

# Elaborato Esame di Stato

Brunello Cesare 5Al A.s. 2019/2020

## Table of Contents

Definizione.....	1
Crescita dei servizi e nascita del fenomeno.....	2
Motivazioni dietro la nascita della pirateria.....	2
Il nuovo petrolio del 21esimo secolo.....	3
Per sconfiggere un pirata.....	3
Il server tracker.....	4
Il sistema informativo.....	4
Altri metodi.....	5
Fallimento del SOPA.....	5
Controversie sull'HADOPI.....	6
La nuova frontiera della pirateria.....	7

## Definizione

Con pirateria intendiamo tutte le azioni che ci permettono di usare programmi o vedere contenuti a pagamento senza spendere un soldo, quindi parliamo di servizi come lo streaming video o il download gratuito di programmi che solitamente richiedono l'acquisto di una licenza.

Si potrebbe pensare che lo scopo di piratare prodotti informatici è quello di condividere software e reperti multimediali gratuitamente, è quindi facile pensare che la pirateria sia in qualche modo affiliata al pensiero del software libero ma non è assolutamente così [1].

Il software libero viene rilasciato con l'intento di essere condiviso a titolo gratuito con lo scopo di trasmettere cultura e aumentare la cooperazione tra le persone [2] mentre la pirateria informatica ha scopi di natura lucrativa. La definizione più generica di pirateria è quella di "varie attività di natura illecita perpetrate tramite l'utilizzo di strumenti informatici, in violazione delle norme giuridiche di uno Stato" [3].

Questa definizione, seppur corretta, non descrive al meglio la pirateria informatica e l'aspetto che ha assunto negli ultimi tempi. La pirateria non è più un taboo digitale, non riguarda più solamente le attività criminose gravi quali attacchi informatici a corporazioni o furto di proprietà intellettuale, ora la pirateria ricade in una zona grigia dell'ambito legale in quanto più o meno tutti gli internauti si sono resi pirati.

I termini più comunemente associati, come "video streaming" o "programmi craccati" ci possono orientare verso una definizione diversa: "la pirateria è l'insieme di tutte le attività mirate alla contraffazione o alla distribuzione illecita di prodotti presenti in rete protetti da copyright e forniti al pubblico dietro una forma di pagamento".

## Crescita dei servizi e nascita del fenomeno

Il fenomeno della pirateria nasce con la progressiva diffusione dell'home computer, nient'altro che un mezzo utilizzato per ottenere un fine, lecito o illecito. Inizia ad essere usato per aggirare sistemi e ottenere dati e/o informazioni solitamente protetti e esclusi dal dominio pubblico.

Queste informazioni sono per esempio, documenti legali, file protetti da copyright, informazioni su persone, credenziali di accesso ad un sistema e altri tipi di dati [3].

Le tecniche per raggiungere questi scopi col tempo si sono diversificate sempre più, alcune, avvicinandosi anche ad attaccare le "vulnerabilità psicologiche" dell'utente umano, allo stesso tempo conservando la loro natura di attacchi informatici fatti di programmi con comportamenti predeterminati da chi li crea.

Ma la pirateria informatica (come la intendiamo noi) non ha come intento quello di creare disservizio o danneggiare proprietà, piuttosto quello evitare di dover pagare per servirsi di servizi web e/o usufruire di software distribuiti sotto licenza.

## Motivazioni dietro la nascita della pirateria

Nella pirateria informatica mancano le motivazioni di alto tasso morale che ci sono negli altri movimenti che mirano a distribuire il software gratuitamente, essa è nata con lo scopo di rivendere software proprietari a prezzi al ribasso.

Col tempo, e con l'avvento di software come eMule e uTorrent, la creazione di reti peer-to-peer è diventata sempre più facile (anche per utenti inesperti) e ciò ha comportato un pesante incremento del bacino di utenza capace di condividere file e programmi senza doversi confrontare con la dura realtà di creare manualmente di una rete di scambio dati.

Più il numero di utenti aumentava e più la pirateria si allontanava dallo scopo di lucro originario in favore di un'altra valuta che col tempo sarebbe diventata sempre più preziosa: **i dati**.

