ISSN 2007-9737

Security Analysis of the Mexican Fiscal Digital Certificate System

Luis Rivera Zamarripa[1], Lil M. Rodríguez[2], Miguel Ángel León Chávez[3]
Nareli Cruz Cortés[1], Francisco Rodríguez Henríquez[4]

[1] Instituto Politécnico Nacional,
Centro de Investigación en Computación,
Mexico

[2] CONACYT Research Fellow-INAOE, Tonantzintla, Puebla,
Mexico

[3] Benemérita Universidad Autónoma de Puebla,
Mexico

[4] CINVESTAV-IPN, Computer Science Department,
Mexico

lrivera_a13@sagitario.cic.ipn.mx , lmrodriguez@inaoep.mx, mleon@cs.buap.mx
nareli@cic.ipn.mx, francisco@cs.cinvestav.mx

Abstract. In 2005 the Mexican National tributary system (SAT) started an ambitious public key infrastructure project with the aim of providing to each Mexican citizen a public/private key pair along with a digital certificate that was issued by SAT itself. As of March 2016, approximately a total of 17 million certificates have been issued. This e-government system permits Mexican citizens to exercise a series of digital on-line services such as: tax declaration, official receipt issuing/verification, contract signing, etc. In particular, all Mexican official invoices became digital by January 2016, effectively going paperless for this service. In this paper, we carefully analyze the Mexican PKI system showing that it has several weak points that can be attacked by malicious adversaries. We report experimental evidence showing that one can launch a simple dictionary attack on SAT's password-based authentication system. We also argue that due to the fact that the hash function SHA-1 has been recently completely broken, an attacker can produce the same signature for two different documents that will verify correctly when using any old FIEL certificate that has the RSA-1204/SHA-1 signature suite.

Keywords. Information security, Mexican public key infrastructure system, digital certificates, RSA.

1 Introduction

Since as early as 2005, there has been a widespread use of electronic invoicing among Latin American countries. In fact, more than half of the roughly 25 billion electronic invoices that were exchanged globally in 2014, were issued in the Latin America region [4]. In particular, Mexico issued roughly 5.8 billion electronic invoices during the 2015 year [28]. As a consequence of these developments, Latin America has become what is arguably the most advanced region worldwide in this field.

Most of the biggest economies in the region, such as, Argentina, Brazil, Chile, Mexico, and Peru, among others, have enacted laws for compulsory issuance and submission of electronic invoices, which has important repercussions for the overwhelming majority of their citizens.

During the last decade, a considerable effort has been done in Mexico to introduce and legislate strong e-commerce and e-government systems. According to *The Economist*, Mexico is the third country in Latin America that has developed an

electronic invoicing model [31]. Since January 2005, the Mexican government has allowed taxpayers to generate fiscal digital invoices, today known as CFDI.[1] These documents are issued through an ambitious public key infrastructure project, which is offered and regulated by means of the Tributary Administration Service, SAT.[2] A previous analysis of the security of this system can be found in [6, 13].

Two of the main objectives of the CFDI service are: (1) To automatize the accounting process of individuals and enterprises and; (2) To thwart the galloping tax evasion, which in Mexico is alarmingly high. Significantly, Fernando Martínez Coss a high-rank officer of the Mexican Tax Administration Service declared that [31]:

> "between 2007 and 2009, the SAT lost $\$3.4$ billion, largely due to what it euphemistically calls *apocryphal invoicing*".

According to the fiscal policies rolled out in 2014, SAT has reported until September 2016 a total of $54,738,719$ taxpayers. From that universe, $20,490,190$ are regular taxpayers,[3] $32,400,350$ are salaried employees, $16,621$ large regular taxpayers, $1,822,870$ are company representatives,[4] and $8,688$ are large taxpayers [29]. The vast majority of that universe of Mexican taxpayers are required to use SAT's CFDI system. As a result of this measure, approximately $5,782,122,364$ CFDI invoices were issued during 2015 along [28].

Currently, there are two types of electronic certificates in use in Mexico: the FIEL and the CSD certificates[5]. Both of them are issued by SAT to the taxpayers. The CSD certificate is specifically used to sign invoices, while the FIEL certificate permits taxpayers to perform some fiscal transactions, such as printing official records, FIEL revocation or renovation, and income tax declaration, among the most important services. Also, the FIEL certificate can be used to sign commercial contracts that are legally-binding in Mexico.

Most if not all of the services that can be carried out by means of the FIEL and CSD certificates, involve the exchange of sensitive information that should offer security assurances against any malicious entity. It is therefore essential that the security guarantees incorporated to the FIEL and CSD certificates meet the highest technological standards. Furthermore, any possible security flaw could seriously impact the financial assets of both, Mexican individuals and institutions operating in Mexico.

One of the most important security services facilitated by a Public Key Infrastructure (PKI) system is the digital signing of documents. A digital signature is the analog of the more familiar autograph signature, which is routinely attached to a written document as a mechanism to accomplish the signee's authentication. However, the digital signature mechanism is in principle more powerful than its autograph counterpart, in the sense that it also offers protection against data modifications. In order to achieve this feature, a signature scheme uses a hash function algorithm. A cryptographic hash function takes a document of arbitrary length as its input, and produces a unique fingerprint of it. This fingerprint is the actual digital object that gets to be signed by the chosen digital signature cryptographic scheme.

As for SAT's FIEL and CSD certificates, RSA-1024 and SHA-1 were chosen as the core primitives for the signature scheme and the hash function, respectively. In May 2015 however, the SAT published in the Official Journal of the Federation [24] that the new version of its signature algorithm should produce from then on, a combination of RSA-2048 with SHA-256. Since apparently the last RSA-1024 FIEL and CSD certificates were issued during the 2015 year, one needs to wait until 2020 to be sure that all the RSA-1024 FIEL certificates have been phased out. Thus, it seems that for a period of time that may last up to four years, the two parametrizations will live together side by side.

