famiglia: mio fratello, i miei zii, le mie cugine e soprattutto i miei nonni Roberto, Cristina e Elda per il supporto e per esserci sempre stati.

Ringrazio i miei coinquili per avermi sopportato durante questi anni, in particolar modo Luca con cui ho passato 3 anni di università e con cui ho condiviso molte esperienze. Sperando di poter coninuare ad andare a molti altri concerti assieme.

Ringrazio la mia amica Anna la quale è riuscita a rendere le giornate più speniserate e divertenti. È grazie a lei se sono riuscito ad apprezzare Padova.

Ringrazio i miei due amici del corso: Marco e Giulio che a volte c'erano e a volte no, ma sono riusciti a rendere le lezioni e gli esami più diverenti.

Ringrazio i miei amici di sempre: Alberto, Alessandro, Elia, Leonardo e Matteo, quelli che mi hanno conosciuto prima di iniziare l'università e che mi sono stati vicini in questi anni.

Ringrazio Lorenzo e Ettore i miei amici di Rorai, quelli con cui sono cresciuto, ma soprattutto mio cugino Mario, che oltre ad essere mio cugino è soprattutto il mio migliore amico.

Ringrazio infine una persona in particolare che ha reso possibile tutto ciò: me stesso.

Padova, Settembre 2024 Davide Maffei

Indice

Elenco delle figure

Elenco delle tabelle

Capitolo 1

Introduzione

In questo capitolo introduttivo verranno presentati un breve riassunto del progetto e la struttura del documento.

1.1 Introduzione del Progetto

L'accesso ai contenuti in *streaming* senza autorizzazione, come nel caso delle dirette sportive trasmesse senza possederne i diritti, viene definito come pirateria *online*. In Italia, questo fenomeno è particolarmente significativo per quanto riguarda le trasmissioni di eventi sportivi, in particolare le partite di calcio. Un gran numero di spettatori riesce a vedere questi eventi tramite siti *web* di *streaming* illegali o utilizzando decoder[g]illegali, che permettono di guardare programmi disponibili solo sulle TV a pagamento senza sottoscrivere alcun abbonamento.
Per contrastare lo *streaming* illecito di contenuti protetti da diritto d'autore, in particolare riguardante gli eventi sportivi e le partite di calcio, lo Stato italiano ha sviluppato una piattaforma nota come "Piracy Shield". Questa iniziativa, che è stata resa operativa dall'Autorità per le Garanzie nelle Comunicazioni (AGCOM) il 1º febbraio 2024, ha come obiettivo quello di combattere la pirateria digitale [3].
Tuttavia, sin dalla sua nascita, il Piracy Shield ha mostrato numerosi difetti e problematiche, risultando facilmente aggirabile e generando anche numerose controversie. Ad esempio, il sistema ha causato il blocco ingiusto ed errato di alcuni servizi a causa della mancanza di controllo da parte di un'autorità pubblica. Le aziende private che detengono i diritti d'autore possono infatti inibire l'accesso a siti presumibilmente pirati senza che vi sia una verifica della legittimità delle loro segnalazioni. Ciò solleva preoccupazioni sulla violazione della net neutrality[g], poiché consente a entità private di influenzare l'accesso ai contenuti *online*. Inoltre, la trasparenza delle operazioni è limitata: nel sito ufficiale dell'AGCOM, infatti, non sono specificati né i soggetti che hanno effettuato le segnalazioni né le motivazioni del blocco, rendendo difficile una valutazione pubblica dell'efficacia e della correttezza del sistema. In questa tesi si propone di analizzare questa nuova piattaforma, il suo funzionamento, le sue criticità e valutare la sua effettiva efficacia. L'analisi è stata condotta esaminando il traffico di rete, classificando e distinguendo i vari flussi per riuscire ad identificare il traffico illegale.

Come prima fase del progetto, sono stati analizzati i dati presenti nel *database* per determinare le caratteristiche del contesto e delle tecnologie coinvolte, analizzando il

traffico di rete. Successivamente, è stato approfondito lo studio dei dati, esaminando le principali porte di rete e i principali ASN (Autonomous System Number) per identificare quali fossero i più utilizzati e trafficati durante le partite di calcio, e per poter ricostruire determinati tipi di traffico, come quello legale e illegale. Si è proseguito con l'identificazione dei flussi di traffico legale relativi allo *streaming* delle partite di calcio, ottenendo così una *baseline* di riferimento per confrontare i dati successivi. Infine, è stato identificato il traffico illegale associato allo *streaming* delle partite di calcio.

1.2 Struttura del documento

Il secondo capitolo descrive il Piracy Shield, la pianificazione del lavoro, gli obiettivi e le tecnologie utilizzate.

Il terzo capitolo approfondisce il lavoro svolto durante lo stage, e di come sono stati implementati gli esperimenti e le analisi effettuate.

Il quarto capitolo descrive l'efficacia di Piracy Shield, le metriche di valutazione e l'analisi delle sue prestazioni.

Nel quinto capitolo vengono riassunti e interpretati i risultati ottenuti.

Nel sesto capitolo vengono presentate le conclusioni finali, gli obiettivi raggiunti, e viene data una breve descrizione delle conoscenze acquisite.

Riguardo la stesura del testo, relativamente al documento sono state adottate le seguenti convenzioni tipografiche:

- gli acronimi, le abbreviazioni e i termini ambigui o di uso non comune menzionati vengono definiti nel glossario, situato alla fine del presente documento;
- per la prima occorrenza dei termini riportati nel glossario viene utilizzata la seguente nomenclatura: *parola*[g];
- i termini in lingua straniera o facenti parti del gergo tecnico sono evidenziati con il carattere *corsivo*.

Capitolo 2

Descrizione del Progetto

In questo capitolo si discuterà di Piracy Shield, di come e perchè è nato, del suo funzionamento e delle problematiche e controversie che ha generato. Inoltre verrà descritta la pianificazione del lavoro, gli obiettivi e le tecnologie e linguaggi utilizzati.

2.1 Piracy Shield

In questa sezione verrà presentata la piattaforma Piracy Shield, spiegando che cos'è, a cosa serve, descrivendone il suo funzionamento e discutendo delle problematiche e delle controversie emerse.
Verrà trattato inoltre anche l'argomento della sua effettiva efficacia, il quale verrà approfondito maggiormente nei capitoli successivi, e di come quindi sia possibie aggirare facilmente tale sistema.

Figura 2.1: Logo di Piracy Sheild [2].

2.1.1 Scopo di Piracy Shield

Piracy Shield nasce dall'esigenza di proteggere i contenuti digitali che vengono trasmessi in *streaming* e che sono vulnerabili alla pirateria, la quale rappresenta, e ha rappresentato per molti anni, una grave minaccia per l'industria dei media e dell'intrattenimento, portando a perdite economiche importanti e intaccando i diritti dei creatori di contenuti. Infatti per quanto riguarda la pirateria audiovisiva di film, serie e sport *live* si stima una perdita di 2 miliardi di euro di fatturato durante il 2023, con circa il 39% della popolazione adulta italiana che ha commesso un atto di pirateria. Per quanto riguarda le trasmissioni sportive, l'incidenza della pirateria risulta essere stabile, crescendo rispetto al 2021, ma diminuendo rispetto al 2022, provocando un danno economico di circa 285 milioni di euro. [6]
A seguito della legge 14 Luglio 2023, n. 93, che è entrata in vigore l'8 Agosto 2023, l'AGCOM ha ricevuto nuovi poteri con il fine di poter contrastare in maniera più efficace ed efficiente le azioni di pirateria *on line* riguardanti gli eventi sportivi trasmessi in diretta, con l'obiettivo di riuscire a bloccarli in maniera tempestiva. Piracy Shield dunque, si propone di tutelare i diritti d'autore tramite il blocco dei Fully Qualified Domain Name (FQDN) e degli indirizzi IP che sono univocamente destinati alla trasmissione illecita dei contenuti protetti. Questo blocco viene effettuato in tempi molto brevi, infatti parliamo di circa trenta minuti a partire dalla segnalazione da parte del titolare che detiene i diritti d'autore, per il tramite di una piattaforma tecnologica dedicata.
La piattaforma Piracy Shield è stata resa attiva a partire dal 1º Febbraio 2024. [2]

2.1.2 IPTV

Prima di iniziare a discutere su come funziona Piracy Shield e come agisce per bloccare i contenuti illeciti, è innanzitutto necessario capire come le persone riescono a guardare i contenuti in *streaming* sulla televisione, come questi vengono trasmessi e come infine vengono piratati.
Per poter vedere i programmi televisivi in *streaming*, si utilizza la tecnologia Internet Protocol Television (IPTV), un sistema che permette di trasmettere sulla rete Internet i segnali televisivi su delle reti informatiche che utilizzano i protocolli TCP/IP[g]anziché utilizzare i segnali TV, TV via cavo o satellitari.
Un servizio IPTV viene generalmente distribuito da un fornitore di servizi come può essere DAZN, Sky o Mediaset, i quali forniscono dei programmi in diretta tramite delle reti Internet Protocol (IP). Tale sistema si distingue dalla Web TV[g], in quanto quest'ultima trasmette i programmi televisivi privilegiando la rapidità (parliamo di comunicazione "*best effort*"). Si distingue anche dalle trasmissioni televisive tradizionali, che effettuano una trasmissione di tipo Multicast[g], dove i segnali dei programmi disponibili fluiscono tutti quanti assieme e sono gli spettatori a seleziore i programmi cambiando il canale TV [28]. Il servizio IPTV, invece, è generalmente usato per ricevere i segnali televisivi tramite connessioni a banda larga. Non privilegia la rapidità, ma la qualità del servizio a favore dell'utente, ed inoltre la modalità di trasmissione varia a seconda dei contenuti: viene utilizzato il protocollo Unicast[g]per i contenuti di tipo *Video-on-Demand* (cioè pre-registrati), dove viene inviato un solo programma alla volta, mentre viene utilizzato il protocollo Multicast per i contenuti in diretta. Il contenuto rimane sulla rete del fornitore e solo il programma selezionato dall'utente finale viene inviato; quando viene cambiato canale, un nuovo flusso viene trasmesso dal *server* direttamente allo spettatore. Per la trasmissione dei contenuti l'architettura di rete

che viene utilizzata è quella distribuita, detta Content Delivery Network (CDN), che permette di distribuire i contenuti su una rete di *server*, in modo da garantire una maggiore qualità di servizio e una maggiore scalabilità. [28] - [25]
È necessario comprendere come avviene la comunicazione tra il fornitore di servizi e l'utente finale. Il servizio viene erogato al cliente tramite l'utilizzo di un modem[g] e di un set-top-box (conosciuto come decoder). Il primo permette di stabilire la connessione con la centrale, gestendo il traffico generale della rete, mentre il secondo riceve lo *stream* digitale con i contenuti video da parte del modem, lo decodifica e lo trasforma in un segnale video che è possibile visualizzare sul televisore. [29] - [30]

(a) Decoder Sky [21] **(b)** Modem Sky [22]

Figura 2.2: Decoder e Modem Sky

Il flusso di comunicazione tra il fornitore di servizi e l'utente viene riassunto così:

1. **Acquisizione dei contenuti:** i contenuti che vengono trasmessi tramite IPTV sono originariamente ottenuti da fonti di trasmissioni legali come per esempio satelliti, cavi oppure torri terrestri.
2. **Conversione e compressione:** i contenuti vengono convertiti in formato digitale e poi compressi per poter essere trasmessi tramite la rete Internet.
3. ***Streaming*:** i contenuti precedentemente compressi vengono inviati sottoforma di pacchetti attraverso una rete a banda larga.
4. **Ricezione:** i pacchetti di dati vengono ricevuti dal modem dell'utente finale, il quale poi lo invia al set-top-box (decoder).
5. **Decodifica:** il dispositivo decodifica i pacchetti ricevuti e li converte in un segnale che può essere interpretato e visulizzato sulla TV.

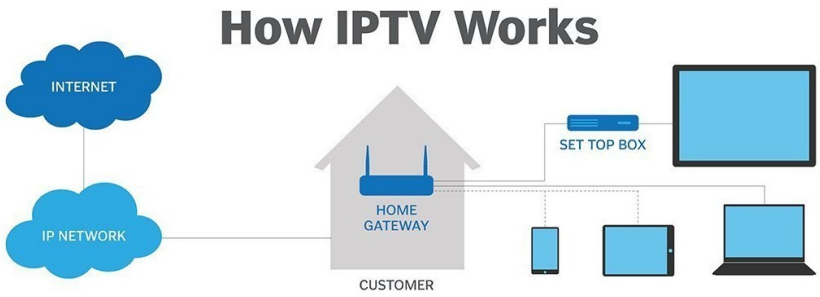


Figura 2.3: Esempio di un flusso IPTV [25]

La descrizione fornita precedentemente illustra il funzionamento di un servizio IPTV. Con queste informazioni, è possibile comprendere come avvenga la pirateria dei contenuti.
Di seguito viene spiegato il processo grazie al quale i pirati digitali riescono ad acquisire, ritrasmettere e ridistribuire i contenuti protetti da *copyright* in maniera illecita:

1. **Acquisizione illegale dei contenuti:** i pirati riescono ad ottenere l'accesso alle trasmissioni o ai contenuti televisivi da delle fonti legali, adottando delle strategie fraudolente, come per esempio:
   - **Abbonamenti legittimi:** da parte dei pirati vengono sottoscritti abbonamenti ai servizi *streaming* come DAZN o Sky, attraverso carte di credito clonate oppure con identità false. Attraverso questo metodo ottengono un gran numero di set-top-box legali che utilizzano poi per ritrasmettere i contenuti.
   - **Acqusizione da fonti legali:** i pirati impiegano dei decoder satellitari che sono stati precedentemente modificati, in modo tale da riuscire a catturare i segnali televisivi direttamente dalle trasmissioni satellitari.
   - **Hacking:** alcuni pirati fanno uso di tecniche di *hacking* che gli permettono di manipolare dei dispositivi di ricezione per *bypassare* le protezioni, così da riuscire ad accedere ai contenuti televisivi.
2. **Ritrasmissione:** dopo aver ottenuto i contenuti, i pirati devono convertirtli e comprimerli per poterli trasmettere su Internet, per farlo possono usare:
   - ***Software* di cattura:** utilizzano dei *software* che catturano il flusso video dal dispositivo sorgente (può essere il decoder oppure un *computer*) i quali riescono a comprimere il video in un formato compatibile alla trasmissione via Internet.
   - **Creazione di flussi RTMP/HTTP:** una volta compressi i contenuti, questi vengono convertiti in flussi Real Time Messaging Protocol (RTMP) oppure Hypertext Transfer Protocol (HTTP), per poi essere trasmessi in *streaming*.
3. **Utilizzo di *server* illegali:** i pirati configurano e fanno uso di *server* illegali, che sostanzialmente fungono da ponte, i quali ospitano e poi ridistribuiscono i flussi dei contenuti. In genere questi *server* sono situati in paesi dove le leggi sulla pirateria sono più permissive. Il modo in cui operano è il seguente:
   - ***Server* di *streaming*:** vengono impiegati dei *server* con elevata capacità di banda, in grado di mantenere e gestire l'accesso di tanti utenti in contemporanea.
   - **Protezione e offuscamento:** per evitare di essere trovati, i pirati fanno uso di Virtual Private Network (VPN) per nascondere e offuscare il traffico.
4. **Distribuzione ai clienti:** gli utenti finali acquistano degli abbonamenti e dei decoder illegali (generalmente sono dei decoder AndroidTV facilmente modificabili) che collegati alla televisione permettono di vedere i programmi in maniera fraudolenta. Questo processo di distribuzione include:
   - **Vendita degli abbonamenti:** i clienti acquistano degli abbonamenti a prezzi ridotti rispetto a quelli ufficiali, i quali gli danno accesso ad un gran numero di canali televisivi e trasmissioni tv.

- ***File* M3U e codici:** gli abbonamenti vengono erogati sotto forma di *file* Moving Picture Experts Group Audio Layer 3 (M3U) o di codici. L'utilizzo che si fa dei *file* M3U è quello di creare delle liste di Uniform Resource Locator (URL) dei flussi video che vengono poi caricati in applicazioni IPTV. I codici, invece, sono delle chiavi o credenzali che vengono inserite dagli utenti nei loro dispositivi per avere accesso ai contenuti.

(a) File M3U **(b)** Decoder AndroidTV

Figura 2.4: Esempio di un file M3U con gli URL e di un decoder AndroidTV

5. **Decodifica:** gli utenti usano questi decoder illegali che sono stati configurati in modo tale da ricevere i flussi IPTV dai *server* dei pirati. Funzionano nel seguente modo:

   - **Decoder modificati:** il loro utilizzo principale sarebbe quello di permettere l'accesso ai *server* illegali. Inoltre spesso vengono eseguite delle applicazioni IPTV che permettono ai clienti di caricare i *file* M3U o i codici per accedere ai contenuti.

La tecnologia IPTV non è di per sé illecita; tuttavia, è l'uso improprio e scorretto di tale tecnologia che costituisce una violazione della legge.

