A new model for forensic data extraction from encrypted mobile devices

Fukami, A.; Stoykova, R.; Geradts, Z.

DOI
10.1016/j.fsidi.2021.301169

Publication date
2021

Document Version
Final published version

Published in
Forensic Science International: Digital Investigation

License
CC BY

Citation for published version (APA):
A new model for forensic data extraction from encrypted mobile devices

Aya Fukami\textsuperscript{a,c,*}, Radina Stoykova\textsuperscript{b,d}, Zeno Geradts\textsuperscript{a,c}

\textsuperscript{a} Netherlands Forensic Institute, The Netherlands
\textsuperscript{b} University of Groningen, The Netherlands
\textsuperscript{c} University of Amsterdam, The Netherlands
\textsuperscript{d} Norwegian University of Science and Technology, Gjøvik, Norway

1. Introduction

Mobile devices frequently contain data relevant to criminal investigations, and forensic analysis of those devices has become an increasingly critical investigative capability for law enforcement agencies. Over the last decades, various forensic science researchers have established methods and processes to extract evidence data from mobile devices in a forensically sound manner (Barmatsalou et al., 2013; Al-Dhaqm et al., 2020; Reedy, 2020). Those methods have been widely used for forensic purposes in real cases, and have tackled general challenges in mobile forensics, such as the lack of standardization within the mobile industry and the rapid rate at which mobile device technology changes. On the other hand, however, new challenges have recently been imposed by the strong security features in modern mobile devices (Chernyshev et al., 2017). Encryption, together with other security guard features has clearly created challenges for forensic investigators seeking to extract data from mobile devices seized at crime scenes. Those security features have disabled many of the data acquisition methods that have been used historically, and new methods to acquire data from modern mobile devices must be explored.

The challenges posed by encryption were publicly highlighted during the 2015 dispute between Apple and the FBI following the widely reported San Bernardino, California, terrorist attack. That case not only sparked an intense legal debate about the regulation of cryptography and governmental access to encrypted devices, but it also brought public attention to issues around the security and privacy of data stored on personal mobile devices. Not surprisingly,
mobile device vendors have been implementing higher levels of security features in their products to address personal data protection. Currently, in modern mobile devices, user data is highly secured from malicious access by unauthorized attackers as long as the user configurations are properly set up.

The impact of encryption on forensic analysis, as well as effective data acquisition processes has been widely researched in the computer forensics domain (Casey et al., 2011; Hargreaves and Chivers, 2008; Kornblum, 2009). It has been suggested that temporary files, data on volatile memory, metadata of encryption scheme, or access to the key management system can decrypt the target data, thereby allowing examiners to extract original data, which can then be used for criminal investigations. Challenges in data acquisition from encrypted mobile devices, however, come from the fact that those pieces of listed data are not accessible by default, requiring modification of the exhibit device. While some of the traditional forensic data acquisition methods are still effective, the target device needs to be directly unlocked and modified for effective data acquisition, which often requires invasive operations.

In this paper, we investigate modern mobile forensic techniques, and compare them with traditional mobile forensic techniques. Looking at the paradigm shift in mobile forensic techniques, it is clear that following the traditional forensic data extraction model is no longer effective. Therefore, a new model for forensic acquisition is proposed, and modern forensic data extraction techniques are evaluated in the context of the controversial, and underdeveloped regulation of encryption and governmental access to encrypted devices.

2. Background: paradigm shift in mobile forensics

Advanced technologies used in modern mobile devices have greatly impacted the effectiveness of mobile forensic techniques. In this section, we provide an overview of traditional mobile device forensic data acquisition techniques, discuss the widespread adoption of encryption and other security features in mobile devices, and then assess the impacts of those security features on traditional mobile forensic techniques.

2.1. Traditional mobile forensic techniques

Forensic data acquisition techniques have been researched for multiple mobile device platforms. Their forensic-soundness are evaluated prior to the implementation, and they are currently available through multiple commercial forensic tools (Barmpatsalou et al., 2013; Al-Dhaqm et al., 2020; Reedy, 2020). The acquisition techniques used in mobile forensics have been categorized using the classification system suggested by National Institute of Standards and Technology (NIST). The classification system includes the following five levels (Ayers et al., 2014; Chernyshev et al., 2017):

- **Level 1: Manual Extraction**
  
  An examiner directly manipulates the target mobile device using the device’s input interface (i.e., keypads and buttons), and records the content shown on the display of the device.

- **Level 2: Logical Extraction**
  
  Data (i.e., files and folders) on the target mobile device is extracted through communicating with its wired/wireless connection interfaces. The extracted data is human-readable since it is in a format that is recognizable by computer applications.

- **Level 3: Hex Dumping/JTAG**
  
  The full or partial raw data (hex dump) stored in the storage media on the target mobile device is acquired if the techniques categorized in this level are used. The debug interface on the target mobile device, such as JTAG (Join Test Action Group), is generally used to perform hex dumping. Techniques that can acquire raw data without hardware destruction are generally categorized into this level.

- **Level 4: Chip-off**
  
  Chip-off requires physical removal of the non-volatile memory chip from the target mobile device. An examiner can obtain an identical copy of the entire raw data of the target mobile device, which possibly contains remnants of deleted data.

