E-MONEY SECURITY DILEMMA: ADVANCED CYBERSECURITY MECHANISMS AND LEGACY MOBILE PAYMENTS IN SUB-SAHARAN AFRICA

Fabrice Neza[1], Anthony Joseph[2] and Mary Joseph[3]
[1]*Mastercard, Inc., New York, New York, USA*
[2]*Pace University, New York, New York, USA*
[3]*Herbert Lehman College, City University of New York, Bronx, New York, USA*

ABSTRACT

Mobile Network Operators (MNO) infrastructures are expanding in many Sub-Saharan African (SSA) countries that are increasingly adopting mobile payment services. The adopted model is outdated and not compliant with today's more secure, dynamic, and application-driven environments. Unstructured Supplementary Services Data (USSD) legacy technology is widely deployed in SSA countries due to several factors including compatibility with handsets, user experience, cost, and ease of deployment for the MNOs. A survey reveals that USSD is used by approximately 65-75% of the population for mobile payments. The MNOs infrastructure was developed for connectivity, rather than security, making the e-money ecosystem prone to cyber-attacks. End-to-end encryption (E2EE) is one of the recommended cybersecurity solution mechanisms. This research shows that tokenization with identity based encryption techniques is a suitable model to secure financial transactions over USSD as it provides an end-to-end encrypted path when combined with Internet Protocol Security (IPSec) and Transport Layer Security (TLS) while lessening changes to the technology itself and the underlying network. This implementation introduces additional cost for maintenance and management, and therefore, the governments need to adopt market-friendly laws, regulations, and policies to develop employable skills and facilitate MNOs' return on investment.

KEYWORDS

Cybersecurity, E-payment Services, USSD, Tokenization, Identity Based Encryption

1. INTRODUCTION

In 2021, 90% of individuals in developed countries were online compared to only 19% of individuals in the least developed countries of the world (ITU, 2021). By the end of 2021, there were more than 4.3 billion mobile internet subscribers (63% of the world's population), and the mobile broadband covered 95% of the world's population (Bahia & Delaporte, 2022). However, this adoption was not evenly distributed: Europe had the highest rate of Internet usage whereas Africa had the lowest rate. Moreover, Sub-Saharan Africa (SSA) had the very lowest mobile internet adoption rate and usage of any region; it has almost half the world's population not covered by mobile broadband networks and its broadband (3G and 4G) connections only exceeded 2G in 2019 (Wyrzykowski, 2020).

Africa's young population is a potential asset for a relatively high growth in Internet usage and its economy (ITU, 2021; Wan, 2022). Chironga et al (2018) ranked Africa second among the world's largest banking markets for growth and profitability, but Africa's growth and profitability are mainly concentrated in five countries, two of which are in SSA: Nigeria and South Africa (Cook, 2019). Because of the increasing adoption of mobile electronic payment (epayment), Mobile Network Operator (MNO) infrastructures are expanding in many SSA countries. This expansion was driven, in part, by the very low availability of bank branches and the improved availability, reliability, and security of electronic channels (Srinivas & Wadhwani, 2019). The two epayment services that have emerged are MNOs' mobile money transfer and banking institutions' mobile banking system (offered through MNOs infrastructures). The SSA region is the current epicenter of mobile money activities in the world with an estimated 70% of the 2021 global total transactions (Onyango, 2022). While mobile money started as a simple payment service in many SSA

countries, it evolved into a platform that allows MNOs to compete in the digital financial services. According to Pazarbasioglu et al (2020), mobile wallets outnumber bank accounts and mobile payment platforms' financial services include credit and savings accounts. Aramé et al (2022) reported on the GSMA's 2021 Global Adoption Survey that 44% of mobile money providers' financial products include savings, insurance or credits.

Most large scale mobile financial services in SSA rely on the Unstructured Supplementary Services Data (USSD), as the key solution for a faster and more economical technology to support mobile money platforms. Though USSD is not the only communication service available for mobile payments, it is estimated that over 90% of mobile money transactions in SSA are driven by USSD (GSMA, 2018). The other options include the Short Messaging Service (SMS), Subscriber Identity Module (SIM) Toolkit (STK), and mobile Internet (Butler et al., 2020). USSD is a session based communication protocol used by the Global System for Mobile (GSM) that is almost seven times faster than SMS. Its operations are simple, do not require access to the Internet access or the user's SIM card and it can be used on basic mobile phones, feature phones, and on smartphones. Most MNOs agree that USSD may be the best available option for SSA's low-income customers because of the security, user experience, compatibility with handsets, real-time interactivity, cost, and ease of deployment (Mosweunyane et al., 2014). The World Bank projected that Africa will house 90% of the world's very poor by 2030 if the current trend of rising extreme poverty continues (Wadhwa, 2018).

Mobile financial services over USSD can be risky if proper cybersecurity measures are not put in place because the USSD packets are transmitted in clear text over the out-of-band signaling channel of the Global System for Mobile (GSM) communications standard (Heine, 1999). While GSM information networks are most secure and immune to destruction when they are most distributed and decentralized, they are also most vulnerable to interception and unauthorized access (Weichbroth & Łysik, 2020). USSD deployments are most vulnerable to interception at the air interface where the consumer mobile phones and MNOs wirelessly interconnect. Whereas developed countries have dropped the use of USSD-like designs in favor of the end-to-end encryption (E2EE) concept, the current mobile network model in SSA countries for financial services is outdated and not compliant with the recommended more secure, dynamic, and application-driven environments. This inherent risk will likely lead to more security breaches as increasingly mobile phones become the effective delivery channels for financial services to poor people and those living in rural and remote areas without proper financial service infrastructures. Moreover, it is typical that in cases of frauds the consumers are the ones who bear the cost, not the providers.

This paper will mainly focus on basic and feature phones because they continue to account for the majority of the African mobile phone market due to their relative affordability and durability, and until they are replaced by smartphones and customers can access the mobile money provider directly via the Internet, they will likely continue to use MNO's USSD platform for financial services (Silver & Johnson, 2018). The main objective of this paper is to develop a prototype that combines identity based encryption with tokenization techniques as a suitable model to secure financial transactions over USSD as it provides end-to-end encryption when combined with Internet Protocol Security (IPSec) and Transport Layer Security (TLS) while lessening changes to the technology itself and the underlying network.

