REFERENCE MATERIAL FOR ASSESSING FORENSIC SIM TOOLS

Paper No. ICCST 2007-74

Wayne A. Jansen
Member, IEEE
National Institute of Standards and Technology
100 Bureau Dr., STOP 8930
Gaithersburg, MD 20899
USA

Aurelien Delaitre
National Institute of Standards and Technology
100 Bureau Dr., STOP 8930
Gaithersburg, MD 20899
USA

Abstract – Subscriber Identity Modules (SIMs) are a fundamental standardized component of most cell phones used worldwide. A SIM can be removed from a phone handset and inserted into another, allowing users to port identity, personal information, and service between devices. All cell phones are expected to incorporate some type of identity module eventually, in part, because of this useful property.

Some of the earliest, general purpose, forensic tools for cell phones targeted SIMs to recover digital evidence. While over time the capabilities and number of such tools have increased, they are not completely free of problems. Validating a forensic SIM tool is an essential quality assurance measure. It allows a forensic specialist to determine how to compensate for any shortcomings identified or whether to use one version of the tool in lieu of another. Tool manufacturers also benefit from rigorously validating their products before releasing them. However, creating reference SIMs that contain comprehensive test data can be time consuming and difficult to accomplish. This paper describes an approach for automating the population of test data onto SIMs to create reference material for use in tool validation. It also covers details of the implementation and explains characteristics of SIMs that pertain to the solution.

Index Terms – Computer Forensics, Cell Phones, Digital Evidence.

I. INTRODUCTION

The Global System for Mobile Communications (GSM) standards for cellular networks were originally developed by the European Conference of Postal and Telecommunications Administrations, continued by the European Telecommunications Standards Institute and then by the 3rd Generation Partnership Project (3GPP), where they are now maintained. Commercial GSM service was started in mid-1991. By 1993, thirty-six GSM networks were operating in twenty-two countries [1]. Although begun in Europe, GSM has become a broader international standard with compliant networks operational in more than 200 countries around the world, including North America [2].[1]

Subscriber Identity Modules (SIMs) are synonymous with mobile phones and devices that interoperate with GSM cellular networks. Under the GSM framework, a cellular phone is referred to as a Mobile Station and is partitioned into two distinct components: the Subscriber Identity Module (SIM) and the Mobile Equipment (ME). As the name implies, a SIM is a removable component that contains essential information about the subscriber. The ME, the remaining radio handset portion, cannot function fully without one. The SIM's main function entails authenticating the user of the cell phone to the network to gain access to subscribed services. The SIM also provides a store for personal information, such as phone book entries and text messages, as well as operational information, such as that involving location.

SIMs are often classified according to the phase of the specifications supported, which is recorded in an element of its file system (i.e., EF_{Phase}).[2] The three phases defined are phase 1, phase 2, and phase 2+, which correspond roughly to first, second, and 2.5 generation network facilities. Another class of SIMs being deployed in third generation (3G) Universal Mobile Telecommunications Service (UMTS) networks is UMTS SIMs (USIMs). USIMs are enhanced versions of present-day SIMs, containing backward-compatible information.

Some of the earliest, general purpose, forensic tools for mobile phones targeted SIMs, not only because of detailed specifications available for them, but also because of the highly relevant and useful digital evidence that could be recovered. A recent assessment of the capabilities of present day forensic tools to recover evidence from SIMs, however, noted discrepancies between the test data placed on a SIM and that recovered and reported in every tool [3]. They include the inability to recover any data from certain SIMs, inconsistencies between the data displayed on screen to the user and that generated in the output reports, missing truncated data in reported or displayed output, errors in the decoding and translation of recovered data, and the inability to recover all relevant data. Furthermore, updates or new versions of a tool, on occasion, performed less capably than a previous version.

Validating a forensic SIM tool is an essential quality assurance measure. The results aid in deciding how to compensate for any noted shortcomings or whether to switch to a new version of the tool. Validation should be carried out

[1] Certain commercial products and trade names are identified in this paper to illustrate technical concepts. However, it does not imply a recommendation or an endorsement by NIST.

[2] The standardized EF names and abbreviations found in the 3GPP TS 11.11 Technical Specification, though sometimes unusual, are used throughout this discussion.

when first choosing a forensic tool to ensure its acceptability and redone when updates or new versions of the tool become available to maintain consistency of results. Validating a tool entails defining a comprehensive test data set, loading it onto the device, and following procedures to acquire and recover the test data [4]. While tool validation is essential, building reference SIMs that contain comprehensive test data can be time consuming and difficult to carry out, requiring the use of various SIM editing tools and handsets to populate the data. In addition, variances exist between SIMs from different manufacturers, such as different file capacities allocated for entries (e.g., the phonebook) and different size data fields supported (e.g., an individual's name in a phonebook entry). Different character encodings may also apply to the various languages of interest. This paper discusses an approach for automating the population of reference test data onto SIMs that attempts to address those types of differences. Details of the implementation are also covered.