Per avere un'idea dell'importanza che hanno assunto i dati basta pensare che nel 2014 sono stati generati 650 Exabyte (650 miliardi di GB) di dati risultati dal traffico sul web e il numero di persone con accesso a internet dal 2014 ha subito un incremento dell'83% (da 2,4 miliardi a 4,4) in 5 anni, ciò significa che chi è in grado di attingere ad una fonte simile di informazioni può prevedere trend, servirsi di marketing avanzato, sviluppare intelligenze artificiali o perfino spiare interi paesi [4].

# Il nuovo petrolio del 21esimo secolo

“Il più semplice scolaro sa oggi verità per le quali Archimede avrebbe sacrificato la vita”

**Ernest Renan**, *Ricordi d'infanzia e di giovinezza*

Questa citazione esplicita come un tempo, dati impossibili da raccogliere ora sono diventati accessibili a moltissimi e inseriti nel giusto ambito analitico identificano trend e statistiche molto allettanti al giusto compratore, una dinamica mai verificatosi prima nella storia umana. Ora la nostra impronta informatica lascia tracce nel web che trovano il loro utilizzo in queste statistiche, prendiamo ad esempio le app mobili: se la maggior parte delle app presenti sul Play Store sono gratuite, come fanno queste a generare introiti?

La risposta è collezionando dati (come informazioni sul calendario, log delle chiamate, posizione del dispositivo, cronologia del browser...) in modo da poterli rivendere [5].

Una volta che questi dati vengono raggruppati in un dataset prendono il nome di Big Data [6]. Ognuno di questi dataset è caratterizzato da tre caratteristiche, dette le “3V”:

- **Volume:** quantità di dati e dimensione del dataset (GB,TB,PB...)
- **Varietà:** differenti tipologie di dati utilizzati, spesso eterogenei perché provengono da device diversi e trattano informazioni differenti (dati non strutturati e semi-strutturati)
- **Velocità:** velocità con cui i nuovi dati vengono generati e rapidità dei dati nell'essere inseriti in un sistema real-time

Una dimensione simile di dati è difficile da ammassare ed è ancora più difficile da analizzare perché le informazioni provengono da innumerevoli fonti e si presentano in formati diversi, la difficoltà nel trattare questo tipo di dati è quella di ricercare un nesso che lega i dati fra di loro e che rappresenti gli “umori” del mercato e che dia un senso a tutte le informazioni che transitano attraverso internet.

## Per sconfiggere un pirata

Ora che sappiamo quali sono le tecnologie coinvolte, come facciamo a combattere la pirateria?

Per poter utilizzare i big data a nostro favore e individuare il traffico dei nostri attori malevoli bisogna disporre di due cose:

1. Un server tracker
2. Un sistema informativo per analizzare i dati

## Il server tracker

Il modello P2P è destabilizzante per il sistema del copyright per una serie di ragioni:

1. Ha provocato una massiccia diffusione di materiale coperto da copyright costringendo i grandi creatori di contenuti ad attacchi legali per tutelare i propri contenuti
2. Essendo una rete decentralizzata non vi è nessuno che detiene la responsabilità di ciò che viene trasmesso
3. La condivisione dei file è anonima, chiunque, qualunque siano le sue intenzioni, può diventare parte del sistema, la sicurezza sta nel buon senso dell'utente e dei file che scarica

Per loro natura, le reti P2P che usano il protocollo BitTorrent necessitano dei cosiddetti server tracker, cioè macchine in grado di mettere in comunicazione i peer che hanno bisogno di trasmettere i dati a vicenda.

Le reti **P2P pure** non dispongono di una figura centrale fissa che serve il resto dei client, quindi avere una sola fonte di log è impossibile a meno che la rete in cui ci troviamo non sia una **P2P con Discovery Server**, in quel caso, si dispone di un server centrale chiamato Discovery che contiene tutte le coordinate degli altri client.

Sia il Server Tracker che il Discovery Server funzionano allo stesso modo: hanno lo scopo di localizzare le altre sorgenti che possiedono il file o parte di esso e si occupano di coordinare i rapporti tra utenti richiedenti un file e quelli che lo offrono [7].

