

Factoring RSA Keys from Certified Smart Cards: Coppersmith in the Wild

Daniel J. Bernstein^{1,2}, Yun-An Chang³, Chen-Mou Cheng³, Li-Ping Chou⁴,
Nadia Heninger⁵, Tanja Lange², and Nicko van Someren⁶

¹ Department of Computer Science, University of Illinois at Chicago, USA
djb@cr.yp.to

² Department of Mathematics and Computer Science
Technische Universiteit Eindhoven, The Netherlands
tanja@hyperelliptic.org

³ Research Center for Information Technology Innovation
Academia Sinica, Taipei, Taiwan
{ghfjdksl,doug}@crypto.tw

⁴ Department of Computer Science and Information Engineering
Chinese Culture University, Taipei, Taiwan
randomalg@gmail.com

⁵ Department of Computer and Information Science, University of Pennsylvania
nadiyah@cis.upenn.edu

⁶ Good Technology Inc.
nicko@good.com

Abstract. This paper explains how an attacker can efficiently factor 184 distinct RSA keys out of more than two million 1024-bit RSA keys downloaded from Taiwan’s national “Citizen Digital Certificate” database. These keys were generated by government-issued smart cards that have built-in hardware random-number generators and that are advertised as having passed FIPS 140-2 Level 2 certification.

These 184 keys include 103 keys that share primes and that are efficiently factored by a batch-GCD computation. This is the same type of computation that was used last year by two independent teams (USENIX Security 2012: Heninger, Durumeric, Wustrow, Halderman; Crypto 2012: Lenstra, Hughes, Augier, Bos, Kleinjung, Wachter) to factor tens of thousands of cryptographic keys on the Internet.

The remaining 81 keys do not share primes. Factoring these 81 keys requires taking deeper advantage of randomness-generation failures: first using the shared primes as a springboard to characterize the failures, and then using Coppersmith-type partial-key-recovery attacks. This is the first successful public application of Coppersmith-type attacks to keys found in the wild.

Keywords: RSA, smart cards, factorization, Coppersmith, lattices.

This work was supported by NSF (U.S.) under grant 1018836, by NWO (Netherlands) under grants 639.073.005 and 040.09.003, and by NSC (Taiwan) under grant 101-2915-I-001-019. Cheng worked on this project while at Technische Universität Darmstadt under the support of Alexander von Humboldt-Stiftung. Heninger worked on this project while at Microsoft Research New England. Permanent ID of this document: 278505a8b16015f4fd8acae818080edd. Date: 2013.09.10.

1 Introduction

In 2003, Taiwan introduced an e-government initiative to provide a national public-key infrastructure for all citizens. This national certificate service allows citizens to use “smart” ID cards to digitally authenticate themselves to government services, such as filing income taxes and modifying car registrations online, as well as to a growing number of non-government services. RSA keys are generated by the cards, digitally signed by a government authority, and placed into an online repository of “Citizen Digital Certificates”.

On some of these smart cards, unfortunately, the random-number generators used for key generation are fatally flawed, and have generated real certificates containing keys that provide no security whatsoever. This paper explains how we have computed the secret keys for 184 different certificates.

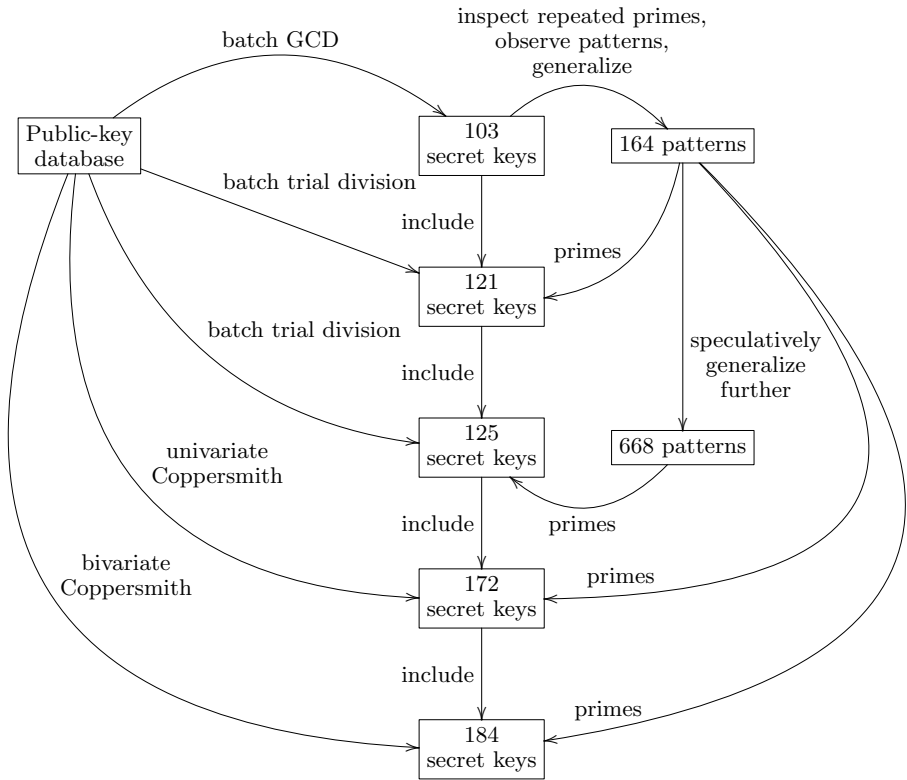


Fig. 1. Retrospective summary of the data flow leading to successful factorizations. After successfully factoring keys using a batch GCD algorithm, we characterized the failures, and used trial division to check for broader classes of specified primes (input on the right) as exact divisors. We then extended the attack and applied Coppersmith’s method to check for the specified primes as approximate divisors.