- **Level 5: Micro Read**
  
  Micro read is a highly-specialized technique, where the stored data in non-volatile memory is extracted in electrical property form through the direct observation of the memory die inside the non-volatile memory chip.
  
  Data acquired through Level 1 and 2 techniques is usually called logical data, while data acquired via Level 3 to 5 techniques is called physical data and has the advantage of including remnants of deleted data. Generally, data parsing is required to present human-readable data after acquiring physical data.

The common understanding in traditional mobile forensic models has been that the higher the acquisition level, the higher the chance of forensic data recovery. As examiners use a higher acquisition level, the accessible range of data becomes wider. Furthermore, physical acquisition can bypass the user authentication mechanisms on smartphones such as pin-codes and passwords in the course of accessing stored data, and it does not require the target device to be in the normal-booting status. Therefore, law enforcement agencies have widely adopted chip-off data acquisition as the highest-level data extraction technique from various mobile devices. Note that even though micro read is ranked as the highest level in the above mentioned classification system, and although past research had proved that reading the data directly from the memory die is possible (Courbon et al., 2017), in practice, it is not regarded as the practical mobile data extraction technique in mobile forensics to the best of the authors’ knowledge.

2.2. Encryption and other security features in modern mobile devices

In order to protect user privacy and provide confidentiality of data, encryption techniques are currently implemented in modern mobile devices by default. Traditionally, in mobile devices, encryption techniques were applied at the application level in order to protect individual user data such as emails and photos. With the growing concerns over security and privacy, however, encryption techniques are now implemented at the system level with hard-coded unique passwords which are not accessible, even by device manufacturers. Therefore, mobile device data at rest is stored in an encrypted manner. Two types of encryption schemes are frequently used in mobile devices. One is Full Disk Encryption (FDE) and the other is File Based Encryption (FBE) (Loftus and Baumann, 2017). FDE is a technique where the entire user data partition is encrypted with a single encryption key, while FBE encrypts data per file bases with different keys, allowing files to be decrypted independently. In Apple devices, FDE was first introduced in iPhone 3 GS with iOS 3.X (Teufl et al., 2013). Apple devices...
with iOS versions higher than 8 use FBE. In Android devices, FDE was introduced in Android 4.4, and was supported up until Android 9. Starting with Android 7.0, FBE has been used as the standard encryption technique. Today, it is reported that more than 80 percent of the Android devices on the market are running on an Android version higher than 6 (Statista, 2013). This means that user data in the Android devices that are seized during the criminal investigation is now mostly encrypted.

In addition to encryption techniques, other “security by design” features are implemented in modern mobile devices. One example is Root of Trust (RoT). When a mobile device boots, each hardware and software component in the boot-chain is validated to ensure that only authorized components are executed on the system. If the validation fails due to unsigned software or for other reasons, the target device does not boot, denying access to the device by malicious users. This makes traditional data acquisition techniques such as those suggested by Vidas et al. (2011) unworkable. The Trusted Execution Environment (TEE), which is also heavily used, provides an isolated environment for security critical components in a system, by separating a normal operating system from a much smaller secure operating system, both running on the same hardware device. Hence a secure world and a normal world can co-exist on a system. ARM’s TrustZone technology is largely used in Android devices. While Apple uses a similar technology called Secure Enclave Processor (SEP) for isolating the cryptographic key and other sensitive information processing. When implementing the TEE, even “rooting”, or acquiring the highest privilege in the system does not allow access to the key data. By including those security features, mobile device manufacturers are protecting not only user data, but also their corporate proprietary data and technologies. As a result, users have little freedom to control their own mobile devices, and they are limited to using them within the device vendor's closed ecosystem.

2.3. Impact of security features on traditional mobile forensic techniques

As discussed, the popular use of encryption, along with complicated security measures on modern mobile devices, is impacting the capability of traditional forensic data acquisition techniques. The effectiveness of the five-level model of mobile forensic extraction techniques which we discussed in section 2.1 can be evaluated as follows in the presence of security features. Note that we assume that the user configurations are set up in a way to enable all the security features on the target device.

- **Manual Extraction**

  In order to perform manual extraction on a modern encrypted mobile device, an examiner needs to know and possess the legitimate user authentication credentials (i.e., pin-codes, passwords, or fingerprints), to properly unlock the target smartphone in a fully operating state. A proper control will display the user data on the target smartphone screen, and the examiner can record its contents using an appropriate recording device. The remaining problems are application security mechanisms for which access codes are needed.

- **Logical Extraction**

  The same requirements for manual extraction can be applied to logical extraction. Once an examiner can take control of the target data with correct user authentication credentials, then the examiner needs to proceed to modifying the system settings such as authorization of the debugging operation, in order to extract logical data through connection interfaces.

  - **Hex Dumping/JTAG**

    While JTAG and other debugging interfaces are still used on modern mobile devices, in many instances, those interfaces are disabled or locked before devices are shipped from the factory. Therefore, examiners may first need to find a way to utilize those debugging interfaces for hex dumping on the target device. Once enabled, hex dumping is still an effective data acquisition method to bypass the device lock. However, as the acquired physical data is in an encrypted state on modern smartphones, decryption procedures are required after data acquisition. The encryption keys are often derived from both the user defined access code, and a cryptographic key stored in the phone which is protected in a secure manner. This makes it only be used by authorized software on the device (Apple, 2020).

