Reliable Identification of Counterfeit Medicine Using Camera Equipped Mobile Phones

Saif ur Rehman, Raihan Ur Rasool, M. Sohaib Ayub, Saeed Ullah, Aatif Kamal, Qasim M. Rajpoot, and Zahid Anwar
School of Electrical Engineering and Computer Science (SEECS),
National University of Sciences and Technology (NUST), Islamabad, Pakistan
Email: (saifur.rehman, raihan.rasool, 08bitsohaiba, saeed.ullah, aatif.kamal, qasim.rajpoot, zahid.anwar)@seecs.edu.pk

Abstract—**The sale of counterfeit medicine is a continuously growing global problem costing billions of dollars and effecting hundreds of thousands of precious human lives annually. Consumers of medicine have no reliable and simple way of distinguishing genuine medicine from counterfeit. In this paper we analyzed existing techniques of counterfeit medicine identification on the bases of usability, security and scalability. Different shortcomings and security weaknesses of these reviewed solutions are identified. Based on these observations, a set of requirements is determined. These requirements address different aspects of a complete solution, such as usability by end users, security of the system against various types of attacks and scalability of the solution. We then propose a framework for reliably ensuring the fidelity of purchased medicine. The proposed framework is simple to use for consumers and does not require specialized equipment, training or technology. The proposed framework will enable the consumers to verify the legitimateness of their purchased medication by using their mobile phones. Data matrix is used to make the verification code machine readable.**

Keywords-**Counterfeit medicine; Mobile Phones; Product Authentication; Data Matrix.**

I. Introduction

Counterfeiting medicine is a multi-billion dollar industry [1, 2], some sources [3–5] quote its annual turnover as high as 70 – 75 Billion USD. The counterfeit medicine market is continuously growing despite the regulatory efforts by governments and international organizations such as World Health Organization (WHO) on international level. It is estimated that 10% of all medicine marketed globally are counterfeit [1], this percentage is as high as 80% in some poor and developing countries. The main reasons for this phenomenon are a high profit margin for counterfeiters and the lack of reliable methods for identification of counterfeit medicine on consumer level. Different pharmaceutics have tried techniques such as holograms [6], special seals, color-variable inks, Ultra violet (UV) inks etc., which had little success because copying these identifiers is not impossible and even inaccurate copies can deceive or confuse consumers. As the market for counterfeit medicine grow, even in European countries due to sales over the internet, counterfeiters are developing more sophisticated techniques for copying medicine packaging and the identification markings used by original manufacturers. There is a need of using such technology that makes counterfeiting financially and temporally impractical. Information technology offers reliable product identification through the use of RFID chips. RFID chips are reliable but infeasible for disposable products such as medicine packing due to their relatively high price of between 15 to 30 cents [7] and the need of RFID readers which cost over USD 500 at least [7]. There are other costs also, such as RFID antennas at USD 250 and above, connection to a computer where the data is to be processed and power requirements. These requirements mean high costs and loss of portability and make a widespread deployment infeasible as major targets of counterfeit medicine are poor and developing countries. Mobile phone infrastructure is deployed in most countries worldwide. A few organizations and government bodies are trying to use this technology for medicine verification; however the existing solutions have weaknesses that can be exploited by counterfeiters to bypass, disrupt or deceive the verification system. We present a solution which mitigates these vulnerabilities and provide a simple to use yet secure technique for verification of medicine. Rather than using encryption for making the verification code secure from copying and replication, we use long random strings as identification codes. Random character strings are inherently secure against analysis based attacks and we show that the computational complexity of a brute force attack makes it temporally and financially infeasible. A central verification database accessible via cellular networks will be used for authentication.

The rest of this paper is organized as follows. Section II covers work done in counterfeit medicine identification. Possible attacks and countering strategies for reviewed solutions are presented in section III. Proposed solution is discussed in section IV. Conclusion and future work is discussed in section V.

