COME WHATSAPP GESTISCE I MESSAGGI

Analisi dettagliata di alcuni funzionamenti dell'applicazione WhatsApp per sistemi Android con illustrazione dei meccanismi di memorizzazione dei messaggi sul filesystem, con particolare riferimento all'implementazione della cifratura e all'eventuale recupero di database corrotti.

Giovanni TESSITORE, laureato in Informatica, è direttore della Sezione Indagini Elettroniche del Servizio Polizia Scientifica con competenza nel campo delle intercettazioni, dell'analisi audio e video, del confronto della voce e del volto, di digital forensics e cybersecurity.

Antonio CASOLARO, laureato in Ingegneria Elettronica, dal 2017 è funzionario addetto nella Sezione Indagini Elettroniche del Servizio Polizia Scientifica ed è referente per le attività di digital forensics e cybersecurity.

WhatsApp Messenger è un'applicazione di messaggistica istantanea gratuita nata nel 2009 e attualmente di proprietà della società Facebook. Tale applicazione, di utilizzo comune tra i possessori di smartphone, ha di fatto soppiantato l'utilizzo dei tradizionali SMS per l'invio di brevi messaggi di testo ed attualmente integra funzionalità di chiamate e videochiamate VoIP, nonché la possibilità di condividere file e posizione GPS.

L'applicazione è disponibile sia su piattaforma mobile sia su PC, sebbene richieda sempre un numero di telefono per poter essere utilizzata. Di fatto si tratta di un client che implementa una versione personalizzata del protocollo XMPP che, attraverso un account utente creato a partire dal numero di telefono seguito dal suffisso @s.whatsapp.net, si connette ai server di proprietà della società per il transito (in tutto o in parte) delle comunicazioni.
Ad inizio 2020 contava oltre 2 miliardi di utenti su tutto il globo e risulta la seconda piattaforma social più utilizzata in Italia.

Stante quindi l'ampio utilizzo dell'applicazione per lo scambio di comunicazioni, l'analisi del suo funzionamento per scopi forensi è divenuto, nel tempo, un fattore rilevante per le attività di polizia giudiziaria.
A tale scopo gli scriventi hanno condotto un'analisi dettagliata di alcuni funzionamenti dell'applicazione WhatsApp per sistemi Android, di cui si espongono di seguito i risultati. In particolare verranno illustrati i meccanismi di memorizzazione dei messaggi sul filesystem, con particolare riferimento all'implementazione della cifratura ed un eventuale recupero di database corrotti. Sarà inoltre fatto riferimento al modo con cui WhatsApp associa l'orario ai messaggi scambiati e come tale informazione viene interpretata sia dai dispositivi mobili sia da uno degli strumenti di analisi forensi maggiormente diffuso quale Ufed.

1. Memorizzazione dei messaggi su Android

L'applicazione WhatsApp memorizza sul dispositivo Android nel quale è installata i messaggi scambiati con gli altri utenti all'interno