2. BACKGROUND

2.1 Terminologies

Table 1

Acronym	Meaning	Acronym	Meaning
AuC	Authentication Center	**BSS**	Base Station Subsystem
BSC	Base Station Controller	**BTS**	Base Transceiver Station
EIR	Equipment Identity Register	**IMEI**	International Mobile Equipment Identities
MoMoAS	Mobile Money Application Server	**CLAE**	Certificate-Less Authenticated Encryption
HLR	Home Location Register	**IMSI**	International Mobile Subscriber Identity
MS	Mobile Station	**MSC**	Mobile Switching Center
NSS	Network SubSystem	**SIM**	Subscriber Identity Module
TMSI	Temporary Mobile Subscriber Identity	**VLR**	Visitor Location Register

Note: A complete list of acronyms can be found in Heine (1999).

2.2 GSM Architecture

GSM developed in the 1980s, deployed in the early 1990s, is perhaps one of the most successful telecommunications technologies of the last few decades (Scourias, 1996). A typical GSM architecture consists of a mobile station, base station system, and a network subsystem. A GSM mobile station (MS) has a Subscriber Identity Module (SIM) protected with a Personal Identification Number (PIN). Each GSM subscriber is identified through the International Mobile Subscriber Identity (IMSI). The MS stores the ciphering algorithm (A5) while the SIM stores the ciphering key generating algorithm (A8), authentication algorithm (A3), authentication key, and the IMSI (Heine, 1999). The base station subsystem (BSS) is where the mobile phone interfaces with the Base Transceiver Station (BTS) via the air interface. BTS houses the radio transceivers that define a cell and handle the radio-link protocols with MS (Heine, 1999). They are controlled by the Base Station Controller (BSC) for radio-channel setup, frequency hopping of calls, and constant monitoring of calls to determine a call handover from one BTS to another (Heine, 1999). Moreover, BSC connects MS via BTS to the Mobile Service Switching Center (MSC); MSC is a central component of the Network Switching Subsystem (NSS) that provides the set of database and control functionalities needed to handle a mobile subscriber (Heine, 1999). The MSC together with the Home Location Register (HLR), the Visitor Location Register (VLR), Equipment Identification Register (EIR), and the Authentication Center (AuC) constitute NSS (Heine, 1999). HLR contains the administrative information of each subscriber registered in a GSM network while VLR contains certain administrative information already in HLR required for call control and allocates the Temporary Mobile Subscriber Identity (TMSI) to support the subscriber identity confidentiality. EIR and AuC registers are databases used for authentication and security purposes.

Although GSM is a relatively secure public wireless standard, it has security weaknesses to which there are proposed solutions (Gardezi, 2006; Srinivas, 2001). Moreover, Srinivas (2001) argued that cellular communications systems are less secure than their wired counterparts and presented solutions to existing weaknesses while Pagliusi (2002) proposed a security model for GSM network architecture that uses the SIM as a security module to encrypt the user authentication on radio interface. Kune et al (2012) demonstrated how attacks against TMSI are possible; they proposed a shorter TMSI allocation time than the time delay between two calls.

2.3 USSD Architecture and Security

Modern mobile telecommunications networks, such as the GSM, allow the transfer of USSD messages between an application and a MS. The USSD session is initiated by either the MSC, HLR or VLR; the USSD initiation process is characterized by a mobile originated USSD transaction (MO-USSD) or a mobile terminated USSD transaction (MT-USSD) (Klinger, 2021). However, in this research the focus will be on the MS originated transactions. The USSD application can reside either in MSC, VLR, HLR or in an Internet Protocol (IP) network using an appropriate interface (ETSI, 1996). When an MS initiates USSD transactions not destined for applications in the MSC, VLR, or HLR in this hierarchical order, or when they are able to decode the service request but cannot support the required application, a USSD handler routes the message to the USSD gateway (USSD GW) using the Mobile Application Part (MAP) protocol (ETSI, 1996). The USSD GW then routes the message to the application server (AS) that handles the subscriber's request. The AS subsequently replies to the USSD GW that forwards the response to MS (ETSI, 1996). In an MS-initiated service request scenario, the session created between the network application and the MS is used for all information transfers until it ends. If the mobile-initiated USSD operation is unsuitable for the network, the service request would be rejected usually with a notification response to the MS (ETSI, 1996).

Because USSD rely on the legacy Signaling System 7 (SS7) protocol to provide MS the capacity to initiate a USSD operation during or outside a call, attacks utilizing the out-of-band signaling channel expose USSD to the same security issues found in GSM networks (Loughney et al., 2004). With the emergence of IP-based networks and broadband technology, IP Signaling Transport (SIGTRAN) was proposed to carry signaling messages over IP. However, it inherited SS7 signaling protocol's vulnerabilities that could allow an intruder to easily intercept the mobile network. Exploiting these vulnerabilities could result in confidential data leaks, financial loss to individual subscribers, or disruption of communication services. Butler et al (2020) identified both passive and active attacks as the two main types of attacks against digital financial services running over USSD on the GSM signaling channel. To lessen these attacks, ITU (Butler et al., 2020)

recommended the use of standardized cryptographic libraries in all applications and end-to-end encryption with standardized protocols, including Transport Layer Security (TLS) v1.2 or higher for the communication between the USSD GW, the Mobile Money Application Server and other back-end services. This paper is about encrypting the data from the mobile user to the Mobile Money Application Server. Furthermore, RFC3788 (Loughney et al., 2004) stipulates that when a network using SIGTRAN protocols involves more than one party, end-to-end security should be the goal and recommends that IPsec or TLS be used to ensure confidentiality of user payload. Additionally, the Central Bank of Nigeria (2019), one major player in SSA's economy, requires "secure transmission of USSD signals between network operator & the USSD aggregators, and between the USSD aggregators & the bank."

While Public Key Cryptography (PKC) is mostly used in TLS certificates that are issued by reputed Certificate Authorities (CA) to provide the confidentiality, integrity, authentication, and non-repudiation required in the end-to-end encryption, the complexity of setting up and using a PKC in basic and feature phones environment could be expensive because it may require additional infrastructure and high computing power, and might not be suitable in processor or bandwidth limited environments. Similarly, the use of STK or java applications to secure USSD transactions adds overhead to the network and requires changes in the network and the mobile devices that may be too costly to customers. To achieve end-to-end encryption on an USSD channel using basic and feature phones, a form of PKC that utilizes some kind of identifier as the basis for the encryption mechanism and has the means to encrypt messages (or verify signatures) absent of any prior key distribution between the communicating entities would be more appropriate. This solution should be able to generate public keys locally and be as transparent to users as practicable.

