DFIR Review

An Alternate Location for Deleted SMS/iMessage Data in Apple Devices

James McGee

Published on: Nov 03, 2022

URL: https://dfir.pubpub.org/pub/yp6efc8q

Synopsis

Forensics Question: Where can recent SMS/iMessage data be recovered when the data has been deleted from the sms.db and overwritten from the database free space?	
OS Version: iOS Versions 13.7, 14.0, 14.6, 14.7.1, 15.0.2	
Tools: Cellebrite UFED Version 7.49.0.2 Cellebrite Physical Analyzer Version 7.49.0.28 DB Browser for SQLite Version 3.12.0 on Windows HxD Hex Editor Version 2.5.0.0 on Windows	

By: James R. McGee, Digital Forensic Examiner, United States Army, james.r.mcgee70.mil@army.mil

Background:

Subjects and persons of interest in criminal investigations are likely to delete data from their devices in an attempt to remove evidence which could be incriminating. With this, the data still present or recoverable on mobile devices can be extremely beneficial to both Law Enforcement Investigators and Digital Forensic Examiners/Investigators. The sms.db should be reviewed when looking for SMS/iMessage data on an Apple mobile device; however, messages deleted by the user of the device can quickly be overwritten within the database. A new area of focus has been found within Apple devices using iOS 14.0 or later, specifically within the private/var/mobile/Library/Biome/streams/public/AppIntent/local file path of a Full File System Extraction. For the intent of being concise, the full file path above will be referred to as “the Biome directory” within this article. The Biome directory is a unique and available resource for the search of SMS/iMessage data present

within the sms.db and/or removed by the user of the device. The files copy message data of SMS/iMessages from the native Chats application and store this data for a finite period of time. Knowing where to locate and how to review the data within the Biome directory can provide data from communications even when the data no longer resides within traditional storage locations.

Location of the Biome Directory:

A request was made to examine a subject's mobile device pertaining to a narcotics crime. A Full File System Extraction of the subject's Apple iPhone 12, iOS 14.6, was obtained and subsequently reviewed in an attempt to locate conversation data between the subject and an undercover agent. A full transcript of the conversation was possessed by the undercover agent to aid in the investigation; however, only one message from the conversation was still present in the sms.db. The Archive File from the Memory Image of the device was searched for specific message body data from the conversation, through both ASCII and Unicode. Messages from the beginning and middle of the conversation were not located within the Archive File. A message from the end of the conversation was searched within the Archive File with two results from within the Biome directory. The Biome directory file was then thoroughly reviewed and another eleven messages from the transcript were located. The Biome directory was fully reviewed and an additional 21 messages from two conversations between the subject and two separate participants were located, which discussed apparent criminal activity involving narcotics. All messages except one had been deleted by the user of the device and were no longer available within the sms.db for review. These messages were only recovered through the Biome directory and manually recorded to provide to the originating office for their investigation.

Structure of the Biome Directory:

The files within the Biome directory are each 1MB in size. The files store data until the size capacity is reached at which point the next file is generated by the device for further storage. Data within the file is not edited or removed by the device until the file itself is overwritten. Throughout all extraction reviews, the average maximum files within the Biome directory at a time was five files. The timespan covered by the five files varies on device to device with the amount of user activity. On the long end, these files can cover up to four weeks of data from low user activity or one and a half to two weeks of activity for higher user activity. These files behave as a knowledge base while the device learns the user's most recent application usages, intents, and patterns of life to present current suggestions through Siri for future application usage. Each file is titled using a 15 digit naming convention reflecting the file creation timestamp in microseconds from midnight on 1 Jan 2001. For example, the "644529678155989" file possessed a creation date and last accessed date of 4 Jun 2021 8:01:18 PM(UTC+0). The creation date and date last accessed for each file are the same while the modified date matches the creation date of the subsequent file. i.e., File A has a creation date of 1 Jan 2021, 12:00:00 PM(UTC+0) and a modified date of 5 Jan 2021, 12:00:00 PM(UTC+0). File B then has a creation date of 5 Jan 2021, 12:00:00 PM(UTC+0).