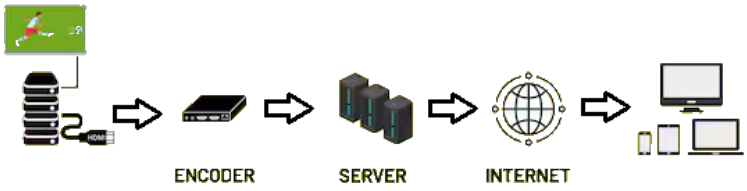


Figura 2.5: Esempio di un flusso IPTV pirata [10]

2.1.3 Funzionamento di Piracy Shield

Piracy Shield è stato sviluppato dall'azienda italiana SP Tech Legal. Come già detto in precedenza, l'obiettivo principale è quello di proteggere i contenuti digitali trasmessi in *streaming*, in particolar modo le dirette sportive. Prima che lo scudo anti pirateria fosse introdotto, il blocco dei siti illegali avveniva in 4-5 giorni, rendendo di fatto inutile la disabilitazione dal momento in cui l'evento sportivo era già terminato. Con Piracy Shield, invece, il blocco avviene in tempi molto brevi, parliamo di circa 30 minuti.
Il funzionamento ed il modo in cui agisce tale piattaforma è abbastanza semplice. Se un determinato sito sta trasmettendo senza autorizzazione una diretta sportiva, l'azienda privata che detiene i diritti d'autore ha la possibilità di inoltrare a Piracy Shield delle segnalazioni contenenti i seguenti dati:

- **FQDN:** nome completo del dominio/FQDN del sito da bloccare (*www.sitopirata.com* è un esempio di FQDN).
- **IP:** gli indirizzi IPv4 e IPv6 del sito da bloccare. (8.8.8.8 è un esempio di indirizzo IPv4 mentre 2001:4860:4860::8888 è un esempio di indirizzo IPv6).
- **Prova:** una prova digitale che attesti la violazione del diritto d'autore e la motivazione per cui stanno chiedendo l'oscuramento. [24]

Dopo che l'azienda privata ha inviato la segnalazione contenente queste informazioni, Piracy Shield creerà un ticket[g]e il sito fraudolento verrà inserito in un'apposita lista. Successivamente il ticket generato verrà inviato ai fornitori di servizi Internet (ovvero i *provider*) che avranno l'obbligo di bloccare il sito illegale entro mezz'ora dalla segnalazione. Siccome questo blocco avviene in maniera totalmente automatica e quindi non c'è nessun controllo di natura umana, per prevenire errori è stata creata una lista autorizzata (detta *whitelist*) in cui sono presenti siti che non possono essere bloccati o oscurati per nessun motivo.
Dopo il blocco del sito, quest'ultimo verrà sottoposto ad un controllo per verificare se ci sia stato effettivamente l'uso di sistemi illegali. In caso di riscontro positivo, chi è dietro la gestione del sito illecito dovrà bloccarlo, altrimenti interverrà direttamente l'AGCOM. Nel caso invece i gestori del sito ritengano ci sia stato un errore, possono presentare un ricorso. [17]

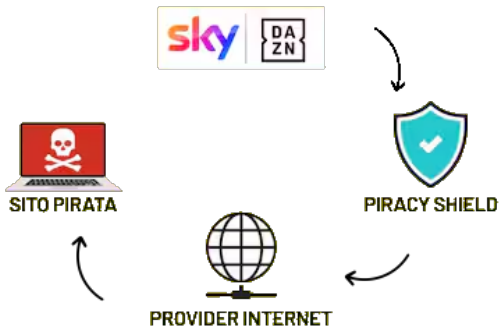


Figura 2.6: Esempio di come agisce Piracy Shield [10]

2.1.4 Problematiche e Controversie

Poco tempo dopo l'implementazione ed entrata in vigore di Piracy Shield, sono emerse varie problematiche e controversie riguardanti le sue modalità di funzionamento e di blocco dei siti, provocando un gran impatto nel panorama digitale italiano.
Uno dei principali problemi è stato quello di come i siti vengono oscurati. Infatti, il blocco avviene attraverso l'inibizione dell'indirizzo IP, dove l'operatore di telecomunicazioni impedisce l'accesso verso uno specifico indirizzo, bloccandone il traffico. Questa soluzione risulta essere difficile da applicare in maniera sicura ed efficiente, dal momento in cui un singolo indirizzo IP può contenere diversi servizi. Attualmente, è comune che su Internet un singolo indirizzo IP possa ospitare diversi siti *web* su un unico *server*. Quindi, quando si decide di effettuare il blocco di un indirizzo IP, è di fondamentale importanza assicurarsi che non sia condiviso da altri servizi legittimi oltre a quello che si intende oscurare. Il problema risiede nell'impossibilità di garantire con assoluta certezza che un indirizzo IP sia utilizzato esclusivamente da un singolo servizio e non sia, quindi, condiviso con altri. Di conseguenza, si rischia di danneggiare ingiustamente quei servizi o siti *web* che, senza saperlo, condividono lo stesso indirizzo IP con il sito illegale.
Oltre a questa problematica nata da questa caratteristica degli inidirizzi IP, c'è anche il problema rappresentato dal crescente utilizzo delle CDN, le quali sono in grado di ospitare un gran numero di siti *web* utilzzando gli stessi indirizzi IP. Questo vuol dire che nel caso in cui accidentalmente dovesse venir bloccata una CDN, c'è il rischio di causare problemi a dei servizi *web* legittimi che la utilizzano. Ed infatti, nei primi giorni di attività di Piracy Shield si sono verificati diversi incidenti dovuti proprio a questo problema. Come per esempio:

- il 15 febbraio 2024 l'indirizzo IP "104.166.170.62" è stato bloccato, provocando il blocco totale all'accesso dei servizi "www.cloud4c.com" e "onsole.zenlayer.com", i quali sono stati costretti a cambiare l'indirizzo IP per poter tornare *online*, dal momento in cui ad oggi il blocco non è stato ancora ritirato. [24]
- il 24 febbraio 2024 la CDN Cloudfare ha subito il blocco del suo indirizzo IP "188.114.97.7". Questo ha provocato oscuramento dei siti e dei servizi che venivano serviti tramite tale IP da parte della CDN su cui si appoggiavano per la loro distribuzione. Il blocco è stato poi successivamente ritirato. [24]
- un altro caso è stato quello dell'Associazione Volontari Carcere di Bologna, il cui corretto funzionamento del sito *web* è stato interrotto.

Questi sono alcuni degli incidenti più noti, ma è possibile che ce ne siano stati altri. Il blocco ingiusto ed errato di questi servizi è dovuto anche alla mancanza di verifiche da parte di enti pubblici sulla correttezza e validità delle segnalazioni dei siti. Questo perché le aziende private che detengono i diritti d'autore degli eventi, utilizzando Piracy Shield, possono inibire l'accesso ai siti presumibilmente pirata, senza che vi sia un monitoraggio da parte di un'autorità competente per verificare la legittimità delle segnalazioni. L'unica verifica effettuata dalla piattaforma è quella di controllare che il ticket inviato dalla società privata sia compilato correttamente, ovvero che i dati richiesti siano stati inseriti in modo adeguato. Non viene, tuttavia, eseguita alcuna valutazione da parte di AGCOM o di altri enti per accertare la fondatezza delle segnalazioni. Il rischio di un errore di natura umana è presente e con uno strumento così potente capace di bloccare un sito o un servizio in pochi minuti, è necessario che ci sia un controllo. In questo scenario è importante anche considerare il concetto

di net neutrality, il quale afferma che tutti i dati in rete devono essere trattati allo stesso modo, e che quindi un azienda o impresa privata non dovrebbe mai avere il potere di poter limitare, attraverso blocchi o oscurazioni, l'accesso ai siti *web*. Infatti i *provider* di servizi Internet non dovrebbero mai effettuare azioni che possono andare a danneggiare la net neutrality. Pertanto, è importante notare che, sebbene Piracy Shield abbia l'intento di proteggere i diritti d'autore, può comunque rappresentare un sistema di censura.
Un'altra problematica riguarda la trasparenza con cui vengono gestite le oscurazioni. Nel sito ufficiale dell'AGCOM è presente soltanto una lista contenente il numero degli indirizzi IP bloccati, ma non viene specificato in alcun modo il soggetto che ha fatto la segnalazione, la motivazione del blocco, né tantomeno l'indirizzo IP bloccato. Oltre alla gestione delle oscurazioni, vi è anche un problema di trasparenza per quanto riguarda i reclami: un'azienda che ha subito un blocco errato può fare ricorso entro 5 giorni, ma, dal momento che non viene notificata l'oscurazione e la lista pubblicata riporta solo il numero degli indirizzi IP bloccati senza ulteriori dettagli, risulta impossibile per chi ha subito un blocco ingiusto poter presentare un ricorso. [24]
Infine per concludere il quadro delle problematiche di Piracy Shield, ci sono alcune vulnerabilità che rendono il sistema facilmente aggirabile. Nel marzo del 2024 il codice sorgente della piattaforma è stato reso pubblico attraverso GitHub[g], permettendo a chiunque di poter visionare e analizzare il codice, ricercando eventuali falle. Infatti, all'interno del codice è stata trovata una funzione che dovrebbe essere ciò che definisce la *whitelist* del sistema, in cui è possibile trovare tutti quei siti che non devono essere bloccati. Di seguito è riportata la funzione citata:

Listing 2.1: Codice di Piracy Shield [11]

```
    def check_unwanteds(self, value):
        result = self.whois.get_text(value)

        result = result.lower()

        if 'cloudflare' in result or 'namecheap' in result or '
            amazon' in result or 'google' in result:
            return True

        return False
```

Questa funzione controlla se nel nome del dominio del sito è contenuta una delle seguenti stringhe, come "cloudflare", "namecheap", "amazon" o "google". Se una di queste stringhe è presente, il sito non verrà bloccato. Questo però implica che un sito pirata può intenzionalmente inserire all'interno del proprio nome di dominio una di queste stringhe per evitare di essere oscurato.

Inoltre, è stato scoperto tramite la analisi svolte durante lo stage, che per poter visionare e navigare un sito o un servizio che è stato oscurato da Piracy Shield è possibile farlo attraverso l'utilizzo di una VPN, uno strumento facile e accessibile a tutti.

2.2 Tecnologie e Strumenti utilizzati

Per il progetto sono stati impiegati diversi strumenti e tecnologie che hanno svolto ruoli importanti nell'analisi dei dati e nella gestione delle informazioni. Gli strumenti

selezionati hanno permesso di sviluppare codice, eseguire *query*, visualizzare dati e garantire accesso sicuro alle risorse.

Visual Studio Code

Visual Studio Code è un *editor* e un ambiente di sviluppo creato da Microsft, risulta essere leggero e potente, adatto alle svariate esigienze degli sviluppatori. Ha un interfaccia intuitiva, personalizzabile e facile da utilizzare che permette agli utenti di scrivere codice in una vasta gamma di linguaggi di programmazione. Risulta avere un gran numero di funzionalità come per esempio il completamento automatico, il *debug* integrato, e tante estensioni che permettono di ampliare le sue capacità di base. Per questi motivi Visual Studio Code è un programma popolare nell'ambito dello sviluppo *software*.
Durante lo stage è servito come ambiente di sviluppo per fare *data analysis* di grandi quantità di dati, in particolare per analizzare i dati provenienti dal *database* ClickHouse. È stato usato come *editor* per scrivere codice Python e per scrivere *query* SQL.

Figura 2.7: Logo di Visual Studio Code [18]

Jupyter Notebook

Jupyter Notebook è una piattaforma *open-source* che da la possibilità agli utenti di creare e condividere in maniera facile dei documenti che contengono codice, equazioni, grafici e testo narrativo come Markdown[g]. È uno strumento molto utizzato nell'ambito scientifico soprattutto per l'analisi dei dati, la visualizzazione e la documentazione. La sua caratteristica e funzionalità più importante è la possibilità di eseguire il codice in dei blocchi, conosciuti come "celle", che possono contenere un'ampia gamma di linguaggi di programmazione.
Durante lo stage è stato usato come estensione integrata dentro Visual Studio Code per poter scrivere codice Python, visualizzare i grafici, e scrivere appunti in markdown.

Figura 2.8: Logo di Jupyter Notebook [16]

ClickHouse

ClickHouse è un sistema di gestione di *database* di tipo colonnare *open-source* che viene largamente usato per l'analisi di grandi quantità di dati in tempo reale. Questo *database* è noto per la sua velocità e scalabilità che lo rendono la soluzione ideale quando si lavora con *query* ad alte prestazioni su dataset di dimensioni molto grandi. ClickHouse utilizza una struttura colonnare, dove i dati sono organizzati a colonne invece che a righe, così da poter ottimizzare le operazioni di lettura e aggregazione. La sua architettura è di tipo distribuita e consente di scalare orizzontalmente su cluster[g] di *server*, garantendo così prestazioni elevate anche con una crescente mole di dati.
È stato scelto come *database* proprio perchè il lavoro da svolgere durante lo stage prevedeva l'analisi di grandi quantità di dati, e ClickHouse si è rivelata la soluzione migliore per poter effettuare *query* complesse, che non richiedevano utilizzo di join[g] tra varie tabelle dal momento in cui si lavorava direttamente su dati presenti in singole tabelle con milioni di righe.

Figura 2.9: Logo di ClickHouse [7]

FortiClient VPN

FortiClient VPN è un'applicazione sviluppata da Fortinet che fornisce delle funzionalità di connessione VPN sicure per dispositivi terminali. Questo *software* viene utilizzato per creare delle connessioni sicure tra utenti remoti e la rete dell'azienda, permettendo di accedere a tutte le risorse in maniera sicura e protetta da qualsiasi luogo. FortiClient VPN garantisce la protezione dei dati che transitano attraverso l'utilizzo di crittografia avanzata permettendo quindi di mettere in sicuro le informazioni che passano da potenziali minacce informatiche.
Durante lo stage è stato utilizzato FortiClient VPN insieme a Putty per poter accedere alla rete e alle risorse di VSIX, in particolare al *database* ClickHouse, contenente tutti i dati su cui bisognava lavorare.

Figura 2.10: Logo di FortiClient VPN [9]

Putty

PuTTY è un *client* terminale *open source* che supporta vari protocolli di rete come Secure Shell (SSH) e Telnet[g]. Questo programma viene utilizzato per creare e stabilire

connessioni sicure e remote a dei *server* e dispositivi di rete, dando la possibilità agli utenti di amministrare sistemi e eseguire comandi e operazioni su delle macchine remote. PuTTY grazie alla sua semplicità e affidabilità fornisce un accesso sicuro attraverso reti non protette.
Durante lo stage è stato utilizzato Putty insieme a FortiClient VPN per poter accedere alla rete e alle risorse di VSIX, in particolare al *database* ClickHouse.

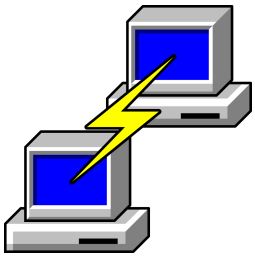


Figura 2.11: Logo di Putty [19]

Python

Durante lo stage come linguaggio di programmazione è stato utilizzato Python, insieme alle sue due librerie principali per *data analisys*, ovvero *Pandas* e *Matplotlib*.
Python è un linguaggio di programmazione ad alto livello, interpretato, famoso per avere una sintassi semplice e facilmente leggibile, che permette uno sviluppo facile e rapido di applicazioni. Questo linguaggio è notevolmente utilizzato per l'analisi dei dati grazie anche alla sua ampia scelta di librerie dedicate.
La liberia Matplotlib è utilizzata per la creazione di grafici e visualizzazioni 2D, con un ampia gamma di personalizzazioni e tipologie di grafici da cui scegliere.
Pandas invece è una libreria fondamentale per fare *data analisys* in Python, dal momento in cui dispone di strutture dati flessibili come i *DataFrame* e le *Series*, che permettono di manipolare e analizzare i dati in maniera efficiente.

Figura 2.12: Logo di Python [20]

Google Drive

Durante lo stage è stato utilizzato Google Drive per poter condividere appunti, documenti, *file* e materiale con i ricercatori che lavoravano sullo stesso progetto.

Google Drive è un servizio di *cloud storage* sviluppato da Google, che permette di salvare, condividere e sincronizzare *online* documenti e *file* in maniera semplice e veloce. Il servizio è accedibile da qualsiasi dispositivo che può connettersi ad Internet, permetendo di poter lavorare in maniera collaborativa, in tempo reale, insieme ad altre persone su documenti, fogli di calcolo e presentazioni le quali sono applicazioni integrate nel Google Workspace, come per esempio Documenti Google e Fogli Google.

Figura 2.13: Logo di Google Drive [14]

Documenti Google e Fogli Google

Documenti Google è un appicazione che permette agli utenti di creare, modificare e condividere documenti in tempo reale con altre persone, permettendo così la collaborazione. Risulta molto comodo e utile grazie alle sue funzionalità come il controllo di versioni, commenti e suggerimenti di modifica.
Durante lo stage è stato utilizzato per scrivere appunti, documenti e per condividere materiale con i ricercatori.

Fogli Google è un applicazione di fogli di calcolo che permete agli utenti di creare, modificare e condividere fogli di calcolo in tempo reale con alre persone, permettendo così la collaborazione e l'analisi dei dati. Risulta molto utile dal momento in cui offre funzionalità come formule, grafici, filtri, *pivot table* e molto altro.
Durante lo stage è stato utilizzato per creare fogli di calcolo, analizzare i dati che erano stati ottenuti dall'interrogazione del *database* e condividere i risultati con i ricercatori.

(a) Logo Documenti Google [12]

(b) Logo di Fogli Google [13]

Figura 2.14: Documenti Google e Fogli Google

2.3 Obiettivi dello studio

Gli obiettivi dello stage erano diversi e riguardavano principalmente l'analisi dei dati raccolti durante un determinato periodo di tempo, con lo scopo di identificare e analizzare i flussi di traffico passanti per lo snodo VSIX. In particolare, l'obiettivo era identificare i flussi di rete riconducibili a trasmissioni in *streaming* di partite di calcio, riuscendo a distinguere tra flussi legittimi e flussi pirata.
Successivamente, in base ai risultati raccolti dalle analisi svolte, è stato necessario valutare l'efficacia di Piracy Shield, ovvero se fosse riuscito a bloccare o meno i siti che trasmettevano illegalmente gli eventi sportivi.
Di seguito vengono riportati gli obiettivi.

ID	Descrizione
Obiettivi Obbligatori	
O01	Conoscenza del dominio dell'analisi di traffico di rete
O02	Conoscenza degli studi effettuati in precedenza nello stesso scenario
O03	Progettazione di un ambiente per l'esecuzione di esperimenti
O04	Implementazione di un ambiente per l'esecuzione di esperimenti
O05	Capacità di implementare i *feedback* forniti durante gli incontri
O06	Esecuzione degli esperimenti e recupero dei risultati
Obiettivi Desiderabili	
D01	Design degli esperimenti necessari allo studio
D02	Analisi approfondita dei risultati degli esperimenti
D03	Documentazione generale degli ambienti di sperimentazione
Obiettivi Facoltativi	
F01	Espansione esperimenti ad altri domini oltre al riconoscimento degli *streaming* illegali
F02	Collaborazione alla redazione di un *paper* scientifico sui risultati ottenuti

Tabella 2.1: Elenco degli obiettivi obbligatori, desiderabili e facoltativi

2.4 Pianificazione del Lavoro

Lo stage si è svolto nel periodo compreso tra il 24 giugno 2024 e il 16 agosto 2024, per un totale di 320 ore di lavoro suddiviso in 8 settimane *full-time*. Le prime due settimane sono state dedicate alla formazione e all'acquisizione delle competenze necessarie per il progetto. Gli esperimenti sono stati condotti nelle ultime sei settimane del periodo di stage. Questi sono stati eseguiti da remoto, principalmente all'interno del laboratorio dei tesisti, ma anche occasionalmente a casa. La tabella 2.2 riporta la pianificazione del lavoro svolto durante lo stage, suddiviso in attività e durata in ore di ciascuna attività.

Durata in ore	Descrizione dell'attività
70	**Background**
25	*Formazione sulle tecnologie adottate*
20	*Formazione sulla metodologia sperimentale*
25	*Studio e ricerca delle tecniche di identificazione e classificazione dei flussi di rete*
120	**Progettazione degli esperimenti**
35	*Analisi del problema e degli esperimenti pregressi*
35	*Progettazione della piattaforma e codice per gli esperimenti*
25	*Consolidamento iterativo della piattaforma*
25	*Stesura documentazione relativa ad analisi e progettazione*
100	**Esecuzione esperimenti**
25	*Preparazione degli ambienti per esperimenti*
50	*Esecuzione esperimenti e raccolta dati*
25	*Analisi dei risultati*
30	**Stesura relazione finale e documentazione**
Totale ore	**320**

Tabella 2.2: Tabella della pianificazione del lavoro

2.5 Variazioni rispetto alla Pianificazione e agli Obiettivi

Dal momento in cui non ci sono state variazioni significative durante lo svolgimento dello stage, non è stato necessario modificare né gli obiettivi né la pianificazione del lavoro. In particolare, tutti gli obiettivi obbligatori sono stati pienamente soddisfatti. Questi obiettivi includevano l'analisi approfondita dei flussi di rete, la valutazione dell'efficacia di Piracy Shield, e la realizzazione di esperimenti pratici per distinguere tra traffico legittimo e pirata. Tra questi, l'analisi dei flussi di rete ha rappresentato l'attività centrale e ha richiesto il maggior impegno temporale. Questo processo è stato cruciale per ottenere una comprensione dettagliata dei dati e per garantire risultati accurati nella classificazione del traffico.

Capitolo 3

Implementazione e Sperimentazione

Questo capitolo approfondisce il lavoro svolto durante lo stage, descrivendo le attività di implementazione degli esperimenti e delle analisi svolte sui dati. Nello specifico verranno presentate le analisi del traffico di rete e dei flussi video, discutendo dell'identificazione del traffico legale e illegale.