Additionally to the aforementioned PKI scheme, SAT has allowed that some fiscal services could optionally be performed by Mexican citizens

[1] "*Comprobantes Fiscales Digitales por Internet*"(fiscal digital invoices via Internet).

[2] After its name in Spanish: "*Servicio de Administración Tributaria*".

[3] SAT uses the Spanish term: *Personas Físicas*.

[4] SAT uses the Spanish term: *Personas Morales*.

[5] Which stand for "*Firma Electronica Avanzada*" and, "*Certificado de sello digital*", respectively.

and enterprises by means of a password-based authentication scheme. This password-based system is named CIECF.[6] The services supported by the CIECF are among others, CSD certificate download, certificate requests, income tax declarations, etc.

Hence, in order to perform a technical analysis of the security level offered by the SAT system, it is necessary to consider both, its cryptographic primitives used in its signature scheme and digital certification system and also, the strength of its password-based access control system. It is easy to see that if somehow an attacker can recover the CIECF (password) and/or obtain the private key corresponding to a given certificate, then the consequences for the legitime user would be dear, as all of the fiscal services discussed above can be exploited by the attacker on behalf of her victim.

Under the assumption that it is possible to attack the password system and/or to break a Mexican digital fiscal certificate, we present in the following several scenarios that describe some of the legal aftereffects for the potential victims:

Scenario 1: Invoice Forgery. The main objective behind a digital invoice infrastructure is that the taxpayer can record her sales and expenditures using digital invoices. Let us consider the situation where an attacker manages to obtain the private key corresponding to a given CSD certificate, which happens to be owned by a store. This action would allow the attacker to forge digital invoices supposedly issued by that store. After that, she could justify any of her expenses using fake invoices created by her.

Scenario 2: Income Tax Declaration. Citizen as well as enterprise income tax declaration can be done through the SAT web site. To this end, SAT has designed two authentication mechanisms available for citizens and enterprises, namely, to input the RFC number [7] along with the CIECF password, or to provide the FIEL certificate. If an attacker can somehow impersonate a legitime citizen/enterprise, she can complete the tax declaration on behalf of the unfortunate victim. Then, in the case that that tax declaration presents a positive balance, the attacker can request that the corresponding refund goes to a bank account of her choice. This brings a direct economic damage to the genuine taxpayer. Unfortunately, this is not at all a hypothetical situation. Recently the SAT along with the PRODECON[8] reported successful fraud attacks as the ones just described, for declarations that were presented during the 2015 fiscal year [2].

Scenario 3: Commercial Contracts. Any commercial contract can be signed using the FIEL certificate and its associated private key. So, an attacker who has compromised a FIEL certificate, can also establish any kind of legal bindings on behalf of the victim. For instance, it is possible to generate a fake contract in which the victim hires at whoever person the attacker selects. Furthermore, if the hash function used in the FIEL certificate is broken, then two different contracts may generate the same signature, which will verify successfully when using the FIEL certificate.

The above scenarios show the serious impacts that any security flaw of the SAT platform can cause to Mexican citizens and enterprises.

1.1 Our Contributions

In this paper, we carefully analyze the Mexican PKI system, and report some of its strong points, but also several of its weaknesses that could be attacked by malicious adversaries. Specifically, we have carried out the following studies:

1. **Harvesting Certificates:** We were able to identify a SAT FTP server in which it was possible to download SAT certificates by being authenticated as an anonymous user. As a result, we managed to collect $9,918,118$ RSA-1024 certificates and $132,535$ RSA-2048 certificates from the SAT infrastructure. Even though the information in a digital certificate is public and anyone can download certificates from the SAT website, some restrictions have

[6] From its name in Spanish: *Clave de identificación electrónica confidencial fortalecida*.

[7] Registro Federal del Contribuyente (Federal Taxpayer Registration Number) see §3.4

[8] *Procuraduría de defensa al contribuyente* in Spanish

been imposed during the last year to avoid that they can be obtained massively.[9]

2. **Batch-GCD Computation:** From our certificate collection, we tried to identify weak RSA-1024 keys by using the Batch-GCD computation recently applied by Bernstein et. al. in [3]. Since within our collection we could not identify any weak key, we concluded that the SAT's pseudo-random number generator shows **no vulnerabilities** against this attack.
3. **Password System:** Aiming to guess CIECF passwords that Mexican citizens and enterprises may have selected, we extracted public data from the FIEL and CSD certificates included in our collection. We found out that a non negligible number of Mexican citizens are predisposed to choose a password with exactly 8 characters, which is generally directly derived from the public information included in their corresponding certificates. Using this premise, we managed to recover 4,969 CIECF passwords out of 133,723 sample trials, i.e., our attack yielded a 3.72% success rate.
4. **SHA-1 Collision:** Because of the work by Stevens et al. in [30], it is now completely obvious that the SAT certificates with the RSA-1024/SHA-1 signature suite are not suitable for security applications such as signing legally binding contracts.
5. **RSA-1024 Factorization:** We have analyzed how much time, money and computational resources an attacker needs to invest to recover an RSA-1024 private key within a year.

