Digital Witness: Remote Method for Volunteering Digital Evidence on Mobile Devices

Nigel Campbell, Trevor Goodyear, Winston Messer, Evan Stuart*, and James Fairbanks
{firstname.lastname}@gtri.gatech.edu
Georgia Tech Research Institute
250 14th Street NW, Atlanta, GA 30319

***Abstract*—Law enforcement requires methods of digital evidence collection from victim or witness devices in a minimally invasive manner. Victims and witnesses are often concerned with minimizing the exposure of data on their phone to authorities. In this paper we describe a system for the secure submission of digital evidence and a micro-service for creating and monitoring chain of custody. These tools minimize device data exposure, encourage cooperation from victims and witnesses, and enforce accountability with regards to handling digital evidence.**

I. Introduction & Motivation

Existing law enforcement tools used to collect evidence from mobile devices are designed around for circumstances where the device in question has been legally taken from the alleged perpetrator of a crime for its potential evidentiary value. As a result, these tools are designed to capture all data held on a mobile device, including data not relevant to the incident in question, or of negligible value. When witnesses and victims are providing evidence to law enforcement, they are often reluctant to share the entire contents of their mobile devices, much as a witness would rather speak to an officer in a public place rather than invite the officer into their own home. Law enforcement requires methods of digital evidence collection from victim or witness devices in a discriminant manner. The current standard operating procedures for digital forensics only allow police to verify images of entire devices, which prevents officers from collecting only the evidence relevant to the case in front of them and leads to overcollection of private data. In a climate of low trust between citizens and the police officers that protect and serve them, it is important to protect the privacy of witnesses and victims that volunteer digital evidence on their mobile devices. Our approach uses public key cryptography for signing and hashing in order to create immutable records that enable defendants, civil society groups, and courts to verify the data collected by police which enables police officers to collect less private data from volunteers.

Desktop computers were typically the means for obtaining and storing information as the Internet initially grew in popularity and usefulness. Technological evolution has afforded people a variety of tools to access the Internet from traditional desktops, to laptops, video game consoles, tablets and cell phones. Today, mobile technology, particularly cell phones, has become a vital tool for personal communications and business relations. Cellular phones have radically changed how society communicates and stays connected. Even the fashion industry now designs clothing with specially placed pockets/spaces to accommodate these devices, including purses with individual holding areas for phones, to special accessories for sports/exercise armbands. The ability to pull targeted, specific information from cell phones is critical in developing a successful case. Courts have struggled to adopt rules for how to treat computers and mobile devices as they do not fit neatly into the predigital paradigm of searches and seizures [1].

Information recovery is of paramount importance in supporting arrests and criminal convictions that are irrefutable. Mobile digital devices have become such a common piece of evidence that police departments across the country are increasingly training officers in how to analyze data/information on phones, especially deleted data. Inevitably, criminals and victims alike use cell phone devices. Law enforcement officers are able to confiscate cell phones from criminals in order to preserve evidence and obtain a warrant to search the digital contents of the digital devices. This mandate however, does not always apply to victims and witnesses. A host of considerations and concerns may prevent victims and witnesses from volunteering to surrender their device for forensics.

Extracting data from mobile devices takes time due to the transfer speed limitations for mobile devices with 100s of gigabytes of storage, and once this data is collected, it must be stored in the custody of law enforcement officers until the conclusion of a trial. Selective collection of pertinent data would ease burdens of time and storage. However, law enforcement must be able to prove the integrity of evidence in a court and current digital forensics tools are designed for validating the integrity of full device captures. By enabling the collection of individual files from mobile devices and tracking the chain of custody over these files at a fine grain level, we are able to provide officers and prosecutors with assurances of the integrity of individual files.

Retrieving information pertinent to the case without accessing all the other miscellaneous information on the phone both increases the chance of successful prosecution while easing witness privacy concerns. There is substantial value in knowing what is on a suspect's, victim's, or witness' phone at the outset of an investigation. Link analysis may prove valuable for connecting data between different devices and entities; however, logical extraction provides a more organized way of finding and examining information, and physical extraction

*Corresponding Author

allows for the rebuilding or re-imaging of deleted texts and photos, resulting in low-risk and high-value data extractions. Once data are isolated in this manner, law enforcement may quickly filter, seek and find evidence without crossing into immaterial or private areas. The *Disclose* system presented is used by the police department of Dekalb County Georgia and is developed under and open source license at https://digitalwitness.github.io.

II. Security and Threat Model

The mobile application users are concerned with minimizing the exposure of data on their phone that is exposed to the authorities.

The authorities are concerned with ensuring that all information collected is accurate and all metadata surrounding that collected information is complete and correct.