3.1 Analisi del Traffico

Durante le prime settimane di stage, per familiarizzare con le tecnologie e gli strumenti utilizzati, nonchè per prendere confidenza con i dati presenti nel *database*, sono state condotte delle analisi sul traffico di rete. Il *database* contiene dati raccolti nel periodo dal 10 al 13 maggio 2024, durante il quale si sono svolte dieci partite di calcio della lega *Serie A*. La base di dati ha al suo interno cinque tabelle contenenti i dati relativi al traffico di rete collezionato durante queste giornate. Tutti gli indirizzi IP, per questioni di *privacy*, sono stati anonimizzati.
Le tabelle sono:

- ***as_info***: contiene informazioni sugli ASN. Questa tabella è utilizzata per associare gli AS ai loro rispettivi proprietari (organizzazioni) e categorie, permettendo così di analizzare il traffico di rete in base all'origine e alla destinazione degli AS. Risulta utile per riuscire a comprendere meglio le dinamiche della rete e per identificare i principali attori al suo interno.
- ***icmp_packets***: registra i pacchetti ICMP (Internet Control Message Protocol). Questa tabella serve a monitorare il traffico ICMP per identificare problemi di connettività e diagnostica di rete. È essenziale per analizzare i messaggi di errore e i controlli di stato della rete.
- ***nf_streams***: memorizza i flussi di dati di rete, rappresentando sequenze di pacchetti tra una sorgente e una destinazione durante un certo intervallo di tempo. Ogni flusso è identificato in modo univoco e include dettagli su tempi di inizio e fine, volume di dati trasferiti, e protocollo utilizzato. Questa tabella è utile per l'analisi dei flussi di traffico di rete, consentendo di identificare *pattern* di utilizzo, rilevare anomalie ed è essenziale per la gestione della rete.

- ***tcp_packets***: raccoglie i dati relativi ai pacchetti TCP (Transmission Control Protocol). Questa tabella è impiegata per analizzare le comunicazioni TCP, monitorare la qualità delle connessioni e diagnosticare problemi di rete come la perdita di pacchetti, la latenza e la congestione.

- ***udp_packets***: contiene dati sui pacchetti UDP (User Datagram Protocol). Questa tabella è utilizzata per monitorare il traffico UDP, analizzare la distribuzione dei pacchetti e diagnosticare problemi di rete specifici per le applicazioni che richiedono una bassa latenza.

Le analisi presentate in questa tesi si sono concentrate principalmente sulla tabella *nf_streams*, poichè contiene i dati relativi ai flussi di rete, i quali rappresentano una solida base per un'analisi più dettagliata del traffico di rete, con l'obiettivo di identificare il traffico di *streaming* video legale e illegale.
È stato innanzitutto realizzato un grafico che mostrasse l'andamento del traffico di rete durante i periodi di registrazione dei dati, in modo da ottenere una visione d'insieme e comprendere la distribuzione del traffico durante le partite di calcio.
Per ottenere il grafico in figura 3.1, è stato interrogato il *database* con la seguente *query* SQL:

Listing 3.1: query SQL per ottenere i dati del traffico di rete

```
SELECT
    sum(in_bytes) AS total_bytes,
    time_first,
    time_last
FROM nf_streams
GROUP BY
    time_first,
    time_last
```

Questa *query* è progettata per aggregare il traffico di rete in *byte* ("in_bytes") per ciascun intervallo di tempo definito dai campi "time_first" e "time_last". Questo permette di ottenere una visione del volume totale di dati trasferiti durante specifici intervalli di tempo. Il campo "time_first" rappresenta il *timestamp* del primo pacchetto del flusso di rete, risultando utile per identificare l'inizio del periodo di attività del flusso, mentre il campo "time_last" rappresenta il *timestamp* dell'ultimo pacchetto del flusso di rete, utile per identificare la fine del periodo di attività.
Questa *query*, costituisce la base per molte delle analisi che sono state svolte durante il lavoro. Dopo l'esecuzione di questo codice SQL, utilizzando le librerie di *Python*, come *Pandas* e *Matplotlib*, è stato creato il grafico di figura 3.1:

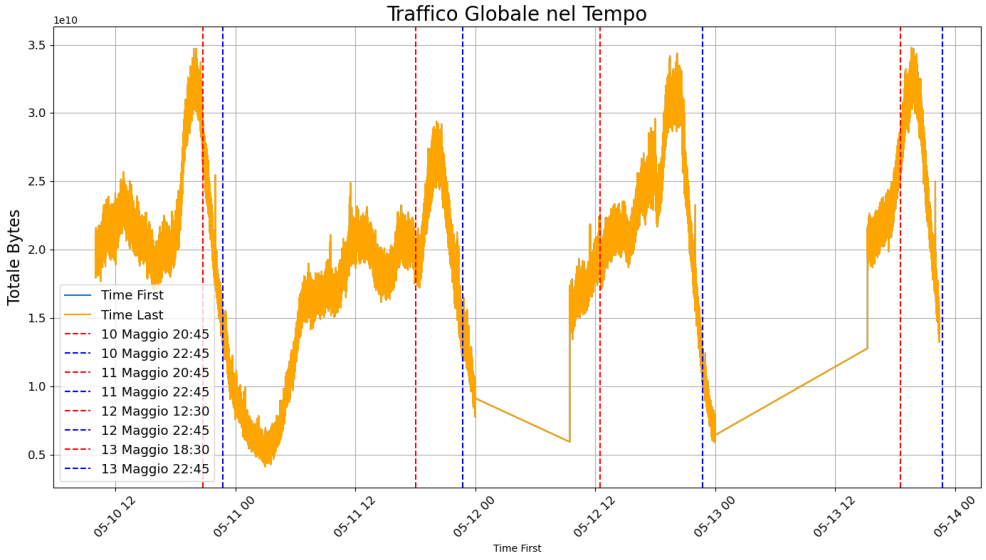


Figura 3.1: Grafico dell'andamento del traffico globale

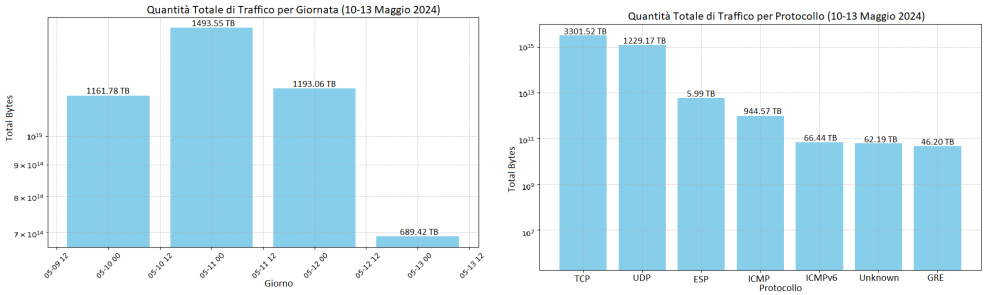
Nel grafico presentato in figura 3.1 l'asse delle ascisse rappresenta il tempo, con date e orari specifici che vanno dal 10 maggio al 13 maggio. È possibile notare come durante le notti del 12 e 13 maggio non ci sia nessuna attività di rete, poichè i dati non sono stati raccolti in quel periodo. L'asse delle ordinate mostra il totale dei *byte* trasferiti, con una scala logaritmica che permette di visualizzare meglio le variazioni significative nel traffico. Inoltre, sono presenti barre verticali tratteggiate di colore rosso e blu, che rappresentano momenti specifici durante il periodo di registrazione: le barre rosse indicano l'inizio delle partite di calcio, mentre quelle blu ne segnano la fine. È evidente che il traffico di rete aumenta significativamente durante gli orari delle partite. Questo incremento può essere attribuito al fatto che molti utenti guardano le partite *online* oppure utilizzano la rete durante questi eventi. Il traffico di rete segue una tendenza giornaliera, con picchi nelle ore serali, un comportamento comune in molte reti, dove l'uso domestico e il tempo libero tendono a concentrarsi nelle ore post-lavorative o scolastiche. Non sorprende, quindi, che i picchi di traffico coincidano con gli orari delle partite di calcio, poichè queste vengono trasmesse in orari serali, quando le persone sono più propense a guardare la TV o a utilizzare la rete. Uno degli obiettivi più importanti dello studio è stato quello di identificare e distinguere i flussi di rete che riguardavano le partite di calcio dal resto del traffico, tema che verrà discusso più avanti.

Uno degli scopi principali è stato quello di analizzare le caratteristiche dei flussi di rete, come la quantità totale di traffico, la durata, la distribuzione dei pacchetti nel tempo e la dimensione dei pacchetti. Questo studio è stato condotto interrogando il *database* e analizzando i dati presenti nella tabella *nf_streams*.

Analisi della quantità di traffico

Le prime analisi condotte riguardano la quantità di traffico di rete globale, ovvero il volume totale di dati trasferiti durante i flussi di rete in tutte le giornate e non solo relativamente agli orari delle partite. Questo è un indicatore importante per valutare

l'attività della rete e per identificare i momenti di picco di traffico.
Lo studio è stato condotto sotto due prospettive principali: identificare la quantità totale di dati trasmessa ogni giorno e analizzare il traffico relativo ai principali protocolli di rete, come ICMP, TCP e UDP.



(a) Quantità traffico per giornata **(b)** Quantità traffico per protocollo

Figura 3.2: Quantità di traffico di rete per giornate e per protocolli

La figura 3.2 mostra per ogni giornata la quantità totale di traffico di rete e la distribuzione del traffico per protocollo. Dal grafico 3.2a risulta che la giornata con il maggior traffico di rete è stata l'11 maggio, con un totale complessivo di 1493,55 TB di dati trasferiti, mentre la giornata con il minor traffico è stata il 13 maggio, con 689,49 TB. Questo risultato non fornisce ancora indicazioni specifiche sul traffico di rete legato alle partite di calcio, ma offre una visione generale dei dati di rete durante i quattro giorni di raccolta.
Mentre dal grafico 3.2b si può notare che il protocollo TCP è quello che trasferisce la maggior quantità di dati, con un totale di 3301,52 TB, seguito da UDP con 1229,17 TB. Sono stati considerati anche altri protocolli, come ICMP, ICMPv6, ESP (utilizzato per la crittografia e l'autenticazione dei pacchetti), e il protocollo GRE (utilizzato per creare tunnel). Il restante dei protocolli è stato aggregato in un'unica categoria chiamata "Unknown" con un traffico complessivo di 62,19 GB.
Il fatto che il secondo protocollo più utilizzato sia UDP suggerisce che una parte significativa del traffico potrebbe essere legata allo *streaming* video (oltre che a quello audio), poichè questo protocollo è comunemente utilizzato per trasmettere *stream* di dati in tempo reale. Questo risultato conferma che i dati collezionati riguardano principalmente flussi video.

Il grafico presente nell'immagine 3.2b mostra l'andamento del traffico di rete per ogni protocollo, evidenziando come questo si distribuisce nel tempo. Nel grafico 3.3, l'andamento relativo al protocollo TCP è molto simile a quello del protocollo UDP, ma il primo risulta essere molto più utilizzato rispetto al secondo e per questo si trova posizionato più in alto nel grafico. Entrambi gli andamenti seguono una tendenza simile al traffico globale presente nell'immagine 3.1, con alcune differenze per quanto riguarda i picchi di traffico. Mentre il traffico del protocollo ICMP risulta completamente diverso rispetto agli altri due protocolli TCP e UDP, questo è dovuto alla minore quantità di dati trasferiti, risultando essere collocato molto più basso nel grafico. Invece il protocollo ESP mostra una somiglianza con l'andamento del traffico globale, anche se meno evidente. Infine per i protocolli sotto la categoria "Unknown", siccome la quantità di dati trasferiti è estremamente bassa, risulta essere quasi invisibile nel grafico.

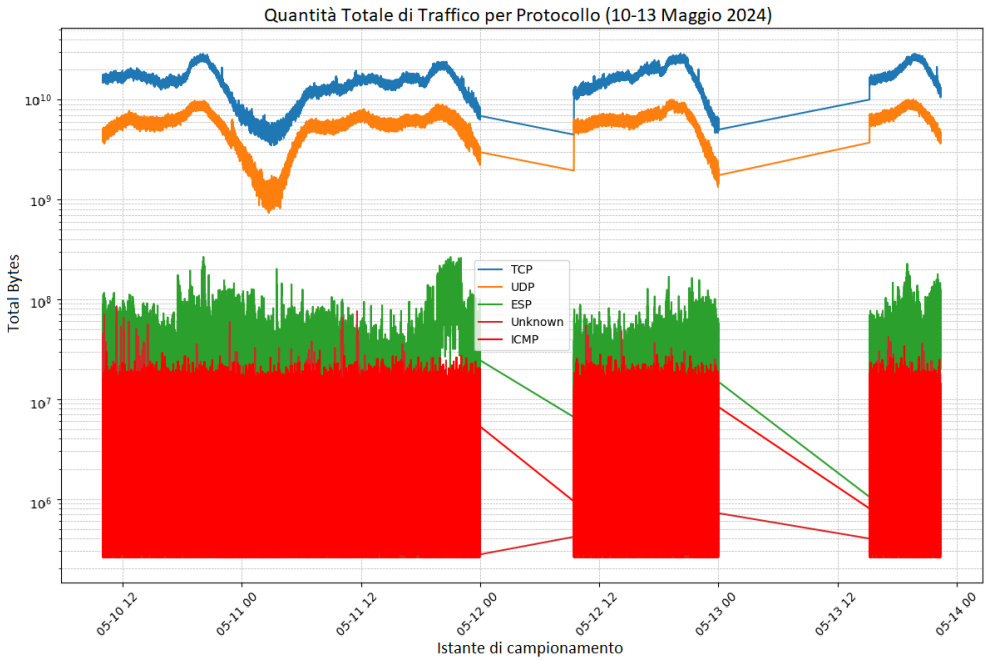
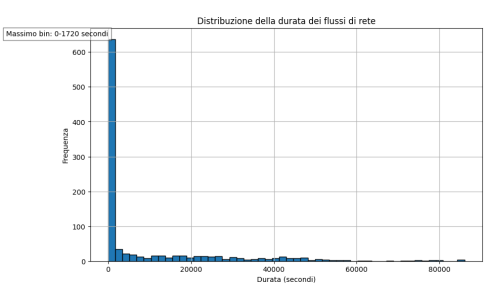


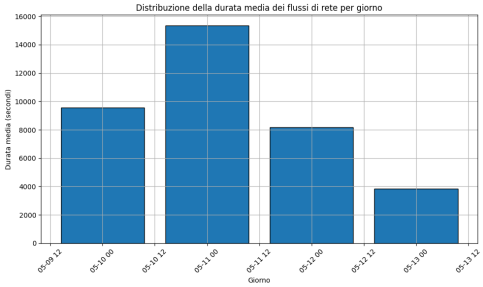
Figura 3.3: Andamento del traffico di rete per protocollo

Analisi della durata di traffico

L'analisi della durata del traffico di rete è un aspetto altrettanto importante per comprendere il comportamento e l'efficienza della rete durante la registrazione dei dati. La durata dei flussi di rete, che rappresenta il tempo durante il quale una connessione è attiva tra due punti o soggetti, può fornire informazioni sulle prestazioni della rete, sui *pattern* di utilizzo e su potenziali problemi di sicurezza. Per questo motivo, l'analisi che è stata condotta si è concentrata nel fornire una panoramica dettagliata sul comportamento dei flussi di rete in termini di durata, evidenziando differenze basate sui protocolli di comunicazione, variazioni giornaliere e distribuzione generale delle durate, prendendo in considerazione la quantità di traffico di rete globale su tutte le giornate e non solo relativamente agli orari delle partite.



(a) Distribuzione della durata dei flussi di rete



(b) Distribuzione della durata media dei flussi

Figura 3.4: Analisi della durata dei flussi di rete

Il grafico 3.4a ("Distribuzione della durata dei flussi di rete") mostra un'istogramma che rappresenta la frequenza delle durate dei flussi di rete. La maggior parte dei flussi ha una durata molto breve, con l'intervallo più popolato tra i 0 e 1720 secondi, il che indica che, generalmente, i flussi di rete tendono a essere brevi. Tuttavia, è presente una coda lunga nella distribuzione, che mostra la presenza di flussi con durate significativamente più lunghe, oltre 60.000 o addirittura 80.000 secondi. Questo fenomeno, noto come "Long Tail", suggerisce che sebbene la maggior parte delle sessioni sia breve, esistono alcuni flussi che si estendono notevolmente nel tempo, potenzialmente a causa di specifiche condizioni di rete o di particolari tipologie di traffico. La durata ridotta dei flussi potrebbe essere influenzata dalla natura delle transazioni o delle sessioni di comunicazione, che spesso sono rapide, come nel caso delle richieste HTTP.
Il grafico 3.4b ("Distribuzione della durata media dei flussi di rete per giorno") illustra come la durata media dei flussi di rete varia nei diversi giorni. È possibile notare una certa variabilità, con alcuni giorni che mostrano durate medie molto più elevate rispetto ad altri. Questa variazione giornaliera può essere dovuta a cambiamenti nel tipo di traffico o nel volume di dati scambiati, influenzati da eventi specifici. Il giorno con la durata media più lunga è stato l'11 maggio, con 15334,13 secondi, mentre il giorno con la durata media più breve è stato il 13 maggio, con 3843,01 secondi, a causa di una minore raccolta dati in quella giornata.

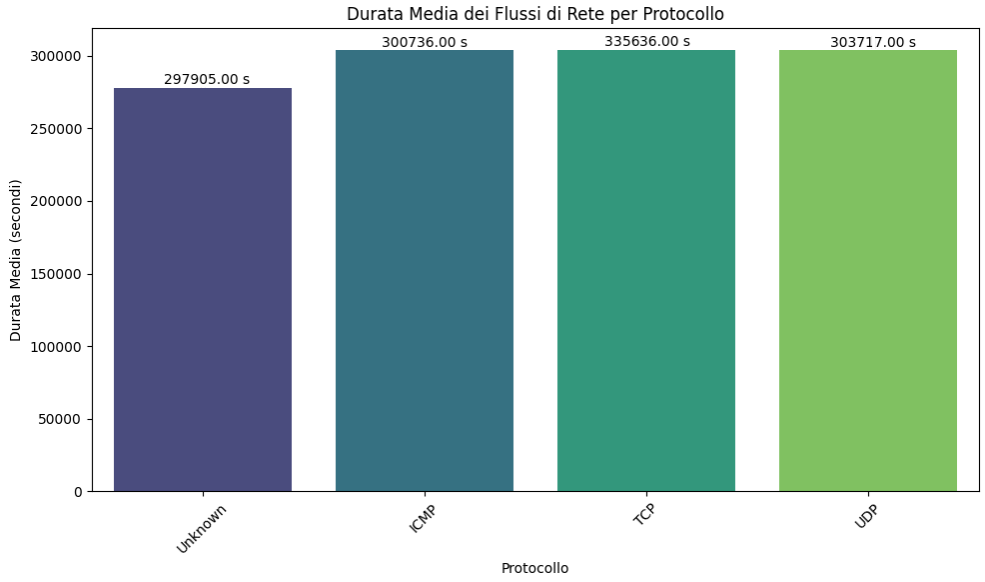


Figura 3.5: Durata media dei flussi di rete per protocollo

Il terzo grafico 3.5 ("Durata Media dei Flussi di Rete per Protocollo") mette a confronto la durata media dei flussi di rete in base al protocollo utilizzato (ICMP, TCP, UDP, "Unknown"). I protocolli mostrano alcune differenze nelle durate medie:

- **ICMP**: protocollo che generalmente viene utilizzato per messaggi di controllo e di errore nella rete, tende ad avere flussi più brevi.
- **TCP**: procotollo noto per il suo meccanismo di controllo della connessione, mostra durate più lunghe, probabilmente a causa della gestione delle sessioni e

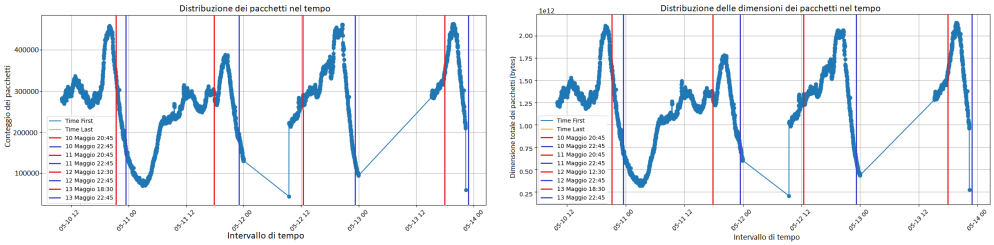
al trasferimento affidabile di grandi volumi di dati.