  - **Chip-off**

    Similar to hex dumping, chip-off lets an examiner acquire the physical data of the target device by bypassing the device lock. As discussed for hex dumping, however, the acquired data is unreadable until it is decrypted.

  - **Micro Read**

    Past research shows that reading memory data at die level is possible (Courbon et al., 2017). However, the miniaturization of the modern semiconductor fabrication process along with its ever-increasing capacities make this procedure impossible. Additionally, even if an examiner can successfully extract the contents of the non-volatile memory from the target mobile device, the data is encrypted. Techniques used in micro read may still allow examiners to extract key materials and analyze hidden security mechanisms from components on the target device, however it remains as an arduous task.

    Contrary to traditional beliefs, going higher in the five-level model is not necessarily more effective in forensic data recovery for modern smartphones. Unless decryption techniques are established, acquiring physical data does not yield meaningful data.

3. **Currently used data extraction techniques from encrypted mobile devices**

   In this section, current major forensic data extraction techniques from modern mobile devices, along with drawbacks with device security features, are introduced. While there are some exceptions in practice where more data extraction methods are available, for example when the target device is already “jailbroken” or “rooted”, we exclude those scenarios in this paper.

3.1. **Manual/logical extraction**

   In cases where an examiner can obtain the user authentication credentials required to unlock the device, or the target device is not locked, the examiner can manually manipulate the device, and perform manual or logical extraction. The user authentication credential required for unlocking the device could be a password, a passcode, pattern-drawing, or a biometric characteristic (fingerprint, voice, face, or other biometric features). If one of the biometric characteristics is used for user authentication, law enforcement investigators in some jurisdictions may be able to spoof the authentication by seizing and copying the fingerprint of the device owner, then use it to unlock the target device. Note that
in most cases biometric authentication only works if the target device is in After First Unlock (AFU) state, and not equipped with other advanced security features such as inactivity-time detection measures. AFU means that the target device is in a state where it has been turned on, and unlocked with user secret at least once after booting, and never turned off since then. When the target mobile device is in Before First Unlock (BFU) state (it has never been unlocked since last booting, or it is turned off), a password, a passcode, or pattern-drawing is required to unlock the device and enable the biometric authentication. Additionally, most biometric authentication methods have a limited timespan (e.g. 48 h for current iOS devices) in which biometric characteristics can be used before the BFU code would be required again. For unlocking the device, examiners should note that there is a “panic” password option available in some modern smartphones. When set up, the panic password can execute a hidden rule, such as wiping data, or disabling some functions of a phone. If the panic password was used instead of the legitimate unlocking password prior to data extraction, manual extraction would fail, and there is a great chance that the data is unrecoverable. Modern mobile devices are also equipped with anti-brute-forcing techniques. After a set number of failed authentication attempts with incorrect user authentication credentials, the device becomes unavailable for a set amount of time. In the worst case, data on the target device can be erased and become unrecoverable.

Once the target device is unlocked successfully, logical extraction can be performed by sending backup commands through user level communication interfaces on the device, such as USB, external storage, Wi-Fi, and Bluetooth. The target phone needs to be configured to accept commands from the connected computer for data extraction. On some modern mobile devices, rooting it (escalating the administrator privilege) is required. Data access management is generally controlled at the application level, and forensic software can use this function to copy selected app-relevant data to a connected storage device. However, in modern mobile devices, applications may choose not to be part of the backup operations supported by the OS. If the user data from an opted-out app is required for extraction, downgrading the app version on the target smartphone may allow examiners to extract the user data. However, since this operation directly modifies the target smartphone, it should be regarded as the last option.

3.2. File system extraction

When basic manual or logical extraction is performed for data acquisition, an examiner can only collect files and folders related to selected apps or communication protocols, and deleted data cannot be recovered. Traditionally, this is where mobile forensic examiners decide whether they proceed to physical acquisition or not. However, since most modern mobile devices use known file systems (i.e. APFS for Apple iOS devices, and ext4 for Android devices), and their data is stored on non-volatile memory in a file system structured format, acquiring full or partial file system data through non-destructive methods is currently a popular data extraction technique for forensic purposes. Compared to traditional logical extraction, file system extraction allows examiners to acquire more data, potentially including deleted data remnants. All data related to the apps is collected, and a forensic tool does not have to communicate and acquire individual data through an app-level API. An examiner can therefore access app-related databases, system files and logs. As long as the deleted data remnants remain in the database, an examiner can recover some deleted data through file system extraction. In order to conduct effective file system extraction, rooting the device is required. Without rooting, examiners can only acquire partial data, and data recovery may be limited.

3.3. Cloud data acquisition

Modern mobile devices store data not only on the physical device, but also on cloud servers provided by manufacturers or OS vendors. Indeed, since the physical device has limited storage capacity, some apps upload old data to the cloud server, and then delete it from the local storage. Once a law enforcement investigator acquires information required to access the cloud server from the target devices (i.e., user credentials) the investigator may access the cloud server, and collect information belonging to the target device. While some forensic tools already have cloud data acquisition capabilities, as this acquisition process requires the use of user credentials, as well as data transfer through the internet from different jurisdictions, court orders and other additional legal procedures are often required. Legal issues regarding this procedure are discussed in Section 5.