1.2 Response to Weaknesses Found

In October 2015, we shared with SAT a preliminary version of our security analysis of its PKI system. We mentioned the main vulnerabilities that we had found at that point on time, including our large collection of SAT certificates, along with a first version of our dictionary attack on the CIECF passwords. Since that meeting, SAT has enforced several measures to avoid or thwart the attacks described in this paper. These measures include the usage of CAPTCHAs for downloading SAT certificates and the announcement of a two-factor authentication which will be based on a password plus a verification code coming from a token. Also, SAT has limited the number of services that are accessible for users that authenticate themselves using the CIECF password. Tax declarations that have a positive balance higher than $10,000 Mexican pesos must be submitted using a digital certificate. Furthermore, SAT estimates that as of March 2017, about 60% of its collection of active FIEL certificates were equipped with the RSA-2048/SHA-256 signature suite [23].

The remainder of this paper is organized as follows. In Section 2 basic cryptographic concepts are presented. In Section 3 the Mexican PKI system is described. In Section 4 the security weakness found in the SAT PKI and password-based systems are discussed in detail. Finally, in Section 5, some concluding remarks are drawn.

2 Cryptographic Background

Hash Functions: A cryptographic hash function is used to construct a short and unique *fingerprint* of some message. If the message is altered, even slightly, then the corresponding fingerprint will change. Let H be a hash function and let x be some message, where x could be a binary string of any arbitrary length. Then the fingerprint, $d = H(x)$, is known as the *m*essage digest of x. A digest could be as short as 160 bits, but for modern applications it is recommended to have a bitlength of at least 256 bits [20]. The primary requirement for a hash function is to avoid collisions. This implies that it should be computationally intractable to find two inputs x, y having the same image, *i.e.*, $H(x) = H(y)$, with $x \neq y$. Due to the birthday paradox, the expected security level of a well-designed hash function is at most half of its digest bitlength. More formally, a hash function is defined as follows:

A Hash function H defined as the mapping, $H : \{0,1\}^* \mapsto \{0,1\}^k$, with k a fixed and small

[9] Recently SAT enforced the policy that if a user wants to download third-party certificates, she must first solve a CAPTCHA challenge.

number, is a computationally efficient function that maps fixed binary chains x of arbitrary length $\{0,1\}^*$ to bit sequences $H(x)$ of a fixed length k. $d = H(x)$, is called the hash value or digest of x.

Signature Schemes: Public key cryptography can be used to generate digital signatures schemes of a message stored in an electronic form. A signature scheme allows a signer who has established a public key pk, to sign a digital message using her associated private key sk in such a way that anyone who knows pk can verify it. The verification procedure must validate that the received document is genuine and that it was created by the owner of the pk. Moreover, the signee must not be able to deny having signed the received document. More formally, a digital signature scheme is defined as follows,

A signature scheme is a tuple of three polynomial-time algorithms $(Gen, Sign, Vrfy)$, satisfying the following properties:

1. The key-generation algorithm Gen takes as input a security parameter 1^n and outputs a pair of keys (pk, sk). These are called the public key and the private key, respectively.
2. The signing algorithm $Sign$, takes as input a private key sk and a message $x \in \{0,1\}^*$. It outputs a signature σ denoted as $\sigma \leftarrow Sign_{sk}(x)$.
3. The deterministic verification algorithm $Vrfy$, takes as input a public key pk, a message x, and a signature σ. It outputs a bit b, with $b = 1$ meaning valid and $b = 0$ meaning invalid. We write this as $b := Vrfy_{pk}(x, \sigma)$.

To generate efficient signature schemes, the Hash-and-Sign paradigm is typically adopted. Basically, this approach consists of applying a hash function on the message to be authenticated. The digest thus produced is then signed by the signee entity using her private key.

The strength of a digital signature scheme depends on both the security level of the asymmetric cryptographic algorithm chosen along with the security level of the hash function. Therefore a careful security balance between these two blocks is highly desirable.

Public Key Infrastructure Public key cryptography needs to rely on an additional infrastructure known as Public Key Infrastructure (PKI), so that it can be used in a robust and reliable way for commercial applications. PKI is a framework consisting of policies about how to generate public/private key pairs and how to issue, publish, and maintain certificates. A certificate is a digital document that binds a public key to an entity, and that it has been signed by its publisher who is commonly referred as a Certification Authority (CA).

The PKI X.509 [7] and the suite of Public-Key Cryptography Standards PKCS, comprise a collection of software, cryptographic technologies and services that allow the protection of the digital transactions that are generated and transfered in a distributed system. This way, PKI X.509 and PKCS standards integrate digital certificates, public key cryptography and Certification Authorities (CA) into a single security architecture.

Pasword-Based Authentication: User authentication is a vital element in modern computer security. According to Pinkas et. al. [21], the most commonly spread mechanism to authenticate a user is through a password-based authentication system. Thus, a recurrent question for this approach is how to assess the security strength provided by a user's password.

The classical unit of measurement used to answer this question is the physical concept of entropy, which was first formalized by Claude Shannon [25]. Entropy measures the amount of information that is unknown due to a random variable, and is quantified as:

$$H(X) = -\sum_{i=0}^{n} P[x_i] \log_2 P[x_i],$$

where X is a discrete random variable that can take values x_i from a finite set $\mathcal{X}$, and $P[x_i]$ corresponds to the associated probability of the value x_i. According to NIST [19], the following heuristic rules can be used to estimate the entropy of human generated passwords:

1. The entropy of the first character is taken to be 4 bits.

2. The entropy of the next 7 characters are 2 bits per character.
3. For the 9th through the 20th character the entropy is taken to be 1.5 bits per character.
4. For characters 21 and above the entropy is taken to be 1 bit per character.
5. A bonus of 6 bits of entropy is assigned for a composition rule that requires both upper case and non-alphabetic characters.
6. A bonus of up to 6 bits of entropy is added for an extensive dictionary check.