- **UDP**: protocollo utilizzato per trasmissioni più rapide e meno affidabili, ha durate varie, usato principalmente in applicazioni relative allo *streaming* video e alla trasmissione di dati in tempo reale. Per questo motivo, la durata media dei flussi UDP è più lunga rispetto a quella di ICMP, ma inferiore a quella di TCP.

Queste analisi evidenziano come vari fattori, inclusi il protocollo di comunicazione e le dinamiche di rete giornaliere, possano influenzare la durata dei flussi di rete.

Analisi dei pacchetti: distribuzione e dimensione

Un'altra analisi importante riguarda la distribuzione temporale dei pacchetti e la loro dimensione. I grafici 3.6, 3.7 e 3.8 offrono una panoramica dettagliata su come i pacchetti di rete variano nel tempo, sia in termini di quantità che di dimensione, fornendo informazioni utili per comprendere meglio il comportamento del traffico di rete. Questa analisi permette di identificare *pattern*, anomalie e tendenze nel flusso di dati, offrendo informazioni sulle risorse di rete.

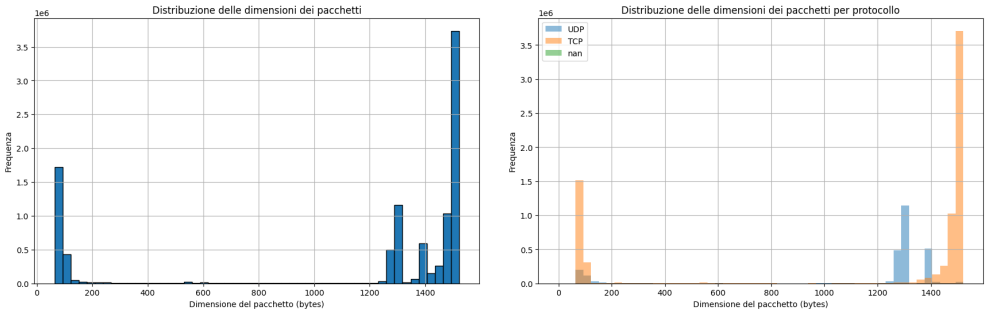


(a) Distribuzione dei pacchetti nel tempo **(b)** Distribuzione delle dimensioni dei pacchetti

Figura 3.6: Distribuzione dei pacchetti e delle loro dimensioni nel tempo

Il grafico 3.6a rappresenta il conteggio dei pacchetti distribuiti nel tempo, calcolando il numero di pacchetti in intervalli di tempo specifici (ogni minuto), mostrando quindi la distribuzione dei pacchetti. Questo tipo di analisi permette di identificare i momenti di elevato traffico di rete e rilevare eventuali *pattern* o anomalie. Dal grafico, si può osservare che il numero di pacchetti varia significativamente nel corso del tempo. Sono presenti picchi evidenti che indicano momenti di elevato traffico, mentre i periodi di bassa attività mostrano un numero ridotto di pacchetti. Come prevedibile, questo grafico mostra un andamento simile a quello del traffico di rete globale dell'immagine 3.1.
Il grafico 3.6b rappresenta come le dimensioni dei pacchetti variano nel tempo, calcolando la somma delle dimensioni dei pacchetti in intervalli di tempo specifici. Questo grafico, similmente all'immagine 3.6a, è importante per identificare i momenti di elevato traffico di rete e capire quando si verificano i picchi di utilizzo. Dall'analisi del grafico, emerge che la dimensione totale dei pacchetti segue un *pattern* ciclico, con picchi significativi durante certi intervalli di tempo. Questi picchi indicano momenti di elevato traffico di rete, probabilmente dovuti a eventi specifici o periodi di utilizzo intensivo. Sebbene i due grafici siano molto simili tra di loro, offrono una visione diversa del traffico di rete, mostrando come i pacchetti si distribuiscono nel tempo e come variano le loro dimensioni.

Lo studio è proseguito con l'analisi della distribuzione delle dimensioni dei pacchetti e della loro frequenza, utile per identificare le dimensioni più comuni dei pacchetti.



(a) Distribuzione delle dimensioni dei pacchetti **(b)** Distrib. dimensioni pacchetti per protocollo

Figura 3.7: Distribuzione delle dimensioni dei pacchetti con protocolli

Il grafico 3.7a mostra la distribuzione delle dimensioni dei pacchetti, considerando i primi dieci milioni di pacchetti nella tabella "nf_streams", poiché la tabella contiene milioni di righe e non è possibile esaminarle tutte. Questa analisi fornisce una visione generale su come variano le dimensioni dei pacchetti nel traffico di rete. La distribuzione delle dimensioni dei pacchetti rivela due principali gruppi: pacchetti molto piccoli e pacchetti di dimensione massima, mentre i pacchetti di dimensione media sono molto pochi. Questo perchè i pacchetti molto piccoli sono spesso utilizzati per messaggi di controllo e segnalazione, come le richieste di connessione o i pacchetti di conferma, mentre i pacchetti di dimensione massima vengono utilizzati per il trasferimento di dati più grandi, come i *file* scaricati o i dati trasmessi durante una sessione di *streaming*. Questa distribuzione è tipica nelle reti, dove i pacchetti di controllo e i pacchetti di dati coesistono, riflettendo la varietà di applicazioni e servizi utilizzati dagli utenti. La presenza di pacchetti di dimensione massima potrebbe indicarte che il traffico di rete raccolto riguarda principalmente flussi video in tempo reale.
Il grafico 3.7b analizza come le dimensioni dei pacchetti variano tra i diversi protocolli, concentrandosi su quelli più importanti ovvero TCP e UDP. Questa analisi permette di capire meglio le caratteristiche specifiche del traffico di rete generato da questi due protocolli. Dal grafico si osserva che i pacchetti TCP tendono a essere più grandi e anche più frequenti rispetto a quelli UDP. Questo è coerente con le caratteristiche dei due protocolli: TCP, essendo orientato alla connessione e garantendo l'affidabilità della trasmissione, tende a raggruppare i dati in pacchetti più grandi per ottimizzare l'efficienza della rete. Al contrario, UDP, essendo un protocollo senza connessione e focalizzato sulla velocità, trasmette spesso pacchetti più piccoli. Tuttavia, il traffico UDP per i pacchetti di piccole dimensioni, risulta essere meno frequente rispetto a quello TCP, coerentemente con i dati del grafico 3.2b, dove si evidenzia che la quantità di dati trasferiti via UDP è circa la metà rispetto a TCP, rendendo sensata questa differenza.

Infine, è stata condotta un'ultima analisi per esaminare come i pacchetti sono distribuiti nel tempo per diversi protocolli, conteggiando il numero di pacchetti per ciascun protocollo in intervalli di tempo specifici. Questa analisi, come le altre fatte fin ora, permette di comprendere quali protocolli di rete sono maggiormente utilizzati e come variano nel tempo. Dal grafico 3.8 emerge chiaramente che i protocolli TCP e UDP sono i più utilizzati, mentre gli altri hanno una quantità di traffico talmente bassa

da risultare quasi invisibili nel grafico. Questo risultato è in linea con le aspettative, poiché TCP è comunemente usato per la maggior parte delle comunicazioni affidabili su Internet, come la navigazione *web* e le *email*, mentre UDP è preferito per applicazioni che richiedono velocità e bassa latenza, come lo *streaming* video o audio.

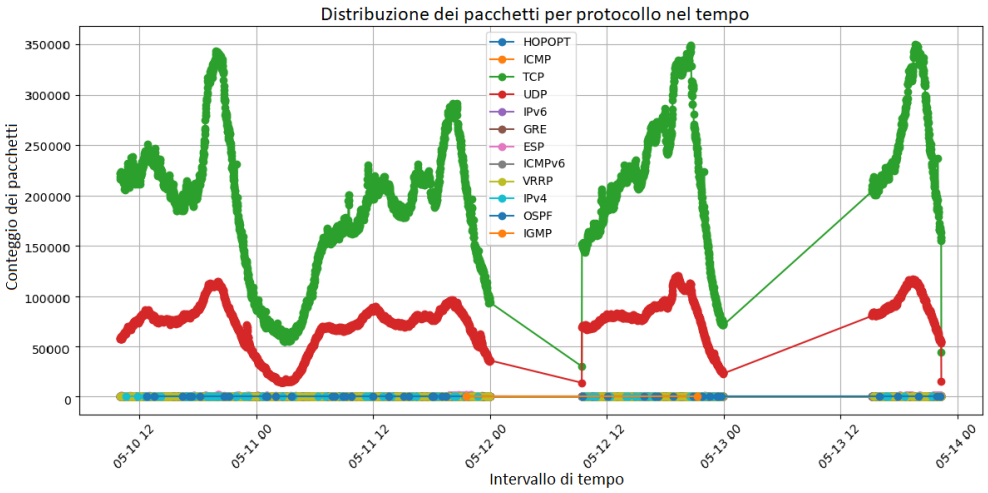


Figura 3.8: Distribuzione dei pacchetti nel tempo per protocollo

Le analisi condotte fin ora, sebbene simili, esaminano casi di studio differenti che, pur fornendo risultati convergenti, offrono una conferma delle osservazioni fatte. È stato esplorato il traffico di rete da molteplici prospettive, analizzandone la quantità, la durata e le caratteristiche dei pacchetti in diverse sfaccettature.

Identificazione dei flussi di rete

Dopo aver acquisito una panoramica generale del traffico di rete nelle analisi precedenti, in cui è emerso che i protocolli TCP e UDP sono i più utilizzati, è ora necessario identificare i flussi di rete specifici.
Infatti, un altro obiettivo principali della tesi è stato quello di identificare e distinguere le varie tipologie di flusso di rete, come ad esempio: flussi di *streaming* video, flussi di *streaming* audio, flussi P2P (Peer-to-Peer) e flussi di *download*. Questo studio è stato condotto interrogando il *database* e analizzando i dati presenti nella tabella *nf_streams*.

L'identificazione dei flussi di rete è stata realizzata principalmente sulla base dei protocolli di comunicazione utilizzati e delle porte di rete coinvolte. Successivamente, per raffinare ulteriormente l'analisi, sono stati considerati anche altri parametri come la durata dei flussi, la loro dimensione e i flag[g] dei pacchetti.
I parametri e i criteri utilizzati per identificare i flussi di rete relativi allo *streaming* video sono stati i seguenti:

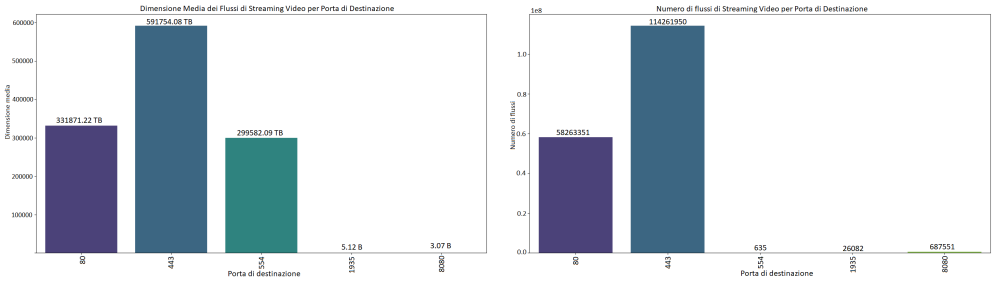
- **Protocollo UDP**: il protocollo UDP è comunemente utilizzato per trasmettere *stream* di dati in tempo reale, come lo *streaming* video e audio. Questo protocollo è usato per le applicazioni che richiedono una bassa latenza e una trasmissione veloce. Nonostante diversi programmi sportivi e servizi di *streaming on-demand* utilizzino il protocollo TCP, la maggior parte delle trasmissioni in tempo reale

sfrutta il protocollo UDP. Pertanto, l'analisi si è concentrata principalmente sui flussi che utilizzano questo protocollo.

- **Porte di rete**: le porte di rete sono numeri di 16 *bit* che identificano i servizi di rete e le applicazioni. Le porte di rete più comuni per lo *streaming* video sono: 80, 8080, 443, 554, 1935.
- **Durata dei flussi**: i flussi di *streaming* video tendono ad avere una durata più lunga rispetto ad altri flussi di rete, dato che le trasmissioni video richiedono un periodo di tempo più esteso per essere completate.
- **Dimensione dei flussi**: i flussi di video in tempo reale tendono ad avere una quantità costante di dati trasferiti, in quanto le trasmissioni video necessitano di un flusso continuo di dati.
- **Flag dei pacchetti**: i pacchetti di *streaming* video tendono ad avere un flag specifico, come il flag "PUSH" o "ACK". Questi flag indicano che il pacchetto è stato inviato o ricevuto con successo.

Sulla base dei criteri sopra elencati, è possibile identificare il traffico di *streaming* video analizzando specifici parametri del flusso di rete. L'utilizzo predominante del protocollo UDP per le trasmissioni in tempo reale, le porte di rete comunemente associate ai servizi di *streaming* video, la durata e la dimensione dei flussi, insieme ai flag specifici dei pacchetti, forniscono una serie di caratteristiche distintive che permettono di isolare i flussi di *streaming* video dagli altri.

I grafici 3.9a e 3.9b mostrano rispettivamente, per porta di destinazione, la dimensione media dei flussi di *streaming* video e il numero di flussi di *streaming* video.



(a) Dim. media flussi video per porta **(b)** Numero flussi video per porta

Figura 3.9: Dimensione media e numero dei flussi video per porta

Da questi due grafici si evince che le porte di rete più utilizzate sono la 443 e la 80. La porta 554 risulta essere la terza per dimensione media dei flussi di *streaming* video, ma è l'ultima per numero di flussi. Questo indica che la porta 554 è utilizzata per flussi di *streaming* video di dimensioni maggiori, ma in numero minore rispetto alle altre porte.

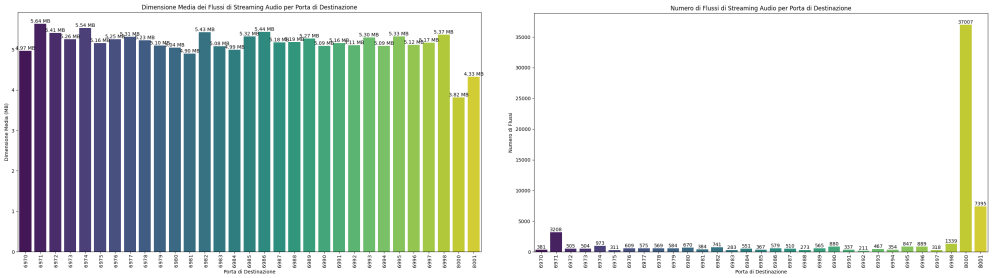
Mentre, per quanto riguarda l'identificazione dei flussi di rete relativi allo *streaming* audio, i parametri e i criteri utilizzati sono stati i seguenti:

- **Protocollo UDP**: il protocollo UDP è comunemente utilizzato per trasmettere flussi di dati in tempo reale, come lo *streaming* video e audio, poichè è adatto alle applicazioni che richiedono bassa latenza e trasmissione veloce.

- **Porte di rete**: Le porte di rete più comuni per lo *streaming* audio sono: 8000, 8001 e, in generale, le porte comprese tra la 6970 e 6999.
- **Durata dei flussi**: i flussi di *streaming* audio spesso hanno una durata prolungata dovuta alla natura continua della trasmissione audio.
- **Dimensione dei flussi**: i flussi di *streaming* audio in tempo reale tendono ad avere una quantità costante di dati trasferiti.

Sulla base dei criteri sopra elencati, è possibile identificare il traffico di *streaming* audio. L'utilizzo prevalente del protocollo UDP per le trasmissioni in tempo reale, l'utilizzo di specifiche porte di rete comunemente associate ai servizi di *streaming* audio, la durata prolungata dei flussi dovuta alla natura continua della trasmissione audio e la dimensione costante dei dati trasferiti sono tutti fattori che, se messi assieme, permettono di isolare il traffico audio dagli altri tipi di traffico di rete.

I grafici 3.10a e 3.10b mostrano rispettivamente, per porta di destinazione, la dimensione media dei flussi di *streaming* audio e il numero di flussi di *streaming* audio.



(a) Dim. media flussi audio per porta **(b)** Numero flussi audio per porta

Figura 3.10: Dimensione media e numero dei flussi audio per porta

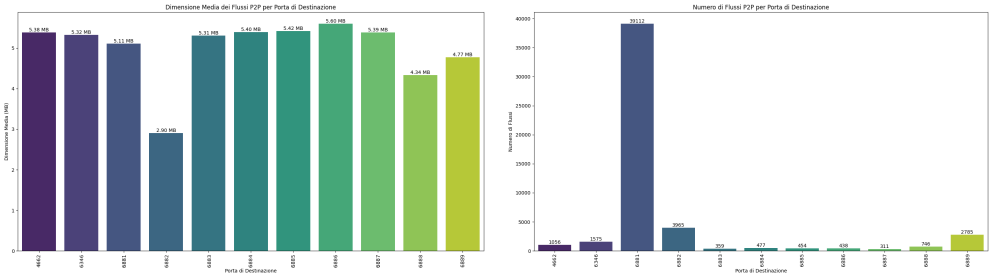
Da questi due grafici emerge che la dimensione media dei flussi audio è molto simile tra le varie porte, mentre il numero di flussi varia notevolmente. La porta 8000 risulta essere la più utilizzata, con 37007 flussi, mentre la porta 8001 è la seconda più utilizzata con 7395 flussi, quasi un quinto in meno rispetto alla porta 8000.

Per quanto riguarda invece il traffico di rete associato ai flussi P2P, questo presenta delle caratteristiche diverse rispetto ad altri tipi di traffico di rete. Il P2P è un tipo di architettura di rete in cui ogni nodo (o "*peer*") è sia un *client* che un *server*, consentendo la condivisione diretta di risorse senza la necessità di un *server* centrale. Questo tipo di traffico viene normalmente utilizzato per condividere *file* di grandi dimensioni, come video, musica, *software* e documenti. Le porte più utilizzate per questo scopo includono la 881, 6882-6889, 4662, e 6346, spesso associate a protocolli P2P famosi come *BitTorrent* ed *eDonkey*. I parametri e i criteri utilizzati per identificare i flussi di rete P2P sono stati i seguenti:

- **Protocolli P2P**: Entrambi i protocolli TCP e UDP possono essere utilizzati per analizzare il traffico. Il protocollo TCP è spesso impiegato per garantire l'affidabilità del trasferimento dei dati, mentre il protocollo UDP può essere utilizzato per velocizzare la trasmissione, accettando un minor grado di affidabilità.

- **Porte di rete**: Le porte comunemente utilizzate per il traffico P2P includono 881, le 6882-6889, la 4662 e la 6346. Queste porte sono specificamente riservate o comunemente usate dai protocolli P2P.
- **Durata dei flussi**: I flussi P2P possono variare notevolmente in durata, spesso rimanendo attivi per periodi prolungati a causa della natura dei trasferimenti di *file* di grandi dimensioni.
- **Volume di dati trasferiti**: Il traffico P2P è caratterizzato da un elevato volume di dati trasferiti, poiché i *peer* scambiano grandi quantità di informazioni tra loro.