3.4. Bypassing device lock/extracting lock-related information

Accessing the user data stored in the internal memory in the locked and encrypted devices typically require unlocking with the correct user authentication credential. However, chances are that user credentials remain unknown to investigators in most cases. Moreover, as mentioned before, brute-forcing all the possible passcodes/passwords/patterns is not realistic due to the preventive technologies implemented on modern mobile devices as discussed in 3.1. Therefore, methods to either bypass or disable device locks of modern mobile devices have been explored by security researchers. Methods such as deleting the lock-related data on the target device, or modifying boot processes to skip the lock operation, have been developed in order to bypass the lock mechanisms and access the user data. In addition to disabling and bypassing the lock, methods to disable the timing restrictions against brute-forcing have also been explored, enabling the brute-forcing directly on the target device (Skorobogatov, 2016). When identifying lock or timing restrictions bypassing procedures, often times, system vulnerabilities are exploited (Fenollosa, 2019; Austinlog and Andro, 2015). Through exploitation, an examiner can brute-force the user authentication credential on the device itself, or extract intermediate information from the device which can be used for recovering the user authentication credential through computation on a designated system off the device. If the intermediate information only resides on the volatile memory on the target device, acquiring required information through vulnerability exploitation is only effective when the device is in AFU state.

3.5. Physical data extraction

Acquiring the physical data of the target mobile device lets examiners bypass its lock mechanism, and allows them to access the internal data directly. Since data decryption procedures are required on modern mobile devices after acquiring physical data, extensive reverse-engineering has been performed by security researchers to identify decryption methods. Through the authors’ experience, data decryption methods have been established for several models of modern mobile devices. For these models, physical data can be acquired through the methods described below.

3.5.1. Physical chip-off

Chip-off analysis (Willassen et al., 2005; Fukami et al., 2017; Breeuwsma et al., 2007) refers to a forensic operation where the memory chip of the target device is physically detached, and then the internal data is dumped for subsequent reconstruction of human-readable data. Detailed chip-off analysis procedures can be found in Breeuwsma et al. (2007). During chip-off, the non-volatile
memory chip is physically removed from the circuit board, and its content is extracted through the specialized reader. Since physical chip-off is a destructive procedure, it is important for an examiner to know if any other component on the board is required for decrypting the data. This is especially important if chip transplant procedures (Heckmann et al., 2018) need to be performed for severely damaged phones.

3.5.2. In-System-Programming (ISP)

While chip-off requires a destructive operation to the target device, if the required device pins for reading the target memory chip are accessible without detaching the chip itself from the circuit board, an examiner can perform In-System-Programming (ISP) for physical data extraction (Silveira et al., 2020). By connecting a memory reader to electrical traces connected to the memory chip on the circuit board, an examiner can access the memory chip and create a bit-by-bit copy of the target memory without damaging the operative state of the target mobile device. In order to successfully acquire data through ISP, the related part of the circuit board of the target device needs to be non-defective. In some cases, where no trace is available on the surface of the circuit board, partial chip decapsulation with laser ablation may be required to perform ISP. When performing ISP, an examiner needs to have a proper understanding of signal integrity and other electrical details. eMMCs (embedded Multi-Media Cards) and eMCs (embedded Multi-Chip Packages), which have been widely used in embedded devices, use single-ended signals, therefore simply connecting the traces may let examiners read the memory data. However, new memory technologies like UFS (Universal Flash Storage) use high speed differential signals (JEDEC, 2020). Performing ISP is therefore becoming challenging as making external connection on a circuit board can greatly disturb the signal integrity.

3.6. Data acquisition with custom boot loaders

If an examiner can load a custom boot loader into the target device during the boot process and run it, there is a great chance that the device can be manipulated by running arbitrary code, making physical data acquisition possible. Traditionally, loading a custom boot loader was enabled by the device manufacturer. Special modes (i.e., download mode or rescue mode) allowed users to run a custom boot loader on the target system during the boot-up. In modern devices, however, in order to maintain system integrity, manufacturers enable boot loaders to run only after they are properly verified to be signed, allowing only their codes to run on the device. The boot loaders are responsible for initializing hardware components and loading the operating system which then starts device operation including encryption. When a modern mobile device is powered on, multiple boot loaders are executed in chain. The first boot loader which is hard-coded in the ROM of the application processor is called bootROM or primary boot loader (PBL), and the one that is loaded by this bootROM is called the secondary boot loader (SBL). The SBL normally loads another boot loader that finally loads the operating system (Hay, 2017). Only when the verification processes are passed, is the boot loader loaded into the system memory, allowing the system to start the normal booting operations. Loading boot loaders through download mode is performed at the SBL level. The verification processes are usually done by checking if each boot loader is properly digitally signed. This process uses the initial verification key, which is stored in the one-time-programmable memory area in the application processor, thereby ensuring the key is never tampered with.