To achieve the objective of end-to-end encryption in USSD-based financial transactions, this research paper combines two models in a prototype: identity-based encryption (IBE) and tokenization. The resulting model uses an improved Certificate-Less Authenticated Encryption (CLAE) method along with Tokenization, Internet Protocol Security (IPSec) and Transport Layer Security (TLS). The features incorporated in this model are a Token Server (TS), a Trusted Module (TM), a Database and a Hardware Security Module (HSM).

2.4 Identity Based Encryption and Tokenization

Identity Based Encryption (IBE) is a solution to the encryption key management problem; it can use an arbitrary string as a public key to enable secure data transfer without certificates. Protection is provided by a key server, Private Key Generator (PKG), (Islam et al., 2015) that controls the dynamic generation of private decryption keys that correspond to public identities. The IBE system design originated with Shamir (1985) and was refined by Boneh and Franklin (2001) after it was further developed by Fiat and Shamir (1986) and Feige et al (1988). Al-Riyami and Paterson (2003) combined IBE with conventional PKC to produce certificate-less PKC (CL-PKC). By separating authentication and authorization from private key generation through the key server, the CL-PKC framework simplifies operation and scaling. CL-PKC authenticates public keys without certificates but instead uses a trusted third party (TTP) that possesses a master key. There are similarities between CL-PKC and identity-based public key cryptography (ID-PKC) but unlike ID-PKC, CL-PKC can be viewed as a model for PKC because it does not have a key escrow property problem (Al-Riyami & Paterson, 2003).

The proposed model in this paper provides a certificate-less authenticated encryption (CLAE) method for sending an encrypted message over a network using identity-based encryption between a sender (MS) having a unique sender identity (idMS) formed by the combination of MSISDN and IMSI, and a recipient (Mobile Money Application Server) having a recipient identity string (idMoMoAS). Prior to encrypting a plaintext message (M) the MS verifies the plurality of public parameters (PK) from the trusted module (TM). During financial transactions, tokenization is used to substitute the actual mobile money account number with non-sensitive placeholders. Tokenization is especially practiced in the payments processing industry and the process involves protecting sensitive data by replacing them with an algorithmically generated number called a token that is by design unrelated to the actual data (Liu et al., 2020; Iwasokun et al., 2018).

2.5 Related Work

Most of the research on mobile payments over USSD security does not describe end-to-end encryption (E2EE) in depth. They focus on the security of the GSM network algorithms, ignoring the plaintext transmitted USSD packets (ETSI, 1996). Nonetheless, a significant number of papers discussed mobile phone communications security solutions using public key cryptography or third-party applications. Wu et al. (2011) proposed an integrated mobile payment framework based on 3G network and the SIM card ToolKit (STK). The research recommended a public key infrastructure to improve mobile payment security wherein the private key would be stored in the phones integrated circuits cards. Nyamtiga et al (2013) and Sekyere et al (2018) proposed a java application installed in the customer's phone to capture the security details, generate secure messages and send USSD commands over the GSM network. Wu & Tan (2009) used RSA-1024 asymmetric algorithm as well as AES and 3DES symmetric algorithms to provide "an end-to-end secure channel between server-side and mobile terminal" for SMS communication. Thomas and Panchami (2014) presented an encryption method that included EasySMS protocol, Blowfish encryption algorithm, and MD5 hashing algorithm for authentication, confidentiality, and integrity respectively to ensure very secure SMS message transfers between mobile users. Agwanyanjaba (2018) explored the challenges of using PIN as the only factor of authentication in USSD-based financial transactions and proposed the use of time bound USSD push augmented with biometric authentication to secure transactions. While evaluating the security challenges of Mobile Money, Castle et al (2016) focused on Android applications and presented a systematic threat model for the challenges associated with mobile money deployments in the developing world. Klinger (2021) proposes the use of a sufficiently powerful quantum computer to run quantum safe symmetric cyphers on simple Universal Integrated Circuit Cards (UICCs) to secure USSD transactions. Moreover, several other identity-based encryption solutions have been proposed to provide security in multiple mobile networks. Islam (2015) proposed the use of symmetric key and identity based techniques for encryption and key management to ensure secure and highly reliable end-to-end SMS communication over GSM networks. Boneh and Franklin (2001) developed a functional IBE scheme, "based on ciphertext security in the random oracle model," with a distributed PKG to disperse the location of the master key. Smith et al (2009) presented an identity-based key agreement system and its implementation for mobile telephony in GSM networks. Additionally, Kumar et al (2006) presented an identity-based cryptography (IBC) mutual authentication and key agreement approach for GSM networks. Unlike the work of this paper, they only used the IMSI number for the public identity key and the protocol security relies on an assumed secure channel to the HLR and VLR. In an extensive search of the research literature published on USSD security, no paper was found involving end-to-end encryption of the USSD sessions in the mobile payments transactions as recommended in different security standards. This paper's contribution is to provide a prototype E2EE that addresses this gap as well as the shortcomings of the existing USSD security in mobile payments.

3. REQUIREMENTS AND SPECIFICATIONS

In the survey conducted to support this study, more than 70% of the 110 SSA respondents stated that they fear their phone getting hacked or data communication intercepted as their biggest concern regarding mobile phone use for financial transactions. Over 70% of the respondents consider mobile money and mobile banking to be safe to very safe, yet these same respondents believe that people's personal information is not safe when using these services: 58% for mobile money and 55% for mobile banking. This apparent contradiction in the respondents' perceptions suggests that there is a general need for stronger security in mobile money and mobile banking services over USSD as the technology has proven to be the most accessible, economical, and practical way to remove traditional barriers to digital financial systems in SSA. According to Aramé et al (2022), mobile money transactions in 2021 were over US $1 trillion and registered mobile money accounts were 1.35 billion, exceeding the 2012 number by a factor of 10.

This research work uses the engineering approach to problem-solving where the problem is identified and the solution is pursued through design. It presents a prototype that ensures end-to-end security of data between the MNO subscriber and the digital financial application server whether located in the MNO infrastructure (mobile money) or in the external financial institution (mobile banking). The model uses an improved certificate-less authenticated encryption (CLAE) method along with Tokenization, Internet

Protocol Security (IPSec) and Transport Layer Security (TLS). The features incorporated in this model are a Token Server (TS), a Trusted Module (TM), a Database and a Hardware Security Module (HSM). The TM is configured for generating a plurality of public parameters and it may be a dedicated server or as part of distributed systems. In this case, it also includes additional software components for the TS. Other security mechanisms might be deployed on the TM hardware as deemed necessary (Malek, 2014).