An Advanced Logical File System Extraction of the test Apple iPhone 8 Plus, iOS 15.0.2, was conducted through Cellebrite UFED 4PC, which did not possess the Biome directory. A Full File System Extraction of the Apple device is required to gain access to the files within the Biome directory. You may be able to review and analyze the files within your extraction depending on the software used or it may be easier to save/export the files for review in another program. Once within the file, data can be narrowed using specific Application Identifiers, contact numbers, or contact entity information. An Application Identifier, also known as a Bundle ID within the Apple App Store, is a unique identifier for a specific application. For example, the Application Identifier for the Apple native Chats application is “com.apple.MobileSMS”. The data can also be searched for a contact number in both the “15551234567” and “1 (555) 123 4567” formats. Lastly, a specific contact entity can be used to narrow the data, such as “Steven”. These different ways of narrowing the data can reduce data analysis time or aid in meeting the specific scope of a Search Warrant, searching for “com.apple.MobileSMS” will yield all the SMS/iMessage sent and received communications within the file while a specific contact number could yield solely communication with one entity.

Testing to Determine Data Population for the Biome Directory:

Full File System Extractions of Apple iPhones possessing different iOS versions were obtained and reviewed for the presence of the Biome directory. It was determined that any iOS 13.7 and older did not have the file location. All iOS versions 14.0 and newer, to include iOS 15.0.2, which was the most recent iOS at the time of the examination, possessed the file location. A review of all iOS 14.0 updates was conducted to attempt to locate a change to account for the Biome directory added to Library which was inconclusive. The list of iOS 14.0 updates appeared too vague to determine the Biome directory addition. The hypothesis was formed that the Biome directory was directly related to Siri Suggestions through additional review of the directory itself and interface with test Apple iPhone devices.

Siri Suggestions “analyzes how you use your devices and apps to provide personalized suggestions and better search results using local, on-device processing, and syncs across your devices with end-to-end encryption using iCloud” [1]. All native applications, or applications preinstalled to the device, and all third-party applications, or applications installed by the user to the device, are by default selected to incorporate usage into Siri Suggestions. “Siri uses local, on-device processing to learn how you use your devices and apps in order to personalize your experience. Using information stored on your device, such as your Safari browsing history, emails, messages, images, notifications, and contacts, as well as information donated or contributed by other installed apps, Siri can suggest shortcuts and provide suggestions in searches, share sheet, calendar, Look Up, Visual Look Up, Safari, apps, and more.” [1]. The user of the device has the option to “see and control the full list of features that Siri personalizes and apps that Siri suggests shortcuts for in Settings > Siri & Search. To stop apps from contributing information to personalize Siri, go to Settings > Siri & Search and tap the app name, then tap to turn off Learn from this App” [1]. The likely unforeseen aspect of this feature is that message content of SMS/iMessage communications are also stored within the file structure of the Biome registry.

Removing the data from the native Chats applications does not impact the data within the Biome directory as the data is written during the initial SMS/iMessage action and stored by the phone until the Biome directory files are overwritten.

Control:

A Full File System Extraction of an Apple iPhone 8 Plus, iOS 14.7.1, test phone was obtained and the Biome directory was reviewed, prior to any additional changes to the device, through Cellebrite Physical Analyzer. A search was conducted for "com.apple" within the file which yielded 86 results. "Com.apple" is the prefix for the common Apple Application Identifier of native applications, or applications pre-installed by Apple on the device at initial startup. Of the 86 results from the "com.apple" search within the file there were zero results for "com.apple.MobileSMS".

Test One:

"Learn from this App" was deselected within "Messages" in "Siri & Search". The message 'Test iMessage' was sent from the Apple iPhone 8 Plus to another Apple iPhone. The second Apple iPhone received the 'Test iMessage' and replied with the message 'Reply', which was successfully received by the Apple iPhone 8 Plus test device.

A Full File System Extraction of the test device was obtained. A review of the corresponding Biome directory file revealed 20 search results for "com.apple.MobileSMS"; however, these results were different in format to those that will be reviewed further within this analysis which contain message body data and other key data values.