Ogni richiedente effettua una richiesta al nostro server tracker il quale tiene traccia di tutte le informazioni contenute nelle richieste che riceve (in particolare l'IP) del richiedente e degli altri server tracker che vengono interpellati.

Ma siccome non tutto quello che passa in una rete P2P BitTorrent è materiale illegale bisogna costruire un database che possa tenere traccia di quali sorgenti sono affidabili e quali no [8][9].

## Il sistema informativo

Per sistema informativo intendiamo l'insieme di strumenti automatizzati o parzialmente automatizzati in grado di ottenere dei dati strutturati ma soprattutto utilizzabili per le operazioni di analisi che andremo ad effettuare. Per assicurarci un sistema informativo solido dobbiamo partire dall'aspetto più basilare, cioè la creazione di una base di dati che possa contenere la grande quantità di dati strutturati e non-strutturati di cui necessitiamo.

Quando si progetta un database che deve gestire big data, percorrere la strada dei Database Relazionali potrebbe non essere l'opzione migliore.

Teniamo a mente quali sono le proprietà basilari di un DB e che dobbiamo cercare di rispettare, esso deve essere [10]:

- Sicuro
- Integro

- Consistente
- Condivisibile
- Persistente
- Scalabile

In sistemi informativi di così alto livello spesso si usano parallelamente sia i database relazionali che i non-relazionali: la differenza principale tra i due è che nel primo viene usato il linguaggio per query strutturate SQL con le sue componenti *DDL* e *DML* (uno usato per definire la struttura delle tabelle, l'altro per modificare i dati contenuti), mentre nei database non-relazionali non viene definita una struttura a priori, bensì essa viene creata via via con l'inserimento dei dati sotto forma di documenti *JSON*, non a caso il nome con cui è conosciuta questa tecnologia è proprio NoSQL cioè Not only SQL [11][12].

Questi due tipi Database non sono mutualmente esclusivi, nel senso che il punto di forza degli RDBMS sta nel gestire le transazioni, grazie al supporto delle proprietà **ACID**:

- **Atomicità**: approccio alla transazione “tutto o niente”
- **Consistenza**: capacità di non violare i vincoli referenziali e di integrità dei dati
- **Isolamento**: più transazioni eseguite allo stesso momento non interferiscono tra di loro
- **Presistenza**: garanzia che l'effetto ottenuto dalle transazioni sia persistente al termine dell'operazione

Mentre un Non-RDBMS è decisamente più efficace quando si parla di gestire grandi quantità di dati non strutturati o se questi dati risiedono su più macchine [13].

## Altri metodi

Per combattere questo sistema negli anni sono stati adottati tantissimi sistemi, diversi tra loro. Un esempio è la casa videoludica Rockstar Games che ha incorporato un sistema anti-pirateria nel loro videogioco GTA IV il quale si attiva solo sulle copie che non disponevano di una chiave di attivazione valida modificando alcune impostazioni e rendendo l'intero prodotto ingiocabile.

## Fallimento del SOPA

Un altro metodo che è stato adottato è quello del SOPA, il Stop Online Piracy Act, una legge promossa dal partito democratico statunitense USA all'inizio del 2012 (ciò che ha causato quella diminuzione nella ricerca del termine streaming nel gennaio 2012) ma mai entrata in vigore [14]. Questa legge fu soggetta a molte controversie in quanto molti esponenti del pensiero del software libero come il co-fondatore di Wikipedia, Jimmy Wales, sostenevano che questo tipo di provvedimento sarebbe stato un ostacolo al primo emendamento e una censura della libertà di parola [15] in quanto alcuni dettagli tecnici di questo decreto sembravano puntare verso una “privatizzazione” di certi settori di internet.

## Controversie sull'HADOPI

L'**HADOPI** (acronimo di Haute Autorité pour la diffusion des oeuvres et la protection des droits sur l'Internet) è un provvedimento simile al SOPA adottato in Francia durante il governo Sarkozy entrato in vigore del 2010. Questi due provvedimenti sono simili nell'obiettivo ma differiscono nel funzionamento [16].

