The Newton Channel

Ross Anderson[1], Serge Vaudenay[2], Bart Preneel[3] and Kaisa Nyberg[4]

[1] Computer Laboratory, Pembroke Street, Cambridge, CB2 3QG
[2] Ecole Normal Supérieure — DMI, 45 rue d'Ulm, 75230 Paris, France
[3] KU Leuven — ESAT-COSIC, Kardinaal Mercierlaan 94, B-3001 Heverlee, Belgium
[4] Finnish Defence Forces, PO Box 919, FIN-00101 Helsinki, Finland

Abstract. Simmons asked whether there exists a signature scheme with a broadband covert channel that does not require the sender to compromise the security of her signing key. We answer this question in the affirmative; the ElGamal signature scheme has such a channel. Thus, contrary to popular belief, the design of the DSA does not maximise the covert utility of its signatures, but minimises them. Our construction also shows that many discrete log based systems are insecure: they operate in more than one group at a time, and key material may leak through those groups in which discrete log is easy. However, the DSA is not vulnerable in this way.

1 Introduction

Many digital signature schemes have the property that the signer of a message can hide some information in the signature that can be recovered by a third party, and that the presence of this hidden information cannot even in principle be detected in any given instance of the signature. These channels were discovered by Simmons, who called them subliminal channels [7].

The problem originally arose in the context of nuclear arms limitation treaty verification. The USA and the USSR had decided to redplace certain sensors in each other's nuclear facilities in order to share certain agreed sensor information, and needed integrity controls to prevent information being manipulated in order to provide false evidence that a test did or did not take place [8]. In addition, both parties wanted to be sure that the integrity mechanisms could not be abused to transmit other, prohibited, information.

This was a special concern with systems used to monitor not just the occurrence of nuclear tests, but the numbers of fielded nuclear weapons. If a Russian sensor designed to relay merely the presence or absence of an American missile in a silo could covertly communicate the silo's location, then this information could have been used to facilitate a first strike. One of the early designs for equipment to verify treaty compliance had just such a weakness: the sensor's location could have been transmitted using a subliminal channel in an early authentication scheme based on discrete logarithms [7].

To see how such channels work, consider the ElGamal signature scheme [3]. Let p be a prime number such that finding discrete logs in F_p^* is hard, let g

be a generator of F_p^*, let $x \in \{1, ..., p-1\}$ be a user's secret signing key, let $y = g^x$ be her published signature verification key, let $k \in \{1, ..., p-1\}$ (with $\gcd(k, p-1) = 1$) be a message key and M the message to be signed. Then the ElGamal signature on M is (r, s) where

$$r = g^k \pmod p \tag{1}$$

$$s = (M + xr)/k \pmod{p-1} \tag{2}$$

The two previously known covert channels in this scheme are:

1. a broadband channel in which the signer shares her signing key with the message recipient, allowing k to be trivially recovered using equation (2). We can thus encode a covert message directly in k;
2. a narrowband channel in which she tries out many values of k until she manages to force a number of bits of r to encode the covert message c. Thus she might wish to encode a ten bit message in the low order bits of r, and try successive values of k until she got lucky. This would take about a thousand tries on average, and in general the covert bandwidth in bits per signature is about the binary logarithm of the number of computations that the signer is willing to perform.

Even the narrowband channel would have been sufficient for a sensor in a missile silo to encode a few bits of information, and over time this information could have revealed its physical location. Narrowband covert channels could also be used to leak cryptographic key material, and in fact any compact secret; a government might, for example, hide a few bits of information about a citizen's arrest record, HIV status or political reliability in the signature on an identity card.

So covert channels are important in a number of applications. However, neither of the above channels is ideal: the signer must either compromise her signing key or accept severe computational limitations on the usable covert bandwidth. This led Simmons to ask whether there is a better scheme — with a broadband covert channel that does not require the sender to compromise the security of her signing key.