Thus mobile devices must generate their own private keys and deliver the public parts to the authorities. The mobile devices must construct unique identifiers for themselves and use this information to sign the information as it is delivered to the authorities.

The authorities must prove that they are collecting only the information that they claim to collect. Since mobile application permissions are not at the file level, care must be taken to convince the user that only the files that they choose will be uploaded. This security model extends to social media applications as public trust in social media companies erodes due to data breaches and malpractice [2].

Evidence Volunteering and *Digital Witness* is a two sided market [3], where authorities are trying to collect evidence and witnesses seek to volunteer information. The key problem with current state of the art is that authorities do not have the technical means of collecting and tracking only a subset of the witness' information. And the witnesses are skeptical of the authorities. Thus an intermediary must broker this transaction and build trust through verifiable proofs of security. The alternative is a completely decentralized system such as blockchain. The fully decentralized model enables transactions to occur with little to no trust. However, you still need to distribute software onto phones users must trust in the privacy guarantees of the software. In this way, blockchain based methods which eliminate the need for a centralized repository of data do not solve the problem in the context of digital evidence volunteering.

This includes cryptographic signing and key management on the part of the submitting devices, and logging, monitoring, and public proof of correctness on the part of the authorities. *Disclose* is a technical means of brokering these information transactions, using open source cryptography and public transparency to build trust on both sides of the market. This approach addresses the human aspects of security and privacy while relying on the mathematical guarantees of existing cryptographic protocols.

III. Methodology

Detailed methods about PKI, Protocols, and Standards that we adopted.

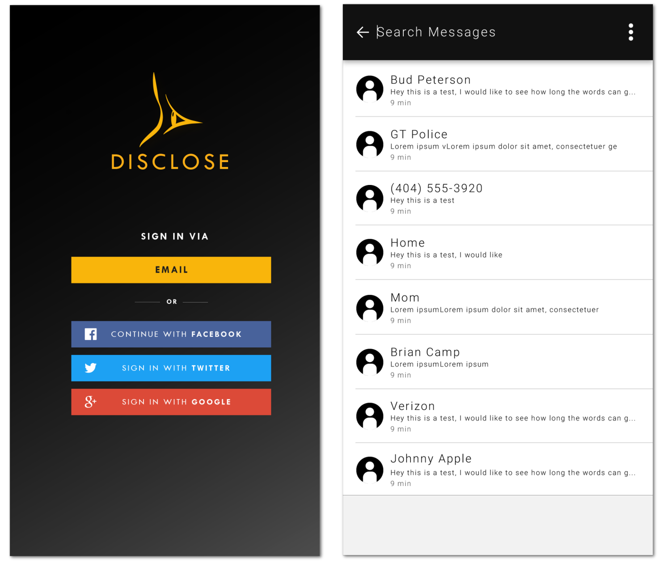

Figure 1. Screenshot showing the login screen (left) and message browsing interface (right)

A. Disclose Mobile App

Disclose is an Android application that allows for the selection, aggregation and submission of digital information to law enforcement. The application allows users to manually curate the evidence that they wish to disclose as opposed to having to submit all information on the device. From the perspective of citizens, this application provides a mechanism which can be used to securely and safely communicate with law enforcement while protecting privacy, which encourages cooperation with law enforcement.

Digital evidence includes anything ranging from photos, videos, text messages, and device logging information. It also includes information from social media applications such as Twitter, Instagram and Facebook. In addition, metadata that is derived from submissions (for instance, EXIF information from photos) can be included as part of submission. While most users are focused on the content of the media on their mobile devices, the metadata can often be more useful to law enforcement, which is concerned with the activities and movements of people as captured by the times and locations associated with images and messages [1].

The workflow for the mobile application is as follows: First, a user account is created. A public private key pairing is generated and tied to the device. Next, the user is presented with several mechanisms for selecting and curating evidence from the device. Finally, the user reviews the selected evidence and submits it to a secure web application which verifies the user and validates the submission. In Figure 1 we see the login

[1]With the US supreme court decision of Carpenter vs the United States, Cell Site Location Information (CSLI) will require a warrant. Thus, volunteering of time and location information of criminal suspects will be more useful to law enforcement over time

screen as well as the interface used for searching and selecting messages to submit as evidence.

In order to maintain a consistent chain-of-custody, evidence should not be modified during submission. The application must also provide mechanismsto ensure that the device (not necessarily the user) is in fact the device submitting the content. Upon account creation, the device generates an x509 ECDSA public-private key pair and is subsequently stored via the Android Keystore [4]. The Keystore API provides assurances that *Disclose* is the only application on the device with access to the public and private keys. Finally, the public key is sent to the central PKI via the Diffie-Helman key exchange algorithm [5]. This key subsequently used in helping to manage chain-of-custody transactions which is discussed later.