II. SIM CHARACTERISTICS

The SIM-ME partitioning of a cell phone stipulated in the GSM standards has brought about a form of portability. Moving a SIM between compatible cell phones automatically transfers with it the subscriber's identity and the associated information and capabilities. In contrast, present-day phones in North America that follow Code Division Multiple Access (CDMA) standards (i.e., TIA/EIA/IS-95-A and B) do not employ a SIM. Instead, analogous SIM functionality is incorporated directly within the device. While SIMs are most widely used in GSM systems, comparable modules are also used in Integrated Digital Enhanced Network (iDEN) phones, which use a proprietary mobile communications technology developed by Motorola, and phones used in UMTS networks (i.e., a USIM). Because of the flexibility a SIM offers GSM phone users to port their identity, personal information, and service between devices, eventually all cellular phones are expected to include (U)SIM-like capability. For example, requirements for a Removable User Identity Module (R-UIM), as an extension of SIM capabilities, have been specified for cellular environments conforming to TIA/EIA/IS-95-A and B specifications, which include Wideband Spread Spectrum based CDMA [5].

At its core, a SIM is a special type of smart card that typically contains a processor and between 16 and 256 KB of persistent electronically erasable, programmable read only memory (EEPROM). It also includes random access memory (RAM) for program execution, and read only memory (ROM) for the operating system, user authentication and data encryption algorithms, and other applications. The hierarchically organized file system of a SIM resides in persistent memory and stores such things as names and phone number entries, text messages, and network service settings. Depending on the phone used, some information on the SIM may coexist in the memory of the phone. Alternatively, information may reside entirely in the memory of the phone instead of available memory on the SIM.

Authenticating a device to a network securely is a vital function performed via the SIM. Cryptographic key information and algorithms within the tamper resistant-module provide the means for the device to participate in a challenge-response dialogue with the network and respond correctly, without exposing key material and other information that could be used to clone the SIM and gain access to a subscriber's services. Cryptographic key information in the SIM also supports stream cipher encryption to protect against eavesdropping on the air interface [6, 7].

Two sizes of SIMs have been standardized, but only the smaller size shown in Figure 1 is broadly used in GSM phones today. The module has a width of 25 mm, a height of 15 mm, and a thickness of .76 mm, which is roughly the footprint of a postage stamp. Although similar in dimension to a MiniSD or MMCmobile removable memory card supported by some cell phones, SIMs follow a different set of specifications with vastly different characteristics. For example, their 8-pin connectors are not aligned along a bottom edge as with removable media cards, but instead form a circular contact pad integral to the smart card chip, which is embedded in a plastic frame. Also, the slot for the SIM card is normally not accessible from the exterior of the phone to facilitate frequent insertion and removal as with a memory card, and instead, is typically found in the battery compartment under the battery.

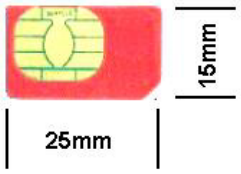

Figure 1: Subscriber Identity Module

When a SIM is inserted into a phone handset and pin contact is made, a serial interface is used for communicating between them. A SIM can be removed from a phone and read using a specialized SIM card reader and software through the same interface. Standard-size smart card adapters are also available for SIMs, which allows them to be inserted into and read with a conventional smart card reader.

A. The SIM File System

Forensic SIM tools extract digital evidence present in the file system of a SIM. The file system is organized in a hierarchical tree structure, as shown in Figure 2. It is composed of the following three types of elements [8]:

- Master File (MF) - the root of the file system that contains dedicated and elementary files.
- Dedicated File (DF) - a subordinate directory to the master file that contains dedicated and elementary files.
- Elementary File (EF) - a file that contains various types of formatted data, structured as either a sequence of data bytes, a sequence of fixed size records, or a fixed set of fixed size records used cyclically.

