User Relationship Classification of Facebook Messenger Mobile Data using WEKA

Amber Umair¹, Priyadarsi Nanda¹, Xiangjian He¹, and Kim-Kwang Raymond Choo²

¹School of Electrical and Data Engineering, University of Technology Sydney, Australia,

²Department of Information Systems and Cyber Security, The University of Texas at San Antonio, San Antonio, TX 78249-0631, USA

amber.umair@student.uts.edu.au

Abstract. Mobile devices are a wealth of information about its user and their digital and physical activities (e.g. online browsing and physical location). Therefore, in any crime investigation artifacts obtained from a mobile device can be extremely crucial. However, the variety of mobile platforms, applications (apps) and the significant size of data compound existing challenges in forensic investigations. In this paper, we explore the potential of machine learning in mobile forensics, and specifically in the context of Facebook messenger artifact acquisition and analysis. Using Quick and Choo (2017)'s Digital Forensic Intelligence Analysis Cycle (DFIAC) as the guiding framework, we demonstrate how one can acquire Facebook messenger app artifacts from an Android device and an iOS device (the latter is , using existing forensic tools. Based on the acquired evidence, we create 199 data-instances to train WEKA classifiers (i.e. ZeroR, J48 and Random tree) with the aim of classifying the device owner's contacts and determine their mutual relationship strength.

Keywords: Mobile Forensics, Social network information forensics, Weka

1 Introduction

Online social networks are a source of information, for example to profile an individual or group, to understand consumer sentiments on a particular topic, to detect an ongoing event (e.g. earthquake), to stay in touch (e.g. Facebook's Safety Check feature), etc [1]. In other words, such information can also be useful in a forensic investigation for both criminal cases and civil litigation. However, mobile device and app forensics is constantly playing catching up due to rapid changes in mobile device technologies [2, 3]. Compounding the challenge is the different formats used to store data on different devices [4, 5]. Unsurprisingly, mobile device and app forensics is an active research area. For example, the authors in [6] forensically examined 20 popular Android instant messaging apps and demonstrated how one can reconstruct message content, in different extent, from 16 of these 20 apps. Other researchers have also shown that a range of

artifacts relating to user activities (e.g. login, uploading, downloading, deletion, and the sharing of files) can be recovered from a mobile forensic investigations [7–9].

Facebook messenger is another popular application (app) where a Facebook user can have text, voice or video conversations with one or more other Facebook users (e.g. one-to-one or one-to-many conversations); thus, this is the focus of this paper.

Contribution 1: Specifically, we seek to demonstrate the artifacts that can be obtained from such an app when installed on an Android device and an iOS device. We use the Digital Forensic Intelligence Analysis Cycle (DFIAC) [10] to guide the forensic investigation and use existing commercial forensic tools (i.e. FTK access data, SQLite, IPhone Analyzer) to acquire the forensic artifacts from both devices. The original DFIAC model comprises the following steps:

- 1. Commence(Scope/Tasking)
- 2. Prepare
- 3. Evaluate and Identify
- 4. Collect/Preserve/Collate
- 5. Analyze
- 6. Inference Development
- 7. Present, Complete / Further Tasks identified

In [10], the authors exported the metadata reports from mobile devices, and the CSV, XLS and SLSX reports were collated and manually combined into a spreadsheet. Then, the spreadsheet was converted in Pajek format for analysis. To highlight the interconnections from the acquired data, a graph (e.g. Fruchterman reingold 2D link chart) can then be created and the information analyzed, for example to identify links between individuals in seemingly disparate cases. In this paper, we limit our investigation scope to messages from only the Facebook messenger app. For example in our iOS case study, the data was acquired from a real-world suicide incident, and we are able to determine the victim's relationship strength with other contacts based on factors such as number of messages exchanged in a day or week, and time and day of the messages.

Contribution 2: We also seek to demonstrate the utility of using machine learning to classify the device owners contacts with respect to relationship strength, from the obtained forensic artifacts. Thus, in step 6 of DFIAC (i.e. Inference Development), we train three WEKA Classifiers (i.e. ZeroR, J48 and Random tree) to efficiently classify the messenger contacts of the phone owner and determine their mutual relationship strength.

Paper's Roadmap: We will now explain how the remaining of this paper is structured. In Section 2, we present our case study, as well as our experimental setup along with the tools used. Section 3 explains how we can use machine learning to determine the device owners closest contacts or friends. The last section concludes this paper.

2 Case Studies

In this section, we will describe our two case studies, namely: an Android device (see Section 2.1) and an iOS device (see Section 2.2). We also remark that our case study Section 2.2 used the backup image from the iPhone of a real-world victim.

2.1 Android Device Case Study

Table 1 summarizes the equipment used in this case study.

Equipment	Version	Purpose
Samsung Galaxy S3	Android Version 4.3	Test device
ADB Android Debug Bridg	e Android Studio 2.3.2.	Android IDE
One Root	Version 1.0	Gain super user access
Root Checker	Version 6.1.7	Verify root access
Forensic Toolkit (FTK)	FTK Imager 3.1.2.0	Disk imaging program
Dell Laptop	Intel Core i7 Windows 10 Ent	Phone images Analysis

 Table 1. Experimental Setup

Device Preparation: To facilitate the creation of a physical image of the Samsung Galaxy S3 device, we root the device to gain super user privileges and verify root access using the freely available One Root and Root Checker software. Android Debug Bridge (ADB) is installed on the laptop so that we can issue shell commands to the device by connecting it using a data cable.