Traditionally in the PKC, public keys are generated by a trusted center (certificate authority) guaranteeing that a public key belongs to a certain recipient. However, those public keys are not protected as they are exchanged over insecure channels and are prone to interception and tampering. In contrast, the CLAE scheme presented with this model, allows the sender (MS) to remove the need for public key certificates by using the identity of the recipient (idMoMoAS) and prior to encrypt the message, locally verify the public parameters of the TM to ensure they were not altered. This verification reduces the risk of an attacker equipped with an IMSI catcher who would replace the public parameters. Instead of using a predetermined parameter to generate the public/private key pair, the MS integrates the identity of the TM (idTM) as well as the identity of the recipient (idMoMoAS). The point of trust is established from the public identity of the TM (idTM) that can be a hostname or a fully qualified domain name. This provides greater flexibility to the proposed model in generating the encryption keys, as the MS can arbitrarily choose any identity of the TM (idTM) and can be assured that its selection will be enforced on the recipient (MoMoAS). Once the system is initialized, any entity in the system (MS or MoMoAS) can self-generate the public key of the other entity and encrypt the sensitive data by using the recipient's locally generated public key derived from the recipient's identity and the plurality of public parameters (PK) from the TM. Equipped with the public parameters, the MS is able to communicate with the MoMoAS independent of the TM over a secure channel by encrypting the message using the MoMoAS's public key (pubMoMoAS), which is generated locally by the MS using the public parameters obtained from TM and the MoMoAS identity string (idMoMoAS). Only the true recipient is then able to decrypt and retrieve the sensitive data using its private key known only to the recipient and derived from the TM (Islam et al., 2015; Malek, 2014).

In GSM network, every user is uniquely identified by the IMSI, a 15-digit number that is assigned to the mobile user by the MNO and is stored on SIM card. Since the IMSI is the SIM card identifier, if a new SIM card is issued it will have a new IMSI. If a SIM card is moved to a new phone, the IMSI moves with the card and doesn't stay with the phone. The Mobile Station International Subscriber Directory Number (MSISDN) is the telephone number with full international prefix used to send a text message or make a call (Heine, 1999). Unlike an IMSI, which is bound to a particular SIM card, MSISDNs can be transferred from phone to phone, and from SIM to SIM. While it might be tempting to use the MSISDNs alone as unique identifiers in mobile money transactions, problems arise when that number is recycled to a new user and that user inadvertently access the previous user's information. Moreover, because MSISDNs are usually public, using them to authenticate users in mobile money transactions as primary identifiers has made the number itself (MSISDN) vulnerable to malicious attacks. Hence, the proposed model will use the combination of MSISDN and IMSI as a unique identity of the MS (idMS) for the identity-based encryption. The key to authenticating the MS securely, reliably and quickly is to use the MSISDN and perform the verification using the SIM card's IMSI number. Identifying the MS based on a combination of MSISDN and IMSI is the strongest and least problematic solution for user identity as it removes the risk of account details being compromised when numbers are recycled, thus keeping users secure and protected. Through this combination, insider attacks will likely be detected as soon as the MS is checked in the HLR and AuC due to the MS identity (idMS) mismatch; the attacker MSISDN and IMSI combination will fail, hence the correct computation of the actual MS public key (pubMS) would fail too since it is derived from the MS identity (idMS). The distribution of the plurality of public parameters (PK), the identity of the financial application server (idMoMoAS), and the idTM can be integrated into the GSM lookup mechanism and carry the information over the SS7 protocol (Dryburgh & Hewett, 2003).

Since the lookup functionality to locate the HLR and the current location of the MS already exists in GSM networks, a flag can be attached to the request message, stating that the public parameters of the TM be sent piggybacked onto the response. A flag is suitable because the public parameters need to be queried only for the very first USSD session request from an MS. All subsequent sessions from the same MS to the MoMoAS do not need any further public parameter lookup. In operation, both the MS and MoMoAS need to communicate with the TM to obtain respective private keys (prvMS and prvMoMoAS) and a plurality of public parameters. However, the MS private key (prvMS) can be embedded on the SIM card similar to the individual subscriber authentication key (Ki) utilized as a secret key to authenticate the MS and the GSM operator that is never transmitted over the radio channel.

The introduction of the Hardware Security Module (HSM) in the design is to offload TS. The HSM stores encryption keys and serves as a secure network computer system of cryptographic operations: key generation, exchange, and encryption that essentially implements algorithms for encryption and hashing for different financial transactions over USSD. TS and TM implement additional logic on top of HSM whereby HSM gets the key generation request from TS through an application programming interface (API). TS receives the transacting MSISDN, and using the HSM capabilities, it encrypts the communication and generates the relative token. Subsequently, if MSISDN needs to be processed later, only the token is available, HSM would be unable to provide the requested data because it does not store all the necessary values. HSM only stores the encryption key; it does not store the token and the encrypted value. This is a critical security measure because in the event where HSM is compromised, it would be unable to provide any suitably meaningful data. In the rare case when an attacker might gain entry into HSM, the attacker would have difficulty matching the held keys with a particular transaction as HSM does not control the key; it is only used to encrypt each value at a given moment. In contrast, TS stores the encrypted values and tokens and keeps track of the encryption keys to be able to rotate them with time. It is able to decrypt any value from the token as it records the key-value relationship. In a real-time encrypted session initiated between MS and the USSD platform, tokenized data are sent back and forth when the service is invoked. The encrypted session remains open over a radio connection until the USSD service is completed, the user terminates the application, an incorrect option is entered from the menu, or a time-out happens. This minimizes the time an attacker would have to intercept and try to decode the random and encrypted token. Each allowable transaction is limited to a specific token. All Internet Protocol (IP) links between different nodes are encrypted. The interface between the MoMoAS, USSD GW and HLR or MSC over SS7 or SIGTRAN is protected using TLS (Loughney et al., 2004; Aramé et al., 2022). All SIGTRAN peers implementing TLS for the security solution authenticate as part of the TLS session establishment and TLS is used on all bi-directional streams; no directional streams are to be used (Loughney et al., 2004). With E2EE being the goal, IPSec and TLS are used to ensure confidentiality of the payload especially for external applications over the Internet. Edge devices (firewalls) are linked with IPSec because TLS only protects the payload. Thus, IPsec Encapsulating Security Payload (ESP) encrypts and protects the TLS payload information in the established tunnel.

