The NSA Back Door to NIST

Thomas C. Hales

Use once. Die once.
—Activist saying about insecure communication

We give a brief mathematical description of the NIST standard for cryptographically secure pseudo-random number generation by elliptic curves, the back door algorithm discovered by Ferguson and Shumow, and finally the design of the back door based on the Diffie-Hellman key exchange algorithm.

NIST (the National Institute for Standards and Technology) of the U.S. Department of Commerce derives its mandate from the U.S. Constitution through the congressional power to "fix the standard of weights and measures." In brief, NIST establishes the basic standards of science and commerce. Whatever NIST says about cryptography becomes implemented in cryptographic applications throughout U.S. government agencies. Its influence leads to the widespread use of its standards in industry and the broad adoption of its standards internationally.

Through the Snowden disclosures, the NIST standard for pseudo-random number generation has fallen into disrepute. Here I describe the back door to the NIST standard for pseudo-random number generation in elementary and mathematically precise terms. The NIST standard offers three methods for pseudo-random number generation [1]. My remarks are limited to the third of the three methods, which is based on elliptic curves.

Random number generators can either be truly random (obtaining their values from randomness in the physical world, such as a quantum mechanical process) or pseudo-random (obtaining their values from a deterministic algorithm, yet displaying a semblance of randomness). The significance of random number generation within the theory of algorithms can be gauged by Knuth's multivolume book *The Art of Computer Programming*. It devotes a massive 193 pages (half of volume two) to the subject! A subclass of pseudo-random number generators are *cryptographically secure*, intended for use in cryptographic applications such as key generation, one-way hash functions, signature schemes, private key cryptosystems, and zero knowledge interactive proofs [3].

Thomas C. Hales is professor of mathematics at the University of Pittsburgh. His email address is hales@pitt.edu.
DOI: http://dx.doi.org/10.1090/noti1078

Elliptic Curves as Pseudo-Random Number Generators

The NIST standard gives a list of explicit mathematical data (E, p, n, f, P, Q) to be used for pseudo-random number generation [1]. Here E is an elliptic curve defined over a finite field $\mathbb{F}_p$ of prime order p. The group $E(\mathbb{F}_p)$ has order n, which is prime for all of the curves that occur in the NIST standard. The elements of the group $E(\mathbb{F}_p)$ consist of the set of points on an affine curve, together with a *point at infinity* which serves as the identity element of the group. The affine curve is defined by an equation $y^2 = f(x)$ for some explicit cubic polynomial f in $\mathbb{F}_p[x]$. Finally, P and Q are given points on the affine curve.

NIST gives a few sets of data, and in each case the prime number p is large. (The smallest is greater than 10^{77}.) No explanation is given of the particular choices (E, p, n, f, P, Q). We are told to use these data and not to question why. The standard stipulates that "one of the following NIST approved curves with associated points shall be used in applications requiring certification under FIPS-140 [U.S. government computer security accreditation]."

When A is any point other than the identity in $E(\mathbb{F}_p)$, we may evaluate the coordinate function x at A to obtain $x(A) \in \mathbb{F}_p$. By further lifting $\mathbb{F}_p$ to a set of representatives in $\mathbb{Z}$, we obtain a function by composition

$$x1 : E(\mathbb{F}_p) \setminus \{0\} \;\to\; \mathbb{F}_p \;\to\; \mathbb{Z}.$$

Write $(n, A) \mapsto n * A$ for the $\mathbb{Z}$-module action of $\mathbb{Z}$ on E. (We write powers of the group element A using multiplicative rather than exponential notation.)

The pseudo-random bit generator is initialized with a random integer seed s obtained by some different process such as a separate random

number generator. What is important for us is that the number s represents the hidden internal state of the algorithm. The hidden state must be kept secret for the pseudo-randomness to be effective. (Once the state is disclosed, a pseudo-random sequence becomes predictable and useless for many cryptographic purposes.)