Test Data Creation: We then create the test data by installing Facebook app on the device. We also proceed to create a test Facebook user ID and undertake the following user activities on the device:

- Sign In. (Login Id and password entered via Facebook application)
- Remove phone number
- Add Henry gray as friend
- Upload post Time is flying
- Message sent to Henry via messenger app Hi Henry, Any Plans for the weekend.
- Comment on own post And I cant do anything about it.

Imaging of phone memory: To examine the device's image, we acquire the physical (i.e. bit-for-bit) image of the device's storage, and we know that the device's memory partitions contain user specific data and are of potential forensic interest.

- /system mmcblk0p9 is where read-only memory (ROM) is installed. Within the '/system' are a number of important folders that a user cannot normally access. For example, Location /system/app all where key ROM applications are located. Things like the device app and the messaging app /system/bin are where important binaries, which allow Android to execute the required commands, etc.
- /data mmcblk0p12 contains information about the installed app, such as SMS and emails. Key directories here are /data/app and /data/data, which are generally wiped when a device is set to the factory default.
- */cache mmcblk0p8* stores the temporary system data for everyday tasks, designed to expedite the system's access to apps.

Example artifacts of what we obtain from using FTK are depicted in Figures 1 to 5.

10fd91790	34	0A	00	00	CO	0A	00	00-C4	AO	00	00	C8	0A	00	00	4 · · · À · · · Ä · · · È · · ·
10fd917a0	98					0E		00-BC				E0	0E		00	·····* ··· 34 ··· · à ····
10fd917b0	5A	D5	FF	FF	04			00-16				42	6F	72	6E	ZÕÿÿ•••••Born
10fd917c0	20	6F	6E	20	41		72	69-6C	20	32		2C	20	31	39	on April 25, 19
10fd917d0	38							86-00	74	74		73	3A	2F	2F	85 ··= · · · https://
10fd917e0	6D	2E	66	61	63	65	62	6F-6F	6B	2E	63	6F	6D	2F	31	m.facebook.com/1
10fd917f0	30			31	37	33	36	30-35	33	33		36	33	2F	70	00017360533063/p
10fd91800	6F	73	74	73	2F	31		36-34	31	31			36	36	31	osts/10641188661
10fd91810	34	31	36	32	2F			00-16				42	6F	72	6E	4162/ · · · · · · · Born
10fd91820	20	6F	6E		41		72	69-6C	20	32		2C	20	31	39	on April 25, 19
10fd91830	38	35	00	00	00	00	8E	09-10	00	0C	00	00	00	00	00	85 · · · · · · · · · · · · · · ·
10fd91840	00	00	00	00	00	00	00	00-00	00	00	00	00	00	00	00	
10fd91850	00	00	00	00	00	00	00	00-00	00	00	00	00	00	00	00	

Fig. 1. User's birthday

04849ddc0	5F	66	69	65	6C	64	22	3A-7B	22	5F	5F	74	79	70	65	_field":{"_type
04849ddd0	5F	5F	22	3A	7B	22	6E	61-6D	65	22	3A	22	4D		73	":{"name":"Mes
04849dde0	73	65	6E	67	65	72	43	6F-6E	74	61	63		4E	61	6D	sengerContactNam
04849ddf0	65	22	7D	2C	22	5F	5F	74-79			6E	61	6D	65	22	e"},"typename"
04849de00	3A	22	4D	65	73	73	65	6E-67	65	72	43	6F	6E	74	61	:"MessengerConta
04849de10	63	74	4E	61	6D	65	22	2C-22		61	6C	75	65	22	3A	ctName", "value":
04849de20	7B	22	74	65		74	22	3A-22	48	65	6E	72	79		47	{"text":"Henry G
04849de30	72	61	79	22		7D		5D-2C	22	62		72			64	ray"}}]],"birthd
04849de40	61	79	44	61	79	22	3A	32-30	2C	22	62	69	72		68	ayDay":20,"birth
04849de50	64	61	79	4D	6F	6E	74	68-22	3A	31	32		22		73	dayMonth":12,"is
04849de60	50	61	72	74	69	61	6C	22-3A	66	61	6C	73	65	2C	22	Partial":false,"
04849de70	6C	61	73	74	46			63-68	54	69		65	22	3A	31	lastFetchTime":1
04849de80	35	30	35	31	39	32	36	37-37	36	31	33	2C	22	6D	6F	505192677613, "mo
04849de90	6E	74	61	67	65	54	68	72-65	61	64	46	42	49	44	22	ntageThreadFBID"