Upon submission, the private key is used to generate a digital signature using the contents of the evidence. This signature is used to verify message integrity at the time of submission.

B. Disclose Web Application

The *Disclose* mobile application works with a web application counterpart. The web application is intended to be used by authorites to consume evidence provided users of the *Disclose* mobile application. Within the web application component, authorities are able to view, search, organize, and export user submitted data as seen in figures 2, 3, and 4.

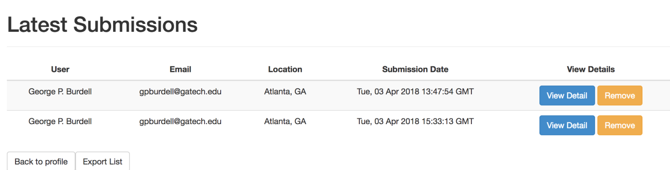

Figure 2. Screenshot of the web application submission feed

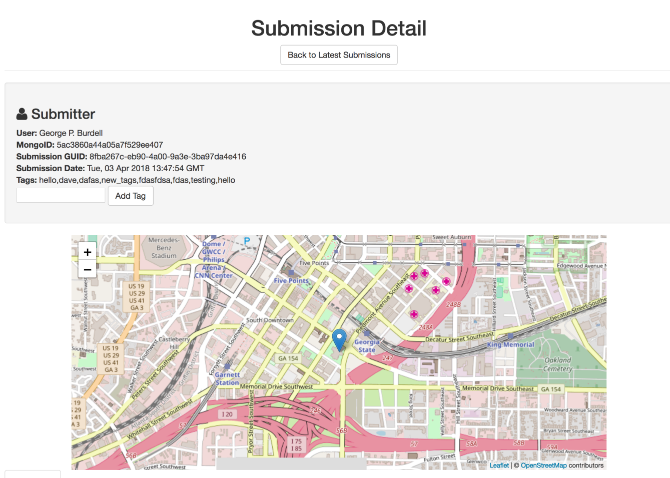

Figure 3. Screenshot of the embedded map view

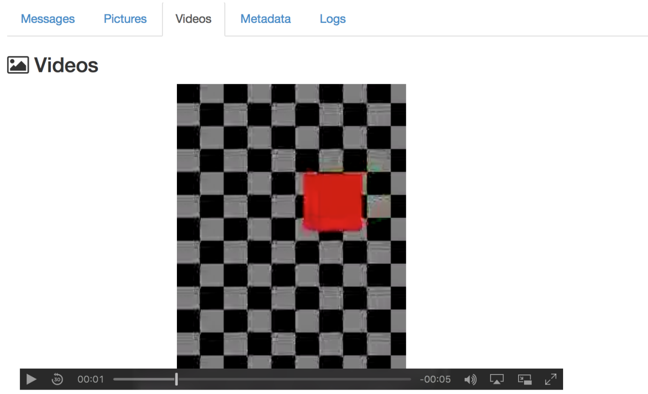

Figure 4. Screenshot showing the web application's embedded video viewer

The web application is intended to serve as a triage tool that will allow investigators to view submitted data and allow them to determine quickly whether or not the submission is of investigatory value and should be exported into other industry standard investigative tools such as EnCase [6] or Autopsy [7]. In order to serve as a sufficient tool for cursory investigations the tool provides viewers for photos, videos, text messages, application logs, exif data, and submission details. Exif data containing GPS coordinates as well as the location of the mobile app user at the time of the submission are plotted within the tool using an embedded OpenStreetMap viewer.

C. Chain of Custody Component

One component of the *Disclose* system is *Custody* which is a microservice for creating and monitoring a chain of transactions that affect data elements once they are collected. *Custody* uses x509 ECDSA public-private key pairs to sign messages and authenticate them. As well as Merkle trees to create publicly verifiable proof of the messages included in the chain [8].

A chain of custody is important in law enforcement applications, the authorities must prove to a court that they have handled the evidence according to the rules of criminal procedure in their jurisdiction[9]. Thus police departments create systems including paper record keeping of who has access to evidence and when. While paper record keeping is not secure in the cryptographic sense, it allows courts to conduct inquiries into the behavior of investigators and determine the answer to two distinct questions.

1) If this evidence was altered, corrupted, or falsified, who is the responsible party?
2) Was this evidence obtained as the "fruit of the poisoned tree" [10]?