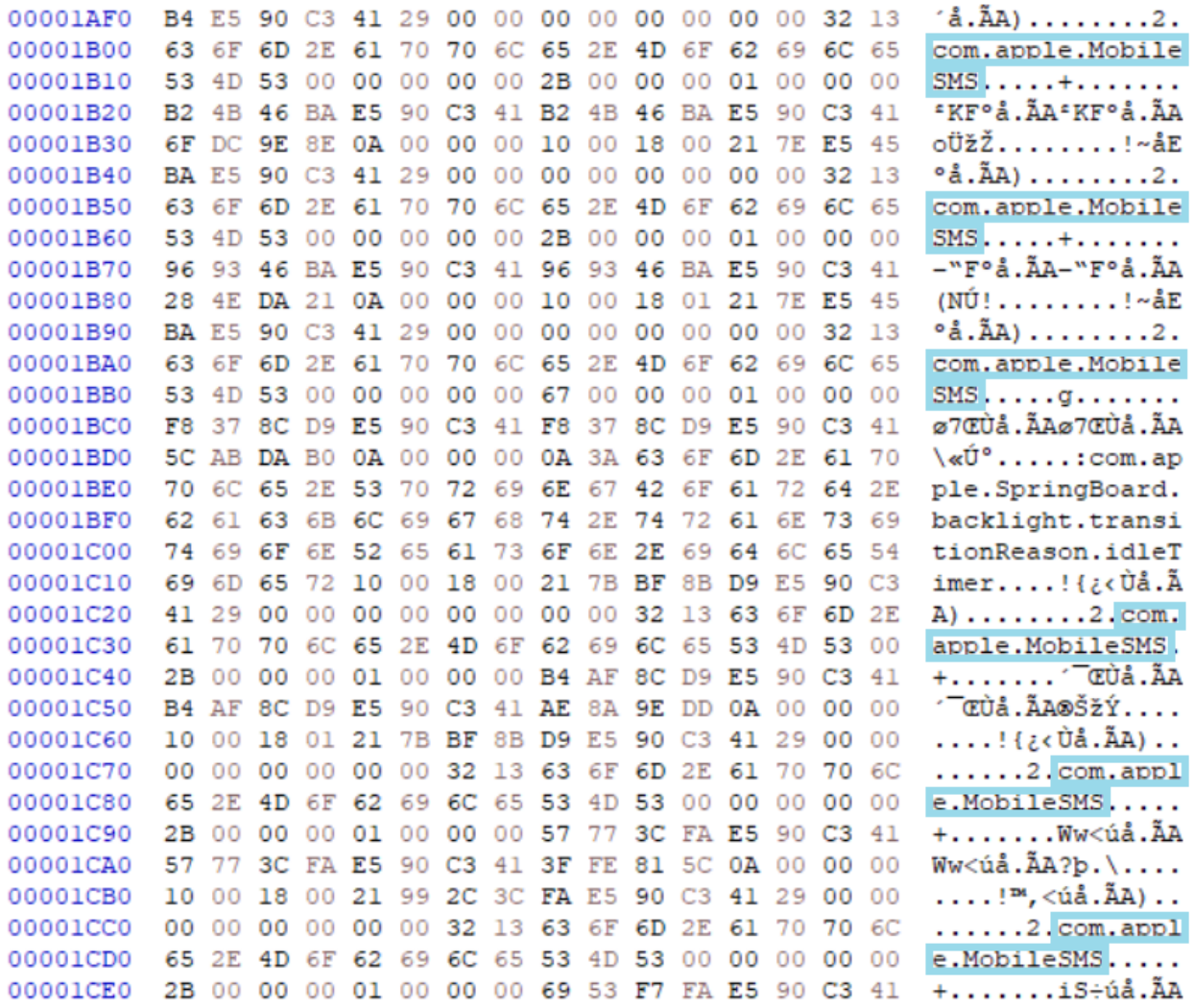


Figure 1 – Depicting the Chats Application Identifier, “com.apple.MobileSMS”, outlined in light blue.