Fig. 2. User contact's birthday

4

0bd593e30 4F 4F 5F 4F																	
Obd593e50 37 33 36 30 35 33 30-36 33 06 5B 7B 22 65 6D 7360533063 · [{"em Obd593e60 61 69 6C 22 3A 6E 75 6C-6C 2C 22 77 73 65 72 5F ail":null, "user_ Obd593e70 6B 65 79 22 3A 22 46 41-43 45 42 4F 4F 4B A3 22 26 61 65 79 22 3A 22 24 44 14-3 45 42 4F 4F 4B 3A 32 22 20 00017360533063", mame": "James Wh Obd593e00 22 6E 61 65 79 22 3A 22 82 33 33 30 30 32 32 30 111"."user_"."James Wh iull:"user_key":" ull."user_key":" 0bd593e0 75 6C 6C 2C 27 79 2C-20 4E 16 65	0bd593e30	4F	4E	45	5F	54	4F	5F	4F-4E	45	3A	31	30	30	30	32	ONE_TO_ONE:10002
Obd593e60 61 69 6C 22 3A 6E 75 6C-6C 2C 22 75 73 65 72 5F ail":null,"user_ Obd593e70 6B 65 79 22 3A 22 46 41-43 45 42 4F 4F 3A 31 key":"FACEBOOK:1 Obd593e90 22 2E E6 61 65 22 3A 22 4A 61 65 73 20 57 6" name":"FACEBOOK:1 Obd593e90 22 2E E6 61 65 22 3A 22 4A 61 65 73 20 57 6" name":"James Wh Obd593e00 69 74 65 22 75 73 65-72 25 FG 65 79 22 3A 22 ull,"user_key":" name":"James Wh Obd593e00 75 6C 6C 22 75 36 51-72 27 55 56 61 65 22 30 2	0bd593e40	32	30	38	31	38	30	31	37-35	31	ЗA	31	30	30	30	31	2081801751:10001
Obd593e70 6B 65 79 22 3A 22 46 41-43 45 42 4F 4F 4B 3A 31 key":"FACEBOOK:1 Obd593e80 30 30 31 37 33 63 0-35 33 33 30 36 32 22 C 00017360533063", Obd593e90 22 6E 61 65 22 3A 22 4A 61 6D 65 73 20 57 68 "name":"James Wh Obd593e00 74 65 22 7D 2C 7B 22-65 6D 61 69 62 23 A 22 3A 22 ull, "user_key":" "name":"James Wh Obd593e00 76 6C 6C 22 27 73 65-72 27 73 23 33 30 30 32 32 30 Obd593e00 31 32 31 33 31 37 35 31-22 22 7D 5D 48 1801751", "na	0bd593e50	37	33	36	30	35	33	33	30-36	33	06	5B	7B	22	65	6D	7360533063 · [{"em
Obd593e80 30 30 31 37 33 36 30-35 33 30 36 33 22 22 00017360533063", "name":"James Wh 0bd593e90 22 6E 61 6D 52 23 A 22-4A 61 6D 65 73 20 57 68 "name":"James Wh 0bd593eb0 75 6C 6C 22 7D 7B 65-72 5F 6B 65 79 22 3A 22 0bd593eb0 75 6C 6C 22 7D 7B 65-72 5F 6B 65 79 22 3A 20 0bd593eb0 75 6C 6C 22 7D 7B 7B 7D 7D 2C 2B 2C 2B 2C 2B 2C 2B 2C 2B 2B 2D 2B 2D 2D 2D 2D 2D 2D 2D 2D	0bd593e60	61	69	6C	22	3A	6E	75	6C-6C	2C	22	75	73	65	72	5F	ail":null, "user_
0bd593e90 22 6E 61 6D 65 22 3A 22-4A 61 6D 65 73 20 57 68 "name":"James Wh 0bd593e00 69 74 65 22 7D 2C 7B 22-65 6D 61 69 6C 22 3A 6E ite"], {"email":n 0bd593e00 75 6C 6C 22 75 73 65-72 5F 6B 65 79 22 3A 22 ull, "user_key":" 0bd593e00 38 31 38 30 31 37 35 31-22 2C 22 6E 61 6D 65 22 3A 22 81801751", "name" 0bd593e00 38 31 38 30 31 37 35 31-22 2C 22 6E 61 6D 65 22 3A 24 56 6E 72 79 20-47 72 61 79 20 70 6C 61 1 herry, Any pla <t< td=""><td>0bd593e70</td><td>6B</td><td>65</td><td>79</td><td>22</td><td>3A</td><td>22</td><td>46</td><td>41-43</td><td>45</td><td>42</td><td>4F</td><td>4F</td><td>4B</td><td>3A</td><td>31</td><td>key": "FACEBOOK:1</td></t<>	0bd593e70	6B	65	79	22	3A	22	46	41-43	45	42	4F	4F	4B	3A	31	key": "FACEBOOK:1
0bd593ea0 69 74 65 22 7D 2C 7B 22-65 6D 61 69 6C 22 3A 6E ite"}, {"email":n 0bd593eb0 75 6C 6C 2C 22 75 73 65-72 5F 6B 65 79 22 3A 22 ull, "user_key":" 0bd593ec0 46 41 43 45 42 4F 4B-3A 31 30 30 32 32 30 BA 22 48 65 6E 72 79 20 70 6C 61 60 52 81801751", "name" "Henry Gray"] H i henry, Any pla 0bd593e10 6E 64 2E 77 20 74 68 20 77 65 65 66 67 72 20 74 68 20 76 66 67 72 20 74 68 62 62 67 66 67 72 20 74 68 62 67 67 <t< td=""><td>0bd593e80</td><td>30</td><td>30</td><td>30</td><td>31</td><td>37</td><td>33</td><td>36</td><td>30-35</td><td>33</td><td>33</td><td>30</td><td>36</td><td>33</td><td>22</td><td>2C</td><td>00017360533063",</td></t<>	0bd593e80	30	30	30	31	37	33	36	30-35	33	33	30	36	33	22	2C	00017360533063",