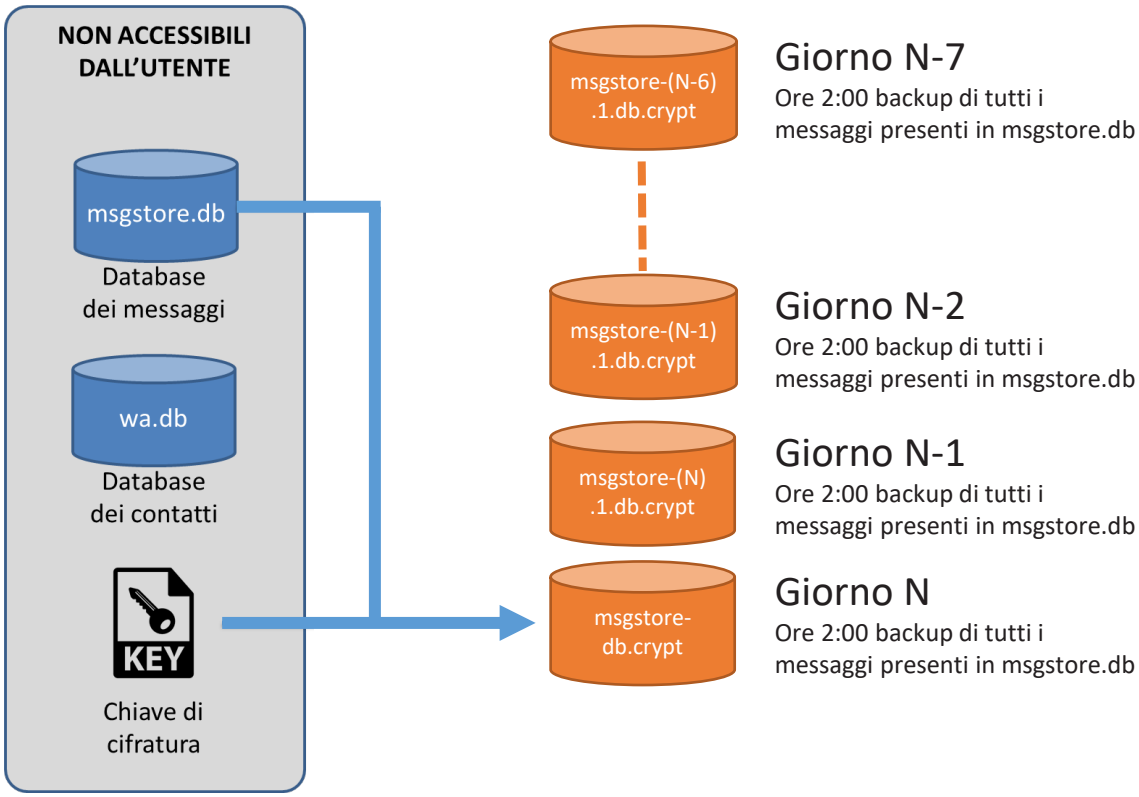

Figura 1: schema riassuntivo del meccanismo di memorizzazione dei messaggi di WhatsApp

di opportuni database *SQLite*. Il principale database dei messaggi è contenuto nel file **msgstore.db** che si trova all'interno di una cartella di sistema non accessibile all'utente (vedi Tabella 1).

L'applicazione WhatsApp effettua giornalmente il backup del file msgstore.db memorizzandolo all'interno di una cartella dell'utente in un file cifrato con suffisso variabile denominato **msgstore-db.crypt***. I files di backup

Contenuto	Posizione	File	Permessi
Database dei contatti	/data/data/com.whatsapp/databases	wa.db (SQLite)	Root
Database dei messaggi	/data/data/com.whatsapp/databases	msgstore.db (SQLite)	Root
Chiave di decifratura	/data/data/com.whatsapp/files	key	Root
Backup dei database messaggi	/SDCard/Whatsapp/Databases	msgstore-db.crypt* msgstore-(data).1.db.crypt*	Utente
File Multimediali	/SDCard/Whatsapp/Media		Utente

Tabella 1: descrizione dei principali database di WhatsApp rispetto ai classici SMS

Nella medesima cartella è contenuto anche il database dei contatti (**wa.db**) che associa il numero di telefono dell'utente al relativo nome del contatto presente in rubrica.
Nella tabella seguente è riportata la posizione, all'interno del filesystem, dei files dell'applicativo WhatsApp con l'indicazione dei relativi permessi necessari per potervi accedere. Laddove sia indicato **root,** l'utente dello Smartphone non ha normalmente possibilità di accesso.

sono cifrati con algoritmo simmetrico AES con chiave a 256bit (per i file crypto8). WhatsApp mantiene sul filesystem i backup dei files msgstore-db dei precedenti 7 giorni nominandoli msgstore-db(**data_corrente**).crypt*.