Figure 1 above displays six of the “com.apple.MobileSMS” results through HxD, outlined in light blue. These identifiers appear to show activity within the Chats application but no pertinent data for the message body, message recipient, or timestamps of the message. The sms.db, sms.db-shm, and sms.db-wal files were exported and opened within DB Browser.

	Message	Account	Date	Message Direction
1	Test iMessage	P:+1706	2021-10-25 03:45:06	Incoming
2	Reply	P:+1706	2021-10-25 03:45:24	Outgoing

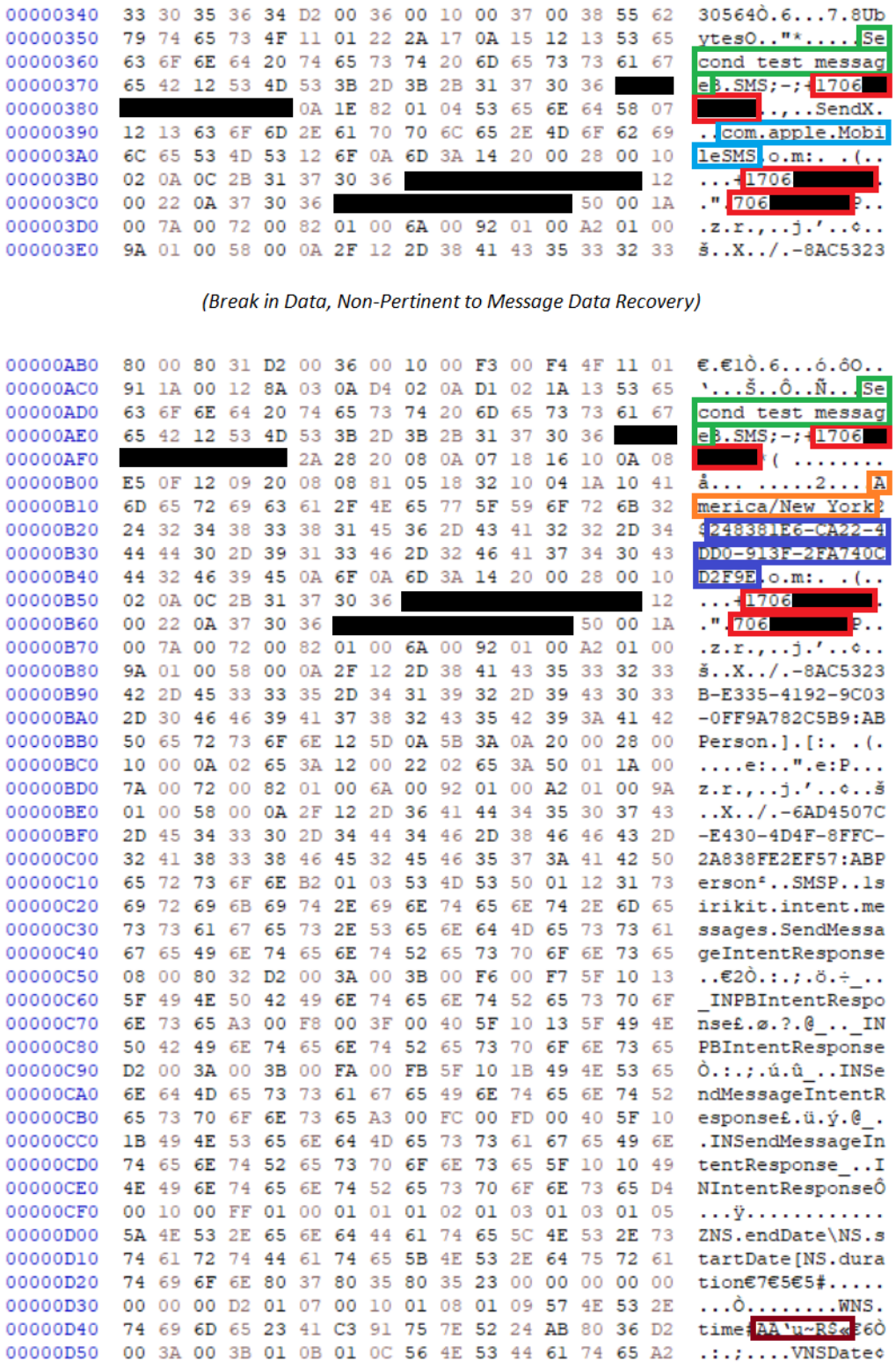
Figure 2 – Depicting all messages present within the sms.db.

