

UNIVERSITA' DEGLI STUDI DI BARI "ALDO MORO"



DIPARTIMENTO DI INFORMATICA

CORSO DI LAUREA MAGISTRALE IN SICUREZZA INFORMATICA

CASO DI STUDIO IN "INFORMATICA FORENSE"

ESTRAZIONE AUTOMATIZZATA DELLE CHAT DI TELEGRAM

Docente:

Ugo Lopez

Componenti del gruppo:

Mirko De Vincentiis

Domenico Gigante

Domenico Picerno

SOMMARIO

1	Introduzione	3
1.1	Storia	3
1.2	Definizione	3
2	L'applicazione.....	5
2.1	Prerequisiti.....	5
2.2	Configurazione del file “config.ini”	5
2.3	Primo avvio	6
2.4	Configurazione del file “configuration.json”	7
2.5	Funzionalità	7
2.5.1	Estrazione delle chat di un singolo utente.....	7
2.5.2	Estrazione delle chat di determinati utenti	8
2.5.3	Estrazione di tutte le chat	9
2.6	Estrazione	9
2.7	Warning	10
2.7.1	This media is not downloadable.....	10
2.7.2	No members into chat <nome_chat>	10
2.7.3	Sleeping for <numero_secondi>s (required by "<metodo_API>")	10
2.8	Sviluppi futuri.....	11

1 INTRODUZIONE

La scienza forense è l'applicazione di tecniche e metodologie scientifiche alle tradizionali investigazioni di carattere giudiziario.

Esistono molte branche della scienza forense. Di queste, l'informatica forense è attualmente la più giovane.

1.1 STORIA

Nel 1984 l'FBI e altre agenzie governative americane iniziano a sviluppare programmi per reperire indizi all'interno dei computer. Nasce, quindi, il Computer Analysis and Response Team (CART) che, però, diventa effettivamente operativo solo nel 1991. Nello stesso periodo, Andrew Rosen realizza per la polizia canadese "Desktop Mountie", il primo strumento software per l'informatica forense.

Nel decennio successivo, vengono sviluppati poi Expert Witness, Encase e Smart.

Si sviluppa molto tra il 1997 ed il 2007, considerati gli anni d'oro" dell'informatica forense

L'informatica forense ha vissuto negli ultimi 20 anni un periodo di straordinario sviluppo, per lo più dovuto al fatto che per lungo tempo le principali fonti di prova sono rimaste tecnologicamente stabili e hanno pertanto consentito di sviluppare e affinare metodologie e strumenti di indagine forense straordinariamente efficaci.

La situazione era ed è tuttavia destinata a mutare:

- già nel 2010 Simon Garfinkel preconizzava la fine della golden age dell'informatica forense di fronte all'incremento esponenziale di capacità dei dispositivi di memorizzazione, al diversificarsi delle fonti di prova digitale e alla diffusione del cloud computing e della crittografia.
- Karie et al. (2015) riassumono, riassumono, in una tassonomia di quasi trenta voci, le principali sfide con le quali l'informatica forense si dovrà confrontare, distinguendo tra sfide tecnologiche, giuridiche, relative al personale e operative. Fra esse si ritrovano le problematiche già evidenziate da Garfinkel e altre legate sia alla difficile interoperabilità tra diversi strumenti utilizzati nelle indagini, sia alla carenza di personale adeguatamente addestrato a svolgere indagini forensi digitali.
- Alcune sfide sono state nel frattempo aggravate dall'ulteriore incremento della capacità dei dispositivi di memorizzazione e da una marcata diversificazione delle fonti di prova.

1.2 DEFINIZIONE

L'informatica forense (o digital forensics, prima nota col nome di computer forensics) è una branca della scienza forense legata alle prove acquisite da computer e altri dispositivi di memorizzazione digitale. Il suo scopo è quello di esaminare dispositivi digitali seguendo processi di analisi forense al fine di identificare, preservare, recuperare, analizzare e presentare fatti o opinioni riguardanti le informazioni raccolte.

L'informatico forense, dunque, è: "Digital forensics, also known as cyber forensics and computer forensics, is generally considered to consist of three roles in one: that of a cyber analyst familiar with

the working of computer devices and networks, a detective with knowledge of investigating crime, and a lawyer with a sound understanding of the law and court procedures” (R. Boddington, “Practical Digital Forensics”)

2 L'APPLICAZIONE

L'applicazione sviluppata permette di estrarre determinate chat a partire da un account Telegram.

Il codice sorgente è consultabile al link: <https://github.com/TheF3n1x/TelegramExporter>

L'estrazione può essere fatta secondo tre modalità:

- Estrazione della chat di una determinata singola chat
- Estrazione di un determinato numero di chat, inserite da riga di comanda
- Estrazione di tutte le chat associate all'account

Terminata l'estrazione, tutte le chat vengono compresse in un file .zip e, di questo, viene calcolato un doppio hash: uno con algoritmo MD5 e l'altro con algoritmo SHA-512, così da creare il sigillo digitale e garantire la catena di custodia.

