Decoding Demographic Information in VeriSign Certificates

Scott G. Renfro
www.securify.com*
srenfro@securify.com

January 17, 2000

Abstract

VeriSign Class 1 certificates include, by default, encrypted personal demographic information (country, date of birth, zip code, and gender) when provided by the subscriber. This information is intended to be read by web sites that have signed a privacy pledge and paid a licensing fee, although VeriSign has never claimed to protect this information. Since the data is poorly encrypted and all Class 1 certificates are published on a public directory server, it is trivial for anyone to recover these personal details. This paper presents details on decoding the One-Step Registration field and includes a short demographic summary of examined certificates. Inclusion of this extension in Class 1 certificates is representative of a more widespread temptation to use certificates as a carrier for all subject attributes. Failure to carefully consider the appropriate certificate profile may allow sensitive information to leak outside the intended scope.

1 Introduction

Since 1997, VeriSign has been including subscriber–entered demographic information including country of residence, date of birth, zip code, and gender, when this information is submitted at registration time. While inclusion of this information is the default, subscribers have the option of opting out during registration. Most subscribers, however, include their personal data. This feature is known as the One-Step Registration field.

When provided, this information is stored encrypted in the subscriber's certificate. Software to read the demographic information is available from VeriSign for a licensing fee. Although VeriSign encrypts the data in the certificate, they have never publicly claimed to protect the data and their privacy statement explicitly states that users should have no expectation of privacy regarding data included in their certificate.

Regardless of the disclaimers, it is unlikely that most subscribers recognize that their personal demographic information is so significantly exposed. Since VeriSign publishes all Class 1 certificates in their public directory server and the encryption used to protect the information is weak, Class 1 subscriber personal demographic information is effectively published for all to read.

This paper provides some background information on the One-Step Registration extension, describes how to decrypt and decode the extension, and provides a demographic summary of certificates examined while researching this subject.

2 Background

In 1997, VeriSign announced an optional One-Step Registration feature that adds an extension containing country, zip code, date of birth, and gender to Class 1 certificates requested by subscribers who

**Kroll-O'Gara ISG*,3600 W. Bayshore Rd Suite 200, Palo Alto, CA 94303 (650) 213–4600

do not opt-out[1]. Although subscriber's are clearly advised that the information is optional, inclusion is the default and most subscribers signing up for Class 1 VeriSign certificates provide the information (see Section 2). As part of the announcement, VeriSign described the availability of an implementation kit available for an annual licensing fee. This kit includes a registration license key to read the One-Step Registration field[1].

To subscribers, the One-Step Registration field has been marketed as a way to ease web site registration, appealing to their desire to gain personalized content and eliminate repetitive data entry. To web sites, however, the feature has always been marketed as a way to anonymously track the demographic make-up of people visiting their sites[3]. This is data is only anonymous in that there is no validation against other data sources to prevent subscribers from entering false information. In reality, it appears many users enter their actual demographics, binding accurate information to real names.

Subscribers might well believe — and incorrectly so — that their information is only available to web sites licensed by VeriSign. Because the information is encrypted, it is not visible when inspecting the certificate in a web browser. Additionally, VeriSign implies limited distribution of demographic information on the registration page by describing the feature as being used "at certain web sites" and "presented to participating web sites"[4]. Additionally, as part of the initial announcement, VeriSign asserted that participating sites must adhere to a consumer privacy pledge[1] and sites which violate the provisions of the pledge will have their reader license key revoked by VeriSign[2].

Other than the implication that the certificates are only presented to certain web sites, VeriSign has apparently never claimed to encrypt or otherwise protect the information from other than licensed web sites. In fact, VeriSign's privacy statement points out that "VeriSign's CPS requires VeriSign to publish all subscriber certificates within the Public Certification Services. Consequently, a subscriber should have no expectation of privacy regarding the content of his or her Digital ID."[5]

The addition of this information in VeriSign certificates generated some discussion within the PKI community and further fueled the debate on appropriate uses of X.509v3 certificates — especially the privacy issues related to binding a public key to the X.500 concept of a unique identity[6]. To date, however, there has been no publicly available information providing further details on the use of this extension in VeriSign Class 1 certificates.