Figure 2 displays the SQL query generated in DB Browser to obtain the message body, contact numbers (redacted), timestamp, and message direction from sms.db to verify the message content was extracted from the device. The messages sent and received on the test Apple iPhone 8 Plus were verified present on the device but the data was not within the Biome directory.

Test Two:

"Learn from this App" was selected within "Messages" in "Siri & Search". The message 'Second test message' was sent from the Apple iPhone 8 Plus to another Apple iPhone. The second Apple iPhone received the 'Second test message' and replied with the message 'Reply', which was successfully received by the Apple iPhone 8 Plus test device.

A Full File System Extraction of the test device was obtained. A review of the corresponding Biome directory file revealed four additional search results for "com.apple.MobileSMS". There were two "com.apple.MobileSMS" results for the outgoing message and the incoming message. The data reflected for Chats application messages solely encompassed the second sent message and second received message, data for the first sent and received messages was not present as the permission to "Learn from this App" was not applied at the time of sending and receiving.



(Break in Data, Non-Pertinent to Message Data Recovery)

Figure 3 – Depicting the Outgoing Message within the Biome File.

Figure 3 displays the pertinent hexadecimal and Unicode data in HxD, used to manually obtain the outgoing message within the file. This method is beneficial if the same message was deleted from the sms.db and could no longer be parsed by forensic software.

Reviewing the data we can see the following:

The message body “Second test message” is outlined in green.

The message recipient is outlined in red, redacted.

The Application ID “com.apple.MobileSMS” is outlined in light blue.

The Time Zone “America/New York” is outlined in orange.

> The chat identifier “248381E6-CA22-4DD0-913F-2FA740CD2F9E” from the sms.db-wal is outlined in purple.
>
> Lastly, the timestamp is outlined in dark red.
>
> This timestamp within the Hex is “41 C3 91 75 7E 52 24 AB”, which is converted to 22 OCT 2021 12:50:04 PM(UTC+0). This timestamp is in Apple Plist Time, which is the number of seconds since midnight, 1 Jan 2001, expressed in Hex. While the seconds since midnight, 1 Jan 2001, for the message was “656599804”, the same conversion expressed through Hex is “41 C3 91 75 7E 52 24 AB”. The conversion of the timestamp will be fully portrayed following all test descriptions within this article.

Manually obtained, this can be documented as “At 12:50:04 PM(UTC+0), 22 Oct 21, the user of the device sent ‘Second test message’ to (redacted).”

```
000013A0  45 43 36 41 45 D2 00 36 00 10 00 37 00 38 55 62  EC6AEÒ.6...7.8Ub
000013B0  79 74 65 73 4F 10 A7 2A 0A 0A 08 12 06 52 65 70  ytesO.§*.....Rep
000013C0  6C 79 20 42 12 53 4D 53 3B 2D 3B 2B 31 37 30 36  ly B.SMS;-;+1706
000013D0  [redacted] 0A 1E 82 01 04 53 65 6E 64  [redacted]..,..Send
000013E0  58 07 12 13 63 6F 6D 2E 61 70 70 6C 65 2E 4D 6F  X...com.apple.Mo
000013F0  62 69 6C 65 53 4D 53 12 2F 0A 2D 3A 14 20 00 28  bileSMS./.-:. .(
00001400  00 10 02 0A 0C 2B 31 37 30 36 [redacted]  .....+1706[redacted]
00001410  [redacted] 22 11 2B 31 20 28 37 30 36 29 20 [redacted]  [redacted]".+1 (706) [redacted]
00001420  [redacted] 50 01 58 00 52 2F 0A 2D 3A 14 20 00  [redacted]P.X.R/.-:. .
00001430  28 00 10 02 0A 0C 2B 31 37 30 36 [redacted]  (.....+1706[redacted]
00001440  [redacted] 22 11 2B 31 20 28 37 30 36 29 20 [redacted]  [redacted]".+1 (706) [redacted]
00001450  [redacted] 50 00 58 00 4A 03 53 4D 53 80 05  [redacted]P.X.J.SMS€.
```