Obd593eb0 75 6C 6C 22 75 73 65-72 5F 6B 65 79 22 3A 22 ull, "user_key":" Obd593ec0 46 41 43 45 42 4F 4F 4B-3A 31 30 30 32 32 30 Obd593ec0 38 31 30 31 37 35 31-22 22 26 61 60 52 810751", "name" Obd593ec0 38 22 48 65 6E 72 79 2C-20 41 67 92 70 6C 61 in henry, Any pla Obd593f00 6E 64 2E 7B 22 65 6D 61-69 6C 22 3A 6E 75 6C 6C Obd593f10 6E 64 2E 7B 22 65 6D 61-69 6C 22 3A 22 46 41 <td>0bd593e90</td> <td>22</td> <td>6E</td> <td>61</td> <td>6D</td> <td>65</td> <td>22</td> <td>3A</td> <td>22-4A</td> <td>61</td> <td>6D</td> <td>65</td> <td>73</td> <td>20</td> <td>57</td> <td>68</td> <td>"name":"James Wh</td>	0bd593e90	22	6E	61	6D	65	22	3A	22-4A	61	6D	65	73	20	57	68	"name":"James Wh
Obd593ec0 46 41 43 45 42 4F 4B-3A 31 30 30 32 32 30 FACEBOOK:1000220 Obd593ed0 38 31 33 31 37 35 31-22 2C 22 66 65 22 Obd593ed0 3A 22 48 65 6E 72 79 20-47 72 61 79 22 7D 5D 48 13.1 % 16 79 20 70 6C 61 16 65 22 81801751", "name" "Henry Gray"}]H 1 henry Gray"]]H 1 henry	0bd593ea0	69	74	65	22	7D	2C	7B	22-65	6D	61	69	6C	22	3A	6E	ite"},{"email":n
0bd593ed0 38 31 38 30 31 37 35 31-22 2C 22 6E 61 6D 65 22 81801751", "name" 0bd593ee0 3A 22 48 65 6E 72 79 20-47 72 61 79 22 7D 5D 43 0bd593e00 69 20 68 65 6E 72 79 20-47 72 61 79 22 7D 5D 43 0bd593e00 62 73 20 66 67 72 20 77 65 65 68 68 67 20 74 66 520 77 65 68 68 67 74 76 65 68 66 67 72 20 74 66 75 65 68 66 67 72 75 76 65 68 67 66 67 72 20 74 66 67 62 62 34 24 74 75 <	0bd593eb0	75	6C	6C	2C	22	75	73	65-72	5F	6B	65	79	22	3A	22	ull, "user_key":"
Obd593ee0 3A 22 48 65 6E 72 79 20-47 72 61 79 22 7D 5D 48 :"Henry Gray"]]H Obd593ef0 69 20 68 65 6E 72 79 2C-20 41 65 79 20 70 6C 61 i henry Gray"]]H Obd593f00 6E 73 20 66 6F 72 20 74 68 65 65 65 65 66 67 20 74 68 65 70 20 74 68 65 66 67 72 20 74 68 65 66 67 72 20 74 68 62 62 73 20 66 67 72 20 74 68 62 62 64 64 22 75 75 65 65 65 65 65 67 62 24	0bd593ec0	46	41	43	45	42	4F	4F	4B-3A	31	30	30	30	32	32	30	FACEBOOK:1000220
Obd593ef0 69 20 68 65 6E 72 79 2C-20 41 6E 79 20 70 6C 61 i henry, Any pla 0bd593f00 6E 73 20 66 6F 72 20 74-68 65 20 77 65 65 66 65 ns for the weeke 0bd593f10 6E 64 2E 7B 22 65 6D 61-69 6C 22 3A 6E 75 6C 6C nd. {"email":null ,"user_key":"FAC 0bd593f20 2C 22 75 73 65 79 22 3A 22 46 41 3<"user_key":"FAC	0bd593ed0	38	31	38	30	31	37	35	31-22	2C	22	6E	61	6D	65	22	81801751", "name"
Obd593f00 6E 73 20 66 6F 72 20 74-68 65 20 77 65 65 68 65 ns for the weeke Obd593f10 6E 64 2E 7B 22 65 6D 61-69 6C 22 3A 6E 75 6C 6C nd. ["email":null 0bd593f20 2C 22 75 73 65 75 6E 6C 6C 6C nd. ["email":null 0bd593f30 45 42 4F 4F 4B 3 31 37 33 33 30 SEDOK:1000173605 0bd593f40 33 33 30 36 33 22 2C 22-6E 61 6D 65 22 3A 22 4A 3063", "name": "J	0bd593ee0	3A	22	48	65	6E	72	79	20-47	72	61	79	22	7D	5D	48	:"Henry Gray"}]H
Obd593f10 GE 64 2E 7B 22 65 6D 61-69 6C 22 3A 6E 75 6C 6C nd. {"email":null Obd593f20 2C 22 75 73 65 72 5F 6B-65 79 22 3A 22 46 41 43 ,"user_key":"FRC Obd593f20 42 4F 4F 4B 3A 31 30 31 37 33 36 30 5E EDOCM:1000173605 Obd593f40 33 33 30 36 32 22 22 22 3A 22 4A 33063", "name":"J	Obd593ef0	69								41						61	i henry, Any pla
Obd593f20 2C 22 75 73 65 72 5F 6B-65 79 22 3A 22 46 41 43 ,"user_key":"FAC 0bd593f30 45 42 4F 4B 3A 31 30-30 30 31 37 33 36 30 35 EBOOK:1000173605 0bd593f40 33 33 30 36 33 22 2C 22-6E 61 6D 65 22 3A 22 4A 33063", "name":"J	0bd593f00	6E			66	6F	72		74-68	65			65	65		65	ns for the weeke
Obd593f30 45 42 4F 4B 3A 31 30-30 31 37 33 36 30 35 EBOOK:1000173605 0bd593f40 33 33 30 36 33 22 2C 22-6E 61 6D 65 22 3A 22 4A 33063", "name": "J	0bd593f10	6E	64	2E	7B	22	65	6D	61-69	6C	22	ЗA	6E	75	6C	6C	nd.{"email":null
Obd593f40 33 33 30 36 33 22 2C 22-6E 61 6D 65 22 3A 22 4A 33063","name":"J	0bd593f20	2C	22	75	73	65	72	5F	6B-65	79	22	3A	22	46	41	43	, "user_key": "FAC
	0bd593f30	45	42	4F	$4 \mathrm{F}$	4B	3A	31	30-30	30	31	37	33	36	30	35	EBOOK:1000173605
Obd593f50 61 6D 65 73 20 57 68 69-74 65 22 7D 01 5E 74 94 ames White"} .^t.	0bd593f40	33	33	30	36	33	22	2C	22-6E	61	6D	65	22	3A	22	4A	33063", "name": "J
	0bd593f50	61	6D	65	73	20	57	68	69-74	65	22	7D	01	5E	74	94	ames White"} .^t.