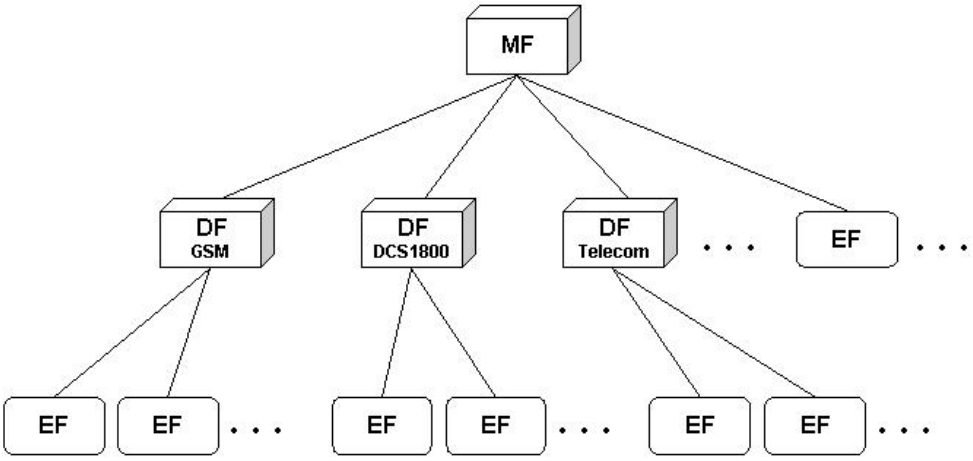

Figure 2: SIM File System

The GSM standards define several important dedicated files immediately under the MF: DF_{GSM}, $DF_{DCS1800}$, and $DF_{TELECOM}$. Several EFs are defined for these DFs and the MF, including many that are mandatory. The EFs under DF_{GSM} and $DF_{DCS1800}$ contain mainly network-related information respectively for GSM 900 MHz and Digital Cellular System (DCS) 1800 MHz band operation. EFs for 850 MHz and 1900 MHz bands used in North America are found respectively under those DFs as well, and typically contain identical information. The EFs under $DF_{TELECOM}$ contain service-related information. The contents of specific EFs that are recovered by most forensic tools and have proved useful in investigations are discussed later in this paper.

Though SIM file systems are highly standardized, the standards allow flexibility such that their content can vary among network operators and service providers. For example, a network operator might not use an optional file system element, might create an additional element on the SIM for use in its operations, or might install a built-in function to provide a specialized service [9].

B. File Access Controls

SIMs, as with smart cards in general, employ a range of tamper resistance techniques to protect their contents. In addition, various levels of rights exist that are assigned to a DF or EF to control the conditions of access [8]:

- Always - Access can be performed without any restriction.
- Card Holder Verification 1 (CHV1) - Access can be performed only after a successful verification of the user's PIN, or if PIN verification is disabled.
- Card Holder Verification 2 (CHV2) - Access can be performed only after a successful verification of the user's PIN2, or if PIN2 verification is disabled.
- Administrative - Access can be performed only after prescribed requirements for administrative access are fulfilled.
- Never - Access of the file over the SIM/ME interface is forbidden.

The SIM operating system controls access to an element of the file system based on its access condition and the type of action being attempted [8]. For example, actions on EFs include searching, reading, and updating the contents. While reading and searching the contents of a particular EF might be allowed without CHV1 verification (i.e., an Always access condition), updating might likely require as a prerequisite CHV1 being correctly verified (i.e., a CHV1 access condition). In general, CHV1 protects core SIM data for the card user against unauthorized reading and updating, while CHV2 protects administrative dialing control data mainly for a card manager (e.g., the parent of a child user), if such a relationship exists. The 4 to 8 digit values of both CHVs can be reset by anyone knowing the PIN values, or their verification completely disabled. So-called ADM Codes are required for Administrative access and are normally kept by the service provider or network operator that issued the SIM.

The SIM operating system allows only a preset number of attempts, usually three, to enter the correct CHV before further attempts are blocked. Submitting the correct Unblock CHV value, also known as a PIN Unblocking Key (PUK), resets the CHV and the attempt counter. If the identifier of the SIM (i.e., its Integrated Circuit Chip Identifier or ICCID) is known, the Unblock CHV for either CHV1 or CHV2 can be obtained from the service provider or network operator. The ICCID is normally imprinted on the SIM along with the name of the network provider. If needed, the identifier can also be read with a SIM tool from an EF, EF_{ICCID}, since the Always access condition applies by definition. If the number of attempts to enter an Unblock CHV value correctly exceeds a set limit, normally ten attempts, the card becomes blocked permanently.

III. DIGITAL EVIDENCE

An assortment of digital evidence from a SIM lies scattered throughout various EFs in the file system. News articles of high profile cases occasionally contain illustrative examples where evidence recovered from a SIM was used successfully in an investigation.