II. Related Work

The use of mobile phones for identification of counterfeit medicine is mostly an under explored domain. Most research on anti-counterfeiting suggests the use of RFID or similar tamper proof chips such as Memory Spots [9] that are specially designed for anti-counterfeiting. A few attempts

978-1-4577-1169-5/11/$26.00 ©2011 IEEE

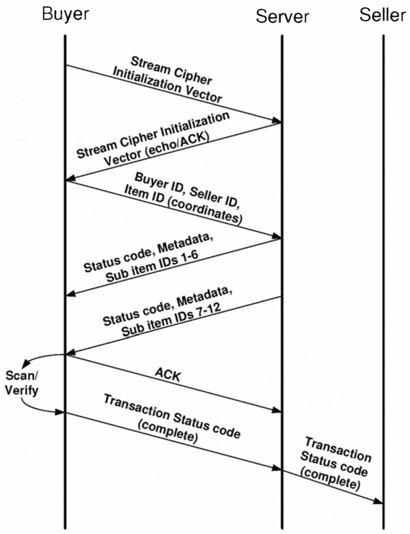
Buyer
Server
Seller
Stream Cipher Initialization Vector
Stream Cipher Initialization Vector (echo/ACK)
Buyer ID, Seller ID, Item ID (coordinates)
Status code, Metadata, Sub item IDs 1-6
Status code, Metadata, Sub item IDs 7-12
Scan/ Verify
ACK
Transaction Status code (complete)
Transaction Status code (complete)

Figure 1. Epothecary's protocol message flow for a successful transaction [8]

have been made recently to develop medicine verification systems using verification codes that can be verified via mobile phones. A brief description and analysis of these follow.

A. Epothecary

M Paik et al. published a drug pedigree tracking and authentication mechanism that uses camera equipped mobile phones, and 2-d barcodes [8]. SMS services of a GSM network are used for communicating with the central server. The proposed system is named Epothecary. It registers all participants in medicine shipment handling such as suppliers, distributors and vendors and each handover requires verification and registration with a central server. The final sale to a consumer is also registered with the central sever. The mechanism is developed such that it is secure against different possible attacks. A major drawback in this system is the mandatory human involvement in the process flow, which creates a serious performance bottleneck and limits the scalability. It should be noted that at higher levels of any product distribution hierarchy, several thousand transactions are expected on daily basis. The system also requires all the participants except consumers to be registered with a Regulatory Body (RB) and that the medicine is packed in such a way that ID tags of all units are visible without disassembling the package. When a medicine package changes hands, the server is contacted multiple times. The purchaser has to scan the ID tags of the seller, the package involved and his own ID tag.

Figure 1 shows the number of messages required for a single successful transaction. The communication from a buyer's end requires them to scan several ID tags. Once the initial verification of participants take place, the buyer subsequently has to scan the ID tags of the smaller units contained in the shipment. The minimum time required for a single successful transaction is calculated as follows, the numbers in brackets indicate seconds. Starting application, entering password $(3 - 5)$, initial tag scans 3 * (3-5), sms send-receive time $2 * (3 - 4)$, scanning of subsidiary tags $12 * (3 - 5)$, sms send- time $2 * (3 - 5)$ This sums up to $60 - 98$ seconds, which is the bare minimum time required, ignoring all network latency and errors both human and electronic. This means that working 24 hours non-stop at ideally fast speed without making any mistakes and without any network latency and errors, a seller or buyer will still only be able to complete 1440 transactions. In this process more than $10,000$ SMS messages will be sent and received and more than 4000 tag scans will be needed. Table I show the time and number of SMS messages required for different number of transactions in best case scenario. As the above calculation clearly shows, while the system is theoretically sound, it is practically un-implementable on large scale due to the human involvement in the process flow and the over complicated procedure.

B. mPedigree

mPedigree is an organization from Ghana using mobile phones for medicine verification. The technique they employ is simple enough; a participant medicine manufacturer assigns a serial number to each pack of medicine. This serial number is stored in a registry which can be accessed by the consumer via free SMS. The buyer can send the serial number to the registry through a text message and the registry then replies usually within a minute, certifying or denying the source of medicine as legitimate. The technique is very simple to use for a literate person, as the verification service is free, a mobile phone can be borrowed for verifying a medicine. A quick response time enables the buyer to ensure the legitimacy of medication in time. This technique poses little obstacle for a counterfeiter once its working gets know. A legitimate original code can be copied on to as many copies as desired for a product and will result in a response verifying the medicine as genuine whereas, it would in fact be counterfeit. The verification code is short i-e only eight numeric characters long [10] and can easily be subjected to even a simple brute force attack.