3 Decoding the demographic extension

The encrypted demographic information included in VeriSign Class 1 certificates is encoded into a private X.509v3 extension. This section will examine the representation of this data within the certificate, how to decrypt it, and how to interpret the resulting plaintext.

3.1 Ciphertext representation

As mentioned above, the encrypted demographic information in VeriSign Class 1 certificates is conveyed in an X.509v3 extension. The ASN.1 description of this extension is shown in Figure 1.

```
id-verisign-demographic OBJECT IDENTIFIER ::=
                         { 2 16 840 1 113733 1 6 3}

verisignDemographicExtension ::= {
    SYNTAX VerisignDemographicExtension
    IDENTIFIED BY id-verisign-demographic }

VerisignDemographicExtension ::= IA5String
```

Figure 1: ASN.1 syntax for the demographic extension

According to the Distinguished Encoding Rules used to encode X.509v3 certificates, `id-verisign-demographic` is encoded as the byte stream shown in Figure 2. Certificates containing this byte stream, therefore, almost certainly contain the extension[1].

```
06 0a 60 86 48 01 86 f8 45 01 06 03
```

Figure 2: DER encoded representation of id-verisign-demographic

The IA5String contains 116 hex characters, which is converted into 58 bytes of ciphertext. Within the sample of certificates reviewed, only 19 bytes of ciphertext varied; the remaining 39 bytes were fixed.

3.2 Decrypting

Using a chosen plaintext attack, the portion of the keystream corresponding to the 19 variable bytes can be easily recovered. Recovering the keystream is trivial since the data appears to be encrypted with an unknown stream cipher using a fixed key and no message key. The result is functionally equivalent to an every-time-pad (i.e. the same keystream is used to encrypt the One-Step Registration field of every VeriSign Class 1 certificate examined, unlike a one-time pad where the keystream is randomly chosen and never reused). It is possible that some portion of the 39 fixed bytes — perhaps the first 32 bytes (i.e. 256 bits) — correspond to either an encrypted message key used in conjunction with RC4 to encrypt the remaining bytes or a key identifier used for a similar purpose. Since those bytes are fixed, even if they correspond to an encrypted message key or key identifier the fact that they never change means the keystream never changes either.

The keystream k in Figure 3 gives the recovered keystream with unknown bytes represented as `00`. The plaintext of the demographic structure can then be recovered as $p = c \oplus k$ where $\oplus$ represents the XOR operation.

```
k = 00000000 00000000 00000000 00000000
    00000000 00000000 00000000 00000000
    0086a100 0000fb0b f2c8b226 9d5bc1e7
    0079ae93 8b72cd00 a700
```

Figure 3: Known bytes of the keystream

3.3 Decoding

After decryption, the One-Step Registration field can be simply decoded to yield the demographic information submitted by the subscriber at the time of registration. There are four pieces of information contained within the structure: country, zip/postal code, date of birth, and gender. The format of the plaintext can be expressed by the ANSI C structure shown in Figure 4. The following sections look at this representation of each item in more detail.

3.3.1 Country

Country is represented as a two byte country code contained in the 34th and 35th byte of the One-Step Registration field. The country code corresponds to countries according to Table 1.

[1] The exception being an incredibly small probability of the byte stream appearing in a public key or signature

```
struct demographics {
        char unknown1[33];
        char country[2];
        char unknown2[3];
        char zipcode[10];
        char unknown3;
        char dob[6];
        char unknown4;
        char gender;
        char unknown5;
};
```

Figure 4: ANSI C structure showing layout of field

Code	Country	Code	Country
AU	Australia	JP	Japan
AT	Austria	MX	Mexico
BE	Belgium	NL	Netherlands
BR	Brazil	NO	Norway
CA	Canada	ZA	South Africa
CN	China	ES	Spain
DK	Denmark	SE	Sweden
FI	Finland	CH	Switzerland
FR	France	TW	Taiwan
DE	Germany	GB	United Kingdom
IN	India	US	United States
IL	Israel	UU	Other Countries
IT	Italy		

Table 1: Country codes used in the One-Step Registration field

3.3.2 Zip code

The zip code is represented as ten characters as entered by the subscriber. The contents are left–padded with spaces (i.e. when the user enters less than ten characters during registration, spaces are inserted from the left until the string totals ten characters). The ten characters are located in bytes 39 to 48 of the One-Step Registration field.