4. PROTOTYPE IMPLEMENTATION AND DISCUSSION

The primary objective of this end-to-end encryption and tokenization security solution is to encrypt the communication so as to prevent eavesdropping and other security breaches from compromising the confidentiality, integrity, and availability of the exchanged data (Quentin et al., 2020). The prototype of the proposed model is represented in Fig. 1. Generic names are used instead of the actual terms (Figure 1) because different network generations introduce different names for each component. For instance, in GSM, cell towers are called “BTS”, in Universal Mobile Telecommunications System (UMTS) they become “NodeB”, in Long-Term Evolution (LTE) “eNodeB”, and in 5G they are termed “gNB” (Gupta & Jha, 2015). Hence, the use of generic terms is more convenient and easier to follow and thereby making the model more applicable. As shown in Fig. 1, when the subscriber initiates a mobile money request (1) through a USSD code comprising an asterisk (*) and hash (#) keys along with a combination of digits (0 to 9), the MS recognizes the code format, and instead of initiating a call it invokes the use of the normal GSM encrypted signaling channel (USSD bearer) to communicate with the USSD infrastructure. The goal is sending an encrypted message by the MS having an identity string (idMS) derived from the combination of the MSISDN and IMSI to a recipient (MoMoAS) having an identity string (idMoMoAS) over a network using identity-based encryption. In this step the MS identifies the TM by idTM (i.e. “mcc.mnc.nameofTM” where MCC is Mobile country Code and MNC is Mobile Network Code). The exchange of the PK (associated with or generated by TM) between MS and TM occurs as part of the normal GSM parameters exchange (2) and are distributed by HLR (3) to the mobile user through the MSC/VLR. Among the flags returned by the HLR in (3) is the identity of the recipient (idMoMoAS). The idMoMoAS is only sent to SIM cards provisioned with mobile money services in HLR. The sender, MS, determines if it has a private key (prvMS), that can be burned directly onto the chip when the SIM is manufactured, and verifies the plurality of public parameters (PK) of the TM, received in (3). The verification process to ensure that the public parameters have not been modified relies on the properties of the PK and prvMS, the known idTM and the idMS (Malek, 2014). The MS then identifies the recipient (MoMoAS) to receive the plaintext message (M) by idMoMoAS received in

(3) from HLR through mobile money user profile provisioning. It is worth noting that idMoMoAS may be the MoMoAS full qualified domain name (FQDN) or hostname. MS generates the identity-based public key encryption of the MoMoAS (pubMoMoAS) using the PK and idMoMoAS (Malek, 2014). This approach allows generating pubMoMoAS locally for offline encryption and communication with TM is not required as a certificate issued by TM is not necessary to authenticate the pubMoMoAS. In this manner, MS is able to send M as an encrypted message to the MoMoAS without accessing the TM, as long as the MS has the required PK and its own prvMS. Furthermore, with the PK and its own prvMS, the MS is able to verify that the PK has not been compromised to ensure that only the MoMoAS having the prvMoMoAS will be able to decrypt C. The MS then encrypts the USSD message (M) as ciphertext (C) including the encrypted message, as well as extra information necessary for decrypting the message. Additional authentication information may also be appended to C (Malek, 2014). In (4), MS transmits the encrypted C to the MoMoAS and safely travels the unsecured GSM channel. An attacker who would intercept it, would require the MoMoAS private key (prvMoMoAS) and the PK of the TM to read the message (Ramadan et al., 2016).