C. EFPIA

The European Federation of Pharmaceutical Industry and Associations has carried out a pilot project in Sweden Stockholm which provides Pharmacy level verification of medication [11]. The packing of medicine carries a Data Matrix which contains a unique identifier. Upon purchase, the package is verified by point of dispense personnel using a data matrix reader, which reads the unique identifier and sends it to the centralized database for verification, if the identifier is present and its status is not marked as

Table I
NUMBER OF POSSIBLE TRANSACTIONS BY A SINGLE USER OF EPOTHECARY

Number of Shipments %	Number of SMS	Minimum Tag Scans	Time taken in hours
100	700	300	1.67
1000	7000	3000	16.7
1440	10080	4320	24

verified it is considered genuine, an alert is generated if the unique identifier has been identified previously, and the particular medicine associated with that identifier is branded as counterfeit. This technique successfully solves the problem of multiple verifications by same identifier. In this solution it is assumed that the point of dispense is trust worthy and not involved in marketing of counterfeit drugs, an assumption that might not be true for many developing and poor countries lacking firm regulatory authorities. The option of user level verification is not considered in this solution. As the verification takes place over the internet, the central verification register is prone to attacks such as Denial of Service DoS and brute force amongst others. A brute force attack trying every possible combination on the registry will not provide the attacker with working verification codes but it will result in crossing off of legitimate codes as counterfeit.

D. NAFDAC

The National Agency of Food and Drug Administration Control of Nigeria has recently launched a project in which medicine packing will include a scratch card containing symbols that, when send via SMS to a given number will verify the medicine for the consumer [12]. Like mPedigree, this is a simple to use technique requiring a common mobile phone set. NAFDAC is using asymmetric encryption for symbol generation to avoid valid symbol generation by counterfeiters. Multiple verification of same authentication code is not countered in this technique. Only six symbols are used, which for the small character set of simple mobile phone does not give a high enough security. This technique is also vulnerable to brute force attack.

III. POSSIBLE ATTACKS AND COUNTERING STRATEGIES FOR REVIEWED SOLUTIONS

In this section we discuss some suggestions for improving the security of existing applications. We show the complexity and cost of attacks mathematically to demonstrate the significance of these suggested changes.

A. Versus Multiple Verifications of a Single Identifier

A problem with mPedigree and NAFDAC medicine verification method is that they do not counter the possibility of reuse of a legitimate identifier. We propose to address this problem by marking a successfully verified identifier as ''Identified'' in the CVR so that a subsequent verification request for the same code would identify the medication as counterfeit. This means counterfeiters will be forced to obtain a separate original sample for each copy they indent to create and even so, the original sample won't be verifiable; this option of course is financially and operationally infeasible for counterfeiters.

B. Versus Brute Force Attack

Brute force is a simple attack, trying all possible combinations to find the required ones. In case of medicine verification techniques the counterfeiters won't need many codes if multiple verification of a single code is not countered. However if multiple verifications of a single code are countered by baring a code once it is used, brute force attack becomes an option for sabotaging the system, thereby damaging the system's reliability. This attack will not give the attacker working codes but it will result in the cancelation of working codes and legitimate medicine will be termed as counterfeit, resulting in false alarm of counterfeit medicine entering the supply chain, financial loss to stakeholders and loss of trust by customers and manufacturers on the dependability of the system. For example, mPedigree uses a string of eight numeric characters as an identifier. This means the maximum possible combinations are one hundred million ($00000000 - 99999999$). Out of these, identifiers would be chosen using a pseudo random number generating algorithm. If chosen randomly, the probability of all numbers being chosen is equal; the probability of a single number is given by:

$$P1 = 1/10000000 = 0.00000001 \tag{1}$$

Considering a realistic example, we can say that a hundred thousand medicine packs are going to be assigned identifiers from the pool of one hundred million possible identifiers. Then the probability of finding a working code P_w is

$$P_w = identifiers/sourcesetsize = 10^5/10^8 = 0.001 \tag{2}$$

This means that in a thousand trials, one working code is likely to be found. A thousand trials manually can take a long time but an automated program can do that in a few minutes, even seconds. Different approaches can be an option for countering this attack.

C. Making the Attack Financially Infeasible

Take nominal service charges from the user, for example one cent/penny etc. This will not harm a genuine user's financial status, but for a brute force attacker, thousand trials would mean at least 10$ per working code and hence

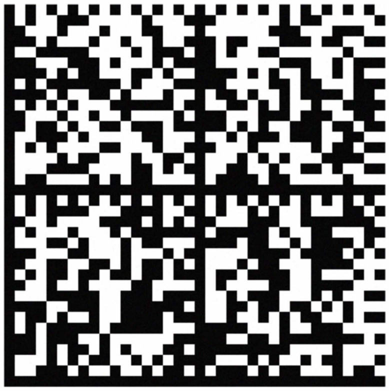

Figure 2. The shown data matrix illustrates the data encoding and capacity of data matrix.

financially infeasible. This approach can potentially cause some users to shy away from verifying their medicine, especially commercial users such as medicine shop and hospital staff.