- Text Message and Call Data [10] – "A pastor of the Pentecostal congregation in the small community of Knutby was sentenced to life in prison for persuading one of his lovers (the au pair) to shoot and kill his

wife and trying to kill the husband of another mistress. Two days after the murder, the pastor's au pair Sarah S. claimed that she did it. Despite her claims … the police believed she had an accomplice."

"The strongest evidence against the pastor was the extensive communication through text messages and voice calls between him and the au pair on the day of the murder and just before that. What they did not know was that their (anonymously sent and) carefully deleted text messages were possible to recover."

- Location Data [11] – "Mr Bristowe told BBC News Online: 'It was mobile phone evidence which made the police look more closely at Huntley. He had been Mr. Useful, helping them to search the college grounds, but when they checked Jessica's phone and discovered when and where it had been switched off alarm bells began to ring… (Jessica's phone) disengaged itself from the network, in effect it says goodbye' at 1846 BST on the Sunday when the girls disappeared. Jessica's phone contacted the Burwell mast when it was turned off."

  ""The police provided us with a map of the route they thought the girls would have taken, and the only place on that route where the phone could have logged on to Burwell (and disengaged itself) was inside or just outside Huntley's house.' It is believed to be that crumb of crucial evidence which forced Huntley to change his story earlier this year and suddenly admit the girls died in his bathroom."

For a reference SIM to be useful in validating forensic SIM tools, its file system must be populated with test data that is normally recovered by such tools. Several general categories of evidence can be identified:

- Service-related Information
- Phonebook and Call Information
- Messaging Information
- Location Information.

A number of EFs from each of these categories whose information is regularly employed by forensic specialists are discussed below as examples of core elements for validation [7, 12].

A. Service-related Information

The Integrated Circuit Card Identification (ICCID) is a unique numeric identifier for the SIM that can be up to 20 digits long. It consists of an industry identifier prefix (89 for telecommunications), followed by a country code, an issuer identifier number, and an individual account identification number [13]. Aside from the prefix, the components of an ICCID are variable, making them sometimes difficult to interpret. The ICCID can "Always" be read from the SIM without providing a PIN and can never be updated. The country code and issuer identifier can be used to determine the network operator providing service and obtain call data records for the subscriber.

The International Mobile Subscriber Identity (IMSI) is a unique 15-digit numeric identifier assigned to the subscriber. It has a somewhat similar structure to the ICCID: a Mobile Country Code (MCC), a Mobile Network Code (MNC), and a Mobile Subscriber Identity Number (MSIN) assigned by the network operator. The MCC is 3 digits, while the MNC may be either 2 or 3 digits, with the MSIN taking up the remainder. The fourth byte of another EF, Administrative Data (AD), gives the length of the MNC [14]. Networks use IMSIs to determine which network a device owner subscribes to and, if not their network, whether to allow those network subscribers to access service.

The ICCID and IMSI can be used reliably to identify the subscriber and the network operator providing service. Since these identifiers can be misinterpreted, however, other SIM data can help confirm a finding.

The Mobile Station International Subscriber Directory Number (MSISDN) is intended to convey the telephone number assigned to the subscriber for receiving calls on the phone. Unlike the ICCID and IMSI, however, the MSISDN is an optional EF. If present, its value can be updated by the subscriber, making it a less reliable data source, since it would then be inconsistent with the actual number assigned.

The Service Provider Name (SPN) is an optional EF that contains the name of the service provider. If present, it can be updated only by the administrator (i.e., Administrator access). Similarly, the Service Dialling Numbers (SDN) EF contains numbers of special services such as customer care and, if present, can help identify to which network the SIM is registered. The Extension3 (EXT3) EF contains additional data for an SDN entry.

B. Phonebook and Call Information

The Abbreviated Dialling Numbers (ADN) EF retains a list of names and phone numbers entered by the subscriber. The type of number (TON) and numbering plan identification (NPI) are also maintained in this EF. The storage permits rudimentary phonebook operation by providing the means to select commonly dialed phone numbers by name and update or call them using a menu or certain keys on the phone. If the ADN storage capacity is insufficient to hold all of the information for an entry (e.g., an unusually long sequence of digits), an index to an Extension1 (EXT1) EF record is used to link to where the additional data is maintained. Most SIMs provide around 100 slots for ADN entries.

The Fixed Dialling Numbers (FDN) EF is similar to ADN insofar as a list of names and phone numbers is involved. However, this list is used only in situations where the user is restricted to dialing just the numbers prescribed by a card manager. If the FDN storage capacity cannot hold all of the information for an entry, an index to an Extension2 (EXT2) EF record is used to indicate where the additional data is maintained.