1.1 Factorization Techniques

Bad randomness is not new. Last year two independent research teams [13,17] exploited bad randomness to break tens of thousands of keys of SSL certificates on the Internet, a similar number of SSH host keys, and a few PGP keys.

Our starting point in this work is the same basic attack used in those papers against poorly generated RSA keys, namely scanning for pairs of distinct keys that share a common divisor (see Section 3). The basic GCD attack, applied to the entire database of Citizen Digital Certificates, shows that 103 keys factor into 119 different primes.

We go beyond this attack in several ways. First, the shared primes provide enough data to build a model of the prime-generation procedure. It is surprising to see *visible* patterns of non-randomness in the primes generated by these smart cards, much more blatant non-randomness than the SSL key-generation failures identified by [13,17]. One expects smart cards to be controlled environments with built-in random-number generators, typically certified to meet various standards and practically guaranteed to avoid such obvious patterns. For comparison, the SSL keys factored last year were typically keys generated by low-power networked devices such as routers and firewalls running the Linux operating system while providing none of the sources of random input that Linux expects.

The next step is extrapolation from these prime factors: we hypothesize a particular model of randomness-generation failures consistent with 18 of the common divisors. The same model is actually capable of generating 164 different primes, and testing all of those primes using batch trial division successfully factors further keys. One might also speculate that the cards can generate primes fitting a somewhat broader model; this speculation turns out to be correct, factoring a few additional keys and bringing the total to 125. See Section 4 for a description of the patterns in these primes.

There are also several prime factors that are similar to the 164 patterns but that contain sporadic errors: some bits flipped here and there, or short sequences of altered bits. We therefore mount several Coppersmith-style lattice-based partial-key-recovery attacks to efficiently find prime divisors close to the patterns. The univariate attacks (Section 5) allow an arbitrary stretch of errors covering the bottom 40% of the bits of the prime. The bivariate attacks (Section 6) allow two separate stretches of errors. The internal structure of the patterns makes them particularly susceptible to these attacks. These attacks produce dozens of additional factorizations, raising the total to 184.

In the end nearly half of the keys that we factored did *not* share any common divisors with other keys; most of these were factored by the Coppersmith-style attacks. This is, to our knowledge, the first publicly described instance of a Coppersmith-style attack breaking keys in the wild.

1.2 Certification

The flawed keys were generated by government-issued smart cards that both the certification authority and manufacturer advertise as having passed stringent standards certifications. See Section 2.1.

It is clear from their externally visible behavior, as shown in this paper, that the random-number generators used to generate the vulnerable keys actually fall far short of these standards. This demonstrates a failure of the underlying hardware and the card's operating system, both of which are covered by certification.

1.3 Response to Vulnerabilities

When we reported the common-divisor vulnerabilities to government authorities, their response was to revoke exactly the certificates sharing common factors and to issue new cards only to those users. See Section 7 for more details.

Our further factorizations demonstrate how dangerous this type of response is. Randomness-generation failures sometimes manifest themselves as primes appearing twice, but sometimes manifest themselves as primes that appear only once, such as the primes that we found by Coppersmith-type attacks. Both cases are vulnerable to attackers with adequate models of the randomness-generation process, while only the first case is caught by central testing for repeated primes.

We endorse the idea of centrally testing RSA moduli for common divisors as a mechanism to detect some types of randomness-generation failures. We emphasize that finding repeated primes is much more than an indication that those particular RSA keys are vulnerable: it shows that the underlying randomness-generation system is malfunctioning. The correct response is not merely to eliminate those RSA keys but to revoke all keys generated with that generation of hardware and throw away the entire randomness-generation system, replacing it with a properly engineered system.

We also emphasize that an absence of common divisors is not an indication of security. If the primes generated by these smart cards had been modified to include a card serial number as their top bits then the keys would have avoided common divisors but the primes would still have been reasonably predictable to attackers. Our work illustrates several methods of translating different types of malfunctioning behavior into concrete vulnerabilities. There are many potential vulnerabilities resulting from bad randomness; it is important to thoroughly test every component of a random-number generator, not merely to look for certain types of extreme failures.

2 Background

2.1 The Taiwan Citizen Digital Certificate Program

Taiwan's Citizen Digital Certificates (CDCs) are a standard means of authentication whenever Taiwanese citizens want to do business over the Internet with the government and an increasing number of private companies.

CDCs are issued by the Ministry of Interior Certificate Authority (MOICA), a level 1 subordinate CA of the Taiwanese governmental PKI. Since the program's launch in 2003, more than 3.5 million CDCs have been issued, providing public key certificate and attribute certificate services. These digital certificates form

a basis for the Taiwanese government’s plan to migrate to electronic certificates from existing paper certificates for a range of applications including national and other identification cards, driver’s licenses, and various professional technician licenses.

Today, Taiwanese citizens can already use the CDC to authenticate themselves over the Internet in a number of important government applications, e.g., to file personal income taxes, update car registration, and make transactions with government agencies such as property registries, national labor insurance, public safety, and immigration. In addition, the CDC is accepted as a means of authentication by a variety of organizations such as the National Science Council, several local governments, and recently some private companies such as Chunghwa Telecom. Overall, the CDC program appears quite successful as a two-sided network, as it has attracted an increasing number of both applications and subscribers.