L'applicazione fa affidamento al client Pyrogram¹, una libreria Python open source che funge da wrapper per le API ufficiali di Telegram.

2.1 PREREQUISITI

Tutti gli applicativi necessari per il funzionamento dell'applicazione sono dichiarati nel file "requirements.txt". La versione di Python, invece, deve essere maggiore o uguale a "3.0.0".

L'installazione di tali applicativi è resa automatica:

- Sistemi **Windows**: avviare il file "startup.ps1" (richiede la shell "Windows Powershell"), che si occuperà di verificare la presenza di una qualunque versione di Python e, se combacia con quella dichiarata nel file "requirements.txt", procede con l'installazione di tutti i restanti applicativi richiesti
- Sistemi **Linux**: nella maggior parte delle distribuzioni Linux è già preinstallato Python; di conseguenza, è sufficiente lanciare il comando "pip3 install -r requirements.txt" da terminale

2.2 CONFIGURAZIONE DEL FILE "CONFIG.INI"

Per il primo avvio, è necessario compilare opportunamente il file "config.ini".

Per farlo, è necessario effettuare il login su "<https://my.telegram.org/auth?to=apps>", inserendo il proprio numero di cellulare. Fatto ciò, si riceverà un codice su Telegram da inserire per autenticarsi.

Dopo essersi autenticati, è necessario creare una nuova applicazione, a cui associare un nome; Telegram si occuperà di generare, invece, i valori "api_id" e "api_hash".

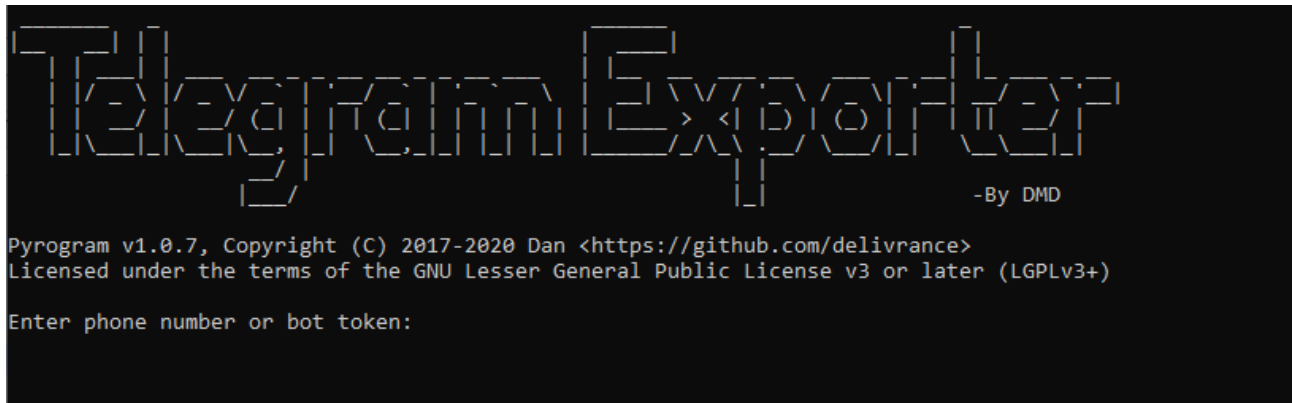
Ottenuti tali valori, bisognerà inserirli nel file "config.ini" al posto dei valori dummy inseriti di default.

¹ <https://github.com/pyrogram/pyrogram>

Fatto ciò, sarà possibile avviare il “run.cmd” (per sistemi Windows” o “run.sh” (per sistemi Linux).

2.3 PRIMO AVVIO

Al primo avvio, verrà richiesto di inserire il numero di telefono (comprensivo di prefisso, ad esempio: +39 per i numeri registrati in Italia) associato all’account da cui si vuole effettuare l’estrazione.

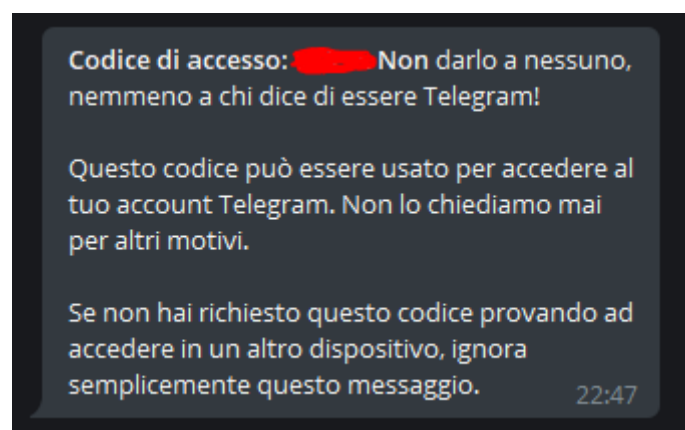


```
Telegram Exporter
-By DMD

Pyrogram v1.0.7, Copyright (C) 2017-2020 Dan <https://github.com/delivrance>
Licensed under the terms of the GNU Lesser General Public License v3 or later (LGPLv3+)

Enter phone number or bot token:
```