Our *Custody* component aims to answer these two questions in a scalable and automatic way. By logging all operations in a structured method, we are able to answer what could have happened to this evidence at this time and "who is responsible?". By tracking all operations with a parent operation we are able to identify chains of evidence operations and identify "who knew about this information, and when did they know it?"

1) Data Model: The *Custody* application tracks data using the following schema.

```sql
create table if not exists identities (
  id integer not null primary key,
  name text not null,
  created_at timestamp not null,
  public_key blob not null -- an x509 cert as ascii
);

-- so we can look users up by their name
CREATE INDEX username_idx
  ON identities (name);

-- so we can look users up by their public key
CREATE INDEX publickey_idx
  ON identities (public_key);

create table if not exists ledger (
  id integer not null primary key,
  created_at timestamp not null,
  identity integer not null,
  message text not null,
  parent blob not null, -- signature of previous
    message
  signature blob not null, -- ecdsa signature of the
     message and parent fields

  foreign key (identity) references identities(id)
);

-- so we can find all messages from a user
CREATE INDEX ledger_identity_idx
  ON ledger (identity);

-- so we can sort all messages by timestamp
CREATE INDEX ledger_createdat_idx
  ON ledger (created_at);
```

Listing 1. The core custody schema creates types for identities and ledger entries

This data layout is mapped into Golang structs [11]. For ease of development and deployment the *Custody* application uses Sqlite, but can target any relational database management system [12]. The service runs as a web services exposing HTTP RPC functions to create identities, sign messages, and audit the ledger.

Identities represent the PKI part of the system, where users are identified and associated with their x509 ECDSA public key. There is a user facing portion of *Custody* to create private keys and share only the public part of the key to the server.

The ledger is the set of messages where every message describes an operation done on the system. It is this ledger that allows an external audit of the system logs.

2) Operations: Every operation conducted by the authorities leads to a message in the ledger describing the operation and the files, cases, or subjects of that operation. These messages are encoded into a plain text format such as JSON and stored in the database along with a parent message representing the last operation and their cryptographic signature for authentication.

The parent field is analogous to the parent commit stored in version control systems such as git [13]. By taking the sequence of signatures from the ledger, we can build a chain of custody. We store these hashes in a chain and use a Merkle Tree to allow third parties to verify that no operations have been forged or forgotten. This allows interest groups that have an interest in the justice system to check the work of the authorities without compromising the privacy of the material. When engaged in a criminal proceeding the messages related to the case will be provided to the defense and they can audit the validity of the messages. This build trust in the correctness of the system and can assist in discovery of Brady Material[14].

The chain of custody problem is similar to validating the integrity of software. The court and defendants want to be sure that no data or records have been forged or erased. Open source software solves this with public repositories and signed commits. In proprietary software development such as the Apple App Store or Google Play, software authors sign their binary artifacts prior to publication, this enables you to ensure the integrity of each version of software, but does not connect the sequence of modifications between two versions of the software. In the case of digital evidence, we want to check the entire sequence of modifications without revealing the content of any data, which must be protected. This requirement is satisfied by storing both the hashes and cryptographic signatures in messages. You cannot forge a message signature without a user's private key, and you cannot forge the hash without the previous message, and no messages contain the data that must be kept private. Messages for data upload contain the hashes of the original data, thus a defendant who is provided access to that original data during the process of defense can verify the integrity, and anyone with an interest in auditing the integrity of the chain of custody can verify the operations on that data.

IV. Summary and Conclusions

The volunteering of digital evidence requires careful consideration of the privacy preferences of users who are least likely to trust the application. The security model requires proving correctness to the authorities and proving privacy to the witnesses. This solution demonstrates that public key cryptography and transparency from the central server is sufficient to build trust in evidence volunteers.

In this paper, we have presented an open-source application that provides the mechanisms necessary for the secure transfer of digital information between citizens and law enforcement. We have also presented a system for creating and monitoring chain of custody transactions involving digital evidence. Together, these tools minimize the exposure of data on the devices of citizens, encourages active cooperation with local law enforcement, and enforces accountability with regards to the handling of digital evidence.

A. Future Work

The *Digital Witness* application must be implemented for all mobile operating systems and platforms, in order to achieve high utilization and deliver evidence to police departments.

This application can achieve greater adoption by generalizing beyond the scope of Police Evidence to applications in cybersecurity events and online criminal activity on darknets like ToR.

V. Acknowledgement

Research reported in this article was supported by The U.S. Department of Justice (DOJ), Office of Justice Programs (OJP), National Institute of Justice (NIJ) under award number 2016-MU-MU-K004 to solicitation NIJ-2016-8976. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Justice.