Certificate registration: In order to generate CDCs, citizens bring their (paper) ID cards to a government registration office. A government official places the (smart) ID card into a registration device. The device prompts the card to generate a new cryptographic key, and the public key is incorporated into a certificate to be signed by MOICA. The certificate is made available in a database online for authentication purposes. In general, an individual will have two certificates: one for signing, and one for encryption, each with distinct keys.

Standards certifications: MOICA states that these cards are “high security”, and “have been accredited to FIPS 140-1 level 2”, and also that “A private key is created inside and the private key can’t export from IC card after key created”. (See [20] or search for “FIPS” on MOICA’s website <http://moica.nat.gov.tw/html/en/index.htm>.) For comparison, the SSL keys factored last year were generated by software-hardware combinations that had never claimed to be evaluated for cryptographic security, such as Linux running on a home router.

2.2 Collecting Certificates

In March 2012, inspired by the results of [13] and [17], we retrieved 3002273 CDCs from the MOICA LDAP directory at `ldap://moica.nat.gov.tw`. Out of these CDCs, 2257569 have 1024-bit RSA keys, while the remaining, newer 744704 have 2048-bit RSA keys, as in 2010 MOICA migrated to 2048-bit RSA and stopped issuing certificates of 1024-bit RSA keys.

The 1024-bit CDCs contain 2086177 distinct moduli, of which 171366 moduli appear more than once. The repeated moduli appear to all be due to expired certificates still contained in the database, which contain the same keys as renewal certificates issued to the same individuals.

2.3 Random Number Generation

While generating high-quality random numbers is critical to the security of cryptographic systems, it is also notoriously difficult to do. Non-deterministic behavior is considered to be a fault in almost every other component of a computer, but it is a crucial component of generating random numbers that an attacker cannot predict. Several national and international standards for random number generation [22,1,11] specify correct behavior for these types of systems. In general, software pseudo-random number generators require a significant amount of entropy before their output is useful for cryptographic purposes.

As we will see later in the paper, the smart cards used in the PKI we examined fail to follow many well-known best practices and standards in hardware random number generation: they appear to utilize a source of randomness that is prone to failing, they fail to perform any run-time testing before generating keys, and they clearly do not apply any post-processing to the randomness stream. The lack of testing or post-processing causes the initial randomness-generation failure to be much more damaging than it would have been otherwise.

Analog RNG circuits: An analog circuit is the standard choice when hardware designers have the luxury of designing dedicated circuits for random-number generation. An analog circuit allows the designer to obtain randomness from simple quantum effects. While the use of radioactive decay is rare in commercial products, the quantum noise exhibited by a current through a suitably biased diode can be amplified and sampled to deliver a high-quality entropy source.

On-chip RNG circuits: Mixing analog and digital circuits on the same die is costly, so chip designers often seek other sources of unpredictability. These sources can include variation in gate propagation delays or gate metastability, which exhibit inherent randomness. Designers can explicitly harness gate-delay variation by building sets of free-running ring oscillators and sampling the behavior at hopefully uncorrelated intervals. To take advantage of randomness in gate metastability, designers build circuits that output bits based on the time it takes for the circuit to settle to a steady state, a variable which should be hard to predict. These designs are often tricky to get right, as the chip fabrication process can reduce or eliminate these variations, and subtle on-chip effects such as inductive coupling or charge coupling between components can cause free-running oscillators to settle into synchronised patterns and metastable circuits to predictably land one way or the other depending on other components nearby on the chip.

Handling entropy sources: Even with a perfectly unpredictable source of randomness, care needs to be taken to convert the raw signal into usable random numbers. Generally, designers characterize circuits in advance to understand the entropy density, test the signal from the entropy source at run time, and run the output through a compression function such as a cryptographically secure hash function. These practices are required by a number of security standards such as FIPS 140 [21].