È possibile illustrare il meccanismo di salvataggio automatico dei files di backup di WhatsApp e relativo meccanismo di *naming dei files* mediante il seguente esempio dove si suppone che il software WhatsApp sia stato in-

Giorno	Data	Ora	File di Backup creati	Riferimento dei file di Backup
0	10/10/2018	16:00	Prima installazione: nessun files	Nessun backup
1	11/10/2018	02:00	msgstore-db.crypt8	Backup del 11/10/2018
2	12/10/2018	02:00	msgstore-db.crypt8	Backup del 12/10/2018
			msgstore-2018-10-12.1.db.crypt8	Backup del 11/10/2018
...				
8	18/10/2018	02:00	msgstore-db.crypt8	Backup del 18/10/2018
			msgstore-2018-10-12.1.db.crypt8	Backup del 11/10/2018
			msgstore-2018-10-13.1.db.crypt8	Backup del 12/10/2018
			msgstore-2018-10-14.1.db.crypt8	Backup del 13/10/2018
			msgstore-2018-10-15.1.db.crypt8	Backup del 14/10/2018
			msgstore-2018-10-16.1.db.crypt8	Backup del 15/10/2018
			msgstore-2018-10-17.1.db.crypt8	Backup del 16/10/2018
			msgstore-2018-10-18.1.db.crypt8	Backup del 17/10/2018
9	19/10/2018	02:00	msgstore-db.crypt8	Backup del 19/10/2018
			msgstore-2018-10-13.1.db.crypt8	Backup del 12/10/2018
			msgstore-2018-10-14.1.db.crypt8	Backup del 13/10/2018
			msgstore-2018-10-15.1.db.crypt8	Backup del 14/10/2018
			msgstore-2018-10-16.1.db.crypt8	Backup del 15/10/2018
			msgstore-2018-10-17.1.db.crypt8	Backup del 16/10/2018
			msgstore-2018-10-18.1.db.crypt8	Backup del 17/10/2018
			msgstore-2018-10-19.1.db.crypt8	Backup del 18/10/2018

Tabella 2: simulazione dei files di backup generati da una ipotetica prima installazione di WhatsApp effettuata il giorno 10 ottobre

stallato il giorno 10 ottobre 2018 alle ore 16:00. Nella Tabella 2, sono riportati i files di backup che saranno generati da WhatsApp. Come si può notare, trascorsi 8 giorni (18 ottobre 2018) verranno salvati tutti i backup dei 7 giorni precedenti (a partire dall'11 ottobre 2018).

In considerazione del meccanismo automatico di memorizzazione dei backup dei messaggi WhatsApp di figura 1 si possono verificare diverse situazioni in relazione ai messaggi cancellati. In particolare:

- **messaggi cancellati dopo che sia avvenuto un backup**: in tal caso i messaggi sono recuperabili integralmente in quanto sono presenti nel file di backup. Tuttavia, dato che WhatsApp mantiene i backup degli ultimi 7 giorni, per poter procedere al recupero non deve essere trascorso un arco temporale maggiore;
- **messaggi cancellati prima che sia avvenuto un backup**: in tal caso il recupero dei messaggi non è garantito, tuttavia è possibile tentare un recupero mediante un'operazione di c.d. carving del database non cifrato **msgstore.db**. È lo stesso principio che si applica per il recupero dei file cancellati e per tale ragione la probabilità di recupero dipende da quanto è stato utilizzato il dispositivo (in particolare l'area di memoria in cui i messaggi erano memorizzati).

Sebbene i files di backup si trovino in una cartella accessibile all'utente, il loro contenuto non è visibile né modificabile a causa della cifratura. L'unico modo per poter accedere al contenuto è di recuperare la chiave contenuta nel file **key** che, tuttavia, si trova all'interno di una cartella non accessibile all'utente (vedi Tabella 2).

L'accesso a tale file può essere ottenuto mediante una procedura di elevazione locale di privilegi (c.d. *rooting*) o in caso di impiego di software di estrazione forense dei dati, nel caso in cui il risultato sia una *copia fisica* o *full filesystem* della memoria del dispositivo.