The Last Numbers Dialled (LND) EF contains a list of the most recent phone numbers called by the device. A name may also be associated with an entry (e.g., a called phonebook entry) and stored with the number. Though a number appears on the list, a connection may not have been successful, only attempted. Most SIMs provide only a limited number of slots (e.g., ten) for these entries. If the LND storage capacity cannot hold all of the information for an entry, an index to an EXT1 EF record indicates where the additional data is maintained. Some phones do not store called numbers on the SIM and instead rely on their own memory for storage.

C. Messaging Information

Text messaging is a means of communication in which messages entered on one cell phone are sent to another via the mobile phone network. The Short Message Service (SMS) EF contains text and associated parameters for messages received from or sent to the network, or are to be sent out as an MS-originated message. SMS entries contain other information besides the text itself, such as the time an incoming message was sent, as recorded by the mobile phone network, the sender's phone number, the SMS Center address, and the status of the entry. The status of a message entry can be designated as unoccupied free space or as occupied by one of the following: a received message to be read, a received message that has been read, an outgoing message to be sent, or an outgoing message that has been sent. Messages deleted via the phone interface are often simply designated as free space and the content retained unchanged on the SIM until they are overwritten. When a new message is written to an available slot, the unused portion is filled with padding, overwriting any remnants of a previous message that might be there.

The capacity for stored messages varies among SIMs. Many cell phones also use their own internal memory for storing text messages. The choice of memory where messages are stored (i.e., SIM or phone) can vary depending on the phone software and user settings [15]. For example, a default arrangement might be for all incoming messages to be stored on the memory of the SIM before using internal phone memory, while outgoing messages are stored only if explicitly requested. Phone models of a particular generation and manufacturer often behave consistently in this respect [15].

The maximum length of a single SMS message entry is 160 characters of text. Messages exceeding that length must be broken down into smaller segments by the sending phone and reassembled by the receiving phone. This feature is especially useful for foreign language character sets such as Chinese or Arabic whose encoding consumes considerably more bits per character than English. A reference number parameter identifies the entries whose segments require reassembly. Such messages are referred to as concatenated messages. SMS messages may originate through other means than a cell phone, such as from an Internet SMS server or through electronic mail.

An SMS message can be coded in different ways. The original and most common encoding scheme is a GSM-specific 7-bit character set packed into a bit stream [16]. Such an encoding cannot be interpreted readily by individuals using a hex editor, nor can it embody all languages. Support for other character sets, such as 16-bit Unicode, was added for languages whose alphabets cannot be represented using the original Western European character set [17].

An Enhanced Messaging Service (EMS) was defined as a way to extend SMS message content to allow simple multimedia messages to be conveyed. EMS messages can contain not only formatted text with different font styles and fonts, but also black and white bitmap pictures and monophonic melodies [17]. EMS message content resides in the SMS EF along with SMS message content. EMS messaging is essentially an application-level content extension to SMS, which conforms to the general SMS message structure and support for concatenated messages. EMS-enabled devices are backward compatible by definition with SMS-enabled devices.

D. Location Information

A GSM network consists of distinct radio cells used to establish communications with mobile phones. Cells are grouped together into defined areas used to manage communications. Phones keep track of the area under which they fall for both voice and data communications. The Location Information (LOCI) EF contains the Location Area Information (LAI) for voice communications. The LAI is composed of the MCC and MNC of the location area and the Location Area Code (LAC), an identifier for a collection of cells. When the phone is turned off, the LAI is retained, making it possible to determine the general locale where the phone was last operating. Because a location area can contain hundreds or more cells, the locale can be quite broad. However, it can nevertheless be useful in narrowing down the region where the event occurred.

Similarly, the GPRS Location Information (LOCIGPRS) EF contains the Routing Area Information (RAI) for data communications over the General Packet Radio Service (GPRS). The RAI is composed of the MCC and MNC of the routing area and the LAC, as well as a Routing Area Code (RAC), an identifier of the routing area within the LAC. Routing areas may be defined the same as location areas or they may involve fewer cells, providing greater resolution.

IV. IDENTITY MODULE PROGRAMMER

The Identity Module Programmer (IMP) was developed as a general purpose tool to populate various types of identity modules with reference test data. The initial implementation works only with SIMs, but the intent is to support other types of identity modules eventually. The remainder of this section discusses the design considerations and implementation aspects of the tool.

A. Design Considerations

The overall data flow of IMP is given in Figure 3. Conceptually the process is straightforward. Reference data is read by the program and used to populate a SIM, once the CHV and ADM codes (i.e., Card Data) have been applied and the appropriate access conditions enabled. Any errors are logged and a summary of the results is reported. The reference test data could be generated manually or automatically. For the latter, a preprocessor would produce the test cases, for example, using combinatorial coverage techniques to generate all two-way combinations of parameter values for each category of evidence [18]. Otherwise, the test data can be created manually.