The essence of the pseudo-random bit generator can be written in the Objective Caml language as follows. In the syntax of this language, each phrase (`let x = a in` ...) defines the value of x to be a. The last line of the block of code gives the output of the function.

```
let pseudo_random s =
  let r = x1 (s * P) in
  let s' = x1 (r * P) in
  let t = x1 (r * Q) in
  let b = extract_bits t in
    (s',b);
```

That is, we successively apply the integer s or r to the point P or the point Q and take the $x1$ coordinate of the resulting point, then extract some bits from the number t. The integer s' becomes the new secret internal state to be fed into the next iteration of the function. The output `b` is passed to the consumer of pseudo-random bits. This output may become publicly known. The function `extract_bits` operates by converting t to a list of bits, discarding the 16 most significant bits (for reasons that do not matter to this discussion), and giving the remaining bits as output. According to NIST standards, by iterating this function, updating the internal state at each iteration, a cryptographically secure stream `b`...of pseudo-random bits is obtained.

The Back Door

This algorithm is fatally flawed, as Ferguson and Shumow have pointed out [5]. Since P and Q are nonidentity elements of a cyclic group of prime order, each is a multiple of the other. Write $P = e * Q$ for some integer e. We show that, once we have e in hand, it is a simple matter to determine the secret internal state s of the pseudo-random bit generator by observing the output `b` and thus to compromise the entire system.

The function `extract_bits` discards 16 bits. Given the output `b`, we take the 2^{16} (a small number of) possible preimages t of `b` under `extract_bits`. For each t, the coordinate x is known, and solving a quadratic, there are at most two possibilities for the coordinate y of a point A on the elliptic curve such that $t = x1(A)$. One such A is $r * Q$. For each A, we compute $e * A$. One of the small number of possibilities for $e * A$ is

$$(1) \qquad e * (r * Q) = r * (e * Q) = r * P.$$

Finally $s' = x1(r * P)$. In short, the internal state s' can be narrowed down to a small number of possibilities by an examination of the pseudo-random output bitstream. Shumow and Ferguson state that in experiments, "32 bytes of output was sufficient to uniquely identify the internal state of the PRNG [pseudo-random number generator]."

The back door to the algorithm is the number e such that $P = e * Q$. To use the back door, one must know the value of e. The NIST standard does not disclose e (of course!), and extensive cryptographic experience suggests that it is hard to compute e from the coordinates of P and Q (unless you happen to own a quantum computer). This is the problem of *discrete logarithms*. But, starting with e, there is no difficulty in creating a pair P and Q. The back door is universal: a single number e gives back door access to the internal state of the algorithm of all users worldwide.

It is a matter of public fact that the NSA was tightly involved in the writing of the standard. Indeed, NIST is required by law to consult with the NSA in creating its standard. According to the *New York Times*, "classified NSA memos appear to confirm that the fatal weakness, discovered by two Microsoft cryptographers in 2007, was engineered by the agency" [4]. The news article goes on to say that "eventually, NSA became the sole editor" and then pushed aggressively to make this the standard for the 163 member countries of the International Organization for Standardization. Further historical and social context appears in [6]. The NSA had facile access to the crown jewel e and motive to seize it. Draw your own conclusions.

Observations

1. The back door to this algorithm is extremely elementary from a mathematical perspective. We wrote the essential algorithm in six lines of computer code, even if more supporting code is needed to make it industrial strength. The algorithm could be explained to undergraduate math majors or sufficiently advanced high school students. The story also has the spy agency intrigue to make a good math club talk or a special lecture in an elementary abstract algebra course. We essentially just need to understand that an elliptic curve is an abelian group whose elements (other than the identity element) are determined by two numbers x and y, that y is the root of a quadratic when x is given, and that every nonidentity element of a cyclic group of prime order is a generator. Easy stuff.

2. Without prior knowledge of the back door, how difficult would it be to rediscover the possible existence of a back door? An analysis of the argument shows the required level of creativity is that of an undergraduate homework problem. We

must think to write the element P as a multiple of the generator Q in a cyclic group of prime order. This a student learns in the first weeks of undergraduate algebra.

The rest of the process of inverting the pseudo-random number generator is determined by the definition of the function itself: simply take each step defining the function and reverse the steps, asking for the preimage of the function at each step of its definition, working from the output back to the secret state s'. Once the question of inverting the function is asked, it is easy to do the group theory, even if it is computationally difficult to write e explicitly.