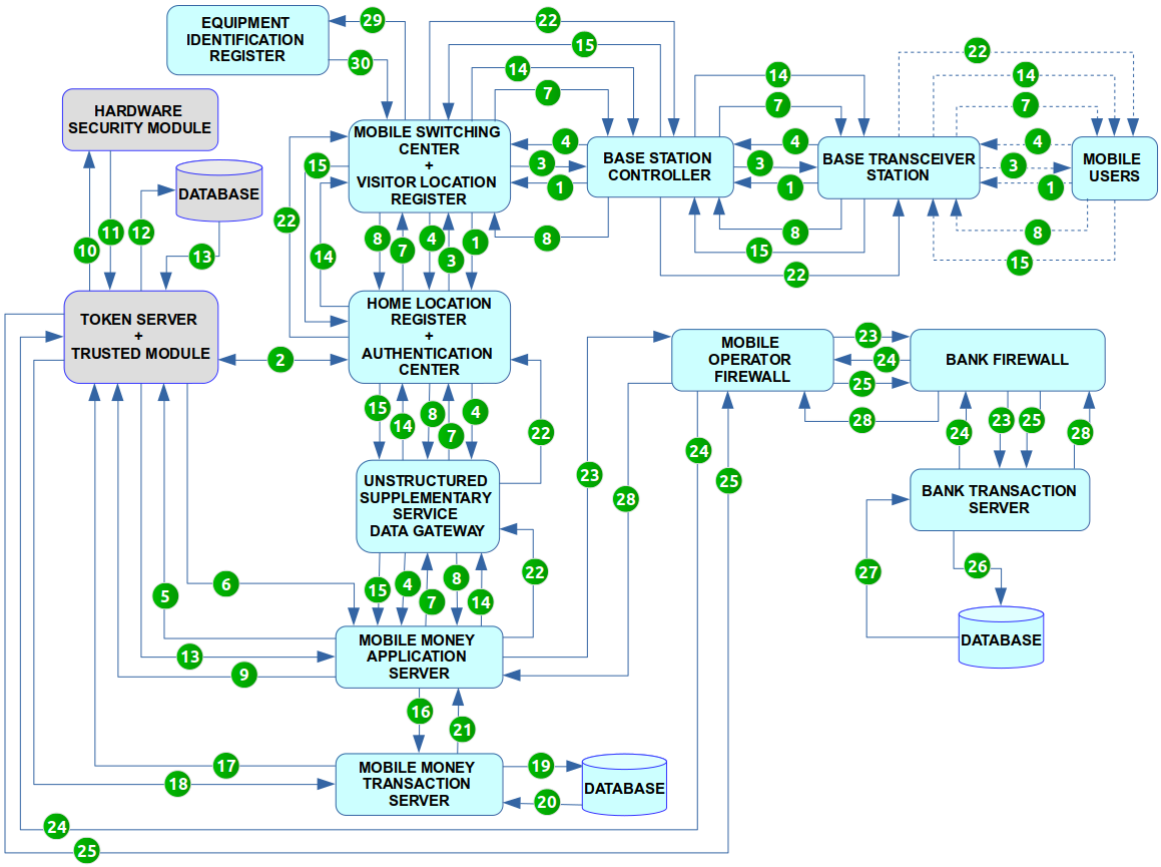

Figure 1. Encryption and tokenization scheme flow diagram

LEGEND

1. MS initiates the USSD session and identifies the trusted module (TM) by the TM identity string (IdTM)
2. Exchange of the plurality of public parameters (PK) between the trusted module (TM) and HLR
3. The PK associated with or generated by TM is distributed by the HLR to the mobile user through the MSC/VLR
4. The MS transmits the ciphertext (C) to the MoMoAS over the mobile network (BTS-BSC-MSC/VLR-HLR/AuC-USSD GW)
5 & 6. The MoMoAS retrieves the PK from TM and its own prvMoMoAS
7. MoMoAS generates pubMS, decrypts (C), and displays a menu at MS along with the request to submit the MoMoAN
8. The MoMoAN is submitted to the MoMoAS
9. The MoMoAS uses the TS's API to securely request a token related to the MoMoAN
10 & 11. The TS uses the HSM to create the token
12. The token, MoMoAN, and possibly the couple MSISDN-IMSI are stored in the token data store (database)
13. The TS securely returns the token to the MoMoAS
14. The token is sent back through USSD GW and HLR to MS where it is securely stored
15. The mobile user securely initiates the action to complete the transaction

16. The MoMoAS securely sends a request to the TrS
17. The TrS securely sends the token to the TS to retrieve the actual MoMoAN
18. The MoMoAN is securely returned to the TrS
19 & 20. The transaction information is securely sent to the Database to be validated and eventually completed
21. The transaction status is securely sent back to the MoMoAS
22. A notification status is sent back to MS through HLR with a flag to terminate the original USSD session
23. If the TrS is located outside the MNO network, the MoMoAS sends requests to the MtrS through an IPSec link
24. The MtrS securely initiates the request to the MNO's TS to retrieve the actual MoMoAN
25. The MoMoAN is securely returned to the MtrS
26 & 27. The transaction information is sent to the bank database to validate and complete the transaction
28. The transaction status is sent back to the MoMoAS and then to the mobile user via (22)
29 & 30. These flows are not relevant to the transactions undertaken between the MS and the MoMoAS.

Upon receipt of C, the MoMoAS determines whether it needs to get a prvMoMoAS from TM to decrypt C and retrieve M. If MoMoAS has not previously stored in memory its prvMoMoAS associated with the idTM, it authenticates itself with the TM to receive its own private key (prvMoMoAS) and the TM's public key (pubTM), as described in (5) and (6). At this time, the MoMoAS may also receive an updated set of PK from the TM. The MoMoAS verifies the sender of the message by inspecting C components, the MS public key (pubMS) generated locally from idMS and the idTM. Once the sender has been verified, the MoMoAS restores M and a menu requesting to submit the Mobile Money Account Number (MoMoAN) is displayed on the MS screen (7). When the mobile user submits this information to the MoMoAS (8), the MoMoAS processes the request and uses the TS's API to request a new token for the MoMoAN (9). The TS in its turn uses HSM (10 and 11) to create the token. HSM is only used to securely manage, process, and store cryptographic keys separately from sensitive data. The token, MoMoAN, and possibly the couple MSISDN-IMSI and the rest of the customer information are stored in the token database (12). The communication path between these back-end servers, MoMoAS, HSM and the TS is performed through TLS (Butler et al., 2020). The copy of the original MoMoAN is overwritten in MoMoAS so that it does not persist in memory and TS returns the token to the MoMoAS (13). In (14) and the token is sent back to the MS through USSD GW and HLR. This process is similar to the authentication scheme in GSM where the RAND is transmitted (via the BSC/BTS) to the mobile station (MS). The token will be used throughout the MNO/MMO's environments, instead of the real MoMoAN and it can't be extracted into anything valuable to fraudsters (Securosis, n.d.; PCI Security Standards Council, 2011). The mobile user performs a transaction in (15) and the MoMoAS sends a request to the Transaction Server (TrS) (16). The TrS sends the token to TS over TLS to retrieve the actual MoMoAN (17) and when TS receives the request, it validates the token using its copy of the token key and de-tokenizes it to obtain the original MoMoAN that is returned to the TrS (18). The transaction information is securely sent to the Database to be validated and eventually completed (19) and (20). The transaction status is sent back to the MoMoAS (21) and a notification status is sent back to MS through HLR with a flag to terminate the original USSD session (22). When the TrS is located outside the MNO network (mobile banking), steps 1-15 apply and MS uses the assigned token to transact with the Mobile Money Operator (MMO). To complete a transaction in this scenario, the mobile user initiates the action (15) and the MNO MoMoAS sends a request to the MMO Transaction Server (MTrS) over a secure connection, through the MNO firewall connected by IPSec to the bank firewall (23). When the MTrS receives the transaction request, it initiates the request to the MNO's TS over TLS to retrieve the actual MoMoAN (24). The communication between the two servers is encrypted by SSL/TLS certificates which are usually signed by trusted public Certificate Authorities (CAs). Another strategy is to issue self-signed SSL certificates and in this case, the certificate is self-signed with its own private key, instead of requesting it from a public or a private CA. If the certificate is generated on the MNO's TS it needs to be securely distributed to the MtrS and applied in its keystore. After the MNO's TS receives the request to retrieve the actual MoMoAN, it validates the token using its copy of the token key and de-tokenizes it to obtain the original MoMoAN which is returned to the MtrS over the secure link (25). The transaction information is sent to the bank database to validate and complete the transaction (26) and (27). The transaction status is sent back to the MoMoAS (28) and a notification status is sent back to the MS through HLR with a flag to terminate the original USSD session (22). Though the flow (29) and (30) are used in normal GSM phone hardware authentication, they are not relevant to the transactions being undertaken between the mobile user and the MoMoAS. For each transaction, a new token is generated and used only once and within a certain valid time period. If a token is used outside of these parameters, it is flagged as invalid to prevent the ability

to replay past transaction requests (replay attack). In the replay attack scenario, the same transaction request is resent, and another transaction can be done with the same data.