Fig. 3. Private Facebook messages

L4c7cf3d0	33	22	2C	22	74	61	67	67-65	64	5F	69	64	73	22	3A	3", "tagged_ids":
L4c7cf3e0	5B	5D	2C	22	73	6F	75	72-63	65	5F	74	79	70	65	22	[],"source_type"
																:"native_timelin
L4c7cf400	65	22	2C	22	72	61	77	5F-6D	65	73	73	61	67	65	22	e","raw_message"
																:"Time is flying
																.", "publish_mode
																":"NORMAL","last
L4c7cf440	5F	65	72	72	6F	72	5F	64-65	74	61	69	6C	73	22	ЗA	_error_details":
																{"message":"","1
L4c7cf460	6F	67	5F	6D	65	73	73	61-67	65	22	3A	22	22	2C	22	og_message":"","

Fig. 4. Facebook status update and Comments

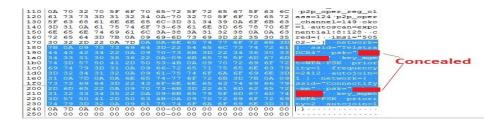


Fig. 5. WIFI and connectify details

2.2 iOS Device Case Study

The device of a teenager who had committed suicide was made available to the researchers for this research, in order to facilitate the determination of the mo-

tive and other factors relevant to the investigation. One of the evidence sources is the victim's iPhone backup files obtained from the victim's laptop. Therefore, artifacts were collected from the victim's iPhone 6 (iTunes version 12.0.1.26) backup. As the data is from an ongoing investigation, we anonymize the information to prevent the identification of the case or the individual(s) involved.

Tools used to obtain the artifacts from the iPhone are FTK Access Data, SQLite Forensic Explorer, Firefox SQLite manager and IPhone Analyzer 2.1.70. Password was not required to extract the personal data from the backup, which included contact numbers, call logs, phone messages, Facebook messenger data/ chats, notes, phone reminders /alarms, pictures, videos and audios. The iPhone Analyzer was able to pull out all details from the backup data, without requiring any passcode. Moreover, it was also able to export all data in the way it was organized on the victims phone. Figures 6 and 7 are a snapshot of what could be obtained from the phone's backup. For example, call logs and messages could be easily seen, browsed and exported. We concealed the phone numbers to protect the identity of phone owner. For the same reason, snapshots from other messages artifacts are not shown.