2 Our Construction

The ElGamal scheme possesses just such a channel. We assume that the modulus $p = qm + 1$ where m is smooth and extracting discrete logarithms is hard in the subgroup of F_p^* of order q that is generated by g^m. If the covert message we wish to convey is c, we can set

$$k \equiv c \pmod m \tag{3}$$

In other words, we set $k = c + k'm$ for some randomly chosen k'. Now, when the recipient gets the signature (r, s), he forms r^q and solves for z the equation

$$(g^q)^z \equiv r^q \pmod{p} \tag{4}$$

This is feasible since the order of the subgroup of F_p^* generated by g^q is smooth. Using the Pohlig-Hellman decomposition [4] in combination with Pollard's rho method [5], this will require time $O(\sqrt{B})$ where B is the smoothness bound (the largest prime factor of m). We will then have

$$c \equiv z \pmod{m} \tag{5}$$

and the covert message can thus be recovered.

Note that the discrete log calculation needs to be done only once. Given z, we can recover the signing key mod m using equation (2), so further messages can be decoded trivially using equation (2).

This channel is a broadcast one, in the sense that anyone may perform the discrete log calculation and recover x mod m. However, we can also create narrowcast channels, in which the covert message c is only available to parties who possess some previously shared secret. In particular, if $p - 1 = mq_1q_2 + 1$, and the discrete logarithm problem is hard in the groups of order q_1 and q_2, then the signer can keep her signing key secret modulo q_1 but reveal its value modulo q_2 to the intended recipient of covert messages. She can now communicate her covert message c as k mod q_2.

In short, when we use the ElGamal signature scheme with g a generator of Z_p^*, we are signing simultaneously in a number of different groups that correspond to the factors of $p - 1$. Our signing key can be secure in some of these, shared with certain parties in others, and will be available to everyone modulo the smooth part of $p - 1$. This smooth part will be at least 2, and where p is a randomly chosen prime, it will be about $\log_2 B$ (see for example [11]).

3 Tailoring the Channel

Of course, the prime p can be chosen so as to provide any desired combination of broadcast and narrowcast channels. A prime that is optimal for broadcast in ElGamal signatures was given in the original specification of the Digital Signature Algorithm: this had $(p - 1)/q = 2^{70}3^{46}5^{30}7^{25}11^{20}$ [1], yielding a broadcast channel of about 352 bits.

The current DSA standard [2] suggests a different pair of primes, namely

$p =$ 1110695048525066847389659955311086494364275721046177400870101023825839678874642481120264311896935336016195066787729193595754779567794960463100584609534872 7

and

$q =$ 1016505658889014629900729618210002584918553821669

The factors of $(p-1)/q$ have been found by Paul Leyland and are:

2
$q_1 = 4196363948260739557$
$q_2 = 4208101743716447893907182873$
$q_3 = 309382440150971553074910129575307384019243127389783508284907$

Using this prime p for ElGamal signatures would provide a secure signature in the groups of order q and q_3; and either of these could be subverted to provide a narrowcast covert channel. There will also be a broadcast channel of somewhat over 160 bits per signature, as the signing key can be recovered in the groups whose orders are 2 (trivially), q_1 (for about 2^{34} computations) and q_2 (for about 2^{47} computations).

It should be clear from this example how to select p for any desired combination of broadcast and narrowcast covert channel capacity. In the case of a randomly chosen prime p, it can be shown that the expected length in bits v of the product of all prime divisors $\leq B$ is approximately equal to $\log_2 B$ [11]. Thus the subliminal channel requires an effort of $O(2^{v/2})$ to communicate v bits, where the previously known narrowband channel needed about 2^v. Moreover, v has a large variance: for one in 100 1024-bit primes, one obtains a value of v which is about four times larger than the expected value [11]. In any case, the values of p and q in the current DSA specification are not particularly out of the ordinary.

Since this channel was discovered by the authors while they were guests of the Isaac Newton Institute of Mathematical Sciences in Cambridge, we hereby name it 'the Newton channel'.

4 Discussion

The Newton channel arises when a digital signature is performed in a composite group with the property that the key in one or more of its subgroups is shared with the recipient. This sharing can be explicit, in the case of the narrowband channel, or implicit in the case where third parties can simple compute the key in the relevant group or groups.