At a concrete level, several factors need to be considered when deciding how to populate a SIM or a similar type of identity module with reference test data. These considerations include the following:

- A method for representing the card and test data
- The reference test data to be populated
- A means to write data to the identity module

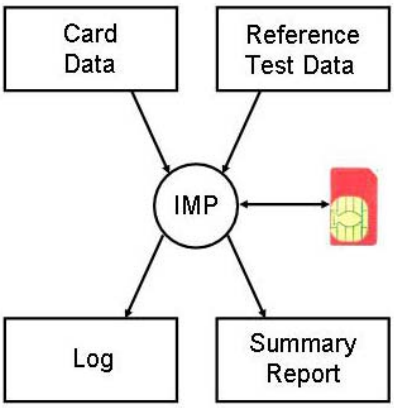

Figure 3: IMP Overview

XML is the method used to represent test data for input to IMP. XML is a popular syntax for representing data, able to be processed by computers and, with some effort, able to be interpreted and understood by humans. Many XML editors exist, as well as tools for defining data type descriptions and schemas against which data representations can be constructed and automatically verified. The latter property is extremely useful in perfecting the syntax of reference data sets, particularly when produced manually. These characteristics motivated its choice. Figure 4 shows an example phonebook entry for an Asian name and an international telephone number encoded in XML.

```
<phonebookentry>
    <description enc="ucs2">阿家里面于</description>
    <address>
        <ton>international</ton>
        <npi>telephone</npi>
        <number>1444412345678</number>
    </address>
</phonebookentry>
```

Figure 4: Example XML Phonebook Entry

Reference data can be populated on a SIM only when the correct access conditions are satisfied for performing update (i.e., write) operations. Table 1 identifies the access conditions needed for the EFs discussed earlier.

Table 1:
Update Access Conditions

CHV1	CHV2	ADM	NEVER
ADN	FDN	IMSI	ICCID
EXT1	EXT2	SDN	
LND		EXT3	
SMS		SPN	
LOCI		AD	
LOCIGPRS		PHASE	
MSISDN			

Two classes of SIMs generally available for use in tool validation are production SIMs and test SIMs. CHV values are usually available for most production SIMs; however, administrator codes are normally kept by the network carrier and not made available. For test SIMs, which are available for development purposes from most SIM manufacturers, the CHV values and ADM codes are usually provided together with the test SIMs. As one can see from Table 1, test SIMs allow a greater range of reference data to be populated. Nevertheless, production SIMs can still form a useful baseline for validation, as long as any EFs not able to be populated by the tool are noted and taken into account during tool validation.

The initial set of reference data was drawn from test scenarios recently used in assessments of forensic SIM tools involving basic, location, EMS, and foreign language data [19]. Basic data includes subscriber (i.e., the IMSI, ICCID, SPN, and LP elementary files), PIM (i.e., the ADN elementary file), call (i.e., the LND elementary file), and SMS message related information. Besides common input data, known problematic input, such as the use of a special character for a phonebook name entry, was included. Foreign language data involves SMS messages and PIM data that are expressed in a language other than English. EMS data includes EMS messages more than 160 characters in length and also containing non-textual content such as black and white bitmap images or monophonic melodies. EMS messages can also contain formatted text with different font styles and fonts. Location data includes location-related information, such as the last location area or routing area where the phone disengaged from the network (i.e., the LOCI and LOCIGPRS elementary files).

A phone handset communicates with a SIM using command directives called Application Protocol Data Units (APDUs) [8]. APDUs used by SIMs have the exact same format as those with smart cards and follow the same protocol. The APDU protocol is a simple command-response exchange, with a single response for each command issued. Each element of the file system has a unique numeric identifier assigned, which can be used to reference the element and perform an operation, such as reading the contents, in the case of a forensic tool [9]. The same APDU protocol can be used to reference an element and perform an update operation to populate an EF with test data.

B. Implementation Aspects

Most forensic SIM tools run under the Windows operating system, which makes that a logical choice for the IMP implementation. To allow other operating systems besides Windows to be supported as well, IMP was programmed in the Java programming language. IMP uses an open source Application Programming Interface (API) called Java Card Communication Access Library (JACCAL) to exchange APDUs with the SIM. A Simple API for XML (SAX) parser is used to interpret the XML encoded, reference test data.