3.3.3 Date of birth

The date of birth, as entered by the subscriber, is represented as six characters from bytes 50 to 55 of the plaintext. The characters are formatted `MMDDYY` so that February 01, 1970 would be entered as `020170`.

3.3.4 Gender

Gender is simply encoded as one character — either `M` for Male or `F` for Female. This character is byte 57 in the plaintext.

4 Survey of demographics

Using the preceding information, a sample set of certificates was retrieved from VeriSign's public directory server at `ldap://directory.verisign.com`[2]. These certificates were analyzed to verify the decryption and decoding techniques and collect some broad statistics on the demographics of the certificates examined.

Of the certificates examined, 77% included the One-Step Registration extension. Of those certificates with the extension, a summary demographic is listed in the table below. These statistics are based on a non-random sample of 16,285 certificates after removing those with either missing data, age less than 10 years, or age greater than 80 years. Certificates with ages greater than 80 and less than ten nearly always represented data entry errors (e.g. listing the current year in the subscriber's date of birth).

Category	**Male**		**Female**		**Total**	
Age 10 – 18	162	(1%)	23	(<1%)	185	(1%)
Age 18 – 24	785	(5%)	156	(1%)	941	(6%)
Age 25 – 34	2,992	(18%)	577	(4%)	3,569	(22%)
Age 35 – 45	4,065	(25%)	746	(5%)	4,811	(30%)
Age 45 – 55	3,796	(23%)	648	(4%)	4,444	(27%)
Age 55 – 65	1,569	(10%)	204	(1%)	1,773	(11%)
Age 65 – 80	495	(3%)	67	(<1%)	5,62	(3%)
Total	13,864	(85%)	2,421	(15%)	16,285	(100%)

Table 2: Demographic summary of examined certificates

5 Conclusion

The inclusion of the One-Step Registration extension in subscriber certificates is representative of a more widespread temptation to use X.509v3 certificates as a carrier for many kinds of subject attributes not needed for the actual purpose of the certificate. Although subscribers have the opportunity to opt–out of the feature, most do not. It is unlikely that many of these users realize that their personal data is not just available to a few participating web sites, but is published on the Internet where it is readable by anyone given a trivial amount of effort.

Organizations planning and deploying both open and closed public key infrastructures must carefully consider the appropriate certificate profile. Failure to do so may allow private data to leak outside the intended scope, compromising valuable information in the process.

[2] The certificate for a user with e-mail address `<email>` (e.g. alice@somewhere.org), can be simply retrieved in base-64 encoded form with the command line: `ldapsearch -h directory.verisign.com -b "" mail=<email> usercertificate;binary`

References

[1] "VeriSign Enhances Digital IDs to Enable Universal Website Login and One-Step Registration." April 14, 1997.
http://www.verisign.com/press/product/isv.html (January 14, 2000).

[2] "VeriSign and eTRUST Team Up to Assure Consumer Privacy on the Internet." April 14, 1997.
http://www.verisign.com/press/partner/etrust.html (January 14, 2000).

[3] "VeriSign Digital ID Usage." 1998.
http://www.verisign.com/about/id_imp.html (January 14, 2000).

[4] "Microsoft Class 1 Enrollment." 1999.
https://digitalid.verisign.com/client/class1MS.htm (January 14, 2000).

[5] "VeriSign, Inc.'s Overall Privacy Statement." February 26, 1998.
http://www.verisign.com/truste/ (January 14, 2000).

[6] "re: [E-CARM] Certificates and privacy and VeriSign." May 8, 1998.
http://www.mail-archive.com/cert-talk@structuredarts.com/mail3.html (January 14, 2000).