To prevent that users generate passwords that are weak, i.e., that they have a low number of bits of entropy, system administrators typically require some sort of policy for users' password selection. Such policy may require that the password exceeds a minimum length, the mandatory usage of upper and lowercase letters, numbers and symbols, etc.

Unfortunately, the enforcement of this kind of policy not necessarily improves the password's strength. If for instance, a policy asks users to include numbers as part of their passwords, and in response to this policy, a user just include them in a predictably way, the password's strength can actually get reduced. As an illustrative example of this situation, in [35] Weir et. al. show that many users have the tendency to just append the numbers at the end of their password selection as in 'pass123'. A hacker that is aware of this tendency will incorporate mechanisms to take advantage of this structure.

Password-based systems usually employ hash functions to generate a digest of it, which is stored for future password verifications. If the digest calculation does not use any source of randomization, e.g. salt, then the output is deterministic an offline attacks are possible. Thus, if an adversary has access to the file where the password digests are stored, then she can test offline as many password guesses as needed until the right digest is produced.

According to several studies [9, 14, 21], a brute force attack can be accelerated by taking advantage of the following observations:

1. Most users choose simplistic passwords such as "123456".
2. Some systems require small length passwords, e.g. passwords with less than 8 characters.
3. Passwords are selected from a small set of alpha-numeric characters.
4. Users compose their passwords based on public information that for them is easier to remember, such as their names, birthdays, birth cities, etc.
5. Users selects common or fashion passwords as monkey, superman, etc.
6. Users choice common words that can be found in a dictionary of their native language.

The above ideas allow the attacker to improve her chances of getting the correct password, since it is possible to generate a set of candidate passwords based on the known information of the user, i.e, instead of performing a blind brute force attack, the attacker applies a dictionary attack.

In the case of on-line attacks against a web system, an attacker relies on the obvious fact that a correct password guess should open a valid session in that system. There exist several countermeasures to thwart this approach, like a server delayed response and/or user's account locking after a given threshold of failed attempts have been produced [21].

3 The Mexican PKI System

SAT is an independent authority that has the responsibility of applying the Mexican Fiscal and Custom laws, having as a main goal that Mexican citizens and companies operating in Mexico pay their respective taxes. Consequently, SAT is also responsible of providing software and administrative tools that allow an efficient tax declaration process [26].

In the remaining of this Section we briefly describe the PKI effort carried out by other countries around the world. Then, the mechanisms that SAT provides for obtaining a Mexican digital

certificate along with some of the services that a taxpayer can perform using her certificate will be summarized.

3.1 PKI Systems by Other Countries

In Canada, U.S. and most of Europe, tax authorities tend to rely on bank records, as opposed to invoices, as a legal proof of commercial transactions.

On the contrary, in the Latin America region, electronic invoices are the main instrument to record sales and purchases among companies and individuals. Around 2005, Chile pioneered the issuance of electronic invoices, although at that time, they were optional and mainly used in businesses transactions. Shortly after, Argentina, Brazil, Costa Rica, and Mexico built on the Chilean model, by making e-invoices compulsory, at first, only for large enterprises, and eventually for most if not all of their firms [33, 8].

In 2017, there is a close race between Mexico and Brazil to establish which one of these two countries produces the highest volume of invoicing, where both countries are in the range of billions of e-invoices issued per year. At the moment, it appears that the Mexican e-invoice system is winning the race, which makes it the largest in its kind worldwide. Moreover, from all the Latin American countries, Mexico's electronic invoice system is the most ambitious, as it is the only country that has imposed a compulsory e-invoicing for both, its citizens and its firms [31].

3.2 How a Mexican Citizen can get a SAT Digital Certificate

As it was described above, SAT issues two kind of certificates, the FIEL and the CSD certificates. If a Mexican citizen is interested in obtaining a FIEL certificate, she needs to complete the following procedure [6, 13]:

1. The citizen asks for an appointment through the SAT web site.[10].
2. The citizen should attend this appointment carrying with her a legal ID card, a certificated record of unique population key (CURP) or birth certificate, supporting home address documentation and a USB storage unit.
3. During the visit the taxpayer creates:
   - A password with at least 8 characters, which should contain upper case and lower case characters and numbers.
   - Biometric data: the citizen must print her fingerprints and iris scan.
   - Hand-written signature.
4. Next the SAT clerk asks the citizen for her pair of public/private key that she should have generated and stored previously in her USB unit. However, if the citizen for some reason has missed this action, the SAT clerk can create the public/private key pair and then the corresponding certificate on the flight. The certificate and private key so created are then stored in the USB storage unit provided by the citizen.

Unfortunately, the taxpayer is occasionally misguided by SAT clerks to choose a password with exactly 8 characters, since in the current procedure, she receives a small piece of paper with exactly 8 spots where the user is supposed to write her password to remember it.

Likewise, if a Mexican citizen wants to get a CSD certificate to issue fiscal CFDI invoices, she needs to perform the following procedure [27]:

1. The taxpayer get the SAT's application named *Certifica*.
2. The taxpayer generates a *CSD request*, which is a file with a .sdg extension. For this step it is mandatory to have a valid FIEL certificate. Then, a second file is generated with an extension .key that corresponds to the private key.
3. Next, the taxpayer sends the CSD request using her FIEL or CIEC password, and thereafter she recovers the corresponding public certificate.