For some models of modern mobile devices, signed boot loaders may be publicly available (Hay, 2017). By flashing those boot loaders with known vulnerabilities into the target smartphone, an examiner may gain the highest privilege in the target phone, which in turn leads to full control of the device, allowing successful acquisition of the memory data. An examiner can also try to downgrade parts of the boot chain to lower versions as long as anti-rollback mechanisms are not implemented on the target mobile device. By doing so, the examiner can exploit known vulnerabilities that are fixed with security updates in the actual version of the boot chain. Nevertheless, the most powerful way of breaking into the boot chain to run the arbitrary code is to exploit the bootROM vulnerability, and this technique has been explored and used for accessing data in modern mobile devices (Katalov, 2019).

While modern mobile devices prohibit users from loading custom boot loaders, it is now widely known that PBL-level flashing is possible by booting the device into the processor-level special boot mode. The name of this boot mode is different by each manufacturer. It is called Emergency Download (EDL) for Qualcomm chipset, Device Firmware Update (DFU) mode for Apple chipset, and Download mode for MediaTek chipset. Those modes allow the phone manufacturers to flash software on their devices. Forensic examiners can thus utilize those modes and flash crafted boot loaders into the target smartphone, which helps them acquire user data without modifying it. Unless any additional authorization mechanism is implemented, a set of commands, a special cable, or hardware modifications make the target devices go into those special modes. Data acquisition using custom boot loaders is becoming popular since the same technique could work on wide range of devices with the same chipset, and it is typically hard for mobile device manufacturers to patch the vulnerabilities at processor level. Research has already proved that vulnerabilities on boot-loader level on popular chipsets can be useful for user data acquisition (Hay, 2017; Alendal et al., 2018).

4. Emerging techniques

In addition to the forensic data acquisition techniques described in the previous section, the following methods have been researched as possible techniques useful for forensic data extraction from modern mobile devices.

4.1. Side-channel analysis

When Integrated Circuits (ICs) operate on a circuit board, information related to these ICs may leak in the form of current flow or electromagnetic (EM) emanations. This information can sometimes be used to extract internal secrets such as cryptographic keys (Sayakkara et al., 2019). This type of analysis is called side-channel analysis (SCA), which has been a popular security research field for smart card and other security technologies. Recent work has proved that SCA can be used to retrieve a cryptographic key from the application processor in a modern mobile device (Vasselle et al., 2019). Although research is required for each application processor since the processors are unique, SCA is a promising technique for acquiring cryptographic keys from modern mobile devices. Once acquired, the key can be used to decrypt bootloaders. Meanwhile, in addition to shrinking technology size, device manufacturers are adding features like heterogeneous operation and voltage frequency optimization in order to minimize SCA vulnerabilities.

4.2. Fault injection

Fault injection is a technique where inputs of the controller device are manipulated for the purpose of causing illegitimate behaviors to the target system. Examples of fault injection techniques are glitching or underfeeding the power supply, transmission of electromagnetic signals, and injecting optical beams.
Research has already been performed to show the efficiency of fault injection for attacking the boot sequence and extract the code with the highest privileges from an Android device (Vasselle et al., 2020). Fault injection may also be useful for disabling the lock of debugging interfaces such as JTAG on the target device.

4.3. SoC reverse engineering

System on a Chip (SoC) die-level reverse engineering physically accesses inside SoCs on mobile devices, and examine the internal circuits using highly specialized lab equipment. Through SoC die-level reverse engineering, one can learn how the system is structured by checking internal circuit connections. A semiconductor die consists of multiple layers interconnected with each other. By delaying each layer, and translating the connection into a circuit, one can retrieve the overall design and try to learn and understand how the target system works. SoC reverse-engineering have been performed for multiple intentions, including piracy or counterfeiting reasons (Quadir et al., 2016). One key motivation for SoC die-level reverse engineering for forensic purposes is to retrieve hardware-bound key information, which is stored in the one-time-programmable memory area in a SoC, as discussed in section 3.6.

5. Legal issues related to modern forensic technologies

Since the data provided through forensic analysis may subsequently be relied upon in court, it is always important for forensic examiners to be aware of the legal framework regulating decryption for digital evidence acquisition. In a historical perspective, there are four legislative approaches for granting decryption powers to law enforcement — (i) exceptional access; (ii) decryption orders; (iii) vulnerability exploitation; and (iv) cloud data access. Details of each approach are discussed in this section.

5.1. Exceptional access

Methods providing law enforcement with exceptional access to encrypted data were proposed in the past and are related to backdoors in hardware and software, key escrow systems, and weak cryptography schema. Key escrow allows covert cooperation of independent parties with law enforcement to facilitate the use of the backdoor to decrypt the communication (ENISA, 2016). Examples of weak cryptographic algorithms are Simon and Speck (Beaulieu et al., 2015), which were rejected by the International Organisation for Standardisation due to discovered NSA-designed backdoors (Schneier, 2018). Currently, exceptional access is rejected both by legislators and security experts as it is imposing a high risk for human rights and civil liberties, especially with respect to data protection and privacy (Liguori, 2020), results in a golden age of surveillance, undermining security globally (Europol and ENISA, 2016), renders the systems vulnerable to attacks by criminals (Koops and Kosta, 2018), and requires significant development costs (Penney and Gibbs, 2017). ENISA and EUROPOL stated that backdoors and key escrow must be prohibited (Europol and ENISA, 2016). This means that technical cooperation between mobile device manufacturers and LEAs is currently unlikely, even for forensic data extraction.