ze Files	Examine Files			
Kmarks File System	Device Infory/call_history.db ×			
in 8	Number	Date	Duration	Call Directi
	3249	Wed May 28 21:11:37 AEST 2014	Û	UNKNOW
	2249			
	69	Fri Aug 00 16:05:55 AEST 2014	02	UNKNOW
cemail		Pri Aug 08 19:25:45 AEST 2014		UNKNOW
	20			
	92	Sat Aug 09 08:19:30 AEST 2014	29	
		Gat Aug 09 08:20:53 AEGT 2014	10	UNKNOW
	7370	Mon Aug 11 05 54 49 AEST 2014	34	UNKNOW
		Mon Aug 11 05 55 44 APST 2014	27	UNKNOW
	77	Mon Aug 11 09:20:12 AEST 2014	5	UNKNOV
	77	Mon Aug 11 09/26/12 Aug 1 2014 Mon Aug 11 09/26/49 AEGT 2014	1	UNKNOV
	59	Wed Aug 13 18:08:23 AEST 2014	21	UNKNOV
nt	0.9		21	UNKNOW
ceived		Wed Aug 13 18:33:27 AEST 2014		
		Fil Aug 15 16:44:14 AEST 2014	88	UNKNOV
	70	Sat Aug 16 09:55:59 AEST 2014	0	UNKNO
0	70	Sat Aug 16 08:57:07 AEGT 2014	0	UNKNO
19 A	0989	Sat Aug 16 09:35:37 AEST 2014	6	UNKNOA
	0989	8at Aug 16 09:36:52 AEST 2014	8	UNKNO
	89			
oming	7270		12	UNKNO
	24			UNKNOA
		Set Aug 16 18 14 29 AEST 2014	10	UNKNO
	01 24 24		35	
ortcuts A		Sat Aug 10 19:19:00 AEST 2014	29	UNKNO
	61	Gat Aug 16 10:21:32 AEGT 2014	11	UNKNO
media	61	Sat Aug 16 18:25:15 AEST 2014	33	UNKNO
10-409C-848C-CDC107	77	Sat Aug 16 18 29 23 AEST 2014	30	UNKNO
	77	Sat Aug 10 18:52:41 AEST 2014	21	UNKNO
			21	UNKNO
	70	Thu Aug 21 08:24:49 AEST 2014	0	
	70	Thu Aug 21 08:24:53 AE5T 2014		UNKNO
		Thu Aug 21 12:35:16 AEST 2014	2	UNKNO
	3248	Sun Aug 24 15:46:13 AEST 2014	0	UNKNO
	77	Tue Aug 25 17:39:05 AEST 2014	60	UNKNO
	77	Tue Aug 25 19:11:27 AEST 2014	0	UNKNO
	69 77		61	UNKNO
		Fil Aug 29 06:27:18 AEST 2014	6	UNKNO
	77			
	70		22	UNKNO
		Pri Aug 29 15 36 32 AE5T 2014		UNKNO
	03	Sat Aug 30 10:55:37 AEST 2014		UNKNO
	77	Sat Aug 30 10 57:04 AEST 2014		UNKNO
	01	Mon Sep 01 12:49:19 AEST 2014	14	UNKNO
		Mon Sep 01 12:50:23 AEST 2014	14	UNKNO
	66 69	Mon Sep 01 12:50:23 AEST 2014 Mon Sep 01 16:34:59 AEST 2014	0	UNKNO
	69	Mon Sep 01 16:38:44 AEST 2014	0	UNKNO
			1949	
	62	Mon Sep 01 16:10:09 AEST 2014	1848	UNKNO
	69	Mon Sep 01 16:42:53 AEST 2014	320	UNKNO
	6135	Tue Sep 02 16:11:18 AEST 2014	2	UNKNO
	7370	Wed Sep 03 17:01:51 AEST 2014	3	UNKNO
	89	Wed Sep 03 17:02:58 AEST 2014	33	UNKNO
			2	

Fig. 6. iPhone Analyzer

Call logs and messages can be easily seen browsed and exported. The phone numbers are concealed to protect the identity of phone owner. Similarly messages artefacts snapshot is not shown.

3 Using Weka

In our case studies presented in the preceding section, one challenge we face is the difficulty in quickly pinpointing the more important evidences due to the different data formats, number of social apps on a device, etc. In addition, a realworld user will have possibly a number of identities for different social network accounts, a significantly larger number of contacts, etc. Thus, an investigation triage phase needs to be sufficiently robust.

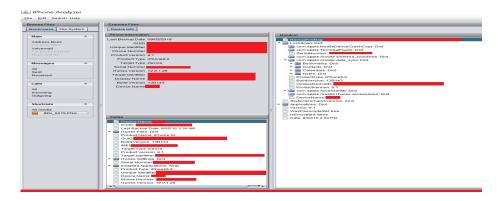


Fig. 7. iPhone Analyzer call details

We posit the importance of identifying strongly connected contacts of the device owner during a triage phase, for example by analyzing the social networking messaging app and their content. Therefore, to classify the contacts with respect to their relationship strength, we extract the data features from Facebook messenger app in our case studies. The focus is on the number of messages exchanged with a certain contact. Moreover, message exchange during certain times of the day / week (e.g. weekend) may be given a higher weight in determining relationship strength, depending on the context. In order to test the effectiveness of our approach, we analyze the message dataset of 199 instances which represent the message communication pattern of a user with his/her contacts.

Weka (Waikato Environment for Knowledge Analysis) [11] is used to determine the best performing classifiers among ZeroR [12], J48 Decision Tree [13] and Random Tree algorithms [14]. Specifically, we evaluate their performance on our dataset, based on the following key performance indicators: number of correctly identified instances, False Positive Rate(FP), Recall and F-measure.