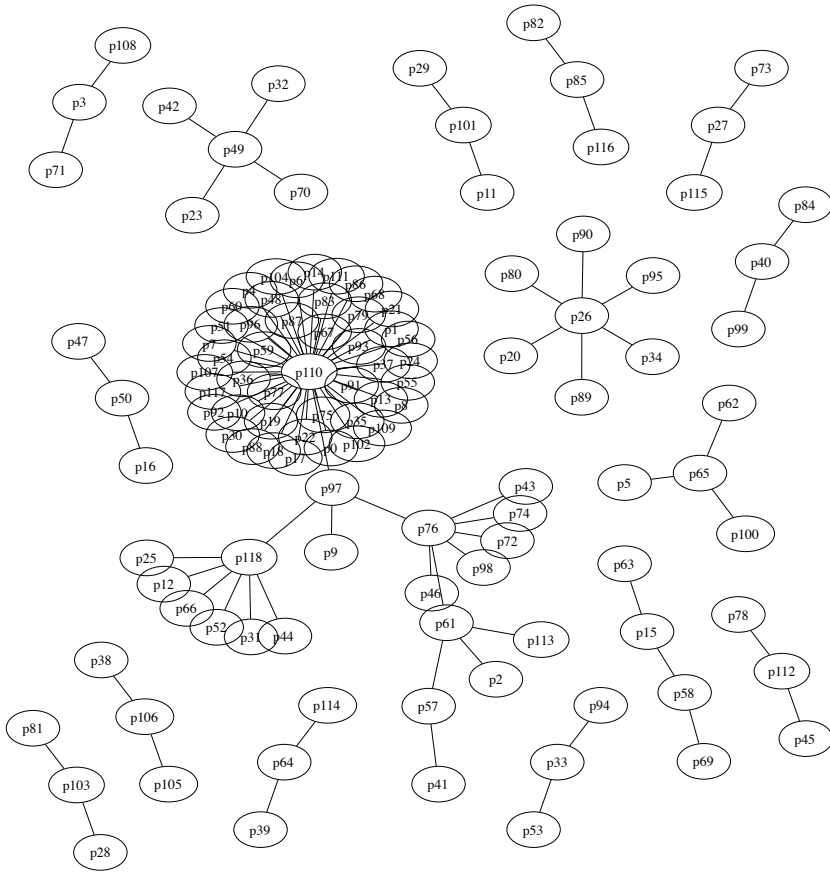


Fig. 2. Relationships between keys with shared factors. Each ellipse represents a prime; edges connect prime factors dividing the same modulus.

Nearly all of the shared prime factors had a similar and immediately apparent periodic structure. We hypothesized that nearly every repeated prime factor had been generated using the following process:

1. Choose a bit pattern of length 1, 3, 5, or 7 bits, repeat it to cover more than 512 bits, and truncate to exactly 512 bits.
2. For every 32-bit word, swap the lower and upper 16 bits.
3. Fix the most significant two bits to 11.
4. Find the next prime greater than or equal to this number.

We generated the 164 distinct primes of this form corresponding to all patterns of length 1, 3, 5, and 7 and tested divisibility with each modulus. This factored a total of 105 moduli, including 18 previously unfactored moduli, for a total of 121.

None of the repeated primes exhibit a (minimal) period of length 9 or larger. On the other hand, the data for period lengths 1, 3, 5, 7 shows that patterns with longer periods typically appear in fewer keys than patterns with shorter periods, and are thus less likely to appear as divisors of two or more keys, raising the question of whether there are primes with larger periods that appear in only one key and that are thus not found by the batch-GCD computation. We therefore extended this test to include length-9 periods and length-11 periods. The length-9 periods factored 4 more keys but the length-11 periods did not factor any new keys, leading us to speculate that 3, 5, and 7 are the only factors of the period length. We then ran a complete test on all length-15 patterns but did not find any further factors. The total number of certificates broken by these divisibility tests, together with the initial batch-GCD computation, is 125.

Sporadic errors: The handful of shared prime factors in our sample of GCD-factored keys that did not match the above form were differing from patterns in very few positions. We experimented with finding more factors using brute-force search starting from $0xc0\dots0$ and found a few new factors, but these factors are more systematically and efficiently found using LLL in Coppersmith's method, as described in the next section.

We also experimented with searching for sporadic errors in factors using the techniques of Heninger and Shacham [14] and Paterson, Polychroniadou, and Sibborn [23]. The main idea is to assume that both of the factors of a weak modulus share nearly all bits in common with a known pattern, with only sporadic errors in each. It is then possible to recover the primes by enumerating, bit by bit, a search tree of all possible prime factors, and using depth- or breadth-first search with pruning to find a low-Hamming weight path through the search tree of all solutions.

Unfortunately, there are a few difficulties in applying this idea to the case at hand. The first is that because the primes are generated by incrementing to the next prime, a single sporadic error is likely to cause the least significant 9 bits of each prime to appear random (except for the least significant bit which is set to 1), so generating a solution tree from the least significant bits necessarily begins with that much brute forcing. Second, there is only a single constraint on the solutions (the fact that $pq = N$), instead of four constraints, which results in a lower probability of an incorrect solution being pruned than in the examples considered by [14,23]. And finally, in order to apply the algorithms, we must guess the underlying pattern, which in our case requires applying the algorithm to 164^2 possibilities for each modulus.

Applying this algorithm using only the all-zeros pattern for both factors to the 45 moduli with 20 bits of consecutive zeros took 13 minutes and factored 5 moduli. All of these moduli were also factored by the GCD method or Coppersmith methods described in the next section.

5 Univariate Coppersmith

Several of the factors computed via the GCD algorithm in Section 3 follow the bit patterns described in Section 4, but are interrupted by what appear to be sporadic errors. Coppersmith's method [6,7] factors RSA moduli if the top bits of the primes are known, which matches our situation if the errors appear in the bottom few bits of a factor. The method uses lattice basis reduction to factor in polynomial time if at least half of the most significant bits of a prime factor are known; however, since the running time scales very poorly as one approaches this bound, we will be more interested in less optimal parameters that are efficient enough to apply speculatively to millions of keys.

This section presents this method following Howgrave-Graham [16] for lattices of dimension 3 and 5 and gives an outlook of how more keys could be factored using larger dimensions. The idea is as follows: we assume that some prime factor p of N is of the form

$$p = a + r$$

where a is a known 512-bit integer (one of the bit patterns described in the previous section) and r is a small integer error to account for a sequence of bit errors (and incrementing to next prime) among the least significant bits of p .

In the Coppersmith/Howgrave-Graham method, we can write a polynomial

$$f(x) = a + x$$

and we would like to find a root r of f modulo a large divisor of N (of size approximately $N^{1/2} \approx p$). Let X be the bound on the size of the root we are searching for. We will use lattice basis reduction to construct a new polynomial $g(x)$ where $g(r) = 0$ over the integers, and thus we can factor g to discover r .