2. Ordine di arrivo e di visualizzazione dei messaggi

I messaggi WhatsApp vengono salvati nei relativi database SQLite insieme a diverse informazioni. Tra esse di particolare interesse è l'orario di invio/ricezione dei messaggi. L'orario mostrato per i messaggi ricevuti fa riferimento al campo *timestamp* della tabella *messages* del database. **Tale orario è stabilito prendendo sempre come riferimento l'orario UTC di presa in carico dei messaggi da parte dei server di WhastApp**. L'applicazione di WhatsApp adatterà a *runtime* tale orario convertendolo in formato GMT ed applicando il fuso orario impostato nel sistema. Se il dispositivo in esame ha un orario non corretto, ad esempio perché ha disabilitata la sincronizzazione via rete internet, la visualizzazione dell'orario dei messaggi sarà influenza da tale discrepanza; tuttavia nel database WhatsApp sarà comunque memorizzato l'orario corretto UTC di presa in carico dei messaggi sul server.
Ciò implica che se l'orario del telefono sarà riposizionato correttamente, ad esempio riabilitando la sincronizzazione internet, WhatsApp correggerà automaticamente tutti gli orari dei messaggi a ritroso.

Effettuando su due cellulari la sequenza di azioni riportata in Tabella 3, come a simulare disservizi di rete, si evince come WhatsApp visualizzi i messaggi nell'ordine di arrivo (figura

Tabella 3: esempio di gestione dell'orario di WhatsApp 2018 alle ore 16:00

Orario	Telefono A	Telefono B
17:10	Invio messaggio a Tel. B: "Msg A ore 17:10"	Rete disabilitata
17:11	Rete disabilitata	Invio messaggio a Tel. A: "Msg B ore 17:11"
17:12		Rete abilitata Ricezione messaggio da Tel.A: "Msg A ore 17:10"
17:13	Rete abilitata Ricezione messaggio da Tel. B: "Msg B ore 17:11"	

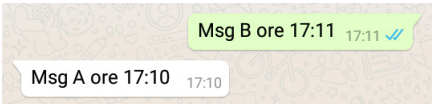

Figura 2: sequenza dei messaggi sul Telefono B così come mostrata da WhatsApp

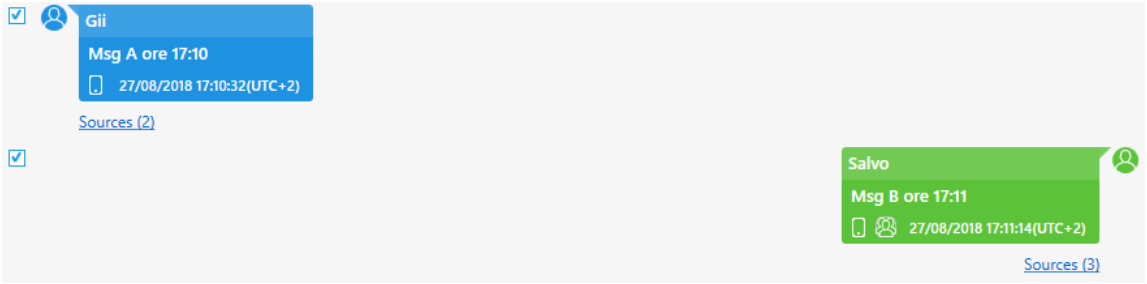

Figura 3: sequenza dei messaggi sul Telefono B così come mostrata da UFED

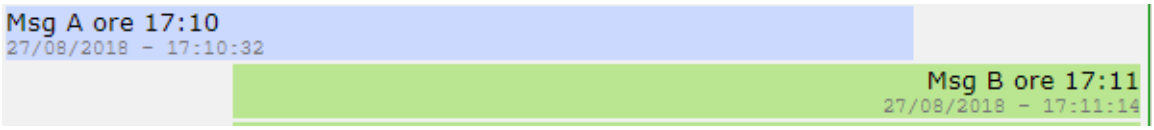

Figura 4: sequenza dei messaggi sul Telefono B così come mostrata da WhatsApp Viewer

2) e non nell'ordine temporale di invio degli stessi. Tale comportamento si verifica altresì con WhatsApp Web.