D. Making the Attack Temporally Infeasible

Set a limit on the number of verifications permitted from a particular user in a given space of time. This limit, if realistically adjusted will permit a genuine frequent user such as a medicine store keeper who verifies medicine for customers, to use the system. A brute force attack will not be feasible as, even an eight character long verification code gives approximately hundred million possible combinations. While a high number, hundred million combinations can be exhausted in a few days using a simple application designed for the purpose. Imposing a time limit will force the attacker to wait and thus repel the attack. Even if a user is allowed to attempt to verify a code every 10 seconds, it will take a single user more than 31 years, a hundred users more than 3 years of 24 hour non-stop work. The down side of this approach is additional complexity on the server's side. All numbers requesting verification will be matched against the list of numbers from the past ten seconds. If the server gets a lot of requests at one time, it might crash due to the high number of threads it would be running for count down timers on currently barred numbers.

E. Shortcomings

The suggested fixes can work against simple attacks if the solution is deployed over GSM networks, but more complex attack are possible and can be used to disrupt the working and credibility of the solution. The existing solutions are developed and implemented for local use and pose potential problems in scalability. For example, assigning of identifiers to imported medicine will be a problem and might allow counterfeit medicine to enter the supply chain. Similarly, if international manufacturers do support the system, verification via mobile phones will become a costly option. We propose a solution that is flexible enough to be customized for a large scale, multi participant international implementation and it also increases the computational complexity for any attack many folds as shown in section IV-C.

IV. Proposed Methodology

Our proposed system utilizes the existing infrastructure of mobile technology for medicine verification. A data matrix will be printed on the packing of medicine which will contain, Manufacturer ID, Product ID, unique ID of the package, the authentication code and optional meta data. Hash value of the authentication code will be stored in a Central Verification Register (CVR). Hash value will be stored rather than the original authentication code to protect the system against a compromised database. Reverse engineering a value from its hash is infeasible. A camera equipped mobile phone will be used to snap shot the data matrix using a data matrix reader specially developed for this purpose. The application will read the data and paste the unique identifier code into a text message i-e SMS. The SMS will be sent to the CVR. The CVR upon receiving the unique identifier will first search for Manufacturer ID, if found, it will check for the Product ID. When both manufacturer and product are identified as valid, the CVR will search the Verified Medicine Log (VML) for the unique ID of package. If the unique ID is listed in the VML, it would indicate reuse and the medicine will be identified as counterfeit. If the unique ID is not present in VML, the CVR will compute the hash value of the authentication code and check its value in the database. Once the authentication code is verified, the CVR will confirm that medicine as genuine and notify the user by replying on the same number also providing additional information for the benefit of the user such as date of manufacture, date of expiry, manufacturer, active ingredient and its dosage. The CVR will also record the verification event in the VML along with relevant information such as unique ID, mobile number requesting verification and date of verification. The CVR will also check the expiry date of a medicine and will notify the users if they try to verify expired medicine. The Manufacturer ID and Product ID are included to serve multiple purposes; they serve as a preliminary check, if their values do not check out, the verification request responds by identifying the medicine as counterfeit. The ID's can be used to stop sales of a particular product when needed and statistical data for research might also be obtained easily.

A. Rationale for Proposed Framework

The question might arise that if existing solutions' security issues can be fixed easily, what is the need to develop a comparatively complex framework? To answer this question we have to consider the following points

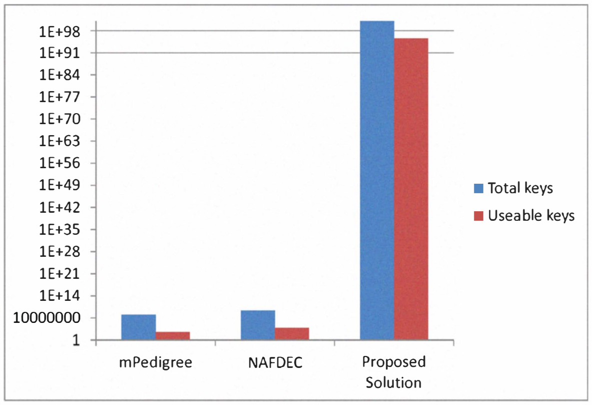

Figure 3. Total vs useable identifiers

- While simple attacks such as brute force and multiple verifications are countered, complex attacks might still be possible.
- Requiring users to enter longer identifiers increases the complexity of usage of any solution.