For IMP to populate a SIM, it must be removed from a phone and placed into an appropriate reader. Either a specialized reader that accepts a SIM directly or a general-purpose reader for a full-size smart card can be used, provided that it is compatible with the PC/SC (Personal Computer/Smart Card) specification [20], a popular general-

```
<xs:group name="groupPhoneBookEntry">
    <xs:choice>
        <!-- Phone book entry -->
            <xs:element name="phonebookentry">
                <xs:complexType>
                    <xs:sequence>
                        <!-- Name of the contact -->
                        <xs:element name="description">
                            <xs:complexType mixed="true">
                                <!-- Encoding of the name -->
                                <xs:attribute name="enc" type="typeEncoding"/>
                            </xs:complexType>
                        </xs:element>
                        <!-- Address of the contact, i.e. the phone number -->
                        <xs:element name="address" type="typeAddress"/>
                    </xs:sequence>
                </xs:complexType>
            </xs:element>
        <xs:element name="empty" type="typeEmpty"/>
    </xs:choice>
</xs:group>
```

Figure 5: XML Schema Excerpt

purpose architecture for smart cards. For full-size card readers, a standard-size smart card adapter is needed to house the SIM for insertion into the reader.

It is possible that the defined reference test data can exceed the capacity of an EF or the size of the field. Attempts to exceed either type of limit are detected by the SIM itself. Out of bounds references are disregarded and overly long data are truncated to the space available. IMP logs any deviations between the populated data and reference data as they occur. A summary of all reference test data populated also appears in the output report. In addition, the contents of certain EFs that IMP is unable to populate are also reported to provide a known definitive baseline for tool validation.

V. CONCLUSION

One consideration in constructing the XML schema for test data is defining ways to represent deleted entries. Because no delete operation exists for SIMs, deletion is accomplished in most cases by updating information in an elementary file with strings of hexadecimal “FF.” The one exception involves SMS message content, whereby a status flag representing a deleted entry is used instead of “FF” overwrite. SIMs use three structures for elementary files, which also affect how information is stored and acted on. Transparent files are a sequence of bytes that can be accessed via an offset. Linear fixed files are a list of records of the same length that can be accessed by absolute record number, via a record pointer, or by seeking a record by pattern. Cyclic files comprise a circular queue of records maintained in chronological order, which are accessible the same as with linear fixed records, with the oldest overwritten if storage is full. For example, for linear fixed files a record number could be used to specify the content of the indicated record, whereby a deleted entry simply does not appear. However, that might lead to errors in the reference data set, such as duplicate entries, which would not be automatically detectable by an XML validation tool. Instead of record numbers, data for record entries are listed sequentially and populated in the order of appearance. Deleted entries can then be designated with a special “empty” tag. An excerpt from the current version of the schema for the phonebook definition, which specified the alternative for designating an empty entry, is shown in Figure 5.

Forensic examination of cell phones is a growing subject area in computer forensics. Identity modules play an important role in this process. Forensic examination tools for phone handsets and identity modules translate data to a format and structure that is understandable by the examiner and can be effectively used to identify and recover evidence. However, tools may contain some degree of inaccuracy. For example, the tool’s implementation may contain a programming error; a specification used by the tool to translate encoded bits into data comprehensible by the examiner may be inexact or out of date; or the protocol used to access the SIM may be incorrect, causing the tool to function improperly in certain situations.

The failure of a forensic tool to correctly recover and report relevant SIM data greatly impedes the ability of the forensics specialist and jeopardizes the credibility of the overall results. While experience with a forensic tool provides an understanding of its limitations, it does not replace the need to validate the tool’s capabilities, particularly for the initial and subsequent versions of a tool selected for use, and when patches or updates are applied to the tool or the tool’s operating environment. Quality measures should always be applied to ensure that results remain consistent and any variations understood. This principle applies both to forensic specialists that use such tools and forensic tool manufacturers who produce them.

Suitable reference materials are essential for validating a forensic tool. However, creating suitable reference materials can be problematic and time consuming. The technique outlined in this paper provides a means to create reference material automatically for SIM tool validation, based on selected sets of test data. The technique was implemented in the Java programming language and the test data represented in XML to provide a high degree of platform independence. The resulting program is relatively straightforward to use and requires only a minimal amount of additional equipment (i.e., a PC/SC compatible card reader) to populate a SIM for use in tool validation.

VI. REFERENCES

[1] C. Dechaux, R. Scheller, What are GSM and DECT?, Electrical Communication, 2nd Quarter, 1993, pp. 118-127.

[2] GSM World, GSM Global Networks on Air, March 2006, <URL: http://www.gsmworld.com/news/statistics/networks_complete.shtml>.