5.2. Decryption orders

In order to address encryption challenges in criminal investigations without exceptional access, multiple countries introduced decryption orders. Such orders allow compelled disclosure or assistance by service providers or manufacturers (Lewis et al., 2017), and the orders are enforced with penalties in some countries. In the United Kingdom and France, refusal to disclose the encryption key can lead to criminal penalty. Similarly service providers carry civil liability, and in Belgium even criminal liability for failing to comply with obligations to assist law enforcement in criminal investigations (Walden, 2018). Norway is one of the first countries to update its legislation in 2017 allowing law enforcement to obtain biometrics for unlocking devices (Koops and Kosta, 2018). Less trivial is the question of compelled disclosure of password by suspects since unlike biometrics, a password does not exist independent of the suspect’s will. Compelled disclosure for suspects is likely to remain controversial and for exceptional cases (Koops and Kosta, 2018) since it creates concerns about the privilege against self-incrimination, right to silence and abuse of state power. Moreover, this solution is unsuitable when the user was unidentified, unable or unwilling to provide the key (Penney and Gibbs, 2017; Shah, 2015). Decryption orders for cooperation with providers or manufacturers also have significant drawbacks for privacy and security. Given that backdoors are forbidden it is hard to understand how providers must comply with a requirement to decrypt communication in transit or at rest. Propositions for in-house digital forensics by providers and manufacturers are also dubious considering that law enforcement agencies will be provided only with the decrypted data without information on the used forensic method, its reliability and the accuracy of the results.

5.3. Vulnerability exploitation

Considering the manifold drawbacks and limitations of exceptional access and compelled disclosure, new legislation regulating “lawful hacking” has already been introduced in several countries (Gutheil et al., 2017). Most types of lawful or governmental hacking are considered exceptional and highly intrusive. We focus here on exploitation of known system vulnerabilities since it proves to be highly useful for mobile evidence acquisition and less-intrusive in comparison to interception or development of malware. Vulnerability exploitation is broadly understood as use of any type of vulnerability, including social engineering and side-channel analysis. According to law, exploiting vulnerabilities must be employed by law enforcement agencies as a last resort after other less intrusive investigation measures have failed (Liguori, 2020). This is understandable since such practice creates an increased risk for privacy violations and data leakage, may undermine security, requires a vulnerability disclosure framework, and has international economic, political and technological effects (Budish et al., 2018; Liguori, 2020). However, as demonstrated in section 3 and Section 4, in practice often times this is the only viable solution to access encrypted mobile phone. Liguori argues that a legal framework for lawful hacking must be developed to address the following key issues: (i) legal concept/scope; (ii) prerequisites for deployment; (iii) development and sharing of hacking tools; (iv) accountability and disclosure of vulnerabilities; and (v) jurisdictional issues (Liguori, 2020).

The existing few national legislations on lawful hacking provide some safeguards for human rights and to prevent abuse of power by law enforcement (Gutheil et al., 2017). Ex ante safeguards include judicial authorization and limiting the measure by crime type and duration. Importantly, ex post control includes strict reporting and oversight of lawful hacking, as well as notification of targets of hacking practices and remedies in case of abuse of powers. However, new lawful hacking regulation in the US, France, Australia, and Germany still faces major challenges related to lack of vulnerability disclosure processes and ensuring transparency and accountability of law enforcement agencies (Liguori, 2020).

A report on government disclosure processes in Europe stated that only a limited number of countries have a transparent...
procedure for vulnerability disclosure (Pupillo et al., 2018). Similar to the US Vulnerabilities Equities Process (VEP), the report recommends adoption of procedures by all law enforcement agencies, where they have: (i) an obligation to report vulnerabilities; (ii) may only temporarily restrict knowledge of a vulnerability; and (iii) an oversight body ensures compliance. In the absence of clear legislation in relation to obligations of law enforcement to disclose vulnerabilities, to whom and under which condition, forensic examiners might be put in a position to take legislative decisions. This inevitably leads to undesirable practices. Law enforcement might be reluctant to disclose vulnerabilities to providers or users in order to exploit them further for evidence acquisition. Even on trial proceedings, investigators might be unwilling to disclose sensitive investigation methods related to security flaws in systems, and unfortunately we have seen the use of alternative explanations for how evidence was found, a practice known as “parallel construction.” (Criminal Legal News, 2018; Human Rights Watch, 2018).

Consequently, exploiting known vulnerabilities for mobile forensics is a justifiable and reasonable approach, as long as it is strictly regulated and assures protection of civil rights and liberties. As examined, very few countries have regulation in place and due to the international effects of such activity a European Regulation or an International treaty might be preferable (Budish et al., 2018). Unlike known vulnerabilities for access to evidence, the use of zero-day exploits might create vulnerability market for law enforcement (Liguori, 2020). Therefore, they are unlikely to meet the requirements of proportionality and subsidiarity, and might be permitted only for serious crimes or terrorism (Koops and Kosta, 2018).