(Break in Data, Non-Pertinent to Message Data Recovery)

```
00001AA0  F4 4F 11 01 16 1A 00 12 8F 02 0A D9 01 0A D6 01  ôO.........Ù..Ö.
00001AB0  1A 06 52 65 70 6C 79 20 42 12 53 4D 53 3B 2D 3B  ..Reply B.SMS;-;
00001AC0  2B 31 37 30 36 [redacted] 2A 28 20 08  +1706[redacted]*( .
00001AD0  0A 07 18 16 10 0A 08 E5 0F 12 09 20 08 08 DE 03  .......å... ..Þ.
00001AE0  18 32 10 26 1A 10 41 6D 65 72 69 63 61 2F 4E 65  .2.&..America/Ne
00001AF0  77 5F 59 6F 72 6B 32 24 37 45 36 35 35 36 38 38  w_York2$7E655688
00001B00  2D 45 34 30 46 2D 38 36 39 46 2D 30 35 30 38 2D  -E40F-869F-0508-
00001B10  36 37 32 39 32 44 32 32 31 43 44 32 0A 2F 0A 2D  67292D221CD2./.-
00001B20  3A 14 20 00 28 00 10 02 0A 0C 2B 31 37 30 36 [redacted]  :. .(.....+1706[redacted]
00001B30  [redacted] 22 11 2B 31 20 28 37 30 36 29  [redacted]".+1 (706)
00001B40  20 [redacted] 50 01 58 00 12 2F 0A  [redacted]P.X../.
00001B50  2D 3A 14 20 00 28 00 10 02 0A 0C 2B 31 37 30 36  -:. .(.....+1706
00001B60  [redacted] 22 11 2B 31 20 28 37 30 36  [redacted]".+1 (706
00001B70  29 20 [redacted] 50 00 58 00 B2 01  ) [redacted]P.X.².
00001B80  03 53 4D 53 50 01 12 31 73 69 72 69 6B 69 74 2E  .SMSP..1sirikit.
00001B90  69 6E 74 65 6E 74 2E 6D 65 73 73 61 67 65 73 2E  intent.messages.
00001BA0  53 65 6E 64 4D 65 73 73 61 67 65 49 6E 74 65 6E  SendMessageInten
00001BB0  74 52 65 73 70 6F 6E 73 65 08 00 80 32 D2 00 3A  tResponse..€2Ò.:
00001BC0  00 3B 00 F6 00 F7 5F 10 13 5F 49 4E 50 42 49 6E  .;.ö.÷_.._INPBIn
00001BD0  74 65 6E 74 52 65 73 70 6F 6E 73 65 A3 00 F8 00  tentResponse£.ø.
00001BE0  3F 00 40 5F 10 13 5F 49 4E 50 42 49 6E 74 65 6E  ?.@_.._INPBInten
00001BF0  74 52 65 73 70 6F 6E 73 65 D2 00 3A 00 3B 00 FA  tResponseÒ.:.;.ú
00001C00  00 FB 5F 10 1B 49 4E 53 65 6E 64 4D 65 73 73 61  .û_..INSendMessa
00001C10  67 65 49 6E 74 65 6E 74 52 65 73 70 6F 6E 73 65  geIntentResponse
00001C20  A3 00 FC 00 FD 00 40 5F 10 1B 49 4E 53 65 6E 64  £.ü.ý.@_..INSend
00001C30  4D 65 73 73 61 67 65 49 6E 74 65 6E 74 52 65 73  MessageIntentRes
00001C40  70 6F 6E 73 65 5F 10 10 49 4E 49 6E 74 65 6E 74  ponse_..ININtent
00001C50  52 65 73 70 6F 6E 73 65 D4 00 10 00 FF 01 00 01  ResponseÔ...ÿ...
00001C60  01 01 02 01 03 01 03 01 05 5A 4E 53 2E 65 6E 64  .........ZNS.end
00001C70  44 61 74 65 5C 4E 53 2E 73 74 61 72 74 44 61 74  Date\NS.startDat
00001C80  65 5B 4E 53 2E 64 75 72 61 74 69 6F 6E 80 37 80  e[NS.duration€7€
00001C90  35 80 35 23 00 00 00 00 00 00 00 00 D2 01 07 00  5€5#........Ò...
00001CA0  10 01 08 01 09 57 4E 53 2E 74 69 6D 65 23 41 C3  .....WNS.time#AÃ
00001CB0  91 75 8F 3D 66 8C 80 36 D2 00 3A 00 3B 01 0B 01  'u.=fŒ€6Ò.:.;...
00001CC0  0C 56 4E 53 44 61 74 65 A2 01 0B 00 40 D2 00 3A  .VNSDate¢...@Ò.:
```