[10] Appointments can be solicited at: https://citas.sat.gob.mx/citsat/home.aspx

3.3 CIECF, FIEL and CSD Transactions

Once that a Mexican taxpayer follows the procedures discussed in the previous section, and obtains a CIECF password, a FIEL or a CSD certificate, she is entitled to perform multiple fiscal transactions through the SAT web system, including:

1. CIECF updating: It permits to update the password.
2. Private taxpayer website access: By only being authenticated with the CIECF password, it is possible to download CFDI invoices issued or received by the taxpayer.
3. CFDI issuing: If the taxpayer is under a special legal regulation named *Régimen de Incorporación Fiscal* in Spanish, then it is possible to issue invoices by using the *Mis cuentas* SAT application. Another option for this action is to directly use the CSD certificate.
4. Income tax declaration: Until the 2014 national tax declaration, it was possible to present an income tax declaration using only the CIECF password, provided that the refund balance was less than $\$40,000$ Mexican pesos. However, for the 2015 national tax declaration this option was only valid for less than $\$10,000$ Mexican pesos refunds.
5. Signing legally-binding commercial contracts.
6. Performing transactions in the domain of other governmental Mexican institutions such as the Mexican Institute of Social Security.

Selected e-government transactions that are allowed for SAT certificate holders are summarized in Table 1. For a complete catalog of transactions, we refer the reader to the SAT web site [27].

Table 1. Allowed e-government transactions that can be performed according to the authentication method and/or certificate presented by the citizen.

Transactions	CIECF	FIEL	CSD
CIECF Updating	√	√	X
CFDI Download	√	√	X
CFDI Issue	√	X	√
Income Tax Declaration	X	√	X
Renovate/Revocate FIEL	X	√	X
Asking for CSD	X	√	X
Signing contracts	X	√	X

3.4 SAT RFC Code Number

User authentication is a crucial element in modern computer systems. In many systems, users are authenticated by querying them about something they should know, most commonly, username/password knowledge. Often, institutions and companies have specific policies that determine the username of their employees.

For example, all Universities have a specific policy for generating usernames, which usually consists of a combination of the user's surname, year of enrollment, etc.

In the case of the SAT system, the user name is the Federal Taxpayer Registration Number (RFC) which is issued to each Mexican citizen by this institution. The RFC number is conformed by 10 alphanumeric characters plus 3 more characters called *homoclave*. Positions 1 and 2 of the RFC come from the first consonant and the first vowel of the first taxpayer's surname. The next two positions come from the first letter of the second taxpayer surname and the first letter of the first taxpayer name. Positions 5 to 10 are the year, month and day of birthday in a two-digit format.

Finally the *homoclave* is assigned by SAT using an algorithm that is publicly available at [22]. This implies that the RFC number can be calculated for any regular taxpayer knowing the personal data mentioned above. The RFC is considered to be a public information and it is included in plaintext as part of a certificate. As an illustrative example, suppose that a person named Juan Pérez Pérez was born in the State of Michoacán in September, 5 1979.

Then, following the aforementioned rules, and by applying the algorithm for the *homoclave*, the corresponding RFC of this person is PEPJ7909055R3.

Hence, from the information included in the SAT certificate of a Mexican citizen, one can trivially determine her RFC. At the same time, having any valid RFC, it is possible to get the corresponding certificate using the web SAT system.

4 Security Weaknesses of the Mexican PKI System

In this Section we briefly outline the results obtained by applying several well-known attacks to the SAT PKI system.

4.1 Batch-GCD Computation Attack

As a part of this work, we collected a total of $9,918,118$ of RSA-1024 certificates and $132,535$ RSA-2048 certificates that we manage to get using a SAT FTP server. Once again, we stress that this collection was recovered by legal means, using SAT's public FTP servers and logging in as legal anonymous users.

By means of the software library available from [3], we launched the Batch-GCD computation attack on our collection of SAT certificates. Roughly speaking, this attack looks for shared primes among the target keys. We extracted the RSA modules used in each of the RSA-1024 certificates and we looked for prime collisions. However, it was not possible to find out any two certificates sharing the same module. Next, we looked for shared primes, but we could not find any weak pair of certificates. Since this attack was not successful, we conclude that the SAT software was equipped with a sound implementation of a pseudo-random number generator for issuing RSA public/private keys.

4.2 Spamming-Like Attacks

As it was mentioned in Section 1, a FIEL certificate binds citizen's personal information (such as name, picture, iris scan, fingerprint, e-mail, calligraphic signature, etc.) to her public key.

As SAT states in [26], citizen personal data is strictly protected by law. However, some of the citizen personal data can be classified as public information (for example, full name, RFC, CURP, etc). Notwithstanding that citizen data can be linked to a digital certificate which by definition is itself a public digital document, this information must be handled with care, for the knowledge of a massive number of certificates can be exploited by mounting spam-like attacks against the owners, such as spamming through email address harvest ing, phishing, etc.

In the case that an attacker collects a significant amount of SAT certificates, it becomes trivial to parse their contents to create a searchable database with the collected information. This way, the attacker has access to the complete name or business name, birthday, age, city, CURP, RFC, and quite often, a valid e-mail of her victims. It becomes then possible to send spam to solicit passwords or any other private information to valid e-mail addresses belonging to real Mexican citizens. Moreover, the attacker can check for the existence of any certificate of his choice by performing a standard query over name or any other content fields. We illustrate this point by showing in Figure 1 the digital certificate of a prominent Mexican politician.

4.3 CIECF Password Dictionary Attack

A natural question would be to know how much strength brings the CIECF password.

The CIECF password has a length of at least eight characters, and it must contain numbers between (0..9) and lower case alphabetical letters between (a..z) and upper case alphabetical letters bettween (A..Z). If we evaluate the strength of a CIECF password according to the NIST rules (cf. §2), we obtain 24 bits of entropy. In other words, using a brute-force approach an attacker should try an estimated average of 2^{24} different guesses in order to obtain the target password.