I grafici 3.10a e 3.10b mostrano rispettivamente, per porta di destinazione, la dimensione media dei flussi di *streaming* P2P e il numero di flussi di *streaming* P2P.



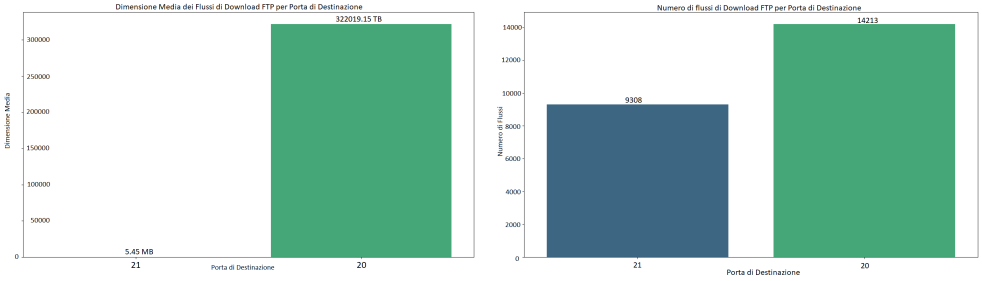
(a) Dim. media flussi P2P per porta **(b)** Numero flussi P2P per porta

Figura 3.11: Dimensione media e numero dei flussi P2P per porta

Infine, per identificare il traffico di rete associato ai *download*, l'attenzione è stata rivolta al protocollo FTP (File Transfert Protocol). Questo protocollo è uno dei più comuni e utilizzati per il trasferimento di *file* su una rete. Esso opera su due porte principali, la porta 20 per il trasferimento dati e la porta 21 per il controllo delle connessioni. La scelta di focalizzarsi sul protocollo FTP è dovuta alla sua diffusione e dalla sua capacità nel gestire trasferimenti di *file* di grandi dimensioni. I parametri e i criteri utilizzati per identificare per identificare questo tipo di traffico sono stati i seguenti:

- **Protocollo FTP**: FTP è progettato per il trasferimento di *file*, rendendolo ideale per analizzare il traffico e identificare i flussi di *download*. Poiché FTP utilizza il protocollo TCP per garantire la consegna affidabile dei dati, l'analisi si è concentrata esclusivamente su questo protocollo.
- **Porte di rete**: Il traffico FTP utilizza principalmente le porte 20 e 21. La porta 21 è destinata al controllo della connessione, mentre la porta 20 è utilizzata per il trasferimento dei dati.
- **Durata dei flussi**: I *download* di file tramite FTP possono variare notevolmente in durata, a seconda delle dimensioni del *file* e della velocità della connessione. Tuttavia, i flussi di *download* tendono a essere più lunghi rispetto ad altri tipi di traffico a causa della quantità di dati trasferiti.
- **Dimensione dei flussi**: I flussi di *download* tramite FTP tendono a trasferire grandi quantità di dati. L'analisi delle dimensioni dei flussi può aiutare a distinguere i *download* di file dai flussi di dati meno intensivi.

I grafici 3.12a e 3.12b mostrano rispettivamente, per porta di destinazione, la dimensione media dei flussi di *download* e il numero di flussi di *download*.



(a) Dim. media flussi di *download* per porta **(b)** Numero flussi *download* per porta

Figura 3.12: Dimensione media e numero dei flussi di *download* per porta

Le informazioni utili che si possono ricavare da questi due grafici evidenziano che la porta 20 risulta essere la più utilizzata, sia per il numero di flussi di *download* sia per la dimensione media dei flussi, con una dimensione media di circa 32 GB. Questo dato è coerente con la funzione della porta 20, destinata al trasferimento dei dati, mentre la porta 21 risulta essere inferiore sia in termini di numero di flussi sia in termini di dimensione media, siccome è adibita al controllo delle connessioni.
Le analisi condotte fin ora hanno rivelato che molte porte mostrano un alto numero di flussi e una dimensione media che si aggira intorno all'ordine dei *megabyte*. Ad esempio, nel grafico 3.10 la porta 8000 presenta 37007 flussi con una dimensione media di 3,82 MB. Al contrario, la porta 21 mostra 14213 flussi con una dimensione media di 32 GB. Questo confronto dimostra chiaramente che i dati sono coerenti con la natura dei flussi di rete. Infatti, i flussi di *streaming* audio, video o P2P tendono ad essere di dimensioni molto più ridotte e più numerosi rispetto ai flussi di *download*, che sono caratterizzati da dimensioni molto maggiori e da un numero inferiore.

Per offire una visione di insieme complessiva delle analisi condotte fino a questo punto, è stato realizzato il grafico 3.13 che sintetizza la quantità di traffico generato dai diversi tipi di flussi di rete analizzati, tra cui lo *streaming* video, lo *streaming* audio, il traffico P2P e i *download* tramite FTP.
Il grafico 3.13 mostra chiaramente come i diversi tipi di flussi di rete contribuiscano al traffico totale:

- **P2P**: Il traffico P2P con un volume di 253.08 GB risulta essere il terzo tipo di traffico più rilevante tra quelli analizzati. Questo tipo di traffico è spesso associato alla condivisione di *file* di grandi dimensioni, come video, musica e *software*.
- **Download**: Il traffico associato ai *download* FTP rappresenta un volume relativamente basso rispetto agli altri tipi di flussi di rete, ma comunque rilevante per specifiche applicazioni professionali e di scambio dati.
- **Streaming Video**: Questo tipo di flusso ha generato un volume significativo di traffico, pari a 89.20 TB. Lo *streaming* video rappresenta una delle principali fonti di consumo di banda, infatti il volume elevato di traffico video conferma che esso rappresenta una parte significativa del traffico di rete.

- **Streaming Audio**: Il traffico generato dallo *streaming* audio è di 273.02 GB. Anche se inferiore rispetto allo *streaming* video, questo tipo di traffico è comunque rilevante.

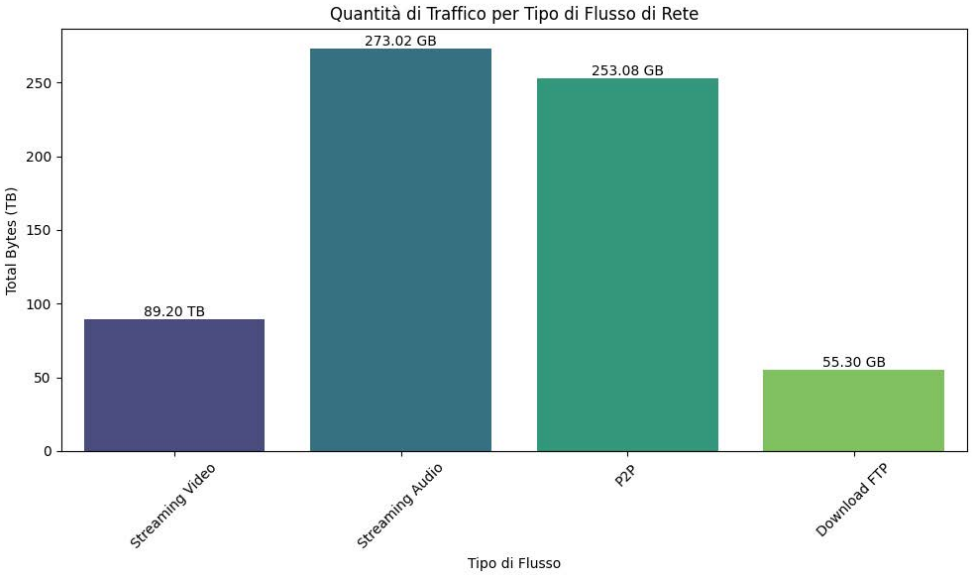


Figura 3.13: Quantità di traffico per flusso di rete

In conclusione, il grafico fornisce una panoramica chiara della distribuzione del traffico di rete, permettendo di identificare i flussi di rete più rilevanti e le relative quantità di traffico generato.

3.2 Identificazione del Traffico Legale

L'identificazione del traffico legale è stata realizzata partendo dall'individuazione delle CDN che detengono i diritti di trasmissione per eventi sportivi, in particolare per le partite di calcio. Le CDN sono infrastrutture molto importanti per la distribuzione efficiente di contenuti in rete. Il loro scopo principale è quello di ridurre la latenza e migliorare la velocità con cui gli utenti finali accedono ai contenuti, attraverso una rete distribuita di *server* e l'ottimizzazione delle risorse di rete. Una CDN è quindi un insieme di *server* distribuiti in maniera strategica in varie località geografiche. Questi *server*, conosciuti come *Edge Servers*, memorizzano in cache[g] i contenuti richiesti dagli utenti, come *file* video, immagini, pagine *web* e altri elementi multimediali. La posizione dei *server* è scelta in modo da essere il più vicino possibile agli utenti finali, riducendo così il tempo di risposta e migliorando la qualità del servizio. Gli *Edge Servers* ottengono i contenuti dai *server* di origine (ovvero gli *Origin Server*), dove questi contenuti vengono originariamente ospitati. Quando un contenuto viene aggiornato o pubblicato, viene inviato dall'*Origin Server* agli *Edge Servers*, che lo memorizzano per le distribuzioni future. I vantaggi principali di utilizzare una CDN sono la riduzione della latenza, l'ottimizzazione delle risorse di rete, la scalabilità e la ridondanza. [27] Siccome le partite di calcio trasmesse durante le giornate di raccolta dati si sono svolte tra il 10 e il 13 maggio 2024, è stato necessario identificare prima di tutto gli emittenti

televisivi che detenevano i diritti di trasmissione per quelle partite, ed in questo caso sono risultate essere Sky e DAZN. Il passo successivo è stato quello di trovare le CDN utilizzate da questi servizi, risultando che Sky e DAZN si affidano principalmente alle CDN dell'azienda Akamai [4] [5].
Per verificare l'accuratezza di queste informazioni, è stato effettuato un controllo sui dati raccolti, più precisamente interrogando la tabella "as_info", che contiene informazioni sugli ASN. Sono stati ricercati tutti quei ASN associati ad Akamai e sono stati identificati 12 ASN che fanno riferimento a questa azienda. Per escludere l'eventualità che Sky e DAZN potessero utilizzare altre CDN oltre ad Akamai, sono stati esaminati anche gli ASN di altre note aziende che offrono servizio di *Contend Delivery Network*, come ad esempio Amazon, Cloudflare e Fastly, riscontrando che ciascuna di queste aziende aveva solo un ASN presente nella tabella "as_info". Questi risultati, coerenti con le fonti citate [4] [5], validano l'ipotesi che Sky e DAZN utilizzino principalmente le CDN di Akamai per poter distribuire le partite di calcio.
Dopo aver identificato le *Contend Delivery Network* utilizzate per trasmettere legalmente le partite di calcio e trovato i corrispondenti ASN, è stato possibile procedere con la ricostruzione del traffico di *streaming* video legale delle partite, ottenendo una *baseline* che successivamente potrà essere utilizzata per poter confrontare il traffico legale con quello presunto illegale, facilitando così l'identificazione di eventuali flussi pirata.
Per ottenere la *baseline* è stato necessario interrogare il *database* e analizzare i dati presenti nella tabella "nf_streams". L'obiettivo principale è stato quello di ottenere gli indirizzi IP che generavano flussi di rete durante le giornate e gli orari delle partite, con le seguenti caratteristiche:

1. La media del traffico degli indirizzi IP deve essere compresa tra 1 e 20 Mb/s. Questo intervallo riflette la gamma di qualità video di una partita di calcio in *streaming*, che può variare da *144p* (che richiede una velocità di connessione circa 1 Mb/s) a *4K* (che richiede una velocità di connessione di circa 20 Mb/s).

2. La media di traffico di per sè non è sufficiente, questo perchè un indirizzo IP potrebbe avere una media di 10 Mb/s generata da un picco di traffico molto alto che non c'entra con lo *streaming* video. Però se dovessimo utilizzare soltanto la media del traffico, poichè un indirizzo IP verrebbe comunque considerato come un dispositivo che stava guardando la partita, è necessario anche considerare la deviazione standard del traffico, così da poter scartare quegli indirizzi IP che non hanno un traffico costante caratteristico dello *streaming* video.

3. Infine, è stato necessario verificare che il traffico degli indirizzi IP durante le partite fosse significativamente diverso dal traffico negli orari al di fuori delle partite. Più precisamente è stato fatto un confronto tra il traffico degli IP durante le partite con il traffico degli stessi IP in due finestre temporali specifiche: due ore prima e due ore dopo le partite. Se il traffico non varia significativamente tra queste finestre, significa che l'attività di rete non è legata alla visione delle partite. Questo perchè se il traffico sia prima che dopo le partire rimane uguale e costante nel tempo, è molto probabile che il dispositivo di rete in questione non fosse sintonizzato per guardare l'evento sportivo, ma piuttosto un'altra attività, come ad esempio il *download* di un aggiornamento.

Il punto 1 è stato facilmente risolto interrogando il *database* con la seguente *query*:

Listing 3.2: query SQL per ottenere gli indirizzi IP che durante le partite di calcio hanno generato un traffico compreso tra 1 e 20 Mb/s

```
    SELECT
    ip, avg(in_bytes) AS avg_traffic_bytes_per_second
FROM
    (
        SELECT
            CASE
                WHEN least(destination_port, source_port) =
                    destination_port THEN destination_ip
                ELSE source_ip
            END AS ip,
        FROM
            nf_streams
        WHERE
            (source_asn = 20940 OR destination_asn = 20940)
            AND (
                (time_first < toDateTime('2024-05-10 22:45:00', '
                    UTC') AND time_last > toDateTime('2024-05-10
                    20:15:00', 'UTC')) OR
                (time_first < toDateTime('2024-05-11 22:45:00', '
                    UTC') AND time_last > toDateTime('2024-05-11
                    17:30:00', 'UTC')) OR
                (time_first < toDateTime('2024-05-12 22:45:00', '
                    UTC') AND time_last > toDateTime('2024-05-12
                    12:00:00', 'UTC')) OR
                (time_first < toDateTime('2024-05-13 22:45:00', '
                    UTC') AND time_last > toDateTime('2024-05-13
                    18:00:00', 'UTC'))
            )
    )
GROUP BY
    ip
HAVING
    avg_traffic_bytes_per_second >= 125000 AND
        avg_traffic_bytes_per_second <= 2500000
```

La *query* 3.2 è la base da cui poi sono state derivate le altre *query* per soddisfare i punti 2 e 3. Il codice 3.2 viene utilizzato per estrarre una lista di indirizzi IP che hanno comunicato con un ASN (in questo caso, quello di Akamai) durante gli orari delle partite di calcio, calcolando il traffico medio per ogni indirizzo. La *query* interna seleziona gli indirizzi IP associati al traffico in cui la porta di destinazione o la porta di origine è la più piccola tra le due. Questo criterio viene utilizzato perhè, generalmente, in una comunicazione tra due porte, quella più piccola rappresenta il *server*, mentre quella più grande rappresenta il *client*. In questo caso ci interessa il flusso *server-client* e non *client-server*, in quanto è quello che trasmette il video. Infine, la *query* esterna raggruppa gli indirizzi IP per eliminare i duplicati, fornendo così una lista unica di IP che hanno avuto una sessione di comunicazione con l'ASN di interesse. La clausola *HAVING* assicura che il traffico medio sia compreso tra 1 e 20 Mb/s (dove 125000 e 2500000 sono rispettivamente 1Mbps e 20Mbps).
Per soddisfare il punto 2 è stato necessario eseguire più passaggi. Dopo aver ottenuto la lista di IP comunicanti con Akamai durante il periodo delle partite, con un traffico medio compreso tra 1-20Mbps (3.2) si è proceduti con il calcolo della media (ovvero μ) del traffico generato da questi indirizzi. Il passo successivo è stato calcolare la varianza

del traffico:

$$\mathrm{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

dove μ è la media dei valori x_i. In questo caso per il calcolo della varianza sono stati usati gli strumenti matematici della libreria *Pandas*.
Infine per ottenere la deviazione standard del traffico, è stato sufficiente applicare la radice quadrata della varianza:

$$\mathrm{std}(X) = \sqrt{\mathrm{Var}(X)}$$

Avendo calcolato la media e la deviazione standard è possibile definire due limiti di traffico:

- Limite superiore: μ + std(X)
- Limite inferiore: μ - std(X)

Se il traffico di un indirizzo IP è all'interno dell'intervallo:

$$[\mu - \mathrm{std}(X)\,;\, \mu + \mathrm{std}(X)]$$

allora possiamo affermare che il traffico è costante e non ha picchi significativi. Questo perché la deviazione standard rappresenta la dispersione dei dati rispetto alla media, e un traffico che rientra in questo intervallo suggerisce una variabilità contenuta intorno alla media, tipica di uno *streaming* video regolare. Se il traffico fosse caratterizzato da picchi significativi, ci aspetteremmo che la deviazione standard sia più alta, indicando una maggiore variabilità e pertanto una distribuzione dei dati più ampia. Pertanto, mantenendo il traffico all'interno di questo intervallo, possiamo affermare con maggiore certezza che l'attività di rete è coerente con lo *streaming* video continuo e non è influenzata da eventi di traffico anomali o sporadici.
Infine per soddisfare il punto 3 è stato necessario confrontare il traffico degli indirizzi IP durante le partite con il traffico degli stessi IP in due finestre temporali specifiche: due ore prima e due ore dopo le partite. Partendo dalla lista degli IP ottenuta, è stato necessario estrarre la quantità di traffico durante le partite e nelle due finestre temporali. Ottenuto questo valore per ogni IP, è stato calcolato il traffico medio durante le partite e confrontato con il traffico registrato due ore prima e due ore dopo l'evento. Questa analisi ha permesso di identificare eventuali variazioni significative nel traffico, che potrebbero indicare un cambiamento comportamentale associato alla visione delle partite. La variazione del traffico è stata espressa in termini percentuali rispetto al traffico medio durante le partite. Gli IP che mostravano una variazione significativa (superiore al 50%) prima o dopo le partite sono stati considerati come traffico potenzialmente correlato alla visione delle partite stesse. In questo modo è possibile capire se un indirizzo IP ha mostrato un comportamento di rete anomalo o consistente con lo *streaming* video, rispetto ai periodi temporali precedenti e successivi all'evento sportivo. In questo modo, è stato possibile raffinare ulteriormente la classificazione del traffico legale, migliorando l'affidabilità dell'analisi.
Sviluppati i tre punti 1, 2 e 3 è stata ottenuta una lista di indirizzi IP che probabilmente stavano guardando le partite di calcio o su Sky oppure su DAZN. Da questa lista è stata ricavata la *baseline* del traffico legale presente in figura 3.14.