Figure 4 – Depicting the Incoming Message within the Biome File.

Figure 4 displays the pertinent hexadecimal and Unicode data in HxD, used to manually obtain the incoming message within the file.

Reviewing the data provides the same and some additional data in comparison to an outgoing message:

The message body “Reply” is outlined in green.

> The message recipient (the Apple iPhone 8 Plus test device – extracted device) is outlined in red, redacted.

The message sender (sending Apple iPhone) is outlined in yellow, redacted.

> “P.X.R”, outlined in purple, represents the message direction and shows this is an incoming message.

The Application ID “com.apple.MobileSMS” is outlined in light blue.

The Time Zone “America/New York” is outlined in orange.

> Lastly, the timestamp is outlined in dark red. This timestamp within the Hex is “41 C3 91 75 8F 3D 66 8C”. Again this is in Apple Plist Time, and a convenient conversion capability for this time stamp is available through Doubleblak Digital Forensics’ website [2].

Manually obtained, this can be documented as “At 12:50:38 PM(UTC+0), 22 Oct 21, the user of the device received the message ‘Reply’ from (redacted).”

Test Three:

“Learn from this App” was again deselected within “Messages” in “Siri & Search”. The message ‘Third test message’ was sent from the Apple iPhone 8 Plus to another Apple iPhone. The second Apple iPhone received the ‘Third test message’ and replied with the message ‘Reply x3’, which was successfully received by the Apple iPhone 8 Plus test device.

A Full File System Extraction of the test device was obtained. A review of the corresponding Biome directory file revealed zero additional search results for “com.apple.MobileSMS”. The sms.db, sms.db-shm, and sms.db-wal files were exported and opened within DB Browser.

	Message	Account	Date	Message Direction
1	Test iMessage	P:+1706	2021-10-25 03:45:06	Incoming
2	Reply	P:+1706	2021-10-25 03:45:24	Outgoing
3	Second test message	E:	2021-10-26 00:10:04	Incoming
4	Reply	E:	2021-10-26 00:10:38	Outgoing
5	Third test message	E:	2021-10-26 03:54:34	Incoming
6	Reply x3	E:	2021-10-26 03:55:19	Outgoing

Figure 5 – Depicting all messages present within the sms.db.

Figure 5 displays the SQL query generated in DB Browser to obtain the message body, contact numbers (redacted), timestamp, and message direction from sms.db to verify the message content was extracted from the device. The messages sent and received on the test Apple iPhone 8 Plus were verified present on the device but the data was not within the Biome directory.

Breakdown of the Apple Plist Timestamp Conversion:

The Apple Plist timestamp is a conversion from a hexadecimal expression into the number of seconds since 1 Jan 2001. Here, we will show the steps converting “41C391757E5224AB” into a readable timestamp value.