USSD transactions are performed using an MS when a mobile money customer initiates requests via the mobile network and terminates at the mobile money application server that can be administered by the MNO or an independent technology vendor such as a bank. Despite USSD being a session-based technology providing relatively stronger security than SMS, it can still be breached because GSM encryption is only applied between the MS and the BTS but not across of the rest of the MNO's network where messages are transferred in plaintext. This research work proposes new technologies for end-to-end encryption of USSD-based financial transactions and demonstrates how these new technologies are integrated and applied into the existing USSD technology. The exchange of public parameters from the TM to the MS through HLR is compatible with that used to authenticate the mobile station (MS) on a GSM network. Moreover, the changes made within the network will not negatively impact it because the benefits of the added tokenization and security features are likely to outweigh the additional cost of the inclusion of the TM and TS subsystem. The proposed prototype model is adaptable. It is likely to be sustained over time because USSD is a device-independent technology that can even be used with smartphones and potentially in newer related devices.

5. CONCLUSION

This paper provided a solution that demonstrated how an improved CLAE scheme coupled with tokenization add a security layer to USSD-based financial services. The solution allows the sender and the recipient to securely communicate with each other with the sender verifying public parameters from a trusted third party and locally generate the recipient's public key derived from its identity prior to encrypting a plaintext message. This potentially low cost solution increases mobile payments security over USSD without having to make deep changes in the GSM network architecture. Future research will pursue simulation and testing of the proposed prototype. Other researchers are also encouraged to do the same so that the potential efficiency and cost effectiveness of the proposed CLAE with tokenization solution can be determined. However, governments will need to adopt market-friendly laws and regulations as well as link talents with adequate MNOs or financial institutions which need specific skillsets to fill the knowledge gap once the solution is deployed. Creating a supportive environment in the initial stages of this solution's implementation could help speed-up the return on investments for MNOs as investments will be needed to support changes in accommodating new equipment or issuing SIM cards supporting the solution. Governments might also need to define guidelines and policies to support the proposed solution including the creation of an environment that safeguards security and privacy, information standards, and a legal framework for citizen privacy enforcement. There are security standards which address advanced cybersecurity mechanisms that are already implemented in many SSA countries in a locally customized form such as the General Data Protection Regulation (GDPR), Control Objectives for Information and Related Technology (COBIT) and ISO 27001 (Yadav, 2019).

REFERENCES

Al-Riyami, S. and Paterson, K., (2003). Certificateless public key cryptography. *Advances in Cryptology - ASIACRYPT 2003: Springer Lecture Notes in Computer Science*, Vol. 2894, pp. 452-473.

Agwanyanjaba, W., (2018). Enhanced Mobile Banking Security: Implementing Transaction Authorization mechanism via USSD Push. University of Nairobi. http://erepository.uonbi.ac.ke/handle/11295/153184

Aramé et al., (2022). State of the Industry Report on Mobile Money 2022. GSMA. https://www.gsma.com/sotir/wp-content/uploads/2022/03/GSMA_State_of_the_Industry_2022_English.pdf

Bahia, K. and Delaporte, A., (2022). The State of Mobile Internet Connectivity 2022. *GSMA.* https://www.gsma.com/r/wp-content/uploads/2022/10/The-State-of-Mobile-Internet-Connectivity-Report-2022.pdf

Boneh, D. and Franklin, M., (2001). Identity based encryption from the Weil pairing. *Advances in Cryptology - Crypto' 2001: Springer Lecture Notes in Computer Science*, Vol. 2139, pp. 213-229.

Butler, K. et al., (2020). Security Testing for USSD and STK Based Digital Financial Services Applications. *ITU and The Financial Inclusion Global Initiative (FIGI)*. https://figi.itu.int/wp-content/uploads/2021/04/Security-testing-for-USSD-and-STK-based-Digital-Financial-Services-applications-1.pdf

Castle, S. et al., (2016). Let's Talk Money: Evaluating the Security Challenges of Mobile Money in the Developing World. 1-10. 10.1145/3001913.3001919

Central Bank of Nigeria, (2019). The Regulatory Framework for the Use of the Unstructured Supplementary Service Data (USSD) in the Nigerian Financial System. *CBN*. https://www.cbn.gov.ng/Out/2020/FMD/CBN%20Rule%20Book%20Volume%201.pdf

Chironga, M. et al., (2018). Roaring to Life: Growth and Innovation in African Retail Banking. https://www.mckinsey.com/~/media/mckinsey/industries/financial%20services/our%20insights/african%20retail%20bankings%20next%20growth%20frontier/roaring-to-life-growth-and-innovation-in-african-retail-banking-web-final.ashx

Cook, M., (2019). This Country Recently Became Africa's Largest Economy: Now it's Too Big for Businesses to Ignore. *Brink*. www.brinknews.com/this-country-recently-became-africas-largest-economy-now-its-too-big-for-businesses-to-ignore/

Dryburgh, L. and Hewett, J., (2003). *Signaling System No. 7: Protocol, Architecture, and Applications*. Cisco Press.

ETSI, 1996. GSM Technical Specification (GSM 09.02), Version 5.3.0. *European Telecommunications Standards Institute (ETSI)*. https://www.etsi.org/deliver/etsi_gts/09/0902/05.03.00_60/gsmts_0902v050300p.pdf

Feige, U. et al., 1988. Zero-knowledge proofs of identity. *Journal of Cryptology*, Vol. 1, pp. 77-94.

Fiat, A. and Shamir, A., (1986). How to prove yourself: Practical solutions to identification and signature problems. *Advances in Cryptology - Crypto' 86: Springer Lecture Notes in Computer Science*, Vol. 263, pp. 186-194.

Gardezi, A., 2006. Security in Wireless Cellular Networks. *WUSTL*. https://www.cse.wustl.edu/~jain/cse574-06/ftp/cellular_security/index.html

GSMA, (2018). State of the Mobile Money Industry in Sub-Saharan Africa. https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2019/07/GSMA-Sub-Saharan-Africa-SOTIR-presentation.pdf

Gupta, A. and Jha, R., (2015). A Survey of 5G Network: Architecture and Emerging Technologies. *IEEE Access*, Vol. 3, pp. 1206-1232.

Hanouch, M. and Chen, G., (2015). Promoting Competition in Mobile Payments: The Role of USSD. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/586781468127790413/promoting-competition-in-mobile-payments-the-role-of-ussd

Heine, G., (1999). *GSM Networks: Protocols, Terminology, and Implementation*. Boston, Mass: Artech House.

Islam, S. et al., (2015). Secure end-to-end SMS communication over GSM networks. *Proceedings of 12th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, pp. 286-292.