References

[1] Josh Goldfoot. The Physical Computer and the Fourth Amendment. *Berkeley Law Scholarship Repository*, 2011.

[2] United State House of Representatives Committee Hearing 443490-1. Facebook hearing on data protection, 2018.

[3] Geoffrey G Parker and Marshall W Van Alstyne. Two-sided network effects: A theory of information product design. *Management science*, 51(10):1494–1504, 2005.

[4] Google. Android keystore system, 2018.

[5] W. Diffie and M. Hellman. New directions in cryptography. *IEEE Trans. Inf. Theor.*, 22(6):644–654, 1976.

[6] Lee Garber. Encase: A case study in computer-forensic technology. *IEEE Computer Magazine January*, 2001.

[7] Dan Manson, Anna Carlin, Steve Ramos, Alain Gyger, Matthew Kaufman, and Jeremy Treichelt. Is the open way a better way? digital forensics using open source tools. In *System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on*, pages 266b–266b. IEEE, 2007.

[8] Johannes Buchmann, Erik Dahmen, Elena Klintsevich, Katsuyuki Okeya, and Camille Vuillaume. Merkle signatures with virtually unlimited signature capacity. In *Applied Cryptography and Network Security*, pages 31–45. Springer, 2007.

[9] United States House of Representatives The Committee on the Judiciary. *Federal Rules of Criminal Procedure*. The National Court Rules Committee, Dec 2017.

[10] G.A. Bartlett. Constitutional law - search and seizure - "fruit of poisonous tree" doctrine - jacobs v. warden, 367 f.2d 321 (4th cir. 1966). *William and Mary Law Review*, 9(1), 1967.

[11] Shiju Varghese. Structs and interfaces. In *Go Recipes*, pages 53–74. Springer, 2016.

[12] Mike Owens and Grant Allen. *SQLite*. Springer, 2010.

[13] L Torvalds, JC Hamano, J King, NTN Duy, J Schindelin, J Nieder, et al. Git: the stupid content tracker. *Git source code README file. From the initial commit of Git's source code into Git itself (revision e83c516)*, 2005.

[14] J. Kaplan, R. Weisberg, and G. Binder. *Criminal Law: Cases and Materials*. Aspen casebook series. Wolters Kluwer Law & Business/Aspen Publishers, 2012.

VI. Author Information

A. Nigel Campbell

1) *Title:* Research Scientist I
2) *Organization:* Georgia Tech Research Institute
3) *Address:* 75 5th St NW Atlanta, GA 30308
4) *Email:* Nigel.Campbell@gtri.gatech.edu
5) *Phone:* 6783503466

B. James Fairbanks

1) *Title:* Research Scientist II
2) *Organization:* Georgia Tech Research Institute
3) *Address:* 75 5th St NW Atlanta, GA 30308
4) *Email:* James.Fairbanks@gtri.gatech.edu
5) *Phone:* 4044076226

6) *Bio:* Dr. James Fairbanks's previous and current research cuts across disciplines from mathematics to engineering applications to use graphs to answer scientific research questions. The focus on streaming data processing is useful for understanding near real-time systems, such as the data collection such as the data collection of a Dementia patient's social network interactions with their environment. His work on graph and network analysis problems in the cyber and social media domains is directly relevant to this problem. At the Georgia Tech Research Institute, his lab specializes in data analysis and high-performance computing as applied to healthcare, social science, and complex network analysis problems. His prior work includes an analysis of the numerical sensitivity of spectral clustering algorithms for graph partitioning. These algorithms including statistical analyses of graph data will support the inference processes used in this work. He am also a core maintainer of the Julia package LightGraphs which is a core package for the commercial product JuliaPro, and was used in several papers about systems biology, mathematical optimization, and other scientific domains.

C. Trevor Goodyear

1) *Title:* Research Scientist I
2) *Organization:* Georgia Tech Research Institute
3) *Address:* 75 5th St NW Atlanta, GA 30308
4) *Email:* Trevor.Goodyear@gtri.gatech.edu
5) *Phone:* 4044076964

D. Winston Messer

1) *Title:* Research Scientist II
2) *Organization:* Georgia Tech Research Institute
3) *Address:* 1575 Northside Drive, Atlanta, GA 30318
4) *Email:* Winston.Messer@gtri.gatech.edu
5) *Phone:* 4044076560

E. Evan Stuart

1) *Title:* Research Scientist I
2) *Organization:* Georgia Tech Research Institute
3) *Address:* 75 5th St NW Atlanta, GA 30308
4) *Email:* Evan.Stuart@gtri.gatech.edu
5) *Phone:* 4044078519