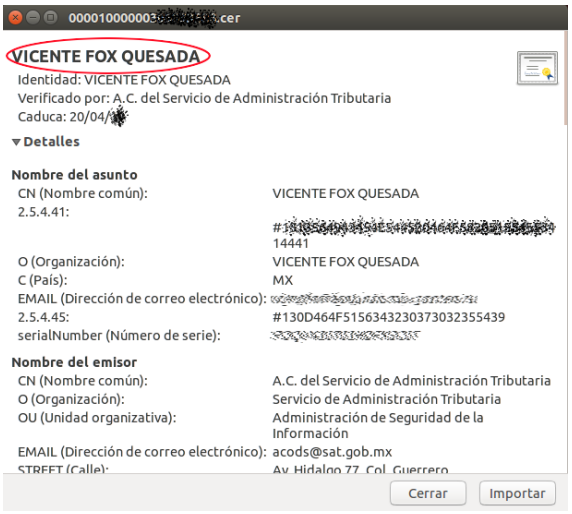

Fig. 1. Digital Certificate of Vicente Fox Quesada (e-mail address and other private information has been blurred on purpose)

This gives us an idea of how large is the difference between the CIECF password security strength and the standard international recommendation of providing a security level of 128 bits for any sensitive exchange of data through the Internet and other digital media.

Currently, the SAT web system allows up to twenty failed attempts before blocking the user's account as a security countermeasure against password dictionary attacks [23]. However, if the attacker happens to have a large collection of certificates, then it is possible to mount an extremely simple dictionary attack, based on the following assumption: a non-negligible number of Mexican citizens choose as their password the first 8 characters of their RFC.

Following this basic strategy of assuming that the first 8 characters of the citizen RFC is her password, we were able to recover $4,969$ passwords from an universe of $133,723$ real citizens' passwords tested. This result shows how easy is for an attacker to recover a citizen CIECF password. The attack was even more powerful because the SAT web site capitalize all the passwords, which reduces even further the CIECF password entropy.

The associated risks of this attack range all the way from just changing the CIECF password to present a fake tax declaration.[11]

From a technical point of view, our attack shows that the existential password recovery of the SAT system is considerably lower than universal password recovery [11, §5].

It is important to point out that organizations should also be concerned about protecting the confidentiality of user identifiers (e.g. usernames) as urged by NIST in [16]. Concealing usernames makes harder for attackers to perform targeted attacks. But perhaps, this policy might be challenging to implement since typically identifiers are based on public information.

4.4 SHA-1 Collisions

As of May 2015 both, FIEL and CSD certificates were using Security Hash Algorithm (SHA-1) as their hash function.

SHA-1 is a family of cryptographic hash functions proposed in 1995 by NIST, as a FIPS standard [15]. Since then, SHA-1 was endorsed by many industry security standards. However, an important weakness of SHA-1 has been publicly known since 2005, when Wang et. al. presented in [34] a collision search attack on the full 80-step SHA-1 procedure. Wang et al. attack has a complexity of less than 2^{69} operations, while its theoretical security bound should be of about 2^{76} steps.

Recently, an ever increasing number of initiatives have started a process for deprecating SHA-1 from the public web altogether. Significantly, the Internet navigator Chrome will show visible errors for SHA-1 certificates starting with the current *Chrome 56* version. Furthermore, NIST deprecated SHA-1 for governmental usage back in 2011 [17]. Specifically, NIST stated that there are some applications where the usage of SHA-1

[11] The SAT PKI system described here is not the unique governmental site that interacts with their users using the RFC number. For example, the Housing Fund of the Institute of Security and Social Services of State Workers (FOVISSSTE) site [5], allows to ask for very sensitive information (such as salary, number of years in service, etc.) about the current status of an employee through an application that only requests the citizen's RFC number.

should be mandatorily avoided, *quite especially*, the process of signing a PKI certificate.

In addition, NIST considers that after 2013 the risk associated with SHA-1 is unacceptable for all information security applications. Moreover, the Certification Authority Browser Forum [12], published on 2011 its baseline requirements for SSL, where it was stated that SHA-1 should not be used by any web browser neither by certificate authorities.

The last nail in the coffin was hammered in by Stevens et al. in [30], where the authors presented the first collision for full SHA-1. Perhaps even more importantly, the authors of [30] carefully selected the prefix of the colliding messages so that they allow to forge two distinct and arbitrary documents with the same SHA-1 hash. Hence, using the tool that the team of [30] will release by the end of May, anyone can submit two different PDF documents, which after some modifications in the submitted documents, will produce the same SHA-1 digest for both of them. Hence, it is just fair to say that the SHA-1 hash function is completely broken.

From the above discussion, it becomes clear that the usage of SHA-1 is another important weak point of the SAT certificates. Indeed, since collision attacks over SAT certificates using SHA-1 are now a certain threat, an attacker can produce the same signature for two different documents that will verify correctly when using any old FIEL certificate that has the RSA-1204/SHA-1 signature suite.

Currently, NIST recommends the usage of the SHA-2 and SHA-3 standards for professional cryptographic applications. These two standards were approved by NIST in 2001 and 2015, respectively.

4.5 How Costly is to Recover an RSA-1024 Private Key within One Year?

As it was mentioned above, the FIEL and CSD certificates that were issued by SAT until May 2015, utilize RSA-1024 as their signature scheme. RSA was proposed in 1977 by Adleman, Rivest and Shamir, and it is still one of the most popular cryptosystems still in use. The security of this algorithm is based on the computational difficulty of the integer factorization problem. A k-bit RSA modulus N is defined as, $N = p \cdot q$, where p, q are prime numbers that need to meet certain security requirements, and are chosen to have a bitlength of about $k/2$ bits. Given a modulus N, The RSA version of the integer factorization problem consists of finding the primes p and q. Finding such RSA prime factorization for a sufficiently large modulus N is considered a hard computational problem.