Analizzando il database msgstore.db del Telefono B si verificherà che l'orario del messaggio "Msg A ore 17:10" ricevuto alle ore 17:12 riporta, coerentemente con quanto detto precedentemente l'orario di invio ovvero le ore 17:10.

La stessa sequenza di messaggi viene invece riportata da UFED nell'ordine mostrato nella figura 3. Come si può notare, in questo caso, UFED mostra i messaggi nell'ordine temporale di invio (presa in carico) e non nell'effettivo ordine di visualizzazione dei messaggi sul dispositivo. Lo stesso comportamento si ottiene utilizzando il visualizzatore di database WhatsApp Viewer come riportato in figura 4.

3. Generazione della chiave di cifratura dei file di backup

Al fine di meglio comprendere il meccanismo di generazione della chiave di cifratura dei file di backup dei messaggi WhatsApp (studio condotto sulla versione 2.18.260) è stata realizzata una sperimentazione su un telefono di test (Telefono Samsumg Galaxy Grand Prime – Android versione 5.1.1) dove erano stati preventivamente abilitati, **tramite procedura di *rooting***, i permessi di root. In tal modo è stato possibile accedere al file contenente la chiave (/data/data/com.whatsapp/files/key) ed al file del database corrente dei messaggi (/data/data/com.whatsapp/databases/msgstore.db), normalmente non visibili dall'utente.
Da tale sperimentazione è emerso che:

1. a parità di numero di telefono, ad ogni installazione di WhatsApp viene generata una nuova chiave di cifratura per i successivi file di backup;
2. a parità di numero di telefono e di database di backup esistente, creato a partire dallo stesso numero di telefono, WhatsApp genera sempre la stessa chiave se l'utente sceglie di ripristinare le chat;
3. i backup WhatsApp, generati dalla stessa chiave, condividono una parte di informazione uguale per tutti. Per tale ragione è possibile verificare, anche senza chiave, se due o più backup WhatsApp sono stati generati dalla medesima chiave (e quindi dallo stesso numero di telefono);
4. a partire da un Backup WhatsApp sconosciuto, è possibile verificare se sia stato generato da uno specifico numero di telefono provando a ripristinarlo in una nuova installazione di WhatsApp su un dispositivo che abbia quello specifico numero di telefono.

In figura 5, è riportato lo schema secondo il quale sono generate le chiavi di WhatsApp. Da tale schema è possibile intuire come sia possibile rigenerare la chiave per un determinato backup di WhatsApp avendo a disposizione il numero di telefono ed il backup stesso.
Relativamente ai file generati dall'installazione WhatsApp è emerso che:

1. il database corrente dei messaggi WhatsApp msgstore.db viene cancellato

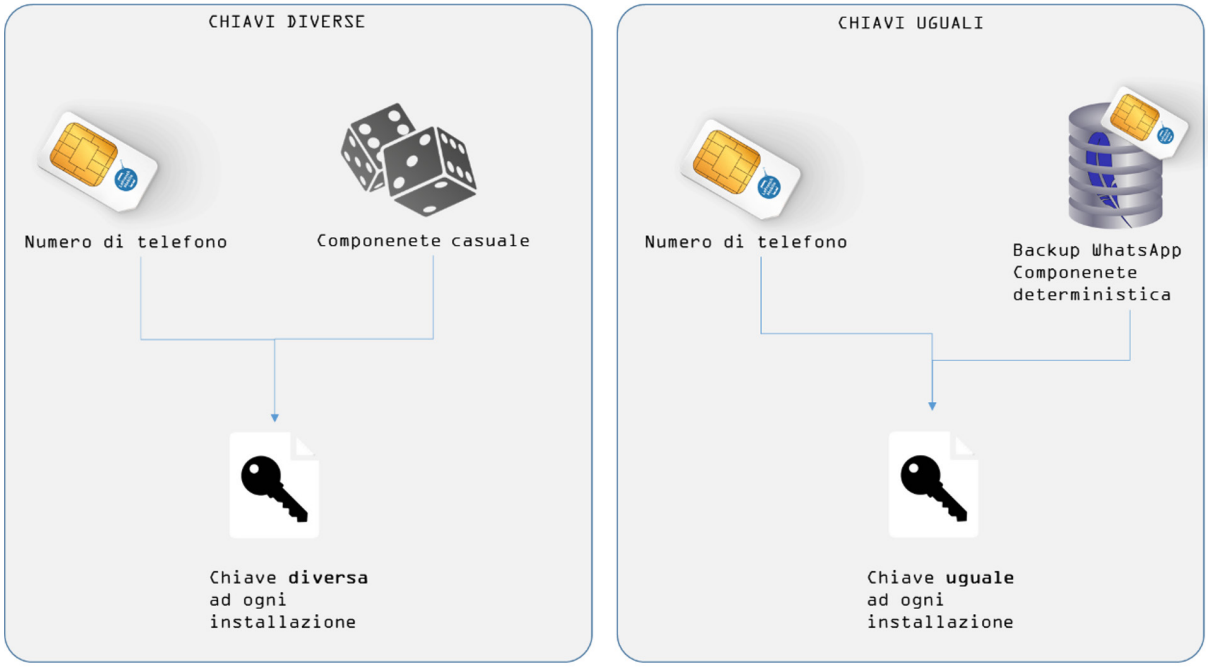

Figura 5: schema di generazione delle chiavi

quando si disinstalla l'applicazione. Ciò comporta la perdita dei messaggi scambiati dopo l'ultimo backup automatico o manuale;
2. il file della chiave key viene anch'esso cancellato al momento della disinstallazione di WhatsApp.

4. Decifrare i file di backup Crypt8

In maniera simile a quanto accade per i backup WhatsApp in formato crypto12, quelli in formato crypto8 sono cifrati con algoritmo a chiave simmetrica AES (Advanced Encryption Standard) a 256bit. La chiave è memorizzata all'interno di una cartella non accessibile all'utente:

/data/data/com.whatsapp/files/key

Viceversa i database di backup cifrati di WhatsApp sono memorizzati all'interno di una cartella accessibile all'utente:

/sdcard/WhatsApp/Databases/msgstore.db.crypt8

La procedura di decifratura è nota in letteratura. I passaggi da effettuare sono i seguenti:
1. estrazione della chiave a 256bit dal file key. La chiave si trova tra il byte 126 ed il byte 157 (32byte => 256bit);
2. estrazione del vettore di inizializzazione (Initialization Vector) che si trova all'interno del database cifrato tra il byte 51 ed il byte 66 (16byte => 128bit);
3. rimozione dell'header dal database cifrato. Prima di procedere alla decifratura del database è necessario rimuoverne l'header (primi 67 byte del file);
4. decifratura del database a partire dagli elementi sin qui creati (database cifrato senza header, chiave e vettore di inizializzazione).

Il database decifrato ottenuto è in formato SQLite.

5. Recuperare i messaggi da database corrotto

Nel caso in cui il database SQLite ottenuto risulti parzialmente corrotto nel suo contenuto è comunque possibile tentare un ripristino eliminando le informazioni non corrette.

Per recuperare il contenuto dei messaggi dal database SQLite si può procedere come di seguito:
1. Dump del contenuto del file msgstore.db in formato standard sql:
   echo .dump | SQLite3.exe msgstore.db > file_dump.sql
2. Rimozione delle righe nel file *file_dump.sql* i cui valori risultano *NULL*. Rimozione dell'ultima riga di *ROLLBACK*;
3. Ricostruzione del database in formato SQLite a partire dal file *file_dump.sql* importandolo all'interno del software DB Browser for SQLite. ©