To be more specific, the hexadecimal expression is a double precision 64-bit hexadecimal expression which is converted to a decimal floating-point value.

The first step is converting the hexadecimal expression into binary.

```
41C391757E5224AB  =  1000001110000111001000101110101011111100101001000100100101010111
  Hexadecimal                                        Binary
```

Figure 6 – Depicting the Hexadecimal to Binary Conversion.

This value is broken down prior to conversion to decimal.

Bit 63 Sign Bit	Bits 62—52 Exponent Value	Bits 51—0 Significand
0	10000011100	1.0011100100010111010101111110010100100010010010101011
	Decimal conversion minus the constant	*Decimal conversion of the significand*
	1052 - 1023 = 29	1.22301244110193851888

Figure 8 – Depicting the Separation of the Binary Value and Conversion into Decimal.

The Bit 63 Sign Bit is “0”, because the hexadecimal value is not negative.

Bits 62 – 52 of the total binary value is “10000011100”, which is also the binary value for the first three digits of the hexadecimal expression, “41C”. The binary value “10000011100” converted into decimal is “1052”. “1023”, a constant value, is subtracted from “1052” to obtain an exponent, which is “29” in this case.

Bits 51 – 0 encompass what is called the “significand”. The significand is composed of the remaining binary value with an added whole number before the decimal point. The remaining “391757E5224AB” of the hexadecimal is converted to a binary value which results in “0011100100010111010101111110010100100010010010101011”. From this, the significand is “1.0011100100010111010101111110010100100010010010101011”. The significand converted to decimal is “1.22301244110193851888”.

Putting it all together, we multiply “2” by the exponent “29” which is then multiplied against the significand.

$$2^{29} * 1.22301244110193851888 = 656599804.64174401759903481856$$

Figure 8 – Depicting the Final Mathematical Expression, Resulting in Seconds since midnight on 1 Jan 2001.

656599804.64174401759903481856 is the number of seconds since midnight on 1 Jan 2001 for our timestamp, or 22 OCT 2021 12:50:04 PM(UTC+0). It should be noted that all numbers after the decimal point in the final solution are not required for a conversion to the final timestamp, the date/time reached is the same with or without them, but these numbers are beneficial when converting the duration of a telephone call. Cellebrite Physical Analyzer expresses this value to seven decimal places is its conversion. Understanding of reaching the value to hundreds and thousands of decimal places is pertinent for verifying time values in the manner of call duration values. A convenient conversion capability for this time stamp is available through Doubleblak Digital Forensics’ website [2].

Conclusion:

The data still present or recoverable on mobile devices can be extremely beneficial to both Law Enforcement Investigators and Digital Forensic Examiners/Investigators as subjects and persons of interest in criminal investigations are likely to delete or remove data from their devices. When the sms.db can no longer provide the best evidence for the investigation the Biome registry can be reviewed for recent SMS/iMessage data. The ability to locate and review the Biome directory can turn an investigation and provide critical data available even when it no longer resides within traditional storage locations.

Sources/References:

1. Apple: https://www.apple.com/legal/privacy/data/en/siri-suggestions-search/

2. Doubleblak Digital Forensics: https://www.doubleblak.com/blogPosts.php?id=7#dosTime

DFIR Review

The data found in this paper is vital to the DFIR community. It offers a different approach to try to recover deleted data from recent SMS or iMessages, somewhere other than the SQLite sms.db database.

One reviewer reported finding different results during their testing. Another reviewer pointed out that there is the potential to recover significantly more data than just strings and dates. Common Data Types identified by Apple for this structure includes Contacts, Images, Monetary, Logical, and Files. Additionally, it is possible to create custom objects.

Future Work

Future work could include documenting the structure for files identified in the AppIntent directory. Additional work could also include determining what other data is being stored with the “com.apple.MobileSMS” label.

Reviewers

Jessica Hyde, David Loveall (subreviewer) (Methodology Review)

Johann Polewczyk (Methodology Review, Validated Review Using Reviewer Generated Datasets)

Cesar Quezada (Methodology Review)