The general number field sieve GNFS [12] is the current state-of-the-art algorithm for factoring large numbers. Using the GNFS method, Kleinjung et. al. [10] established in December 2009, a new factorization record of an RSA modulus of a size of 768 bits. According to Kleinjung et. al. in [10], the total effort of this factorization in a single core 2.2 GHz AMD Opteron with 2GB RAM processor would have taken a little less than $1,700$ core-years. However, the authors in [10] mentioned that an optimized version of their attack would have taken around $1,100$ core-years to attack an arbitrary 768-bit RSA modulus. Hence, there is a high chance that this problem can be solved within one year using a rather small-size cluster composed by around 1,100 CPU cores.

In [10], the authors estimate that breaking a $1,024$ bits RSA modulus is possibly thousand of times harder than factorizing a 768 bits number. Further, Adrian et. al. in [1, Table 2] reported cost estimates of how many core-years are required for factorizing an RSA-1024 number based on an extrapolation of the costs reported for RSA-768 in [10], concluding that the effort would require $1,120,000$ core-years.[13] Thus, if we want to perform the whole computation within one year, one needs to use a super-computer having at least $1,120,000$ cores. According to the Top500 [32] super-computer list, there are several clusters in the world that have this number of cores. For example, the *Sequoia-BlueGene* super computer has the required number of cores, with an estimated building cost and one-year power costo of around $208 million dollars.

[12] A voluntary consortium that promulgates industry security guidelines for SSL certificates

[13] This has to be compared with the effort of finding a collision for full SHA-1, which took about $6,500$ CPU time plus 100 GPU years [30]. This computation took two calendar years of research and development.

Note that even when RSA-1024 can only be attacked by well funded adversaries, it has been deprecated by NIST [18] since 2013.

5 Conclusions

In this paper we described several security vulnerabilities on the SAT PKI system. They can be summarized as follows:

— SHA-1 that is still used in about seven million SAT certificates, has been completely broken in February 2017 by Stevens et al. in [30]. Among other consequences, this weakness disqualifies the usage of those SAT certificates for signing legally binding contracts.

— RSA-1024 that is still used in in about seven million SAT certificates, can be subject of attacks by well-funded adversaries.

— Despite the fact that SAT has announced and partially implemented a migration process to RSA-2048 and SHA-256, that transition phase will likely take two more years to be completed.

— RSA-2048 provides 2^{112} bits of security whereas SHA-256 provides 2^{128} bits of security. Ideally SAT should be migrating to RSA-3072 or elliptic curve cryptography, in order to avoid its current cryptographic disbalance.

— The minimum taxpayer password length is of only eight characters, and the manner in which many Mexican citizens choose it is quite predictable.

— The SAT system allows excessively many important processes to be done by citizens using the CIECF password authentication mechanism.

— The SAT web system allows an arbitrary number of trials for a taxpayer's password.

5.1 SAT Reaction

In October 2015, we communicated a preliminary version of these results to SAT. Since then, SAT has enforced several measures to reduce the vulnerabilities found in our study. These measures include the usage of CAPTCHAs for downloading SAT certificates and the announcement of a two-factor authentication procedure which will combine a password-based system along with the usage of a soft token. Also, SAT has limited the number of services that are accessible for users that authenticate themselves using the CIECF password. Tax declarations that have a positive balance higher than $10,000 Mexican pesos must be submitted using a digital certificate. Furthermore, SAT estimates that as of March 2017, about 60% of its collection of active FIEL certificates were equipped with the RSA-2048/SHA-256 signature suite [23].

Acknowledgements

The authors Lil M. Rodríguez and Francisco Rodríguez-Henríquez would like to thank CONACyT (project number 180421) for partially funding this research.

References

1. **Adrian, D., Bhargavan, K., Durumeric, Z., Gaudry, P., Green, M., Halderman, J. A., Heninger, N., Springall, D., Thomé, E., Valenta, L., VanderSloot, B., Wustrow, E., Béguelin, S. Z., & Zimmermann, P. (2015).** Imperfect forward secrecy: How diffie-hellman fails in practice. **Ray, I., Li, N., & Kruegel, C.**, editors, *Proceedings of the 22nd Association for Computing Machinery (ACM) SIGSAC*, ACM, pp. 5–17.
2. **Aristegui Noticias (2015).** Alerta a contribuyentes: fraudes mediante suplantación de identidad. http://tinyurl.com/nwk5ca8.
3. **Bernstein, D. J., Chang, Y., Cheng, C., Chou, L., Heninger, N., Lange, T., & van Someren, N. (2013).** Factoring RSA keys from certified smart cards: Coppersmith in the wild. **Sako, K. & Sarkar, P.**, editors, *Advances in Cryptology - ASIACRYPT 2013 Part II*, volume 8270 of *Lecture Notes in Computer Science*, Springer, pp. 341–360.