Let L be the lattice generated by the rows of the basis matrix

$$\begin{bmatrix} X^2 & Xa & 0 \\ 0 & X & a \\ 0 & 0 & N \end{bmatrix}$$

corresponding to the coefficients of the polynomials $Xxf(Xx), f(Xx), N$. Any vector in L can be written as an integer combination of basis vectors, and, after dividing by the appropriate power of X , corresponds to the coefficients of a polynomial $g(x)$ which is an integer combination of f and N , and is thus divisible by p by construction. A prime p is found by this method if we can find g such that $g(r_i) \equiv 0 \pmod{p}$ holds not only modulo p but over the integers. The latter is ensured if the coefficients of g are sufficiently small, which corresponds to finding a short vector in L .

To find such a short vector, we apply the LLL lattice basis reduction algorithm [18]. To finish the algorithm, we regard the shortest vector in the reduced basis as the coefficients of a polynomial $g(Xx)$, compute the roots r_i of $g(x)$, and check if $a + r_i$ divides N . If so, we have factored N .

The shortest vector v_1 found by LLL is of length

$$|v_1| \leq 2^{(\dim L - 1)/4} (\det L)^{1/\dim L},$$

which must be smaller than p for the attack to succeed.

In our situation this translates to

$$2^{1/2} (X^3 N)^{1/3} < N^{1/2} \Leftrightarrow X < 2^{-1/2} N^{1/6},$$

so for $N \approx 2^{1024}$ we can choose X as large as 2^{170} , meaning that for a fast attack using dimension-3 lattices up to the bottom third of a prime can deviate from the pattern a . In the following we ignore the factor $2^{(\dim L - 1)/4}$ since all lattices we deal with are of small dimension and the contribution compared to N is negligible.

5.1 Experimental Results

A straightforward implementation using Sage 5.8 took about one hour on one CPU core to apply this method for one of the 164 patterns identified in Section 4. Running it for all 164 patterns factored 160 keys, obviously including all 105 keys derived from the patterns without error, and found 39 previously unfactored keys.

It is worth noting that the 160 keys included all but 2 of the 103 keys factored with the GCD method, showing that most of the weak primes are based on the patterns we identified and that errors predominantly appeared in the bottom third of the bits. The missing 2 keys are those divisible by `0xe0000...0f`. Including `0xd0000...0`, `0xe0000...0`, `0xf0000...0` as additional bit patterns did not reveal any factors beyond the known ones, ruling out the hypothesis that the prime generation might set the top 4 bits rather than just 2. Instead this prime must have received a bit error in the top part.

5.2 Handling More Errors

Coppersmith’s method can find primes with errors in up to 1/2 of their bits using lattices of higher dimension. Getting close to this bound is prohibitively expensive, but trying somewhat larger dimensions than 3 is possible. For dimension 5 we used basis

$$\{N^2, Nf(xX), f^2(xX), xXf^2(xX), (xX)^2f^2(xX)\}$$

which up to LLL constants handles $X < N^{1/5}$, i.e. up to 204 erroneous bottom bits in p for N of 1024 bits. The computation took about 2 hours per pattern and revealed 6 more factors.

We did not use higher dimensions because the “error” patterns we observed are very sparse making it more profitable to explore multivariate attacks (see Section 6).

5.3 Errors in the Top Bits

The factor $0xe000 \dots f (2^{511} + 2^{510} + 2^{509} + 15)$ appeared as a common factor after taking GCDs but was not found by the lattice attacks described in this section applied to the basic patterns described in Section 4. We can apply Coppersmith's attack to search for errors in higher bits of p by defining the polynomial f as $f(x) = a + 2^t x$. Here t is a bit offset giving the location of the errors we hope to learn. The method and bounds described in this section apply as well to this case.

However, since we hypothesize that the prime factors are generated by incrementing to the next prime after a sequence of bits output by the flawed RNG, we will not know the least significant bits of a because they have been modified in the prime generation process. This problem might speculatively be overcome by brute forcing the m least significant bits of each pattern: for each application of the algorithm to a single pattern a , we would apply the algorithm to the 2^{m-1} patterns generated by fixing a and varying the bottom m bits, with the least significant bit always fixed to 1. This will find factors if finding the next prime from the base string with errors did not require incrementing by more than those bottom m bits.

The following rough analysis suggests that for this attack to have a 50% chance of success, we need to apply the algorithm to 128 new patterns for every old pattern. Recall that the chance that a number around z is prime is approximately $1/\log z$, where \log is the natural logarithm. In particular, each number around 2^{512} has about a $1/355$ chance of being prime. Since $1 - (1 - 1/355)^{256} \approx 0.5$, trying 128 patterns for the bottom eight bits for odd patterns has a 50% chance of covering a sufficiently large interval to find a prime. See [12] for more precise estimates. Applying this to our 164 base patterns, our implementation would require 20992 core hours, or close to 2.5 core years. It is fairly likely that more factors would be found with this search but the method presented in the following section is more efficient at handling errors in top and bottom positions unless a very large portion of the top bits are erroneous.

6 Bivariate Coppersmith

The lattice attacks described in the previous section let us factor keys with unpredictable bits occurring in the least significant bits of one of the factors, with all of the remaining bits of the factor following a predictable pattern. In this section, we describe how we extended this attack to factor keys with unpredictable bits among the middle or most significant bits of one of the factors, without resorting to brute-forcing the bottom bits.