Il grafico presentato mostra l'andamento del traffico legale durante diverse partite di calcio tra il 10 e il 13 maggio 2024. Le linee verticali tratteggiate indicano i momenti

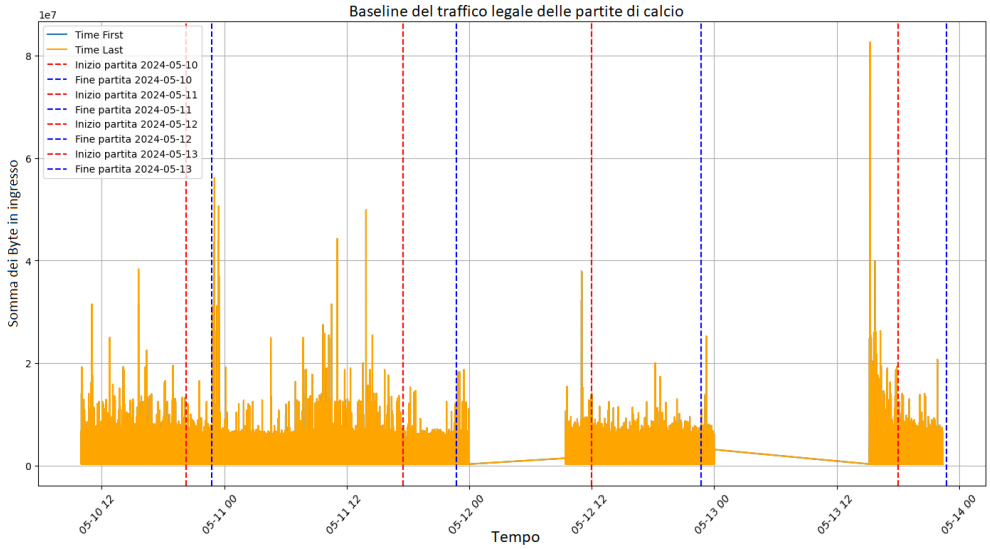


Figura 3.14: *Baseline* del traffico legale

di inizio e fine delle partite nelle date specifiche: le linee rosse rappresentano l'inizio delle partite, mentre quelle blu indicano la loro fine. Dal grafico emerge un *pattern* ricorrente del traffico, correlato agli orari delle partite di calcio. In particolare, si può osservare che il traffico tende ad aumentare nei momenti precedenti e durante le partite, con picchi all'inizio di ogni evento sportivo. Durante i periodi tra una partita e l'altra, il traffico presenta un calo, riflettendo la diminuzione dell'attività di *streaming*. Questo andamento ciclico suggerisce che il traffico legale è fortemente influenzato dagli eventi sportivi.

Per verificare l'accuratezza dell'implementazione dei criteri adottati per ottenere il traffico legale fosse corretta, è stato prelevato un sample[g] di 30 indirizzi IP dalla lista e analizzato il loro traffico. I dati ottenuti sono risultati coerenti con le aspettative, come è possibile notare nei grafici 3.15 e 3.16. Si evince che il traffico degli indirizzi IP campionati è costante e non presenta picchi significativi, come ci si aspetterebbe da un flusso di *streaming* video regolare. Dunque è possibile affermare che l'implementazione per identificare il traffico legale è stata corretta e che la *baseline* ottenuta è affidabile e rappresentativa del traffico legale.

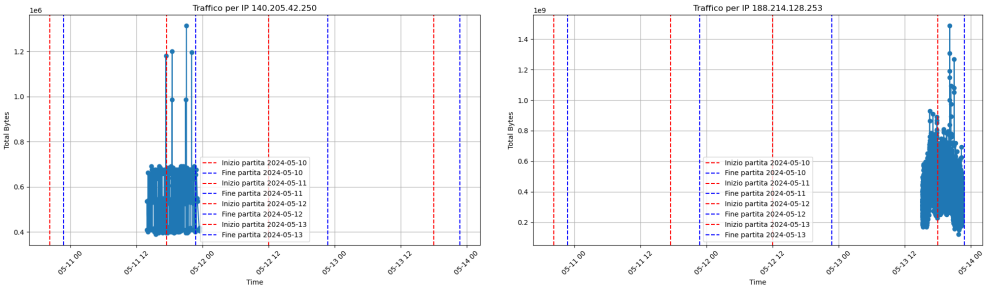


Figura 3.15: Traffico di 2 inidirizzi IP campionati

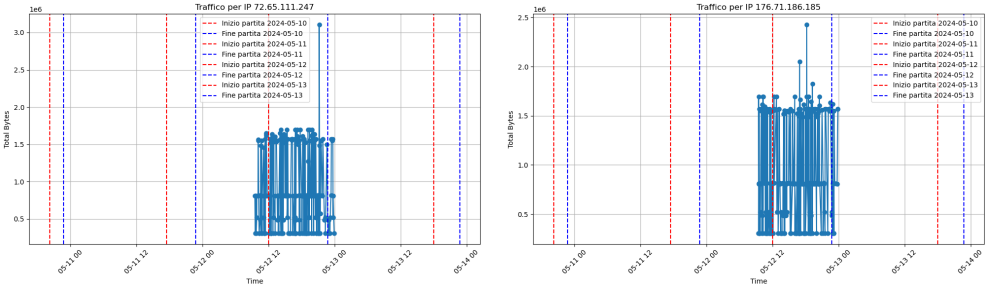


Figura 3.16: Traffico di 2 inidirizzi IP campionati

3.3 Identificazione del Presunto Traffico Illegale

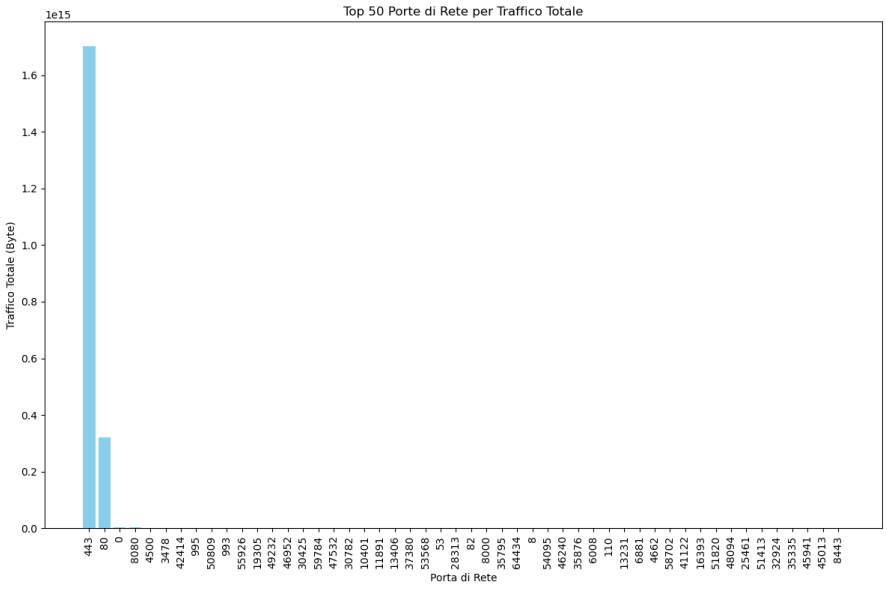


Figura 3.17: Top 50 porte di rete più utilizzate

L'identificazione del presunto traffico illegale è iniziata con l'individuazione delle porte di rete più utilizzate durante il periodo delle partite di calcio. I risultati ottenuti sono illustrati nel grafico 3.17. Dopo aver identificato le porte di rete con il maggior traffico, è stato necessario analizzare i dati generati per rilevare eventuali flussi di rete simili a quelli associati allo *streaming* video, seguendo il *pattern* del traffico legale precedentemente stabilito con la *baseline* nel grafico 3.14. Per questo motivo si è deciso di generare un grafico del traffico per ciascuna porta, sovrapponendo a ciascuno la *baseline* del traffico legale. Tra tutte le porte analizzate, la porta 41122 è risultata essere quella più interessante, in quanto il traffico generato (3.18) è molto simile a quello legale, come evidenziato nel grafico 3.19.

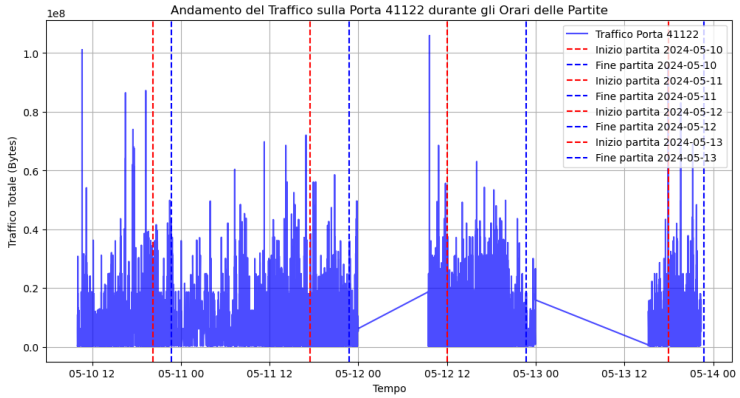


Figura 3.18: Traffico porta 41122 durante le partite

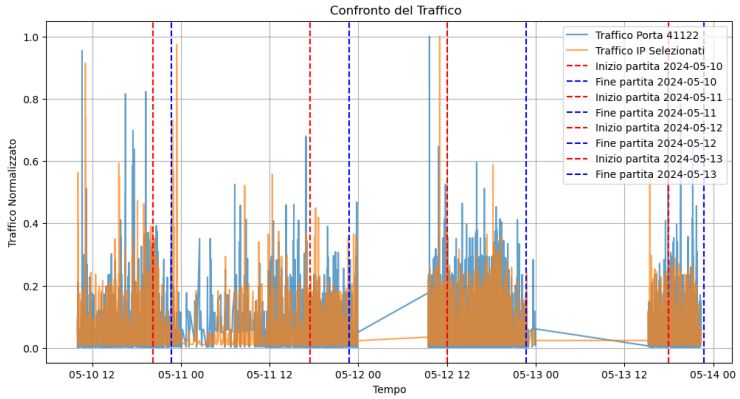


Figura 3.19: Confronto tra *baseline* e traffico porta 41122

La somiglianza tra il traffico della porta 41122 e la *baseline* del traffico legale suggerisce che il traffico generato da questa porta potrebbe essere associato allo *streaming* video illegale. Tuttavia, prima di confermare questa ipotesi e procedere oltre con l'analisi, è stato necessario verificare che il traffico della porta 41122 fosse effettivamente legato ad un flusso di *streaming* video. Per fare ciò, è stato necessario analizzare i pacchetti di rete associati a questa porta, in particolar modo i flag dei pacchetti TCP. I flag TCP forniscono informazioni importanti sullo stato della connessione e sul tipo di dati trasmessi, permettendo di distinguere tra differenti tipi

di traffico. Nello specifico, i flag ACK (*Acknowledgment*) e PSH (*Push*) sono molto frequenti nei flussi di *streaming* video. Il flag ACK segnala che un pacchetto è stato ricevuto correttamente e che il ricevitore è pronto a ricevere ulteriori dati. Quando questo flag è combinato con il PSH, che ordina al ricevitore di elaborare immediatamente i dati senza attendere ulteriori pacchetti, si rafforza l'indicazione che il traffico potrebbe essere legato allo *streaming* video, poiché tali flussi richiedono una trasmissione continua e rapida dei dati in tempo reale. Al contrario, il flag SYN (*Synchronize*), utilizzato all'inizio di una connessione TCP per stabilire la comunicazione tra due nodi, è meno indicativo di traffico di *streaming* video. Un'elevata presenza di pacchetti con flag SYN potrebbe invece suggerire tentativi di connessione o attività come *test* di rete e scansioni delle porte, che non sono necessariamente collegati a flussi di video in tempo reale, ma piuttosto a operazioni di *port scanning* o altri tipi di verifica della connettività. Pertanto, se l'analisi rivela una prevalenza di pacchetti con i flag ACK e PSH, è altamente probabile che il traffico sulla porta 41122 sia effettivamente correlato allo *streaming* video, il che rafforzerebbe l'ipotesi di un traffico illegale.
Di fatti l'analisi riguardante la porta 41122 ha confermato che il traffico di rete generato è effettivamente legato allo *streaming* video, come è possibile vedere nel grafico 3.20.

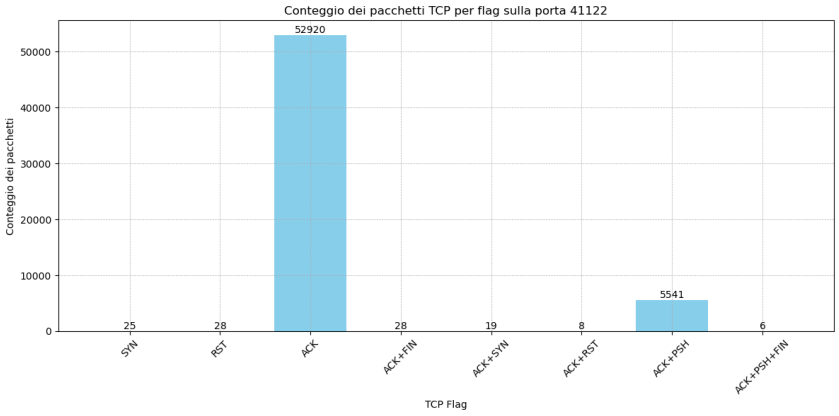


Figura 3.20: Flag dei pacchetti per la porta 41122

Guardando i dati rappresentati nel grafico, è evidente che la maggior parte dei pacchetti è caratterizzata dal flag ACK, con un numero significativo anche di pacchetti contenenti la combinazione di flag ACK+PSH. D'altronde, la bassa frequenza dei pacchetti con il flag SYN, indica che la maggior parte delle connessioni è stata stabilita per la trasmissione continua dei dati piuttosto che per l'avvio di nuove connessioni. Questo elemento è coerente con il comportamento atteso per il traffico di *streaming* video, dove la connessione, una volta stabilita, viene mantenuta attiva per lunghi periodi, piuttosto che essere frequentemente riavviata.

Un'altra analisi è stata condotta per verificare ulteriormente che il traffico della porta 41122 fosse effettivamente legato allo *streaming* video illegale, confrontando questo traffico con quello generato da degli ASN che sono stati identificati come legati allo *streaming* video illegale. Questi ASN sono stati ottenuti da una lista di siti internet bloccati in Italia da parte dell'AGCOM, con l'utilizzo di Piracy Shield. Questo flusso di rete è stato ricostruito con lo stesso metodo utilizzato per costruire la *baseline* del traffico legale, dove in quel caso l'ASN di interesse era quello di Akamai, ovvero il

20940, mentre in questo caso sono stati utilizzati gli ASN presenti nella lista di siti oscurati. Sono stati applicati i punti 1, 2 e 3 per poter ottenere una lista di indirizzi IP coinvolti in questo traffico e successivamente generare il flusso di rete associato a questi indirizzi IP. Il grafico 3.21 mostra tale flusso, che nel complesso risulta essere un traffico generale. È difficile identificare un *pattern* specifico relativo allo *streaming* video, fatta eccezione per la terza giornata di partite, ovvero il 12 maggio, in cui emerge una forma a campana tipica dello *streaming*. Per il resto, il grafico sembra rappresentare un traffico di rete generico. Questo potrebbe essere dovuto al fatto che gli ASN bloccati non erano coinvolti esclusivamente nella trasmissione di flussi di *streaming*, ma anche in altre attività di rete.

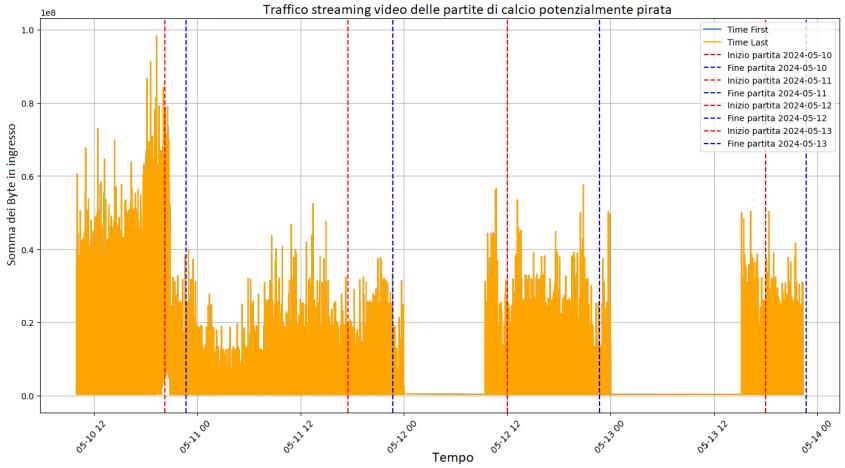


Figura 3.21: Traffico di rete generato da indirizzi IP associati allo *streaming* pirata

Il grafico 3.22 confronta, durante le partite, il traffico della porta 41122 con il traffico generato dagli indirizzi IP associati allo *streaming* pirata.

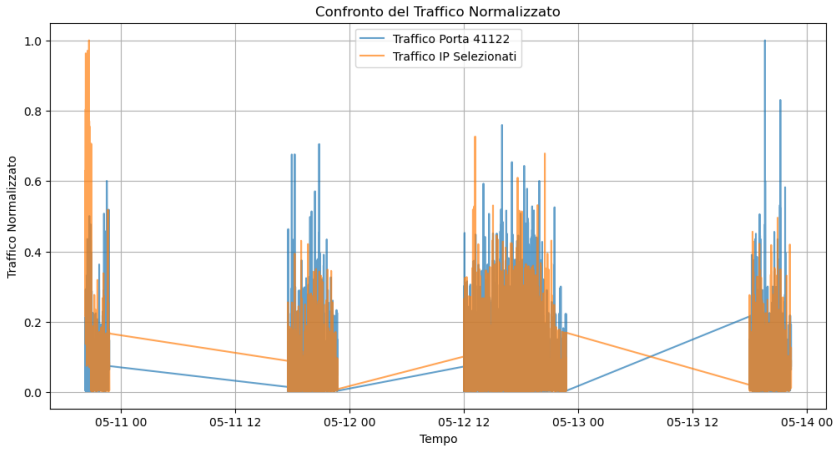


Figura 3.22: Confronto tra traffico della porta 41122 e traffico generato da indirizzi IP associati allo *streaming* pirata

Il confronto tra i due flussi di rete mostra una certa somiglianza tra il traffico della

porta 41122 e il traffico illegale, soprattutto durante la giornata del 12 maggio. Tutti questi risultati ottenuti fino ad'ora non fanno altro che rafforzare l'ipotesi che il traffico generato dalla porta 41122 sia effettivamente associato allo *streaming* video illegale. Tuttavia, per ottenere una risposta più formale e concreta, è stato necessario utilizzare degli algoritmi che hanno permesso di analizzare la somiglianza tra il traffico della porta 41122 e la *baseline* del traffico legale in modo più rigoroso. Tra gli strumenti utilizzati ci sono la ***Cross-Correlation***, la **Distanza *Dynamic Time Warping (DTW)***, la **Deviazione Standard** e la ***Cosine Similarity***. Ciascuno di questi metodi è stato selezionato perchè riescono ad evidenziare vari aspetti della somiglianza tra i due dataset, contribuendo a una comprensione complessiva più chiara.
Nelle sezioni seguenti, verrà spiegato il funzionamento di questi algoritmi e il loro ruolo nell'analisi.

Cross-Correlation

La Cross-Correlation è uno strumento statistico che misura la somiglianza tra due serie temporali come funzione di un ritardo temporale applicato a una di esse. Questo strumento è particolarmente utile quando si cerca di identificare se e come due serie temporali sono correlate nel tempo, ovvero se i cambiamenti in una serie si riflettono in cambiamenti nell'altra, e a quale distanza temporale avviene questa correlazione.
Nel contesto dell'analisi, la Cross-Correlation è stata utilizzata per determinare la relazione temporale tra il traffico della porta 41122 e il traffico associato agli indirizzi IP selezionati. Un picco significativo della correlazione incrociata in corrispondenza di un *lag* pari a zero (cioè senza ritardo temporale) suggerisce che i due flussi di traffico sono correlati quando non c'è alcun ritardo tra di loro. Questo risultato è indicativo di una stretta relazione temporale tra i due flussi di rete, suggerendo che le attività rilevate sulla porta 41122 avvengono simultaneamente con quelle associate agli IP selezionati, supportando così l'ipotesi di uno streaming video illegale. Il grafico 3.23 mostra la Cross-Correlation tra il traffico della porta 41122 e il traffico generato dagli indirizzi IP associati allo *streaming* pirata.