5.4. Access to cloud evidence

Since data from phones is often copied to cloud storage and duplicated in multiple back-ups, lawful access to cloud data is another alternative for law enforcement to obtain information, by directing a search order to the cloud provider (Pell, 2016; Walden, 2018). Cloud data from mobile devices is a rich source of evidence, however the legal and technical challenges for law enforcement are not trivial. So far, only the United States (US) has introduced legislation to regulate lawful access to cloud storage. According to the CLOUD Act (The CLOUD Act, 2018) foreign governments can compel US-based Cloud Service Providers (CSPs) to directly disclose stored data or intercept communications in real time, if they have entered a bilateral agreement with the US government.1 The CLOUD act explicitly states that it shall “not create any obligation that providers be capable of decrypting data (Walden, 2018). Therefore, in case of ‘zero knowledge privacy’, meaning that the provider never knows the plain text content of the data being stored, law enforcement must rely on other techniques to decrypt the data themselves. In the European Union (EU), there is a pending proposal for a Regulation on European Production and Preservation Orders for electronic evidence in criminal matters, that would allow law enforcement access to service provider datam including in encrypted form, but the proposal has been on hold for the past two years (Sippel, 2021). The European data protection board criticized it for lack of sufficient safeguards (Board, 2020). Despite the struggles to establish EU-based e-evidence regime, the EU Commission entered further negotiations with the United States to reach an agreement on cross-border access to electronic evidence for judicial cooperation in criminal matters (Council of the European Union, 2019), that might result in deepening the existing legislative loopholes. In the US, the CLOUD act imposes further challenges since there is no clear procedure to ensure that data disclosure to a foreign government meets the requirements laid out in the bill while service providers are inappropriately empowered to mediate between their business interests, human rights, and law enforcement interests (Abraha, 2019). The legislation is broadly criticized on the grounds that it (i) fails to clarify who should be subject to a search warrant in a layered cloud service arrangement; (ii) does not define digital evidence, categories of data, and types of “serious crimes” where cloud access is justified; (iii) lacks judicial review; and (iv) has weak protection of privacy and procedural rights.2 Moreover, from a cloud forensics perspective, there are specific risks to the reliability of cloud evidence related to remote acquisition, reliance on CSP assistance, loss of volatile data in virtual machines, and encryption (Zawoad and Hasan, 2013; Pichan et al., 2015). It should also be noted that due to synchronization issues cloud back-ups might not contain all the data available in mobile phones (Jacobsen, 2017). Currently, neither the EU nor the US legislator mention any requirements for reliability of digital evidence or digital forensics procedure. Same gap is identified in the proposed Second Additional Protocol to the Cybercrime Convention (Council of Europe, 2018). As of today, a EU–US consensus has not been reached, and the patchwork legislation shows the need for an international treaty for regulating encryption, access to cloud data, and digital evidence exchange according to internationally-agreed digital forensic standards.

5.5. Alternative solutions

Some authors discuss alternatives to the existing types of legislation. Proposals include (i) restricting the design, use and sale of encryption; (ii) improving law enforcement data analytics capabilities, or regional decryption labs (Lewis et al., 2017); and (iii) criminalization of the supply, possession or use of cryptographic technologies for criminal conducts. However, most of them suffer unfavorable limitations. Law enforcement agencies need to develop new decryption methods continuously, which is time and resource consuming. Moreover, law enforcement agencies can hardly compete with new security by default solutions included in mobile devices and operating systems. As discussed in section 4, law enforcement examiners need to keep performing reverse engineering to access encrypted phones. Reverse engineering is an indispensable method for law enforcement in order to correctly interpret the system structure, security features, file systems, and other software details for the purpose of evidence acquisition and tool testing. However, current legislation insufficiently addresses the tension with vendors’ intellectual property and trade secrets protection and the need of law enforcement to perform reverse engineering to collect digital evidence. Moreover, legislation often does not addresses issues with reverse engineering techniques for evidence acquisition in relation to obligations for data protection, security and vulnerabilities disclosure, and procedural obligations like cross-examination in court.

The new model for mobile acquisition proposed in this paper includes vulnerability exploitation capabilities and contributes for the standardization and minimisation of forensic hacking techniques in evidence collection. It will provide a clear understanding regarding the intrusiveness of each level in the model, and when it is justified to exploit vulnerabilities for mobile forensics purposes. The model also accommodates cloud and reverse engineering acquisition.

---

1 CLOUD Act s 102(1–6) and s 103 (b).
2 Ibid.
6. New mobile forensic model

As we have seen through section 3 and section 4, current approaches for accessing user data in modern mobile devices have changed greatly from traditional ones. Traditionally, forensic data extraction techniques have focused on acquiring physical data, which when subsequently parsed can recover deleted data. This approach used to be effective because the data was stored in clear-text on non-volatile memory on mobile devices. As a result, the five-level data extraction model has been followed as a standard model. However, with the implementation of encryption and other complex security features, simply acquiring raw data does not help recover user data any more. Worse, destructive procedures such as chip-off may destroy key components needed to decrypt acquired data. Moreover, secure deleting features on mobile devices can effectively delete data remnants on the system, and recovering deleted data from physical data is becoming almost impossible. Additionally, without user authentication credentials, acquiring user data, be it logical or physical, is becoming a great challenge, regardless of the acquisition level. Therefore, categorizing the mobile data extraction method by the extracted data type is becoming less effective. Currently, either extracting the data in clear-text, or extracting the encryption key is the major objective in forensic data extraction. Without the right user authentication, this can only be achieved either by exploiting system vulnerabilities on the target device or by identifying and accessing the stored cryptographic keys. However both methodologies require extensive reverse-engineering prior to working on the target mobile device. Taking this current situation into account, we propose a new mobile forensic data extraction model as follows:

- User secret based acquisition
  
  If an examiner can unlock the phone with the correct user authentication, the target smartphone can be manually operated, and can be set up in a way that it authorizes data extraction through its user interfaces. Manual and logical extraction introduced in section 3.1 fall into this category. As discussed in section 5, compelled disclosure of the password from the device owner is not regarded as an appropriate method. However acquisition may be available through seizing biometric information of the device owner. After unlocking the device, an examiner can modify the device setting, and extract either logical, file system, or physical data by rooting the device.

- Reverse-engineering based acquisition
  
  Reverse-engineering of modern mobile devices is essential in forensic study. Reverse-engineering can be done both in software and hardware. Once an examiner learn the internal structure of the target mobile device operation through reverse-engineering, the examiner may be able to reconstruct the original user data. One example is to identify encryption mechanism and to retrieve the encryption key. Once those information can be retrieved, an examiner can acquire the physical data from the target smartphone with methods discussed in section 3.5, and then decrypt the data off-device.

- Vulnerability exploitation based acquisition
  
  When the target device is locked and encrypted, these features need to be either bypassed or disabled for data extraction. Bypassing or disabling the device lock, encryption, and other security features generally require exploiting system vulnerabilities. The vulnerability exploitation may require the combination of hardware and software attacks. Once those features are bypassed, examiners can choose to acquire either full or partial logical, file system, or physical data. As discussed in section 5, use of open and unpatched vulnerabilities is justified from a legal perspective. However, in many cases zero-day vulnerabilities found through extensive reverse-engineering are required for effective data extraction. Multiple works have already shown the effectiveness of vulnerability exploitation in digital forensic domain (Alendal et al., 2018; Hay, 2017; Shwartz et al., 2017).

Fig. 1 shows a simple flowchart for choosing a proper data extraction technique. Each technique is categorized according to the above mentioned model.

Essentially, without the proper user authentication credentials, system vulnerability exploitation needs to be performed. On the other hand, once the user secret is available, an examiner can use it to manually operate the target phone. Some mobile forensic tool vendors already provide automated versions of those vulnerability exploitation and data extraction procedures shown in Fig. 1. When testing and evaluating those tools, the acquisition level can be categorized using this new model.

7. Conclusions and recommendations

Due to growing security and privacy concerns by mobile device users, manufacturers are aggressively implementing encryption and other complicated security mechanisms. This trend is greatly affecting traditional forensic data acquisition capabilities. Traditionally, acquiring raw data from non-volatile memory on a mobile device would yield meaningful data – including deleted info – which could then be used for criminal investigations. Therefore chip-off and micro read have long been regarded as the highest level of effective technologies in forensic data acquisition. However, as we discussed in this paper, current physical data acquisition practices cannot provide human-readable data due to encryption. Also, effective data erasing functions at the OS level make it difficult to find data remnants in physical data. At the same time, other security features are making it difficult for forensic examiners to acquire even live data on the target device. Therefore, bypassing or disabling device lock and encryption while keeping user data integrity is becoming the most important forensic technique for modern mobile devices. Extensive reverse-engineering, as well as exploiting vulnerabilities, is therefore becoming essential for forensic examiners when performing mobile forensics. Vulnerabilities found through reverse-engineering have already been used for acquiring evidence data from locked and encrypted mobile devices.

In the meantime, however, the use of backdoors and vulnerabilities in forensic analysis has generated controversy and sparked policy discussions by lawmakers and human rights organizations. While exceptional access is less likely to be granted by manufacturers, the use of known vulnerabilities can be justified in the absence of less intrusive investigative measures to access evidence. Currently, there is no clear legislative rule about the use of zero-day exploits for acquiring data from encrypted devices. Responsible disclosure may, however, provide a reasonable ground rule for forensic examiners to follow.

In order to standardize and validate mobile forensic data extraction techniques, further research and efforts are needed. This may be performed by organizations such as a multi-disciplinary EU commission to evaluate the methodology, along with its proportionality and reliability. Additionally, the legislative debate must be enriched by including forensic examiner subject matter expertise. National legislation on lawful exploitation of vulnerabilities will have negative extraterritorial political, economical, and human
Fig. 1. New mobile forensic data extraction model.

right effects. Preferably, strong protection of privacy and system security with encryption must be codified in international treaty, which explicitly regulates exceptions for investigative purposes and implements universal safeguards for human rights.

As suggested in our new mobile forensic data extraction model, exploiting mobile device system vulnerabilities is essential in extracting evidence data from modern encrypted mobile devices for forensic investigation. Proper understanding of the technical details, together with properly following the legal requirements is essentially important for forensic examiners when conducting forensic data acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was conducted under the EU EXFILES project. The EXFILES project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 883156.

References


ENISA, 2016. Opinion paper on encryption strong encryption safeguards our digital