In the basic setup of the problem, we assume that one of the factors p of N has the form

$$p = a + 2^t s + r$$

where a is a 512-bit integer with a predictable bit pattern (as described in Section 4), t is a bit offset where a sequence of bit errors s deviating from the

predictable pattern in a occurred during key generation, and r is an error at the least significant bits to account for the implementation incrementing to the next prime.

To apply Coppersmith's method, we can define an equation $f(x, y) = a + 2^t x + y$ and try to use lattice basis reduction to find new polynomials $Q_i(x, y)$ with the property that if $f(s, r)$ vanishes modulo a large unknown divisor p of N and s and r are reasonably small, then $Q_i(s, r) = 0$ over the integers. In that case, we can attempt to find appropriate zeros of Q_i . The most common method to do this is to look at multiple distinct polynomials Q_i and hope that their common solution set is not too large.

These types of bivariate Coppersmith attacks have many cryptanalytic applications, perhaps most prominently Boneh and Durfee's attack against RSA private key $d < N^{0.29}$ [3]. Our approach is very similar to that described by Herrmann and May for factoring RSA moduli with bits known [15], although for the application we describe here, we are less interested in optimal parameters, and more in speed: we wish to find the keys most likely to be factored using very low dimensional lattices.

Algebraic independence: Nearly all applications of multivariate Coppersmith methods require a heuristic assumption that the attacker can obtain two (or several) algebraically independent polynomial equations determined by the short vectors in a LLL-reduced lattice; this allows the attacker to compute a finite (polynomially-sized) set of common solutions. Most theorem statements in these papers include this heuristic assumption of algebraic independence as a matter of course, and note briefly (if at all) that it appears to be backed up experimentally.

Notably, in our experiments, this assumption *did not* hold in general. That is, most of the time the equations we obtained after lattice basis reduction were not algebraically independent, and in particular, the algebraic dependencies arose because all of the short vectors in the lattice were polynomial multiples of a single bivariate linear equation. This linear equation did in fact vanish at the desired solution, but without further information, there are an infinite number of additional solutions that we could not rule out. However, we were often able to find the solution using a simple method that we describe below.

Herrmann and May [15] describe one case where the assumption of algebraic independence did not hold in their experiments, namely when X and Y were significantly larger than the values of s and r . Similar to our case they observed that the polynomials of small norm shared a common factor but unlike in our case this factor was the original polynomial f . Note that the linear polynomial in our case vanishes over the integers at (s, r) while f vanishes only modulo p .

We experimented with running smaller dimensional lattice attacks in order to generate this sublattice more directly. The attack worked with smaller degree equations than theoretically required to obtain a result, but when we experimented with lattices generated from linear equations, this sublattice did not appear. Note that we specify a slightly different basis for the lattice, in terms of monomial powers rather than powers of f , which may have an effect on the

output of the algorithm compared to the examples in [15] and might explain why we find a useful linear equation in the sublattice instead of the useless factor f .

6.1 Implementation Details

Lattice construction: Let X and Y be bounds on the size of the roots at x and y we wish to find. Our lattice is constructed using polynomial multiples of $f(x, y) = a + 2^t xX + yY$ and N up to degree k vanishing to degree 1 modulo p . Our lattice basis consists of the coefficient vectors of the set of polynomials

$$\{(Yy)^h (Xx)^i f^j N^\ell \mid j + \ell = 1, 0 \leq h + i + j \leq k\}$$

$$= \{N, xXN, f, (xX)^2N, (xX)f, \dots, (yY)^{k-2}(xX)f, (yY)^{k-1}f\},$$

using coefficients of the monomials $\{1, x, y, x^2, \dots, y^{k-1}x, y^k\}$. The determinant of this lattice is

$$\det L = N^{k+1} (XY)^{\binom{k+2}{3}}.$$

and the dimension is $\binom{k+2}{2}$. Omitting the approximation factor of LLL, we want to ensure that

$$(\det L)^{1/\dim L} < p$$

$$\left(N^{k+1} (XY)^{\binom{k+2}{3}}\right)^{1/\binom{k+2}{2}} < N^{1/2}.$$

So for $N \approx 2^{1024}$, setting $k = 3$ should let us find $XY < 2^{102}$ and $k = 4$ should let us find $XY < 2^{128}$. The parameter choice $k = 2$ results in a theoretical bound $XY < 1$, but we also experimented with this choice; see below.

Solving for solutions: After running LLL on our lattice, we needed to solve the system of equations it generated over the integers to find our desired roots. The usual method of doing this in bivariate Coppersmith applications is to hope that the two shortest vectors in the reduced basis correspond to algebraically independent polynomials, and use resultants or Gröbner bases to compute the set of solutions. Unfortunately, in nearly all of our experiments, this condition did not hold, and thus there were an infinite number of possible solutions.

However, a simple method sufficed to compute these solutions in our experiments. In general, the algebraic dependencies arose because the short vectors in the reduced basis corresponded to a sublattice of multiples of the same degree-one equation, with seemingly random coefficients, which vanished at the desired roots. (The coefficient vectors were linearly independent, but the underlying polynomials were not algebraically independent.) The other polynomial factors of these short polynomials did not vanish at these roots. This linear equation has an infinite number of solutions, but in our experiments our desired roots corresponded to the smallest integer solution, which we could obtain by rounding.

Let

$$ux + vy - w = 0$$

be an equation we want to solve for x and y . If u and v are relatively prime, then we can write $c_1u + c_2v = 1$, and parametrize an integer family of solutions

$$\begin{aligned}x &= c_1w + vz \\y &= c_2w - uz\end{aligned}$$

with $z = c_2x - c_1y$.