ITU, (2021). *Measuring Digital Development Facts and Measures, Development Sector*. Geneva: International Telecommunication Union. https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2021.pdf

Iwasokun, G. et al., (2018). Encryption and Tokenization-Based System for Credit Card Information Security. *International Journal of Cyber-Security and Digital Forensics*, Vol. 7, pp. 283-293.

Klinger, A., (2021). Low resource requirement, quantum resistant, encryption of USSD messages for use in financial services. ITU. https://www.itu.int/dms_pub/itu-t/opb/tut/T-TUT-PROTO-2021-PDF-E.pdf

Kumar, P. et al., 2006. Mutual Authentication and Key Agreement for GSM. *Proceedings of the 2006 International Conference on Mobile Business*. https://ieeexplore.ieee.org/document/771073

Kune, D. et al., (2012). Location Leaks on the GSM Air Interface. https://www.researchgate.net/publication/267805153_Location_Leaks_on_the_GSM_Air_Interface

Liu, W. et al., (2020). Review of State of the Art: Secure Mobile Payment. *IEEE Access*, Vol. 8, pp. 13898-13914.

Loughney, J. et al., (2004). Security Considerations for Signaling Transport (SIGTRAN) Protocols. *IETF*. https://tools.ietf.org/html/rfc3788

Malek, B., (2014). Method and System for a Certificate-less Authenticated Encryption Scheme Using Identity Based Encryption, US Patent No. 8, pp. 694 771.

Mosweunyane, G. et al., (2014). Design of a USSD System for TB Contact Tracing. *Health Informatics*, Vol. 8, pp. 15-26.

Nyamtiga, B. et al., (2013). Security Perspectives for USSD versus SMS in Conducting Mobile Transactions: A Case Study of Tanzania. *International Journal of Technology Enhancements and Emerging Engineering Research*, Vol. 1, pp. 38-43.

Onyango, S., (2022). Africa accounts for 70% of the world's $1 trillion mobile money market. Quartz Africa. https://qz.com/africa/2161960/gsma-70-percent-of-the-worlds-1-trillion-mobile-money-market-is-in-africa/

Pagliusi, P., (2002). A Contemporary Foreword on GSM Security. *Infrastructure Security*, pp. 129-144.

Quentin, C., (2020). Towards a Triad for Data Privacy. 10.24251/HICSS.2020.535.

Scoping SIG & Tokenization Taskforce, August (2011). PCI Data Security Standard (PCI DSS) Information Supplement: PCI DSS Tokenization Guidelines, version 2. *Payment Card Industry (PCI) Security Standards Council*. https://www.pcisecuritystandards.org/documents/Tokenization_Guidelines_Info_Supplement.pdf

Pazarbasioglu et al., (2020). Digital Financial Services. World Bank. https://pubdocs.worldbank.org/en/230281588169110691/Digital-Financial-Services.pdf

Ramadan, M. et al., (2016). EEE-GSM: End-to-end encryption scheme over GSM system. *International Journal of Security and its Applications*, Vol. 10, pp. 229-240.

Scourias, J., (1996). Overview of GSM: The Global System for Mobile Communications. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.24.3034&rep=rep1&type=pdf

Securosis, (n.d.) Understanding and Selecting a Tokenization Solution. https://securosis.com/assets/library/reports/Securosis_Understanding_Tokenization_V.1_.0_.pdf

Sekyere, B. et al., (2018). Improving Electronic Banking in Ghana using USSD. *International Journal of Computer Applications*, Vol. 180, pp. 8-13.

Shamir, A., (1985). Identity-based cryptosystems and signature schemes. *Advances in Cryptology - Crypto' 84: Springer Lecture Notes in Computer Science*, Vol. 196, pp. 47-53.

Silver, L. and Johnson, C., (2018). Internet Connectivity Seen as Having Positive Impact on Life in Sub-Saharan Africa. *Pew Research Center's Global Attitudes Project*. https://www.pewresearch.org/global/2018/10/09/internet-connectivity-seen-as-having-positive-impact-on-life-in-sub-saharan-africa/

Smith, M. et al., (2009). Securing Mobile Phone Calls with Identity-Based Cryptography. *Proceedings of the Advances in Information Security and Assurance Third International Conference and Workshops*, pp. 210-222.

Srinivas, S., (2001). The GSM Standard (An Overview of its Security). *SANS Institute*. https://www.sans.org/reading-room/whitepapers/telephone/gsm-standard-an-overview-security-317

Srinivas, V. and Wadhwani, R., (2019). Recognizing the Value of Bank Branches in a Digital World. *Deloitte*. www2.deloitte.com/us/en/insights/industry/financial-services/bank-branch-transformation-digital-banking.html

Thomas, M. and Panchami, V., (2014). An encryption protocol for end-to-end secure transmission of SMS. *IEEE Transactions on Information Forensics and Security*, Vol. 9, pp. 1157-1168.

Wan, H., (2022). Increases in Africa's Older Population Will Outstrip Growth in Any Other World Region. United States Census Bureau. https://www.census.gov/library/stories/2022/04/why-study-aging-in-africa-region-with-worlds-youngest-population.html

Wadhwa, D., (2018). The Number of Extremely Poor People Continues to Rise in Sub-Saharan Africa. *World Bank*. https://blogs.worldbank.org/opendata/number-extremely-poor-people-continues-rise-sub-saharan-africa

Weichbroth P. and Łysik, L., (2020). Mobile Security: Threats and Best Practices. *Mobile Information Systems, 2020*, pp. 1-15.

Wu, H. et al., (2011). Mobile Payment Framework Based on 3G Network. *Proceedings of the Third International Symposium on Electronic Commerce and Security Workshops (ISECS '10)*, pp. 172-175.

Wu, S. and Tan, C., (2009). A High Security Framework for SMS. *Proceedings of the 2nd International Conference on BioMedical Engineering and Informatics (BMEI)*. https://ieeexplore.ieee.org/document/5305796

Wyrzykowski, R., (2020). Mobile connectivity in Sub-Saharan Africa: 4G and 3G Connections Overtake 2G for the First Time. *GSMA* https://www.gsma.com/mobilefordevelopment/blog/mobile-connectivity-in-sub-saharan-africa-4g-and-3g-connections-overtake-2g-for-the-first-time/

Yadav, N., (2019). ISO 27001 vs. COBIT: A Comparison. *Advisera*. https://advisera.com/27001academy/blog/2019/05/06/cobit-vs-iso-27001-how-much-do-they-differ/