Dopo averlo inserito e premuto “Enter”, verrà richiesto l’inserimento di un PIN numerico; tale PIN verrà inviato dall’utente “Telegram” tramite messaggio; il messaggio avrà la struttura che segue:



Se attiva l’autenticazione a due fattori, sarà anche richiesto l’inserimento della password per accedere al cloud dell’utente.



```
Telegram Exporter
-By DMD

Pyrogram v1.0.7, Copyright (C) 2017-2020 Dan <https://github.com/delivrance>
Licensed under the terms of the GNU Lesser General Public License v3 or later (LGPLv3+)

Enter phone number or bot token: +39[redacted]
Is "+39[redacted]" correct? (y/N): y
The confirmation code has been sent via Telegram app
Enter confirmation code: [redacted]
Two-step verification password required
Password hint: [redacted]
Enter password (empty to recover): [redacted]
Do you want to clean extraction folder from previous extractions files? (y/N):
```

Inserita la password, verrà completato il primo avvio e l'applicazione sarà pronta per effettuare le estrazioni. Essendo il primo avvio, inserire “y” oppure “N” non comporterà alcuna eliminazione di file.

2.4 CONFIGURAZIONE DEL FILE “CONFIGURATION.JSON”

Il file “configuration.ini” è pensato per la configurazione contestuale del funzionamento dell'applicazione.

Al momento, possiede solo un parametro, di nome “export_media”. Se a questo è associato il valore 1, allora verranno scaricati anche tutti i media recuperati nelle chat; in caso valga 0, non verranno scaricati (in ogni caso, nelle chat estratte sarà mantenuta traccia del momento in si sia inviato/ricevuto tale media).

2.5 FUNZIONALITÀ

Le funzionalità dell'app sono esposte nel menu principale, il quale compare subito dopo aver scelto se cancellare o meno i file relativi alle eventuali estrazioni precedenti.

```
Enter:
[1] to extract the chats for a single user
[2] to extract the chats for multiple users
[3] to extract all chats
Please enter your choice:
```

Tutte le chat estratte saranno organizzate all'interno del percorso
<locazione_del_file_py>\extraction\<cartella_relativa_alla_singola_estrazione>.

2.5.1 Estrazione delle chat di un singolo utente

Inserendo 1 nel menu principale, si atterrà sulla schermata di inserimento dell'utente scelto:

```
Enter:
[1] to extract the chats for a single user
[2] to extract the chats for multiple users
[3] to extract all chats
Please enter your choice: 1
You can enter one of the following information:
- Book name
- Telegram username
- Channel name
- Group name
- Phone number (in this case remember to indicate also the phone prefix):
- Or press enter if you want to see a list of the chats
Please enter your decision:
```

Come descritto, è possibile effettuare la ricerca per nome con cui il contatto è salvato in rubrica, per username associato all'account scelto, nome del canale (solo per i canali), nome del gruppo (solo per i gruppi), numero di telefono associato all'account scelto; nel caso in cui non venga inserito alcun carattere, verrà mostrata la lista di tutte le chat identificate e l'utente potrà sceglierne una da estrarre.

```

Enter:
[1] to extract the chats for a single user
[2] to extract the chats for multiple users
[3] to extract all chats
Please enter your choice: 1
You can enter one of the following information:
- Book name
- Telegram username
- Channel name
- Group name
- Phone number (in this case remember to indicate also the phone prefix):
- Or press enter if you want to see a list of the chats
Please enter your decision: XXXXXXXXXX

[get_contact] Retrieving all matching contacts

[get_contact] Person chat match found

[get_chat_ids_by_dialogs] Retrieved chat with username: XXXXXXXXXX
[write_all_chats_logs_file] Processing chat with XXXXXXXXXX
[write_all_chats_logs_file] Processing members chats

[compress_and_hash_extraction] Creating extraction zip archive...
[compress_and_hash_extraction] Extraction zip archive created successfully
[compress_and_hash_extraction] Creating zip hashes...
[compress_and_hash_extraction] Zip hashes created successfully