In experiments with the already-factored moduli, we observed that the solution was often the minimum integer value of x or y among the solution family. So we searched for z among the rounded values of $-c_1w/v$ and c_2w/u . This solution successfully factored the moduli in our dataset whenever the shortest-vector polynomial returned by lattice basis reduction was not irreducible.

For the handful of cases where the lattice did result in independent equations, we computed the solutions using a Gröbner basis generated by the two shortest vectors.

6.2 Experimental Results

We ran our experiments using Sage 5.8 [24] parallelized across eight cores on a 3.1GHz AMD FX-8120 processor. We used fpLLL [4] for lattice basis reduction, and Singular [8] to factor polynomials and compute Gröbner bases. For each lattice, we attempted to solve the system of equations either by factoring the polynomial into linear factors and looking for small solutions of the linear equations as described above or using Gröbner bases.

We attempted to factor each of the 2,086,171 1024-bit moduli using several different parameter settings. For $k = 3$, we had 10-dimensional lattices, and attempted to factor each modulus with the base pattern $a = 0$ using $Y = 2^{30}$, $X = 2^{70}$, and $t = 442$. We then experimented with $k = 4$, $Y = 2^{28}$, and $X = 2^{100}$, which gave us 15-dimensional lattices, and experimented with a base pattern $a = 2^{511} + 2^{510}$ and five different error offsets: $t = 0$ with $Y = 2^{128}$ and $X = 1$, and $t = 128, t = 228, t = 328$, and $t = 428$ with $Y = 2^{28}$ and $X = 2^{100}$. Finally, we experimented with the choice $k = 2$, $X = 4$, $Y = 4$ and the choices of t and a used in the $k = 4$ experiments, which used 6-dimensional lattices and theoretically should not have produced output, but in fact turned out to produce nearly all of the same factorizations as the choices above. We ran one very large experiment, using $k = 2$, $t = 1$, $Y = 2^{28}$, $X = 2^{74}$, $t = 438$, and running against all 164 patterns, which produced 155 factored keys, including two previously undiscovered factorizations. The choice $k = 1$ with the same parameter choices as $k = 2$ did not produce results.

6.3 Handling More Errors

From these experimental settings, it seems likely that many more keys could be factored by different choices of parameters and initial pattern values; one is limited merely by time and computational resources. We experimented with iterating over all patterns, but the computation quickly becomes very expensive.

Table 1. Experimental results from factoring keys using a bivariate Coppersmith approach, using the parameters listed in the text. Where we collected data, we noted the very small number of cases where the lattice produced algebraically independent polynomials; all of the other cases were solved via the heuristic methods described above.

k	$\log_2(XY)$	$\# t$	$\#$ patterns	$\#$ factored keys	$\#$ alg. indep. eqns.	running time
2	4	5	1	104	3	4.3 hours
2	4	1	164	154	21	195 hours
3	100	1	1	112	-	2 hours
4	128	5	1	108	4	20 hours

Patterned factors: Mysteriously, using the base patterns $a = 0$ and $a = 2^{511} + 2^{510}$, the algorithm produced factorizations of keys with other patterned factorizations. This is because the product of the bit pattern of the relevant factor multiplied with a small factor produced an integer of the form we searched for, but we are as yet unable to characterize this behavior in general.

Higher powers of p : Similar to the univariate case we can construct higher-dimensional lattices in which each vector is divisible by higher powers of p , e.g. using multiples of N^2 , Nf , and f^2 for divisibility by p^2 . However, this approach is successful in covering larger ranges of XY only for lattices of dimension at least 28, which would incur a significantly greater computational cost to run over the entire data set of millions of keys.

More variables: More isolated errors can be handled by writing $p = a + \sum_{i=1}^c 2^i s_i$ with appropriate bounds on the $s_i < X_i$ so that the intervals do not overlap. The asymptotically optimal case is described in [15] and reaches similar bounds for $\prod_{i=1}^c X_i$ as in the univariate and bivariate case. However, the lattice dimension increases significantly with c . For $c = 3$, i.e. two patches of errors together with changed bottom bits to generate a prime, the condition $(\det L)^{\dim 1/L} < p$ holds only for lattices of dimension at least 35 at which point $X_1 X_2 X_3 < N^{1/14}$ can be found. A lattice of dimension 20 leads to the condition $X_1 X_2 X_3 < 1$. A sufficiently motivated attacker can run LLL on lattices of these dimensions but we decided that factors found thus far were sufficient to prove our point that the smart cards are fatally flawed.

6.4 Extension to Implicit Factoring

Ritzenhofen and May [19] and Faugère, Marinier, and Renault [9] give algorithms to factor RSA moduli when it is known that two or more moduli have prime factors that share large numbers of bits in common. Unfortunately, these results seem to apply only when the moduli have prime factors of unbalanced size, whereas in our case, both prime factors have 512 bits.

7 Hardware Details, Disclosure, and Response

Around 2006–2007, MOICA switched card platforms from their initial supplier and began to use Chunghwa Telecom’s HiCOS PKI smart cards, specifically Chunghwa Telecom HD65145C1 cards (see [5]), using the Renesas AE45C1 smart card microcontroller (see [10]). We have confirmed these details with MOICA.