- The correctly identified instances are the accurately classified instances, which indicate the precision of a classifier.
- FP measure denotes the number of examples predicted positive that are actually negative.
- Recall / sensitivity is the fraction of relevant instances that have been retrieved over the total amount of relevant instances.
 - Recall=True Positive / (True Positive + False Negative)
- F-measure is a measure of a test's accuracy. It is the harmonic mean of recall and precision.
 - F-measure = 2 * Recall * Precision / (Recall + Precision)

The features / attributes of our dataset are presented in Table 2. The J48 decision tree is the Weka implementation of the standard C4.5 algorithm. It starts from the training data, builds a predictive model in a tree structure. Its goal is to achieve optimal classification with a minimal number of decisions. The end nodes are the targets/classes.

 Table 2. Attribute Details

Attribute	Description
User	Phone owners Facebook contact/ friend id.A,B,C,D,E
Wavg	Weekly messages exchanged. Can be less than or greater than
	320. (64 msgs/day X 5 days = 320)
Weekend	Messages exchange on weekends 0-No messaging 1-Messaging on
	Saturday or Sunday 2-Messaging on both Saturday and Sunday
Relationship	Relationship type with phone owner W-Weak M-Medium S-
-	Strong

Random Tree Classifier is a supervised machine-learning classifier based on constructing a multitude of decision trees, choosing random subsets of variables for each tree, and using the most frequent tree output as the overall classification. We use this classifier, as it is known to correct for the J48 decision tree classifier over-fitting issue. In this method, a number of trees are grown (i.e. a forest). Variation among the trees is introduced by projecting the training data into a randomly chosen subspace before fitting each tree. Testing this algorithm on test data resulted in reduced correctly classified instances but the tree structure revealed more detailed decisions on the data attributes as shown in Figure 8.

wavg < 319.5 : W (92/0)	weekend <= 1
wavg >= 319.5	wavg <= 320: W (78.0)
weekend < 1.5	wavg > 320
weekend < 0.5 : W (42/0)	weekend <= 0: W (35.0)
weekend >= 0.5	weekend > 0: M (24.0)
user = A	weekend > 1
wavg < 320.5 : W (1/0)	wavg <= 319: W (25.0)
wavg >= 320.5 : M(12/0)	wavg > 319: S (37.0)
user = B : M (7/0)	
user = C : M (1/0)	
user = D	
wavg < 321 : W (3/0)	
wavg >= 321 : M(4/0)	
user = E : W (0/0)	
weekend >= 1.5 : S (37/0)	
Size of the tree : 16	Size of the tree : 9
Size of the tree. 16	Size of the tree. 9
Random Tree	J48 Tree

Fig. 8. Random Tree and J48 Tree

To evaluate performance of J48 decision tree classifier and random tree classifier, we compare their outputs to that of the ZeroR Classifier. ZeroR is the simplest classification algorithm and is based on frequency table. This classifier relies on the target/class only and ignores the features. It is useful for determining the baseline of a model. We analyze the data by using the following three test options using ZeroR, Decision Tree and Random Tree.

- Option 1: With K- fold cross validation(K=199)
- Option 2: With 66% Split data
- Option 3: With test data

3.1 Option 1: Classifiers with K- fold cross validation(K=100, 150, 199):

For K-fold, data is decomposed into K-blocks. Then, for K = 1 to X, the Kth block is made the test block and the rest of the data become the training data. Classifier is trained, tested, and then K is updated. Theoretically, the higher the number of folds, less biased results are achieved [15]. It is important that $K_i=X$, where X=no. of instances. In our dataset analysis, we use three different values of K=X=100, 150 and 199 to achieve unbiased results. ZeroR provides the baseline 69.3% accuracy for the model when used with K-fold cross validation for all three values of K (100, 150, 199). J48 classifier outperforms with a perfect correctly identified instances. Moreover, J48 classifier results remain consistent for all three values of K. The results with J48 also appears optimistic, therefore the same data are used with the random tree classifier, which results in 98.9% correctly identified instances with K=199. Similarly, other performance indicators like FP, Recall and F-measure are more realistic when using Random Tree. The changes in K value vary between the results of Random Tree classifier from 0.5% to 1%.

Table 3 summarizes the results with K-fold cross validation for all three classifiers.

Classifier	Κ	Correctly	FP	Recall	F-measure
		classified			
ZeroR	100	69.30%	0.693	0.693	0.568
	150	69.30%	0.693	0.693	0.568
	199	69.30%	0.693	0.693	0.568
J48	100	100%	0	1	1
	150	100%	0	1	1
	199	100%	0	1	1
Random Tree	100	98.40%	0.024	0.985	0.984
	150	99.40%	0.011	0.995	0.995
	199	98.90%	0.023	0.99	0.99

Table 3. Test Option 1: With K- fold cross validation(K=100, 150, 199)

3.2 Option 2: Classifiers With Split Data (50%, 66%, 80%)

Initially, we tested the classifiers on Weka default split value of 66%. By splitting the data of 199 instances in 66% means that 66% of data (131 instances) were used as training and 34% (68 instances) as test.

In this test option, our classifiers show significantly decrease in precision as compared to the K-fold cross validation, but J48 and Random tree still performs with an above 90% accuracy rate. We also analyze the behavior of all three classifiers by splitting the data in 50% and 80%. J48 and Random tree achieve accuracy rates of 100% and 97.50% respectively, at 80% of data splitting. However, ZeroR achieves the highest accuracy (69.30%) at 66% data splitting.

and lowest accuracy (62.50%) at 80% split data. Table 4 summarizes the results of all three classifiers with 50%, 66% and 80% split data.