It is clear that the Newton channel can be avoided by operating in a group of prime order. In the example above, we could replace g by $g^{(p-1)/q}$ (or $g^{(p-1)/q_3}$), and indeed this is the approach taken by DSA (see [6] for more information on this algorithm and its background). When the DSA was first proposed, claims were made that it appeared to have been designed to maximise the covert channel capacity [9]. This claim was denied at the time by a senior NSA official [10], and we can now see that he was right: the DSA does not maximise the covert utility of a signature, but minimises it — by eliminating the Newton channel.

Our methods have implications for security as well as covertness: anyone can recover both the message key k and the signing key x modulo m. In fact, in a typical discrete log based cryptosystem, we would expect to be able to recover all key material modulo the smooth component of the group order. It would be imprudent of a designer to allow such severe key leakage, as many random number generators regularities due to resonances, implementation bugs and the like. An ElGamal scheme using the p and q originally proposed with the DSA would be leaking over two thirds of each key. In many implementations, this could be enough to mount an attack.

This, and other, security considerations will be discussed in more detail in a future paper. For the meantime, we recommend that designers of ElGamal and Diffie-Hellman type systems should always use groups of prime order unless there are good reasons not to.

5 Conclusions

We have answered Simmons' question by demonstrating that the ElGamal signature scheme has a broadband covert channel, the Newton channel, that does not require the sender to compromise the security of her signing key. However, Simmons' conjecture that such schemes did not exist was not entirely mistaken, since the bandwidth of the Newton channel in bits per signature is exactly equal to the number of bits that the signer is prepared to compromise of her signing key. So it may well be that Simmons' conjecture holds with a more precise formulation; we express no opinion on this.

We have also established that, contrary to popular belief, the design of the DSA does not maximise the covert utility of its signatures, but minimises them. Our construction also shows that many discrete log based systems are insecure, while the DSA is not vulnerable in this way. Given the authorship of the DSA, these are perhaps the results that one might have expected.

Acknowledgement: The authors are grateful to the Isaac Newton Institute, 20 Clarkson Road, Cambridge, for hospitality while this research was being conducted; to the National Fund for Scientific Research (Belgium), which sponsored the second author; and last but not least to Paul Leyland for factoring $p - 1$.

References

1. 'A Practical RSA Trapdoor', R Anderson, in *Electronics Letters* v 29 no 11 (27 May 1993) p 995
2. 'Digital Signature Standard,' *Federal Information Processing Standard (FIPS) Publication 186*, National Institute of Standards and Technology, US Department of Commerce, Washington D.C., May 1994
3. 'A Public Key Cryptosystem and a Signature Scheme based on Discrete Logarithms', T ElGamal, *IEEE Transactions on Information Theory, v 31, no 4 (1985) pp 469–472*
4. 'An Improved Algorithm for Computing Logarithms over $GF(p)$ and its Cryptographic Significance', SC Pohlig, ME Hellman, *IEEE Transactions on Information Theory, v 24, no 1 (Jan 78) pp 106–110*

5. 'Monte Carlo Methods for Index Computation (mod p), ' JM Pollard, *Mathematics of Computation, v 32 no 143 (Jul 78) pp 918–924*
6. *'Applied Cryptography'*, B Schneier (2nd edition), Wiley 1995
7. 'Subliminal Channels; Past and Present', GJ Simmons, *European Transactions on Telecommunications v 5 no 4 (Jul/Aug 94) pp 459–473*
8. 'How to Insure That Data Acquired to Verify Treaty Compliance are Trustworthy', GJ Simmons, *Contemporary Cryptology* (IEEE, 1992) pp 617–630
9. 'Subliminal Communciation is Easy Using the DSA', GJ Simmons, *Advances in Cryptology - EUROCRYPT 93*, Springer LNCS v 765 pp 218–232
10. Comment made from the floor at Eurocrypt 93, B Snow
11. 'On Diffie-Hellman Key Agreement with Short Exponents', PC van Oorschot, MJ Wiener, *Advances in Cryptology - EUROCRYPT 96*, Springer LNCS v 1070 pp 332–343