B. *Benefit of Using Data Matrix*

Data matrix presents data in machine readable format; this saves a user the trouble and time of manually typing in difficult to remember data such as long URLs and random number strings. The Data matrix shown in Figure 2 is encoding:

$$FexofenadineHCL60mgTab.Mfg : 03/09Exp : 03/12ID : Rd\#9 \wedge L@5 - (beKEp\#, . * @$$

The data matrix is shown in the size it was read by a 2 Mega Pixel camera phone. As shown, the identifier can contain capital and small alphabets, numbers and special characters.

If the medicine verification code in existing solutions is increased from six or eight characters to even fourteen or sixteen, it will increase the difficulty for the user and yet, for today's processing power, the number is still low enough to mount attacks on the system.

C. *Strength of Proposed Solution in Terms of Mathematical Complexity*

The number of maximum unique possible combinations given by any character set is given by

$$S^n$$

Where S represents the total individual symbols in the character set; 0 to 9 in case of numeric, and n represents the number of symbols to be used. For existing solutions using 8 digits the total possible symbols are $108 = 100000000$

This number will increase ten times for every added digit. The proposed solution uses data matrix which can utilize a much larger character set, namely the GSM 03.38 07bit-character set which has a total of 127 alphanumeric and special characters that are supported by all mobile phones. The special characters might be hard to find and recognize for a user but machine readable data matrix saves the user from facing that complexity and the user would just have to take a photo of the data matrix with a camera phone. The number of possible combinations using the full character set with eight digits is shown by

$$127^8 = 67675234241018881 \tag{3}$$

While this possible number of identifier is many folds larger than the one provided by 108, due to the machine readability of the data matrix all 160 characters can be used for identification. The maximum possible size of identifier set is using a single SMS text message is given by

$$127^{160} = 4.06 \times 10^{336} \tag{4}$$

Here, we compare this number to the computational complexity of Advance Encryption Standard (AES), which is the international standard in secret key cryptography. The computational complexity of a brute force attack on AES is given by

$$2^{256} = 1.16 \times 10^{77} \tag{5}$$

As it can be clearly seen, the computational complexity of an attack on the proposed system is several folds higher than AES and therefore completely infeasible. To further elaborate this point, we consider the example of one billion identifiers selected from the pool size calculated. The probability of any number to be selected as one of the billion is given by

$$Pn = 10^9/127^{160} = 2.46 \times 10^{-328} \tag{6}$$

The probability as shown is extremely low. This number means that at the rate of one billion attempts per seconds by one billion individual attacking bots will still take hundreds of billions of years to find a single working code from a pool of one billion randomly selected identifiers.

D. *Comparison to Reviewed Solutions*

To further elaborate the capacity for scalability and protection against attacks of the proposed solution, it is compared to the reviewed solutions. To compare the different solutions, the numbers of useable identifiers possible in each solution are compared. The number of useable identifiers is derived by the following formula. Let the pool size = x Let the required failed attacks rate = n The permissible error/successful attack rate = 1-n The maximum number of permissible errors/successful attacks in the pool size is given by

$$x * (1 - n)$$

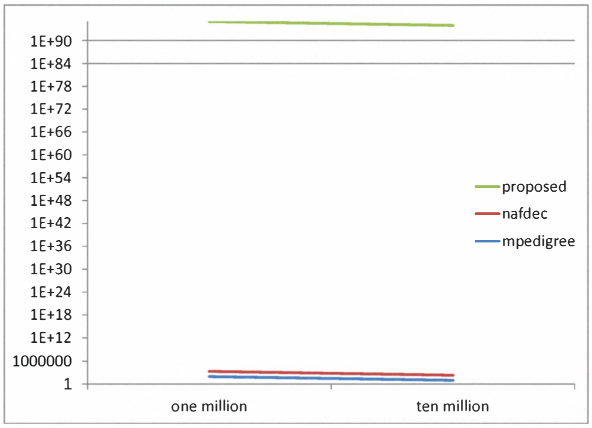

Figure 5. Brute force attacks needed for one success in one and ten million active identifiers. NAFDEC & mPedigree vs proposed solution