3.3 Option 3: Classifiers With Test Data

In the third test option, we provide a separate test data to Weka, to check the performance of our dataset. In this test option, Random tree classifier results improves by 0.5% as compared to option 1 (K-folds) and 6.8% as compared to option 2 (split data). Therefore, on an average the performance of the Random Tree classifier improves by 3.65% when a new/unknown test data is introduced. The performance of ZeroR and J48 is almost identical to the first test (K-folds) – see Table 5.

Classifier	% split	Correctly classified	FP	Recall	F-measure
ZeroR	50%	67.70%	0.677	0.677	0.546
	66%	69.30%	0.693	0.647	0.49
	80%	62.50%	0.625	0.625	0.481
J48	50%	95.95%	0.085	0.96	0.957
	66%	94.12%	0.101	0.941	0.937
	80%	100%	0	1	1
Random Tre	e 50%	95.95%	0.085	0.96	0.957
	66%	92.60%	0.105	0.926	0.922
	80%	97.50%	0.042	0.975	0.974

Table 4. Test Option 2: With Split Data (50%, 66%, 80%)

Table 5. Test Option 3: With Test Data

Classifier	Correctly classified	FP	Recall	F-measure
ZeroR	69.3%	0.693	0.693	0.568
J48	100%	0	1	1
Random Tre	e 99.4%	0.001	0.995	0.995

4 Conclusion and Future Work

In this paper, we studied the potential of using machine learning classifiers to facilitate mobile forensics, specifically in terms of Facebook messenger artifact triaging. Specifically, after acquiring forensic artifacts from an Android device and an iOS device, we created 199 data-instances and trained three WEKA Classifiers (i.e. ZeroR, J48 and Random tree). This was done so that we were able

to classify the device owner's contact classification into weak, medium and strong (i.e. determine their mutual relationship strength). Our analysis with the three test options and three different classifiers revealed that J48 appeared to highly biased or overfitted to the provided dataset, and Random tree achieved optimal performance in all three test options with increased accuracy when tested with a different test dataset.

Future work includes extending this work to other classifiers as well as using a broader range of datasets.

Acknowledgments

The first author is supported by the Australian Government Research Training Program Scholarship.

References

- Umair, A., Nanda, P., He, X.: Online social network information forensics: A survey on use of various tools and determining how cautious facebook users are? In: 2017 IEEE Trustcom/BigDataSE/ICESS. (Aug 2017) 1139–1144
- Barmpatsalou, K., Cruz, T., Monteiro, E., Simoes, P.: Current and future trends in mobile device forensics- a survey. ACM Computing Surveys 51 (2018) 46
- Dezfouli, F.N., Dehghantanha, A., Eterovic-Soric, B., Choo, K.K.R.: Investigating social networking applications on smartphones detecting facebook, twitter, linkedin and google+ artefacts on android and ios platforms. Australian Journal of Forensic Sciences 48(4) (2016) 469–488
- Anglano, C., Canonico, M., Guazzone, M.: Forensic analysis of telegram messenger on android smartphones. Digital Investigation 23 (2017) 31–49
- Marturana, F., Me, G., Berte, R., Tacconi, S.: A quantitative approach to triaging in mobile forensics. In: 2011IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications. (Nov 2011) 582–588
- Walnycky, D., Baggili, I., Marrington, A., Moore, J., Breitinger, F.: Network and device forensic analysis of android social-messaging applications. Digital Investigation 14 (2015) S77 – S84 The Proceedings of the Fifteenth Annual DFRWS Conference.
- Daryabar, F., Dehghantanha, A., Choo, K.K.R.: Cloud storage forensics: Mega as a case study. Australian Journal of Forensic Sciences 49(3) (2017) 344–357
- Cahyani, N.D.W., Ab Rahman, N.H., Glisson, W.B., Choo, K.K.R.: The role of mobile forensics in terrorism investigations involving the use of cloud storage service and communication apps. Mobile Networks and Applications 22(2) (2017) 240–254
- Yang, T.Y., Dehghantanha, A., Choo, K.K.R., Muda, Z.: Windows instant messaging app forensics: Facebook and skype as case studies. PLOS ONE 11(3) (03 2016) 1–29
- Quick, D., Choo, K.K.R.: Pervasive social networking forensics: intelligence and evidence from mobile device extracts. Journal of Network and Computer Applications 86 (2017) 24–33
- 11. Azuaje, F.: Witten ih, frank e: Data mining: Practical machine learning tools and techniques 2nd edition. BioMedical Engineering OnLine 5(1) (Sep 2006) 51

- Lee, K., Palsetia, D., Narayanan, R., Patwary, M.M.A., Agrawal, A., Choudhary, A.: Twitter trending topic classification. In: 2011 IEEE 11th International Conference on Data Mining Workshops. (Dec 2011) 251–258
- Patil, T.R., Sherekar, S.: Performance analysis of naive bayes and j48 classification algorithm for data classification. International Journal of Computer Science and Applications 6(2) (2013) 256–261
- 14. Breiman, L.: Random forests. Machine learning 45(1) (2001) 5-32
- Refaeilzadeh, P., Tang, L., Liu, H.: Cross-validation. In: Encyclopedia of database systems. Springer (2009) 532–538

12