```

2.5.2 Estrazione delle chat di determinati utenti

Inserendo 2 nel menu principale, si atterrerà sulla schermata di inserimento degli utenti scelto:

```

Telegram Exporter
-By DMD

Pyrogram v1.0.7, Copyright (C) 2017-2020 Dan <https://github.com/delivrance>
Licensed under the terms of the GNU Lesser General Public License v3 or later (LGPLv3+)

Do you want to clean extraction folder from previous extractions files? (y/N): y
[clean_extraction_folder]
Removing files from folder extraction
[clean_extraction_folder] Folder cleaned successfully

[create_extraction_folders] Creating extraction folders
[create_extraction_folders] Extraction folders created successfully

Enter:
[1] to extract the chats for a single user
[2] to extract the chats for multiple users
[3] to extract all chats
Please enter your choice: 2
User separator ';' to select multiple name.
Enter your decision: XXXXXXXXXX;XXXXXXXXXX

[get_contact] Retrieving all matching contacts

[get_contact] Person chat match found
[get_contact] Person chat match found
[*] 0 Username: XXXXXXXXXX First Name: XXXXXXXXXX Last Name: XXXXXXXXXX Telephone number: XXXXXXXXXX
[*] 1 Username: XXXXXXXXXX First Name: XXXXXXXXXX Last Name: XXXXXXXXXX

[get_chat_ids_by_dialogs] Retrieved chat with username: XXXXXXXXXX

[get_chat_ids_by_dialogs] Retrieved chat with username: XXXXXXXXXX
[write_all_chats_logs_file] Processing chat with XXXXXXXXXX

```


Le chat saranno estratte all'interno del percorso

<locazione_del_file_py>\extraction\Extraction_<data> <ora>.

2.5.3 Estrazione di tutte le chat

Inserendo 3 nel menu principale, si atterrerà sulla schermata di inserimento degli utenti scelto:

```
Telegram Exporter
-By DMD

Pyrogram v1.0.7, Copyright (C) 2017-2020 Dan <https://github.com/delivrance>
Licensed under the terms of the GNU Lesser General Public License v3 or later (LGPLv3+)

Do you want to clean extraction folder from previous extractions files? (y/N): y
[clean_extraction_folder]
Removing files from folder extraction
[clean_extraction_folder] Folder cleaned successfully

[create_extraction_folders] Creating extraction folders
[create_extraction_folders] Extraction folders created successfully

Enter:
[1] to extract the chats for a single user
[2] to extract the chats for multiple users
[3] to extract all chats
Please enter your choice: 3






[get_chat_ids_by_dialogs] Retrieved chat with username: [REDACTED]
```

2.6 ESTRAZIONE

Le chat saranno estratte all'interno del percorso

<locazione_del_file_py>\extraction\Extraction_<data> <ora>.

Nel suddetto percorso saranno presenti le seguenti entità:

	chats	07/12/2020 16:19	Cartella di file	
	media	07/12/2020 16:19	Cartella di file	
	members	07/12/2020 16:19	Cartella di file	
	extraction.zip	07/12/2020 16:19	Archivio WinRAR ...	1 KB
	extraction_archive_hash.txt	07/12/2020 16:19	Documento di testo	1 KB

- chats: cartella che contiene il file <username>_<nome>_<numero_telefonico>.csv
La presenza dei suddetti attributi non è sempre garantita. Tale file contiene tutti messaggi individuati nella chat, in ordine cronologico discendente.
- media: contiene tutti i media individuati nella chat
- members: contiene un file .csv che indica tutti i partecipanti alla chat
- extraction.zip: archivio contenente tutte le cartelle menzionate finora
- extraction_archive_hash.txt: contiene i valori di hash associati al file extraction.zip

extraction_archive_hash.txt - Blocco note di Windows

File Modifica Formato Visualizza ?

MD5: e6532ed4f8996b7f533383ce4e1076d8

SHA512: 9f46c07b6131ea1ead48bacc05f557dd18113fdd3f731dfb02dabd6b25afc623c130fea52a49b66e944ef47314a9ec43acb5d8679bdacc2e4a14c63196477eb

Esempio di digest calcolati sul file zip relativo ad una estrazione

2.7 WARNING

In alcuni casi, potrebbero comparire dei messaggi di warning nei log a video.

Di seguito è spiegata la loro motivazione.

2.7.1 This media is not downloadable

Questo warning compare nel caso in cui il media individuato non è scaricabile (ad esempio: sticker o sondaggi)

2.7.2 No members into chat <nome_chat>

Questo warning compare nel caso in cui non sia possibile recuperare i membri della chat (ad esempio: chat relative a canali).

2.7.3 Sleeping for <numero_secondi>s (required by "<metodo_API>")

Questo warning compare quando è necessario rallentare con le richieste per evitare di sollevare un alert di flood nei sistemi di Telegram.

2.8 SVILUPPI FUTURI

Come possibili sviluppi futuri, si potrebbe:

- Sviluppare una interfaccia grafica per il tool
- Far sì che il path di salvataggio degli artefatti sia inseribile come parametro nel file `configuration.json` (e non forzatamente `.\extraction`)