Dynamic Time Warping (DTW)

La Distanza Dynamic Time Warping (DTW) è un metodo di misurazione della similarità tra due serie temporali che possono variare nel tempo. A differenza di altre tecniche, DTW è in grado di gestire le variazioni di velocità nel tempo tra le due serie, permettendo un allineamento "elastico". Questo è particolarmente utile quando si analizzano flussi di dati che potrebbero non essere perfettamente sincronizzati ma che, in sostanza, seguono lo stesso *pattern*.
L'applicazione della DTW nell'analisi ha consentito di confrontare in modo più flessibile i flussi di traffico della porta 41122 con quelli associati agli IP selezionati, anche se i due flussi non sono perfettamente sincronizzati nel tempo. Il valore basso della distanza DTW ottenuto, ovvero di circa 8, suggerisce che nonostante eventuali differenze temporali, i *pattern* di traffico delle due serie sono notevolmente simili, rafforzando ulteriormente l'ipotesi di traffico illegale associato allo *streaming* video.

Deviazione Standard

La Deviazione Standard è una misura statistica che quantifica la quantità di variazione o dispersione di un insieme di dati. In un contesto di serie temporali, la deviazione standard misura quanto i valori di una serie temporale si discostano in media dal loro

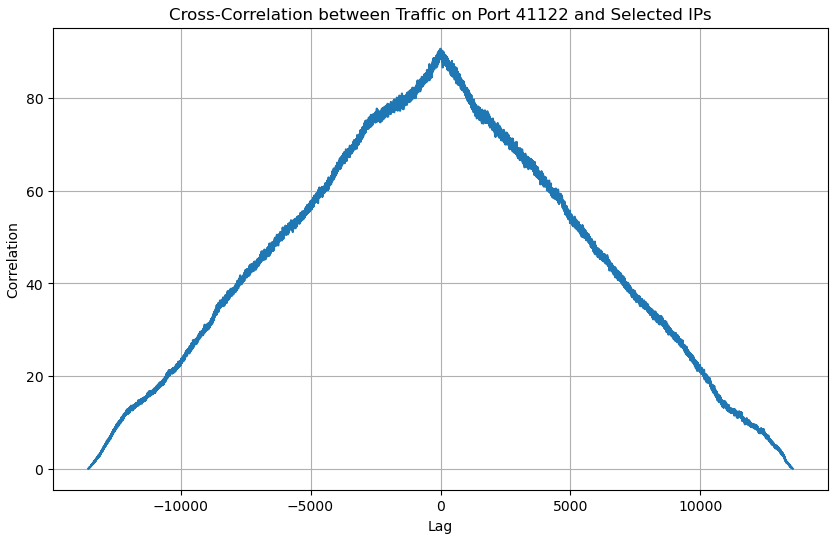


Figura 3.23: Cross-Correlation tra traffico della porta 41122 e traffico generato da indirizzi IP associati allo *streaming* pirata

valore medio. Una deviazione standard bassa indica che i valori sono vicini alla media, mentre una deviazione standard elevata indica che i valori sono più dispersi.
In questo studio, la deviazione standard è stata utilizzata per confrontare la variabilità del traffico della porta 41122 con quella del traffico degli IP selezionati. Valori di deviazione standard simili tra i due dataset suggeriscono che i *pattern* di variazione del traffico sono comparabili, indicando una probabile somiglianza strutturale tra i due flussi di dati, infatti i valori sono:

- Deviazione Standard del traffico della porta 41122: 0.078
- Deviazione Standard del traffico degli IP associati allo *streaming* pirata: 0.065

Cosine Similarity

La Cosine Similarity è una misura che quantifica la similarità tra due vettori nello spazio delle caratteristiche, basandosi sull'angolo tra di essi. Essa varia tra -1 e 1, dove 1 indica una perfetta similarità angolare (cioè i vettori puntano nella stessa direzione), 0 indica nessuna similarità e -1 indica che i vettori sono diametralmente opposti.
In questa analisi, la Cosine Similarity è stata impiegata per determinare la similarità tra il traffico della porta 41122 e quello degli IP selezionati. Un valore positivo e vicino a 1, in questo caso di circa 0.563, indica che i due vettori hanno una forte similarità direzionale, il che implica che i *pattern* di traffico delle due serie temporali sono simili in termini di direzione e magnitudine.

In sintesi, l'utilizzo combinato di Cross-Correlation, DTW, Deviazione Standard e Cosine Similarity ha permesso di confrontare in modo approfondito i flussi di traffico della porta 41122 con quelli associati agli IP selezionati. I risultati ottenuti indicano una somiglianza tra i due flussi, supportando l'ipotesi che il traffico rilevato sulla porta 41122 sia correlato allo *streaming* video illegale. Inoltre, il numero di indirizzi IP associati al traffico della porta 41122 è pari a 2177, mentre gli indirizzi IP correlati

allo *streaming* pirata ammontano a 4401. Tra questi, 32 indirizzi IP risultano essere presenti sia nel traffico della porta 41122 che in quello associato allo *streaming* pirata. Questo ulteriore dettaglio contribuisce a rafforzare ulteriormente l'ipotesi formulata inizialmente.

Capitolo 4

Valutazione dell'Efficacia di Piracy Shield

Questo capitolo valuta l'efficacia di Piracy Shield nel bloccare e identificare il traffico illegale, sulla base delle analisi e studi effettuati durante lo stage. Verranno presentate le metriche di valutazione utilizzate, i risultati ottenuti dall'analisi dei dati e un confronto tra il traffico legale e quello illegale

4.1 Metriche di valutazione

Per valutare l'efficacia di Piracy Shield, è essenziale definire chiaramente le metriche di valutazione. Queste metriche devono riflettere la capacità del sistema di identificare e bloccare il traffico pirata senza compromettere l'accesso ai contenuti legittimi. Le principali metriche considerate includono:

1. **Tasso di rilevamento:** rappresenta la percentuale di traffico pirata che il sistema riesce a identificare correttamente. È una metrica cruciale, poiché misura la capacità del sistema di distinguere tra contenuti legali e illegali.

2. **Tasso di falsi positivi:** questa metrica valuta la quantità di traffico legittimo che viene erroneamente classificato come pirata e quindi bloccato. Un alto tasso di falsi positivi potrebbe indicare degli errori nei blocci, causando disagi agli utenti legittimi.

3. **Tempo di risposta:** misura il tempo necessario affinché il sistema rilevi e blocchi un sito pirata dopo la sua identificazione. Un tempo di risposta rapido è fondamentale per minimizzare la distribuzione di contenuti illegali.

4. **Resilienza ai *bypass*:** è una metrica che valuta quanto sia difficile per gli utenti pirata aggirare le misure di blocco imposte da Piracy Shield. Questo può includere l'utilizzo di VPN o altre tecniche di mascheramento.

5. **Efficienza della gestione degli indirizzi IP:** con la crescente scarsità di indirizzi IPv4, è importante che Piracy Shield utilizzi gli indirizzi IP in modo efficiente, evitando di bloccare inutilmente IP che potrebbero essere riutilizzati per scopi legittimi.

Riguardo al tasso di rilevamento, il tasso di falsi positivi e, di conseguenza, l'efficienza nella gestione degli indirizzi IP, non è possibile stimare queste metriche poiché non si ha accesso ai dati reali di Piracy Shield. Le uniche informazioni disponibili provengono da un documento fornito dall'AGCOM sul loro sito web, che contiene un elenco di indirizzi IP bloccati tramite la piattaforma Piracy Shield dal 2 febbraio al 26 luglio 2024. Tuttavia, questo elenco è incompleto, poiché non specifica il motivo del blocco di ciascun indirizzo IP, né chi è il proprietario di tali indirizzi, rendendo impossibile distinguere tra quelli bloccati per contenuti illegali e quelli bloccati per altri motivi. L'unico dato disponibile è il numero totale di indirizzi IP bloccati, che ammonta a 4377 [1]. Questo dato è tuttavia poco significativo, poiché non consente di stimare il tasso di rilevamento, il tasso di falsi positivi o l'efficienza nella gestione degli indirizzi IP. Per quanto riguarda il tempo di risposta, sappiamo che Piracy Shield è in grado di bloccare un sito pirata in circa 30 minuti, come dichiarato dall'AGCOM. Infine, per quanto concerne la resilienza ai *bypass*, è stato effettuato un test per verificare la possibilità di aggirare le misure di blocco imposte da Piracy Shield utilizzando una VPN, e il risultato ha confermato che è effettivamente possibile eludere il blocco. Questo test è stato condotto utilizzando una VPN gratuita.

4.2 Comparazione tra traffico legale e illegale

Per valutare l'efficacia di Piracy Shield, è stata condotta un'analisi comparando il traffico legale con quello presunto illegale. L'immagine 4.1 rappresenta un grafico che illustra questa comparazione, utilizzando i dati ricostruiti dal traffico generato dagli ASN associati agli indirizzi IP bloccati da Piracy Shield, contrassegnati in arancione, e il traffico legale identificato separatamente, contrassegnato in blu. L'identificazione del traffico legale viene discussa nel paragrafo §3.2, mentre l'analisi del traffico presunto pirata viene discussa nel paragrafo §3.3. Il grafico evidenzia che il traffico illegale presenta significativi picchi di volume in corrispondenza degli orari delle partite di calcio, indicati dalle linee tratteggiate verticali. Questi picchi coincidono temporalmente con l'inizio e la fine degli eventi sportivi, suggerendo una forte correlazione tra il traffico presunto pirata e la trasmissione non autorizzata di contenuti sportivi. È d'obbligo precisare che il flusso di colore arancione presente in figura 4.1, sembra rappresentare un traffico di rete generico, risultando quindi difficile identificare un *pattern* specifico relativo allo *streaming* video, fatta eccezione per la terza giornata di partite, ovvero il 12 maggio, in cui emerge una forma a campana tipica dello *streaming*. Al contrario, il traffico legale, rappresentato in blu, mostra fluttuazioni minori. Inoltre, è interessante notare che il traffico legale e illegale non sempre sono facilmente distinguibili. In alcuni periodi, i volumi dei due tipi di traffico si sovrappongono parzialmente, complicando ulteriormente il compito di distinguere il traffico legale da quello illegale. Questa sovrapposizione suggerisce che potrebbero essere necessari algoritmi di rilevamento più sofisticati o tecniche di filtraggio avanzate per migliorare l'accuratezza dell'identificazione dei due tipi di traffico.

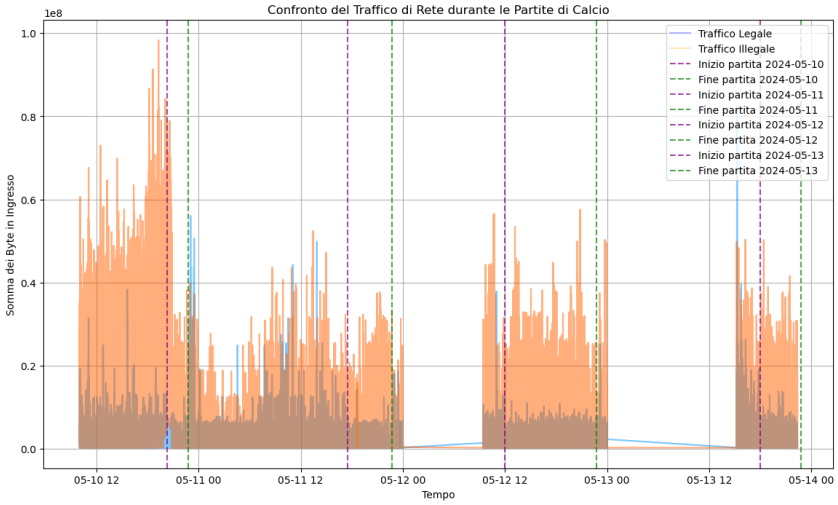


Figura 4.1: Traffico di rete generato da indirizzi IP associati allo *streaming* pirata

4.3 Analisi delle prestazioni di Piracy Shield

L'analisi delle prestazioni di Piracy Shield ha rivelato diversi punti di forza e debolezza che necessitano di attenzione per migliorare l'efficacia complessiva del sistema. Nonostante il sistema sia riuscito a bloccare con successo alcuni siti che trasmettevano contenuti piratati, come partite di calcio senza licenza, ci sono ancora molti problemi da affrontare. Alcuni dei principali punti di forza e debolezza emersi dall'analisi includono:

- Blocchi di contenuti legittimi: uno dei problemi principali emersi è il tasso di falsi positivi, dove siti *web* non correlati alla pirateria sono stati erroneamente bloccati. Questo non solo causa problemi agli utenti legittimi ma può anche danneggiare la reputazione del sistema. È quindi fondamentale cambiare oppure migliorare la metodologia di rilevamento per ridurre questi errori.
- Evasione tramite VPN: un altro problema significativo è rappresentato dalla facilità con cui gli utenti possono aggirare i blocchi utilizzando VPN. Questa vulnerabilità riduce drasticamente l'efficacia del sistema, in quanto gli utenti possono facilmente mascherare la loro attività e accedere ai contenuti piratati senza restrizioni. Questo è dovuto al fatto che l'oscuramento dei siti avviene soltanto a livello nazionale. Infatti sono gli operatori italiani a bloccare l'accesso ai siti pirata, mentre gli utenti stranieri possono accedervi senza problemi.
- Gestione degli indirizzi IP: la gestione degli indirizzi IP è un'altra area critica. Bloccare un IP che appartiene a una Content Delivery Network (CDN) può risultare in blocchi di massa di contenuti leciti. Inoltre, una volta bloccato, un indirizzo IP viene mantenuto inutilizzato per un periodo prolungato, causando un uso inefficiente delle risorse IP, già limitate. Dato che la disponibilità di indirizzi IPv4 è sempre più scarsa, è importante migliorare la precisione dei blocchi e garantire che gli indirizzi IP siano rapidamente rilasciati quando non più necessari.
- Risultati dell'analisi del traffico: malgrado le criticità riscontrate, l'analisi del traffico condotta in questa tesi suggerisce che Piracy Shield ha avuto un impatto

positivo nella riduzione del traffico illegale. Tuttavia, la presenza di siti pirata attivi e la capacità degli utenti di aggirare i blocchi indicano che è necessario un ulteriore impegno per migliorare l'efficacia del sistema.

In conclusione, Piracy Shield ha mostrato un potenziale significativo per contrastare la distribuzione illegale di contenuti, ma richiede miglioramenti e cambiamenti continui per affrontare i problemi emersi durante l'analisi delle prestazioni e per far fronte alle tecniche di elusione utilizzate dagli utenti pirata.

Capitolo 5

Discussione dei Risultati e Conclusioni

In questo capitolo vengono presentati i risultati ottenuti dalle analisi e sperimentazioni, mettendo in evidenza le implicazioni per il funzionamento di Piracy Shield. Si discuerà anche delle criticità riscontrate e delle possibili soluzioni future per migliorare l'efficacia del sistema.

5.1 Interpretazione dei risultati

L'analisi dei dati raccolti attraverso l'utilizzo di Piracy Shield mostra risultati contrastanti. Da un lato, la piattaforma ha dimostrato una certa efficacia nel ridurre il traffico illegale, suggerendo che possa essere un valido strumento nella lotta contro la pirateria digitale. Tuttavia, sono emerse alcune criticità che limitano la sua efficienza complessiva.
Uno dei problemi principali riguarda un discreto numero di falsi positivi, che ha portato al blocco ingiustificato di indirizzi IP associati a contenuti legittimi. Questa problematica non solo ha creato disagi per gli utenti, ma ha anche evidenziato la necessità di migliorare gli algoritmi di rilevamento per distinguere più accuratamente tra traffico pirata e legale. Inoltre, la gestione degli indirizzi IP, in particolare l'inclusione di IP condivisi con servizi legittimi, ha causato interruzioni non necessarie e ha sollevato preoccupazioni legali e di net neutrality. In questo contesto, la net neutrality afferma che tutti i dati su Internet devono essere trattati allo stesso modo, senza discriminazioni o restrizioni da parte dei fornitori di servizi Internet, indipendentemente dal loro contenuto o provenienza. La possibilità che un'azienda privata, attraverso strumenti come Piracy Shield, possa bloccare o limitare l'accesso a specifici contenuti o siti *web*, solleva il rischio di violazioni di questo principio fondamentale. Questi incidenti indicano il bisogno di un utilizzo più preciso ed efficiente delle risorse IP per evitare blocchi errati.
Un'altra criticità significativa è la vulnerabilità della piattaforma alle tecniche di elusione come l'uso di VPN. Gli utenti possono facilmente aggirare i blocchi imposti da Piracy Shield, rendendo meno efficace la sua capacità di prevenire l'accesso ai contenuti pirata. Questo suggerisce la necessità di integrare tecnologie avanzate che possano rilevare e contrastare l'uso di strumenti di elusione per aumentare la robustezza del sistema.
Alla luce di queste osservazioni, emerge chiaramente che per migliorare il funzionamen-

to di Piracy Shield è necessario implementare soluzioni più avanzate per la gestione degli indirizzi IP e per l'identificazione del traffico pirata. L'adozione di algoritmi di rilevamento più sofisticati e una maggiore collaborazione con i *provider* di servizi internet potrebbero contribuire a ridurre i falsi positivi e a prevenire efficacemente l'accesso ai contenuti pirata, garantendo un equilibrio tra la protezione dei diritti d'autore e il rispetto dei diritti degli utenti legittimi.
Questi miglioramenti saranno cruciali per assicurare che Piracy Shield possa diventare uno strumento affidabile ed efficiente nella lotta contro la pirateria digitale, migliorando non solo la precisione del sistema ma anche la sua accettabilità tra gli utenti e le autorità di regolamentazione.

5.2 Il futuro di Piracy Shield

Per questo motivo, il commissario AGCOM Massimiliano Capitanio ha deciso di migliorare la piattaforma Piracy Shield, riconoscendo la necessità di apportare modifiche significative per renderla uno strumento più efficiente e adeguato nella lotta contro la pirateria digitale. Il piano di AGCOM prevede un aggiornamento normativo che permetta alla piattaforma non solo di bloccare, ma anche di sbloccare i domini, superando così il limite imposto dall'attuale legislazione che impedisce la rimozione di blocchi erronei. Questa revisione normativa è stata proposta in risposta alla saturazione delle risorse della piattaforma e all'accumulo di domini e indirizzi IP bloccati, che hanno superato i limiti tecnici e logistici stabiliti dagli accordi con i fornitori di servizi internet. [31]
Oltre alla modifica legislativa, AGCOM ha annunciato l'introduzione di "Piracy Shield 2.0", una versione migliorata della piattaforma, prevista entro la fine dell'anno. Questa nuova versione sarà progettata per gestire meglio la mole crescente di segnalazioni e operatori accreditati, migliorando la capacità di rispondere rapidamente senza generare *timeout* non sostenibili. La necessità di una piattaforma più robusta e scalabile è stata evidenziata dalle numerose critiche ricevute per i problemi tecnici riscontrati, inclusa la pubblicazione non autorizzata del codice su GitHub e la gestione inefficace degli indirizzi IP [8].
Il commissario Capitanio ha sottolineato anche l'importanza di un cambiamento culturale nella lotta alla pirateria. Non basta sanzionare le piattaforme illegali; è necessario sensibilizzare il pubblico sui danni economici e culturali della pirateria. Pertanto, AGCOM sta lavorando a un protocollo che coinvolga anche la Guardia di Finanza e le Procure per sanzionare gli utenti finali che usufruiscono dei servizi pirata, mirando a dissuadere ulteriormente l'uso di contenuti illegali. Questa misura, che prevede sanzioni amministrative tra i 150 e i 500 euro, rappresenta un passo deciso verso un approccio più rigoroso e diretto nella regolamentazione del consumo di contenuti online [8].
Inoltre, è prevista una revisione della gestione degli indirizzi IP per evitare blocchi eccessivi e non necessari che potrebbero compromettere l'accessibilità alle risorse internet lecite. L'obiettivo è di implementare un sistema di blocco temporaneo, che renderebbe nuovamente disponibili gli indirizzi IP dopo alcuni mesi, minimizzando così l'impatto sulla rete e limitando le opportunità per i pirati di cambiare facilmente indirizzo IP e continuare le loro attività illegali [31].
In sintesi, il futuro di Piracy Shield si prospetta come un'evoluzione tecnologica e normativa, volta a rafforzare la lotta contro la pirateria digitale in Italia, migliorando l'efficacia e l'accettabilità della piattaforma attraverso soluzioni tecniche avanzate e un

rinnovato impegno culturale contro l'uso illegale di contenuti [15].