L'HADOPI prevede un sistema di 3 avvertimenti per chiunque sia responsabile di un sito di download di materiale protetto da copyright. Al terzo avvertimento, se l'utente è recidivo nel rimuovere i contenuti specificati negli avvertimenti precedenti, verrà notificato dei procedimenti penali che inizieranno a suo carico, tra cui una multa di 1500 euro e una denuncia per negligenza grave.

Questo sistema però presenta ancora dei gravi punti di debolezza, tre in particolare:

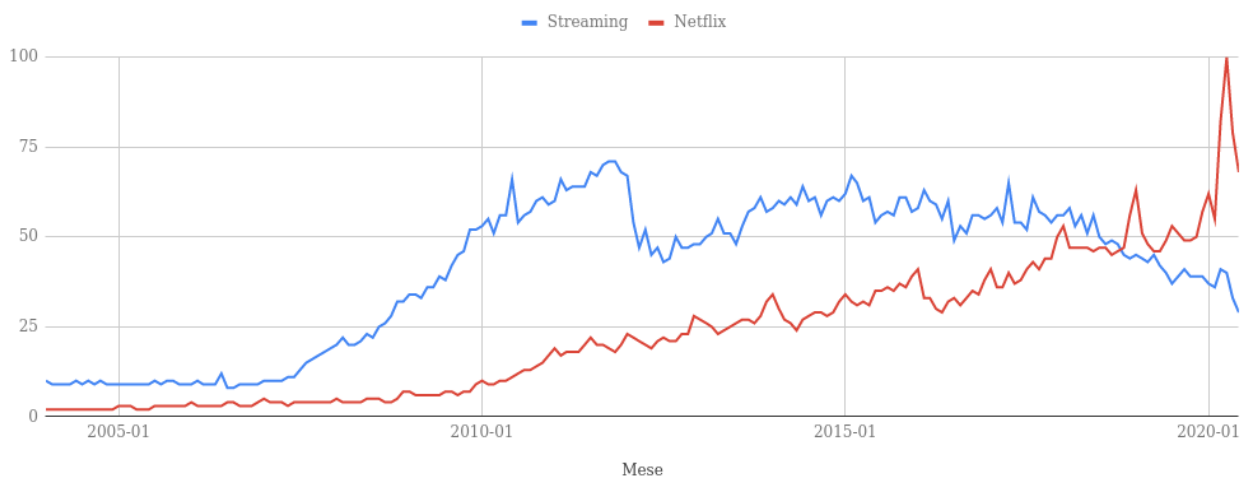
- l'indirizzo IP è il solo modo per tracciare il comportamento degli utenti (problema bypassabile con un proxy)
- la natura sproporzionata della sanzione
- la sanzione si applica al proprietario della linea internet e non all'utente che ha scaricato il materiale

# La nuova frontiera della pirateria

I metodi applicati sono tanti e sono diversi, dai decreti legge ad un approccio più tattile da prima linea. Forse il modo migliore di combattere il fenomeno è scegliere le opzioni legali, evitando così di alimentare un mercato, che sistematicamente si ripete.