Unfortunately, the hardware random-number generator on the AE45C1 smart card microcontroller sometimes fails, as demonstrated by our results. These failures are so extreme that they should have been caught by standard health tests, and in fact the AE45C1 does offer such tests. However, as our results show, those tests were not enabled on some cards. This has now also been confirmed by MOICA. MOICA’s estimate is that about 10000 cards were issued without these tests, and that subsequent cards used a “FIPS mode” (see below) that enabled these tests.

The random numbers generated by the batch of problematic cards obviously do not meet even minimal standards for collecting and processing entropy. This is a fatal flaw, and it can be expected to continue causing problems until all of the vulnerable cards are replaced.

The AE45C1 chip was certified conformant with Protection Profile BSI-PP-0002-2001 at CC assurance level EAL4+ [10]. The HD65145C1 card and HiCOS operating system were accredited to FIPS 140-2 Level 2 [5]. The CC certification stated “The TOE software for random number postprocessing shall be implemented by the embedded software developer”, and the FIPS certification was limited to “FIPS mode” (see <http://www.cryptsoft.com/fips140/out/cert/614.html>). However, neither certification prevented the same card from also offering a non-FIPS mode, and neither certification caught the underlying RNG failures. We recommend that industry move to stronger certifications that prohibit error-prone APIs and that include assessments of RNG quality.

In April 2012 we shared with MOICA our preliminary list of 103 certificates compromised by GCD. We announced these results in a talk in Taiwan in July 2012. We provided an extended list of compromised certificates to MOICA and Chunghwa Telecom in June 2013, along with an early draft of this paper. MOICA and Chunghwa Telecom subsequently confirmed our results; asked the cardholders to come in for replacement cards; revoked the compromised certificates; and initiated the task of contacting 408 registration offices across Taiwan to manually trace and replace all of the vulnerable cards from the same batch.

Acknowledgements. We thank J. Alex Halderman and Mark Wooding for discussion during this project. We also thank Kenny Paterson and the anonymous reviewers for helpful comments and suggestions, and in particular the encouragement to experiment with sparse key recovery methods.

References

1. ANSI. ANSI X9.31:1998: Digital Signatures Using Reversible Public Key Cryptography for the Financial Services Industry (rDSA). American National Standards Institute (1998)
2. Bernstein, D.J.: How to find the smooth parts of integers (May 2004), <http://cr.yp.to/papers.html#smoothparts>
3. Boneh, D., Durfee, G.: Cryptanalysis of RSA with private key d less than $n^{0.292}$. In: Stern, J. (ed.) EUROCRYPT. LNCS, vol. 1592, pp. 1–11. Springer (1999)
4. Cadé, D., Pujol, X., Stehlé, D.: fpLLL (2013), <http://perso.ens-lyon.fr/damien.stehle/fp111/>
5. Ltd. Chunghwa Telecom Co. Hicos pki smart card security policy (2006), <http://www.cryptsoft.com/fips140/vendors/140sp614.pdf>
6. Coppersmith, D.: Finding a small root of a bivariate integer equation; factoring with high bits known. In: Maurer, U.M. (ed.) EUROCRYPT. LNCS, vol. 1070, pp. 178–189. Springer (1996)
7. Coppersmith, D.: Small solutions to polynomial equations, and low exponent RSA vulnerabilities. *J. Cryptology* 10(4), 233–260 (1997)
8. Decker, W., Greuel, G.-M., Pfister, G., Schönemann, H.: SINGULAR 3-1-6 — A computer algebra system for polynomial computations (2012), <http://www.singular.uni-kl.de>
9. Faugère, J.-C., Marinier, R., Renault, G.: Implicit factoring with shared most significant and middle bits. In: Nguyen, P.Q., Pointcheval, D. (eds.) Public Key Cryptography. LNCS, vol. 6056, pp. 70–87. Springer (2010)
10. Bundesamt für Sicherheit in der Informationstechnik. Certification report BSI-DSZ-CC-0212-2004 for Renesas AE45C1 (HD65145C1) smartcard integrated circuit version 01 (2004), https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Zertifizierung/Reporte02/0212a_pdf.pdf?__blob=publicationFile
11. Bundesamt für Sicherheit in der Informationstechnik. Evaluation of random number generators (2013), https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Zertifizierung/Interpretation/Evaluation_of_random_number_generators.pdf?__blob=publicationFile and https://www.bsi.bund.de/DE/Themen/ZertifizierungundAnerkennung/ZertifizierungnachCCundITSEC/AnwendungshinweiseundInterpretationen/AISCC/ais_cc.html
12. Granville, A.: Harald Cramér and the distribution of prime numbers. *Scand. Actuarial J.* 1995(1), 12–28 (1995)
13. Heninger, N., Durumeric, Z., Wustrow, E., Alex Halderman, J.: Mining your Ps and Qs: Detection of widespread weak keys in network devices. In: Proceedings of the 21st USENIX Security Symposium (August 2012)
14. Heninger, N., Shacham, H.: Reconstructing rsa private keys from random key bits. In: Halevi, S. (ed.) CRYPTO. LNCS, vol. 5677, pp. 1–17. Springer (2009)
15. Herrmann, M., May, A.: Solving linear equations modulo divisors: On factoring given any bits. In: Pieprzyk, J. (ed.) ASIACRYPT. LNCS, vol. 5350, pp. 406–424. Springer (2008)
16. Howgrave-Graham, N.: Approximate integer common divisors. In: Silverman, J.H. (ed.) CaLC. LNCS, vol. 2146, pp. 51–66. Springer (2001)