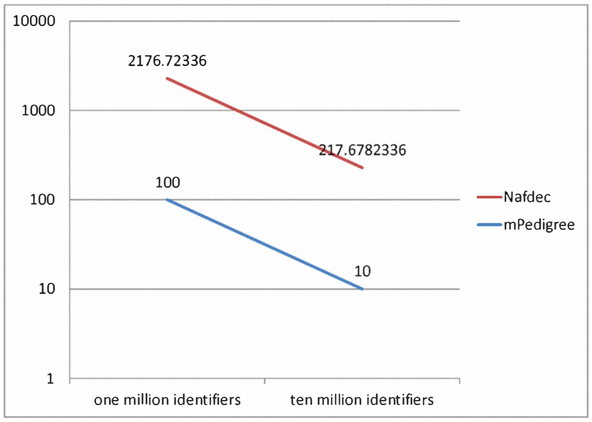

Figure 4. Brute force attacks needed for one success in one and ten million active identifiers. NAFDEC vs mPedigree

The values are calculated for one success out of one million attempts, one out of ten million attempts, and six sigma standard of quality which is represented by the sixth movement around the mean on a standard normal bell curve. The six sigma standard of quality has been developed by Motorola and several others have adopted it. The six sigma standard is used here because it measures quality in terms of error rate. The permissible error rate by six sigma standard is approximately 0.00034%.

It should be noted that for ease of illustration, the graphs presented for the proposed solution are based on string length of 48 characters, and not 160 that constitute a single SMS.

Figure 3 shows the total number of identifiers possible for each solution and the number of useable identifiers based on six sigma standard. Figures 4 and 5 show the number of brute force attempts needed to find one of the active identifiers. It can be seen that for ten million active identifiers the required number of brute force attempts is extremely low. The number of medicine in a single medicine store/pharmacy can range from a few to several thousands. A city may have millions of medicine present in its hospitals and pharmacies. These results show that the reviewed solutions are infeasible for large scale implementation due to weak security.

V. Conclusions and Future Work

The extremely high number of possible identifiers will make an attack infeasible and inefficient on the proposed solution. The proposed solution neither requires any additional infrastructural investment nor any technical training for consumers, and therefore it can be deployed quickly and with little expense.

This paper presents a solution which is resistant against different attacks and will allow reliable verification of medicine using a simple camera phone and a single SMS over a GSM network. As different GSM networks are deployed in different areas, issues such as securely and reliably accessing CVR, verification of internationally manufactured medicine and verification of medicine over the internet are open areas of research.

References

[1] "World Health Organization (WHO) fact sheet N275 "substandard and counterfeit medicine". <http://www.who.int/mediacentre/factsheets/2003/fs275/en/>."

[2] "WHO article "General Information on Counterfeit Medicine". <http://www.who.int/medicines/services/counterfeit/-overview/en/>."

[3] "Megget Katrina, "Drug theft costs industry up to USD 1 BN a year" online report published Oct 2, 2007. <http://www.outsourcing-pharma.com/Contract-Manufacturing/Drug-theft-costs-industry-up-to-1bn-a-year>."

[4] "Pitts Peter, "Counterfeit Drugs to reach USD 75BN by 2010" Published Nov 2005. <http://www.heartland.org/policybot/results/17948/-Counterfeit_Drug_Sales_to_Reach_75_Billion_by-_2010_Report_Says.html>."

[5] "Hirschler Ben, report on Pfizer survey regarding "counterfeit sales in Europe" published at Reuters UK website Feb 16, 2010. <http://uk.reuters.com/article/idUKLDE61E16A20100216-?sp=true>."

[6] "Aldhous Peter, "Hologram Wars". <http://www.nature.com/nature/journal/v434/n7030/box/-434132a_BX1.html>."

[7] "FAQ from RFID journal. <http://www.rfidjournal.com/faq/20>."

[8] M. Paik, J. Chen, and L. Subramanian, "Epothecary: cost-effective drug pedigree tracking and authentication using mobile phones," in *Proceedings of the 1st ACM workshop on Networking, systems, and applications for mobile handhelds.* ACM, 2009, pp. 13–18.

[9] H. Balinsky, E. McDonnell, L. Chen, and K. Harrison, "Anti-counterfeiting using memory spots," *Information Security Theory and Practice. Smart Devices, Pervasive Systems, and Ubiquitous Networks*, pp. 52–67, 2009.

[10] "mPedigree home page. <http://mpedigree.net/>."

[11] "EFPIA "Coding and Identification of products: Towards safer medicines supply" Published May 2009. <http://www.efpia.org/Content/Default.asp?PageID=566>."

[12] "Okoyo Chidi, "NAFDAC to fight drug counterfeiting by SMS" published March 5, 2011. <http://dailytimes.com.ng/article/nafdac-fight-drug-counterfeiting-sms>."