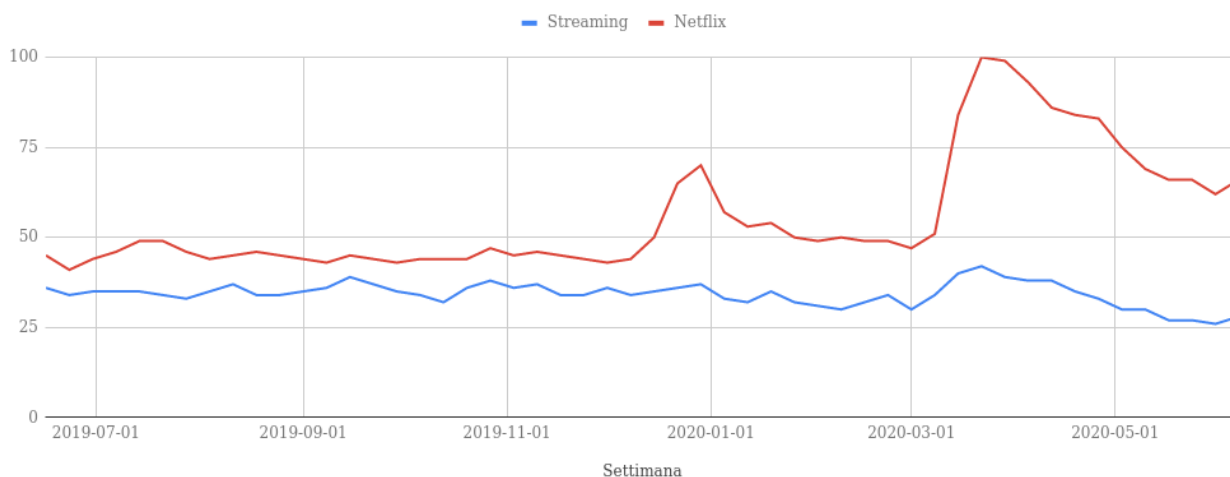
E negli ultimi anni questo trend si è confermato, secondo una stima, entro il 2025 Netflix conterà tra i suoi abbonati il 18,5% della popolazione mondiale (1,43 miliardi di utenti) [17].

Streaming e Netflix



Forse il modo migliore di combattere la pirateria è affidarsi al buon senso degli utenti e sperare che facciano la scelta giusta, un esempio, è quello che è successo con l'inizio della quarantena di quest'anno.

Streaming e Netflix negli ultimi 12 mesi



## Bibliografia

- 1: Linux didattica sulla pirateria - [http://linuxdidattica.org/docs/linuxmagazine/bernardi\\_25.html](http://linuxdidattica.org/docs/linuxmagazine/bernardi_25.html)
- 2: Filosofia del software libero - Wikipedia - [https://it.wikipedia.org/wiki/Software\\_libero#Aspetti\\_filosofici\\_e\\_sociali](https://it.wikipedia.org/wiki/Software_libero#Aspetti_filosofici_e_sociali)
- 3: Pirateria informatica - Wikipedia - [https://it.wikipedia.org/wiki/Pirateria\\_informatica](https://it.wikipedia.org/wiki/Pirateria_informatica)
- 4: Dati generati a confronto passato e presente - <https://blog.microfocus.com/how-much-data-is-created-on-the-internet-each-day/>
- 5: Come le app generano revenue - <https://www.consumer.ftc.gov/articles/0018-understanding-mobile-apps>
- 6: Big Data - [https://it.wikipedia.org/wiki/Big\\_data](https://it.wikipedia.org/wiki/Big_data)
- 7: Peer to Peer - Wikipedia - <https://it.wikipedia.org/wiki/Peer-to-peer>
- 8: Protocollo BitTorrent - <https://it.wikipedia.org/wiki/BitTorrent>
- 9: BitTorrent Specification - <https://wiki.theory.org/index.php/BitTorrentSpecification>
- 10: Sistemi informativi e basi di dati - Cloud Informatica V Anno
- 11: MySQL e NoSQL: cosa sono - <https://www.html.it/articoli/sql-e-nosql-a-documenti-il-confronto/>
- 12: Esempi di NoSQL - <https://www.rackone.it/big-data-con-database-nosql-unintroduzione-pratica/>
- 13: Confronto tra MySQL e NoSQL - <https://lorenzogovoni.com/database-relazionali-e-big-data/>
- 14: Il SOPA Stop Online Piracy Act - [https://it.wikipedia.org/wiki/Stop\\_Online\\_Piracy\\_Act](https://it.wikipedia.org/wiki/Stop_Online_Piracy_Act)
- 15: Co-fondatore di Wikipedia Jimmy Wales sul SOPA - <https://www.youtube.com/watch?v=5X7GwsIERbo>
- 16: HADOPI - <https://it.wikipedia.org/wiki/HADOPI>
- 17: Previsioni sulla crescita della userbase di Netflix - <https://www.corrierecomunicazioni.it/media/netflix-co-in-volata-ma-reggeranno-nel-lungo-periodo/>