5.3 Consuntivo finale

Al termine dello stage, è possibile effettuare un bilancio complessivo del lavoro svolto e dei risultati ottenuti. Tutti gli obiettivi obbligatori fissati all'inizio del progetto sono stati raggiunti. In particolare, è stata acquisita una conoscenza approfondita del dominio dell'analisi del traffico di rete e sono stati implementati con successo gli esperimenti previsti. Per quanto riguarda gli obiettivi desiderabili, essi sono stati parzialmente soddisfatti. È stato infatti possibile condurre un'analisi approfondita dei risultati degli esperimenti, ma non tutte le aree previste sono state esplorate in modo esaustivo. Tuttavia, il lavoro svolto rappresenta una base solida per ulteriori sviluppi e approfondimenti futuri. Infine, gli obiettivi facoltativi, che includevano l'espansione degli esperimenti ad altri domini oltre al riconoscimento degli *streaming* illegali e la collaborazione alla redazione di un *paper* scientifico, non sono stati raggiunti. Questo è principalmente dovuto alla complessità degli stessi e alle limitazioni temporali incontrate durante lo svolgimento del lavoro. In conclusione, lo stage ha permesso di acquisire competenze tecniche specifiche, di approfondire la conoscenza del dominio e di sviluppare capacità di *problem solving* applicate a un contesto reale.

5.4 Conoscenze Acquisite

Durante il tirocinio presso il gruppo SPRITZ, è stato possibile approfondire le conoscenze nel campo della sicurezza informatica, con particolare attenzione all'analisi di grandi dataset relativi al traffico di rete. È stata dedicata particolare attenzione all'analisi dettagliata dei dati e allo studio delle metodologie utilizzate dai pirati informatici per eludere i sistemi di sicurezza. Inoltre, sono stati esplorati aspetti tecnici legati all'implementazione e sperimentazione di soluzioni per il riconoscimento di flussi video illegali, impiegando strumenti per l'analisi dei dati di rete. Questo percorso ha fornito una solida base teorica e pratica per affrontare problemi complessi nel campo della sicurezza informatica. Sono state acquisite competenze tecniche essenziali e approfondita la comprensione delle dinamiche del traffico di rete, con un focus sull'identificazione e classificazione dei flussi. È stato possibile progettare e implementare ambienti sperimentali, sviluppando le competenze necessarie per condurre esperimenti. Infine, sono stati appresi nuovi concetti e elementi, come ASN e CDN, che hanno permesso di comprendere meglio il funzionamento delle infrastrutture di rete.

5.5 Raggiungimento degli Obiettivi

Al termine del progetto, tutti gli obiettivi prefissati sono stati raggiunti, come riportato nella tabella 5.1. L'analisi condotta ha consentito di identificare e classificare i flussi di traffico di rete, permettendo così di distinguere chiaramente tra traffico legale e illegale. Di seguito viene riportata la tabella 5.1 che riassume gli obiettivi prefissati e il loro stato di raggiungimento:

ID	Descrizione	Soddisfatto?
Obiettivi Obbligatori		
O01	Conoscenza del dominio dell'analisi di traffico di rete	●
O02	Conoscenza degli studi effettuati in precedenza nello stesso scenario	●
O03	Progettazione di un ambiente per l'esecuzione di esperimenti	●
O04	Implementazione di un ambiente per l'esecuzione di esperimenti	●
O05	Capacità di implementare i *feedback* forniti durante gli incontri	●
O06	Esecuzione degli esperimenti e recupero dei risultati	●
Obiettivi Desiderabili		
D01	Design degli esperimenti necessari allo studio	❍
D02	Analisi approfondita dei risultati degli esperimenti	❍
D03	Documentazione generale degli ambienti di sperimentazione	❍
Obiettivi Facoltativi		
F01	Espansione esperimenti ad altri domini oltre al riconoscimento degli *streaming* illegali	❑
F02	Collaborazione alla redazione di un *paper* scientifico sui risultati ottenuti	❑

Tabella 5.1: Elenco degli obiettivi obbligatori, desiderabili e facoltativi con stato di soddisfazione.
● Soddisfatto ❍ Parzialmente soddisfatto ❑ Non soddisfatto

Di seguito viene riportata la tabella 5.2 che riassume il lavoro svolto durante lo stage, con delle differenze rispetto alla pianificazione iniziale presente in tabella 2.2

Durata in ore	Descrizione dell'attività
80	**Background**
30	*Formazione sulle tecnologie adottate*
20	*Formazione sulla metodologia sperimentale*
30	*Studio e ricerca delle tecniche di identificazione e classificazione dei flussi di rete*
110	**Progettazione degli esperimenti**
35	*Analisi del problema e degli esperimenti pregressi*
30	*Progettazione della piattaforma e codice per gli esperimenti*
20	*Consolidamento iterativo della piattaforma*
25	*Stesura documentazione relativa ad analisi e progettazione*
110	**Esecuzione esperimenti**
25	*Preparazione degli ambienti per esperimenti*
55	*Esecuzione esperimenti e raccolta dati*
30	*Analisi dei risultati*
20	**Stesura relazione finale e documentazione**
Totale ore	**320**

Tabella 5.2: Tabella della pianificazione del lavoro

Acronimi e abbreviazioni

ASN Autonomous System Number. 2, 17, 31, 32, 37, 43, 48

CDN Content Delivery Network. 5, 9, 30, 31, 44, 48

Fully Qualified Domain Name Fully Qualified Domain Name. 4, 8

FTP File Transfer Protocol. 28, 29

HTTP Hypertext Transfer Protocol. 6, 22

ICMP Internet Control Message Protocol. 17, 20, 22

M3U Moving Picture Experts Group Audio Layer 3. 7

P2P Peer-to-Peer. 25, 27–29

RTMP Real Time Messaging Protocol. 6

SSH Secure Shell. 12

TCP Transmission Control Protocol. 18, 20, 22, 24, 25, 27, 28, 36

UDP User Datagram Protocol. 18, 20, 22, 24, 26, 27

URL Uniform Resource Locator. 7

VPN Virtual Private Network. 6, 10, 12, 42–44, 46

AGCOM Autorità per le Garanzie nelle Comunicazioni. 1, 4, 8–10, 37, 43, 47

IP Internet Protocl. 4

IPTV Internet Protocol Television. 4–7

Glossario

ASN è un numero univoco assegnato a ciascun sistema autonomo (AS) all'interno della rete Internet. 51

cache è una memoria temporanea che memorizza i dati frequentemente utilizzati, permettendo di accedervi in maniera più veloce rispetto ad una memoria principale. La cache permette di ridurre i tempi di accesso ai dati e di migliorare le prestazioni del sistema. 30

CDN è una rete di server distribuiti che fornisce in maniera rapida ed efficiente dei contenuti agli utenti, riuscendo a ridurre la latenza e migliorando le prestazioni di caricamento del contenuto stesso. 51

cluster è un insieme di *computer* o *server* collegati tra loro. Essi lavorano in maniera coordinata per eseguire un'operazione in maniera più efficiente e veloce rispetto ad un singolo *computer* o *server*. 12

dataset in informatica si intende una collezione di dati, in genere di grandi dimensioni e spesso organizzati in tabelle. I dataset vengono utilizzati per analisi, addestramento di modelli di apprendimento automatico o altre applicazioni computazionali. iii, 12, 39, 40, 48

decoder dispositivo elettronico che converte un segnale digitale in un segnale analogico, o viceversa, permettendo l'interpretazione e l'utilizzo dell'informazione originaria contenuta nel segnale. 1, 5, 6

flag è un campo presente nell'intestazione dei pacchetti di dati che permette di specificare lo stato del pacchetto stesso. I flag vengono utilizzati per indicare informazioni come l'inizio o la fine di un pacchetto, l'invio di dati o la richiesta di connessione. 25, 26, 36

FQDN è un indirizzo *web* completo che specifica univocamente la posizione di un sito *web* in Internet e all'interno della gerarchia dei domini. "www.esempio.com" è un esempio di FQDM dove per "www" si intende il sottodominio, "esempio" è il dominio vero e proprio e "com" rappresenta il dominio di primo livello. 51

FTP è un protocollo di rete che permette di trasferire *file* tra due dispositivi connessi in rete. FTP permette di inviare e ricevere *file* in maniera semplice e veloce, permettendo di organizzare e condividere i *file* tra i dispositivi. 51

GitHub è una piattaforma di sviluppo collaborativo di *software* che fornisce funzionalità di collaborazione, di gestione del codice e del controllo di versione per i progetti *software*. 10, 47

HTTP è un protocollo di comunicazione che definisce il modo in cui i dati tra il *client* e il *server* vengono trasmessi su Internet. Viene utilizzato per il trasmerimento delle informazioni sul *web*, permetteno la visualizzazione di pagine *web*. 51

ICMP è un protocollo di comunicazione che permette ai dispositivi di rete di scambiarsi messaggi di controllo e di errore, permettendo così di monitorare e diagnosticare la rete. 51

join è un'operazione nel linguaggio SQL che permette di combinare due o più tabelle di un *database*, permettendo così di ottenere un'unica tabella contenente i dati desiderati. 12

M3U è un formato di file relativo a *playlist* audio e video che elenca percorsi di altri file multimediali o URL da riprodurre in streaming oppure attraverso un lettore multimediale. 51

Markdown è un linguaggio di markup che permette di scrivere dei testi formattati in maniera semplice e veloce, fornendo una visualizzazione del testo intuitiva e immediata. 11

modem è un dispositivo elettronico che permette di stabilire una connessione tra il *computer* e la rete Internet. Si occupa di modulare e demodulare i segnali elettrici in segnali digitali e viceversa, permettendo quindi la trasmissione dei dati. 5

Multicast è una modalità di trasmissione di dati che permette di inviare un singolo flusso di dati a più destinatari contemporaneamente permettendo di ridurre così il carico sulla rete. 4

net neutrality è un principio secondo il cui tutti i dati trasmessi su Internet debbano essere trattati allo stesso modo, senza discriminazioni o preferenze basate sul contenuto, applicazione o piattaforma da parte dei *provider* di servizi Internet. 1, 10, 46

P2P è un modello di architettura logica di rete informatica che permette ai dispositivi di comunicare e scambiarsi dati direttamente tra loro, senza la necessità di un *server* centrale. I dispositivi non sono né *client* né *server*, infatti non c'è una gerarchia, ma sono tutti uguali e possono agire sia da *client* che da *server*. 51

RTMP è un protocollo di comunicazione sviluppato da Adobe Systems che permette la trasmissione in tempo reale di dati di tipo audio e video su Internet, comunemente utilizzato per lo *streaming live*. 51

sample è un insieme di dati rappresentativo di un fenomeno o di un'entità più grande. I campioni vengono utilizzati per analizzare e studiare un determinato fenomeno di interesse, permettendo di trarre delle conclusioni e delle previsioni. 34

SSH è un protocollo di rete che permette di stabilire una connessione sicura tra due dispositivi e permette quindi di trasferire dati in maniera sicura e criptata. 51

TCP è un protocollo di comunicazione che permette di trasmettere dati in maniera affidabile e ordinata su Internet. TCP garantisce che i dati vengano trasmessi in maniera corretta e che vengano ricevuti dall'altro dispositivo. 51

TCP/IP è un insieme di protocolli di comunicazione che definiscono la modalità di trasmissione dei dati su Internet. È composto da due differenti protocolli: il protocollo di controllo della trasmissione (TCP) che si occupa di garantire che i dati vengano trasmessi in maniera corretta, e dal protocollo Internet (IP), il quale invece si occuppa di instradare i dati tra i dispositivi connessi alla rete. 4

Telnet è un protocollo di rete che permette di stabilire una connessione remota con un altro dispositivo e permette quindi di eseguire comandi e trasferire dati. Telnet non usa nessun tipo di crittografia durante la trasmissione di dati, risultando facile intercettare e leggere i dati. 12

ticket nell'ambito informatico viene definito come un documento o un entità digitale che viene utilizzato per registrare e gestire delle richieste di assistenza, segnalazioni di problemi o richieste di supporto tecnico. 8, 9

UDP è un protocollo di comunicazione che permette di trasmettere dati in maniera veloce e senza ritardi su Internet. UDP non garantisce la consegna dei dati e non si preoccupa dell'ordine di arrivo dei pacchetti. Questo protocollo è generalmente utilizzato per applicazioni che richiedono una bassa latenza come lo *streaming* video o audio. 51

Unicast è una modalità di trasmissione di dati che permette di inviare un singolo flusso di dati ad un solo destinatario, il quale riceve contenuti personalizzati, ottenendo *privacy* e sicurezza dei dati. 4

URL è una stringa di testo, più precisamente un indirizzo standardizzato, che specifica la posizione di una risorsa su Internet, permettendo di poterla identificare e accederci. Un URL è composto da diversi elementi, come per esempio il protocollo di comunicazione, il dominio e il percorso della risorsa. 51

VPN è una tecnologia che permette di creare una connessione sicura e criptata tra due o più dispositivi attraverso Internet, (ovvero una rete pubblica, quindi meno sicura) garantendo la protezione dei dati degli utenti. 51

VSIX si riferisce al Centro di Ateneo per la Connettività e i Servizi al Territorio dell'Università di Padova gestisce il Neutral Access Point del Nord Est. Questo centro promuove l'uso di Internet nel Veneto cooperando con Internet Service Provider locali, nazionali e internazionali. Gestisce l'Internet Exchange di Padova, collegando reti pubbliche, private e della ricerca, facilitando l'interscambio di dati e migliorando la connettività e la sicurezza per cittadini, aziende e amministrazioni pubbliche [26]. iii, 12, 13, 15

Web TV è un servizio di trasmissione televisiva che utilizza la rete Internet per trasmettere i programmi televisivi. Questo servizio permette di guardare i programmi televisivi in diretta o in differita. 4

Bibliografia

Siti web consultati

[1] *AGCOM; PDF Elenco Indirizzi IP Bloccati.* URL: https://www.agcom.it/sites/default/files/media/allegato/2024/_SITO_totali%20blocchi_2febbrai_26%20luglio__0.pdf (cit. a p. 43).

[2] *AGCOM; Piattaforma Piracy Shield.* URL: https://www.agcom.it/competenze/antipirateria-e-piracy-shield/piattaforma-piracy-shield (cit. alle pp. 3, 4).

[3] *Agenda Digitale; Piracy Shield, tutte le falle dell'anti-pirateria di Stato.* URL: https://www.vs-ix.org/architecture (cit. a p. 1).

[4] *Akamai; DAZN Group.* URL: https://www.akamai.com/resources/customer-story/dazn-group (cit. a p. 31).

[5] *Akamai; Sky Italia.* URL: https://www.akamai.com/site/en/documents/case-study/akamai-sky-italia-content-delivery-case-study.pdf (cit. a p. 31).

[6] *Ansa; Pirateria audiovisiva.* URL: https://www.ansa.it/sito/notizie/politica/2024/06/24/pirateria-audiovisiva-persi-2-miliardi-euro-di-fatturato_00d124ad-31df-4bde-8a8e-bc3d856649ac.html#:~:text=Una%20perdita%20stimata%20di%20fatturato,pari%20a%20circa%2011.200%20unit%C3%A0. (cit. a p. 4).

[7] *ClickHouse.* URL: https://clickhouse.com (cit. a p. 12).

[8] *DDay; Piracy Shield, a fine anno una nuova piattaforma.* URL: https://www.dday.it/redazione/49579/piracy-shield-a-fine-anno-una-nuova-piattaforma-capitanio-piratate-anche-partite-in-chiaro (cit. a p. 47).

[9] *Fortinet; FortiClient VPN.* URL: https://www.fortinet.com/it/support/product-downloads (cit. a p. 12).

[10] *Geopop; Streaming Illegale.* URL: https://www.geopop.it/streaming-illegale-come-funziona-liptv-e-lo-scudo-anti-pezzotto/ (cit. alle pp. 7, 8).

[11] *GitHub; Fuck Piracy Shield.* URL: https://github.com/fuckpiracyshield/variations/blob/main/variations.py (cit. a p. 10).

[12] *Google; Documenti Google.* URL: https://www.google.com/docs/about/ (cit. a p. 14).

[13] *Google; Fogli Google*. URL: https://workspace.google.com/products/sheets/ (cit. a p. 14).

[14] *Google; Google Drive*. URL: https://www.google.com/drive/ (cit. a p. 14).

[15] *HDblog; Piracy Shield addio, nuova piattaforma antipirateria in arrivo entro fine 2024*. URL: https://www.hdblog.it/mercato/articoli/n585145/piracy-shield-agcom-nuova-piattaforma-antipirata/ (cit. a p. 48).

[16] *Jupyter; Jupyter Notebook*. URL: https://jupyter.org (cit. a p. 11).

[17] *Lexplain; Piracy Shield*. URL: https://www.lexplain.it/piracy-shield-cose-e-come-funziona/ (cit. a p. 8).

[18] *Microsoft; Visual Studio Code*. URL: https://code.visualstudio.com (cit. a p. 11).

[19] *Putty*. URL: https://www.putty.org (cit. a p. 13).

[20] *Python*. URL: https://www.python.org (cit. a p. 13).

[21] *Sky; Decoder Sky*. URL: https://assistenza.sky.it/decoder (cit. a p. 5).

[22] *Sky; Modem Sky*. URL: https://assistenza.sky.it/sky-wifi-fibra/sky-hub/guida-configurazione-sky-hub (cit. a p. 5).

[23] *Spritz Group*. URL: https://spritz.math.unipd.it/index.html.

[24] *Stop Piracy Shield*. URL: https://stop-piracy-shield.it (cit. alle pp. 8–10).

[25] *TechTarget; IPTV*. URL: https://www.techtarget.com/searchnetworking/definition/IPTV-Internet-Protocol-television (cit. a p. 5).

[26] *VSIX*. URL: https://www.vs-ix.org (cit. a p. 54).

[27] *Wikipedia; Content Delivery Network*. URL: https://it.wikipedia.org/wiki/Content_Delivery_Network (cit. a p. 30).

[28] *Wikipedia; IPTV*. URL: https://it.wikipedia.org/wiki/IPTV (cit. alle pp. 4, 5).

[29] *Wikipedia; Modem*. URL: https://it.wikipedia.org/wiki/Modem (cit. a p. 5).

[30] *Wikipedia; Set-top box*. URL: https://it.wikipedia.org/wiki/Set-top_box (cit. a p. 5).

[31] *Wired; Piracy Shield, la piattaforma nazionale antipirateria, sta esaurendo il potere di oscurare siti*. URL: https://www.wired.it/article/piracy-shield-limite-blocco-domini-governo-sblocco-indirizzi-ip/ (cit. a p. 47).