[3] Wayne Jansen, Rick Ayers, Forensic Tools for Mobile Phone Subscriber Identity Modules, Journal of Digital Forensics, Security and Law, April 2006.

[4] Amanda Goode, Forensic Extraction of Electronic Evidence from GSM Mobile Phones, IEE Seminar on Secure GSM & Beyond, Digest No. 2003/10059, February 11, 2003.

[5] Removable User Identity Module for Spread Spectrum Systems, 3rd Generation Partnership Program 2, 3GPP2 C.S0023-0, Version 4.0, June 15, 2001.

[6] Klaus Vedder, Security Aspects of Mobile Communications, in Computer Security and Industrial Cryptography - State of the Art and Evolution, Lecture Notes in Computer Science, Vol. 741, 1991, pp. 193-210.

[7] Svein Willassen, Forensics and the GSM Mobile Telephone System, International Journal of Digital Evidence, Volume 2, Issue 1, Spring 2003 <URL: http://www.utica.edu/academic/institutes/ecii/publications/articles/A0658858-BFF6-C537-7CF86A78D6DE746D.pdf>.

[8] Specification of the Subscriber Identity Module - Mobile Equipment (SIM - ME) interface, 3rd Generation Partnership Project, TS 11.11 V8.13.0 (Release 1999), Technical Specification, June 2005.

[9] Fabio Casadei et al., SIMbrush: an Open Source Tool for GSM and UMTS Forensics Analysis, First International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE'05), November 7-9, 2005, pp. 105-119.

[10] Robert Burnett, Ylva Hård af Segerstad, The SMS Murder Mystery: the dark side of technology, Safety & Security in a Networked World: Balancing Cyber-Rights & Responsibilities, September 2005, <URL: http://www.oii.ox.ac.uk/microsites/cybersafety/extensions/pdfs/papers/robert_burnett.pdf>.

[11] Chris Summers, Mobile phones - the new fingerprints, BBC News Online, December 18, 2003, <URL: http://newsvote.bbc.co.uk/mpapps/pagetools/print/news.bbc.co.uk/1/hi/uk/3303637.stm>.

[12] Tony Dearsley, Mobile Phone Forensics – Asking the Right Questions, New Law Journal, July 29, 2005, pp. 1164-1165.

[13] The International Telecommunication Charge Card, International Telecommunications Union, Telecommunication Standardization Sector (ITU-T), Recommendation E.118, May 2006.

[14] Numbering, Addressing and Identification, 3rd Generation Partnership Project, TS 23.003, V6.9.0 (Release 6), Technical Specification, March 2006.

[15] Svein Willassen, Forensic Analysis of Mobile Phone Internal Memory, IFIP International Conference on Digital Forensics, National Center for Forensic Science, Orlando, Florida, February 13-16, 2005, in Advances in Digital Forensics, Vol. 194, Pollitt, M.; Shenoi, S. (Eds.), XVIII, 313 p., 2006.

[16] Alphabets and Language-specific Information, 3rd Generation Partnership Project, TS 03.38, version 7.2.0 (Release 1998), Technical Specification, July 1999.

[17] Technical Realization of the Short Message Service (SMS), 3rd Generation Partnership Project, TS 23.040 V6.6.0 (Release 6), Technical Specification, December 2005.

[18] David Cohen et al., The Combinatorial Design Approach to Automatic Test Generation, IEEE Software, September 1996, pp. 83-87, <URL: http:///www.research.telecordia.com/papers/gcp/AETGissre06.shtml>.

[19] Rick Ayers et al., Cell Phone Forensic Tools: An Overview and Analysis Update, NIST Interagency Report 7387, <URL: http://csrc.nist.gov/publications/nistir/nistir-7387.pdf>.

[20] PC/SC Workgroup (2005) Interoperability Specification for ICCs and Personal Computer Systems, Part 1 - Introduction and Architecture Overview, Revision 2.01.00, June 2005, URL: http://www.pcscworkgroup.com/specifications/files/pcsc1_v2.01.0.pdf>.

VII. VITA

Wayne Jansen, BS and MS Computer Science, MS Engineering Management, is a Principal Computer Scientist at the National Institute of Standards and Technology, Computer Security Division, in Gaithersburg, Maryland. His research interests are in communications, distributed applications, and computer security. Currently, he leads the Mobile Security Project, focused on mobile software and handheld devices.

Aurélien Delaitre is a Guest Researcher in the Information Technology Laboratory at the National Institute of Standards and Technology (NIST) in Gaithersburg, MD. He graduated with an M.S. in Computer Science from ESIAL, France. His current research focus is on mobile device forensic tools and reference materials for tool validation.