4. **Edicom**, **.** Latam e-invoicing, electronic billing platform. https://tinyurl.com/le4k2jv. Visited on March, 10 2017.
5. **Fondo de la vivienda del instituto de seguridad y servicios sociales de trabajadores del estado (2018).** Simuladores. http://tinyurl.com/gwbtafz.
6. **González-García, V., Rodríguez-Henríquez, F., & Cruz-Cortés, N. (2008).** On the security of mexican digital fiscal documents. *Computación y Sistemas*, Vol. 12, No. 1, pp. 25–39.
7. **Internet Engineering Task Force (2013).** Updates to the Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile. http://tinyurl.com/h97fn5n.
8. **Jiménez, C. E., Criado, J. I., & Gascó, M. (2011).** Technological e-government interoperability. an analysis of iberoamerican countries. *IEEE Latin America Transactions*, Vol. 9, No. 7, pp. 1112–1117.
9. **Klein, D. V. (1990).** "Foiling the Cracker": A survey of and improvements to password security. *Proceedings 2nd Usenix Security Workshop*, pp. 5–14.
10. **Kleinjung, T., Aoki, K., Franke, J., Lenstra, A. K., Thomé, E., Bos, J. W., Gaudry, P., Kruppa, A., Montgomery, P. L., Osvik, D. A., te Riele, H. J. J., Timofeev, A., & Zimmermann, P. (2010).** Factorization of a 768-bit RSA modulus. **Rabin, T.**, editor, *Advances in Cryptology - CRYPTO 2010*, volume 6223 of *Lecture Notes in Computer Science*, Springer, pp. 333–350.
11. **Koblitz, N. & Menezes, A. (2006).** Another look at "provable security". II. **Barua, R. & Lange, T.**, editors, *Progress in Cryptology - INDOCRYPT 2006*, volume 4329 of *Lecture Notes in Computer Science*, Springer, pp. 148–175.
12. **Lenstra, A. K. & Hendrik W. Lenstra, J.**, editors **(1993).** *The development of the number field sieve*, volume 1554. Springer-Verlag.
13. **León-Chávez, M. & Rodríguez-Henríquez, F. (2015).** Security vulnerabilities of the mexican digital invoices by internet. *International Conference on COmputing Systems and Telematics (ICCSAT 2015)*, IEEE, pp. 1–5.
14. **Morris, R. & Thompson, K. (1979).** Password security: A case history. *Communications of the ACM*, Vol. 22, pp. 594–597.
15. **NIST (1995).** FIPS PUB 180-1 Secure Hash Standard. http://tinyurl.com/jyh2l52.
16. **NIST (2009).** SP 800-118 Guide to Enterprise Password. http://tinyurl.com/ccbkby.
17. **NIST (2011).** 800-131A Transitions: Recommendation for Transitioning the Use of Cryptographic Algorithms and Key Length. http://tinyurl.com/jxyppsy.
18. **NIST (2011).** Transitions: Recommendation for transitioning the use of cryptographic algorithms and key lengths.
19. **NIST (2013).** SP 800-63-2 Electronic Authentication Guideline. http://tinyurl.com/qf5ruxf.
20. **NIST National Institute of Standards and Technology (2015).** NIST's Policy on Hash Functions. http://tinyurl.com/ydjec7w.
21. **Pinkas, B. & Sander, T. (2002).** Securing passwords against dictionary attacks. *Proceedings of the 9th Association for Computing Machinery (ACM) Conference on Computer and Communications Security*, CCS '02, ACM, New York, NY, USA, pp. 161–170.
22. **SAT (2006).** Algoritmo para generar el RFC con homoclave para personas fisicas y morales 0610100135506. http://tinyurl.com/zo6ql77.
23. **SAT (2017).** Recent security countermeasures put in place by SAT. Personal communication.
24. **SEGOB (2015).** Diario oficial de la federación. http://tinyurl.com/hyegq8h.
25. **Shannon, C. E. (1948).** A mathematical theory of communication. *Bell system technical journal*, Vol. 27, pp. 379–423.
26. **Sistema de Administración Tributaria (2015).** ¿Qué es el SAT? http://tinyurl.com/zq5x7nk.
27. **Sistema de Administración Tributaria (2015).** Sitio del SAT. http://www.sat.gob.mx/Paginas/Inicio.aspx.
28. **Sistema de Administración Tributaria (2016).** Número de comprobantes. http://tinyurl.com/hbmyzjb.
29. **Sistema de Administración Tributaria (2016).** Padrón por tipo de contribuyente. http://tinyurl.com/gp9w2on.
30. **Stevens, M., Bursztein, E., Karpman, P., Albertini, A., & Markov, Y. (2017).** The first collision for full sha-1. Cryptology ePrint Archive, Report 2017/190. Http://eprint.iacr.org/2017/190.
31. **The Economist (2014).** Electronic arm-twisting. http://tinyurl.com/k6733gb.
32. **TOP500 Supercomputing Sites. (2016).** TOP 10 Sites for November 2015. available at: http://tinyurl.com/q6cncrf. Accessed December 23, 2016.

33. Villalón-Fonseca, R., Mora-Castro, A., Bartels-González, R., Carballo-Chavarría, M., & Raventós, G. M. (2016). Promoting quality e-government solutions by applying a comprehensive information assurance model: Use cases for digital signature. *ICT for Promoting Human Development and Protecting the Environment - 6th IFIP World Information Technology Forum, WITFOR 2016*, volume 481 of *IFIP Advances in Information and Communication Technology*, Springer, pp. 223–234.

34. Wang, X., Yin, Y. L., & Yu, H. (2005). Finding collisions in the full SHA-1. **Shoup, V.**, editor, *Advances in Cryptology - CRYPTO 2005*, volume 3621 of *Lecture Notes in Computer Science*, Springer, pp. 17–36.

35. Weir, M., Aggarwal, S., Collins, M., & Stern, H. (2010). Testing metrics for password creation policies by attacking large sets of revealed passwords. *Proceedings of the 17th Association for Computing Machinery (ACM) Conference on Computer and Communications Security*, CCS '10, ACM, pp. 162–175.

Article received on 04/08/2018; accepted on 14/11/2018.
Corresponding author is Luis Rivera-Zamarripa.