One-way functions are a standard tool in the cryptographer's bag. Every professional who has been trained to analyze cryptographic algorithms knows to ask the question of invertibility. It is unsettling that NIST and others do not seem to have asked this basic question.

Diffie-Hellman Key Exchange

In what follows, let us assume that someone, whom we will call *the Spy*, has access to the back door e. How is it possible for the Spy and the end user (*the User*) of the NIST algorithm to come into possession of the same shared secret (the internal state of the pseudo-random number generator) when all communication between them is public? Information flows from the Spy to the User through the published NIST standard, and from the User back to the Spy through the public output of the pseudo-random generator. The back door must have a remarkable cryptographic design to permit a secret to pass across these public channels yet prevent the secret from becoming known to a third party.

As we now explain, the design of the back door to NIST is based on a well-known algorithm in cryptography called the Diffie-Hellman key exchange [2]. This is an algorithm to share a secret between two parties when there is a possibility that the channel of communication is being monitored. In the current context, the Spy has full knowledge of the Diffie-Hellman key exchange for what it is. However, the User participates in the exchange innocently and unwittingly by blindly following the rules of the NIST protocol.

The Diffie-Hellman key exchange requires a group, which we will take to be a cyclic group E of order n (to preserve notation). The group E, its order n, and a generator Q are made public. To share a secret, the first party (the Spy) picks a random number e, which is kept secret, and publishes $P = e * Q$ to the world. The second party (the User) picks a random number r, which is kept secret, and publishes $r * Q$. Then, by equation (1), the Spy, who knows e and $r * Q$, and the User, who knows r and $e * Q$, can both compute $(re) * Q = r * P$, which is the shared secret. (In our context, the shared secret determines the internal state s' of the pseudo-random number generator.) If E is a group in which the public knowledge of E, n, Q, $P = e * Q$, $r * Q$ does not allow the easy computation of $(re) * Q$, then the shared secret is protected from public disclosure by the difficulty of the computation. In this way, the only two who learn the internal state of the pseudo-random number generator are the Spy and the User.

What we have described here is not an imaginary scenario: NIST documents do in fact publish the data E, n, Q, and P needed to initiate the Diffie-Hellman exchange. A user, when making public the output from the pseudo-random number generator, does in fact complete the exchange. Diffie-Hellman is Diffie-Hellman, whether it has been advertised as such or not.

To say that the Diffie-Hellman key exchange algorithm is well known is a vast understatement. This algorithm is a significant lesson in virtually every first course in cryptography everywhere in the world. Building on Merkle, the Diffie-Hellman paper, by starting the entire field of public key cryptography, is one of the most important papers in cryptography ever written.

What is the significance of all this? It is no secret that the NSA employs some of the world's keenest cryptographic minds. They all know Diffie-Hellman. In my opinion, an algorithm that has been designed by NSA with a clear mathematical structure giving them exclusive back door access is no accident, particularly in light of the Snowden documents.

References

1. E. Barker and J. Kelsey, Recommendation for random number generation using deterministic random bit generators, *NIST Special Publication 800-90A* (2012), `http://csrc.nist.gov/publications/nistpubs/800-90A/SP800-90A.pdf`.
2. W. Diffie and M. Hellman, New directions in cryptography, *IEEE Transactions on Information Theory* **22** (1976), 644–654.
3. M. Luby, *Pseudorandomness and Cryptographic Applications*, Princeton University Press, 1996.
4. N. Perloth, J. Larson, and S. Shane, N.S.A. able to foil basic safeguards of privacy on Web, *New York Times*, September 5, 2013, `http://www.nytimes.com/2013/09/06/us/nsa-foils-much-internet-encryption.html`.
5. D. Shumow and N. Ferguson, On the possibility of a back door in the NIST SP800-90 dual Ec prng, `http://rump2007.cr.yp.to/15-shumow.pdf`, 2007.
6. K. Zetter, How a crypto "backdoor" pitted the tech world against the NSA, *Wired* (2013), `http://www.wired.com/threatlevel/2013/09/nsa-backdoor/`.