Cryptography

Jim Carlson

Science Day
November 17, 2001

The Problem: For Alice to send a message to Bob that only Bob can read. They have to worry about Eve, who is a snoop.

- 450 BC – 1977: Secret codes depend on secrecy of the keyword

- 1977 – ??: the RSA code (Rivest-Shamir-Adelman): **unbreakable even if the key is known**.

Public key codes (like RSA):

– *Essential* for internet commmerce.

– Depend on some beautiful mathematics ... ... $a^{p-1} \equiv 1 \ mod \ p$ ....

Mathematics of RSA.

- It is easy to find large prime numbers.
- It is hard to factor large integers into primes.
- Number Theory: Fermat, Euler, ... , Lenstra

Number theory is applied math!*

* Despite what G.H. Hardy (1877–1947) said.

A simple code: (Julius Caesar).

Plain text: attack at dawn

Cipher text: CVVCKMCVFCYP

Key = C: shift right by two letters

Problem: decipher "XJSIRTSJD."

Cipher text: XJSIRTSJD .... *scrambled*

Stategy: try different keys

B $\Rightarrow$ WIRHQSRIC

C $\Rightarrow$ VHQGPRQHB

D $\Rightarrow$ UGPFOQPGA

E $\Rightarrow$ TUOENPOFZ

F $\Rightarrow$ SENDMONEY .... *unscrambled*

Bingo!

The code is **weak** because the set of keys is **small** — only 25.

An improved code (more keys).

Example A. Key = JARGON

Plain text: attack at dawn

```
   JARGON = 9   0  17  6  14  13
+ attack = 0  19  19  0   2  10
-------------------------------
            9  19  36  6  16  23
(mod 26)    9  19  10  6  16  23
-------------------------------
            J   T   K  G   Q   X
```

```
a  b  c  d  e  f  g  h  i  j  k  l  m
0  1  2  3  4  5  6  7  8  9 10 11 12

n  o  p  q  r  s  t  u  v  w  x  y  z
13 14 15 16 17 18 19 20 21 22 23 24 25
```

Comments on the improved code.

Strengths:

- a is encoded both as J and G, ...

- The "key space" (all six letter words) is LARGE: $26^6 > 3$ billion (9 years at one per second).

Weaknesses:

- Attack by FREQUENCY ANALYSIS (Al-Kindi, 850)

- Problem of KEY EXCHANGE

INSERT AL-KINDI SLIDE

Towards RSA:

- Strings of letters $\Rightarrow$ blocks of numbers:

  PQR $\Rightarrow$ 15, 16, 17 $\Rightarrow$ 151617 ...

- Encryption and decryption rules $\Rightarrow$ mathematical formulas based on modular arithmetic ...

Modular arithmetic is clock arithmetic.

$9 + 5 \equiv 2 \ (mod\ 12)$

$3 \times 5 \equiv 3 \ (mod\ 12)$

More modular arithmetic

Easy: compute, divide by 26, and take the remainder ...

$$15 + 19 \equiv 8 \ (mod\ 26) \quad (1)$$
$$4 - 7 \equiv 23 \ (mod\ 26) \quad (2)$$
$$4 \times 7 \equiv 2 \ (mod\ 26) \quad (3)$$
$$2^5 \equiv 6 \ (mod\ 26) \quad (4)$$

More difficult:

$$2^{-1} \equiv ??\ (mod\ 26) \quad (5)$$
$$7^{-1} \equiv ??\ (mod\ 26) \quad (6)$$

$2x \equiv 1 \ (mod\ 26)$ — NO SUCH x

$7x \equiv 1 \ (mod\ 26)$ — 15 works.

So 2^{-1} doesn't exist, but $7^{-1} = 15$. Why? 2 has a factor in common with 26, but 7 doesn't.

Encryption and decryption

For each key K, a pair of functions, one to scramble, the other to unscramble:

- $e(x)$ encrypts the plaintext x
- $d(y)$ decrypts the ciphertext y

Require: d is the inverse function of e:

$$d(e(x)) = x$$

Some families of functions:

Example 1.

- $e(x) \equiv x + K \ mod \ N$
- $d(y) \equiv y - K \ mod \ N$

Example 2.

- $e(x) \equiv Kx \ mod \ N$
- $d(y) \equiv Ly \ mod \ N$,

where $KL \equiv 1 \ mod \ N$

Example 3: RSA!

- $e(x) \equiv x^K \ mod \ N$

- $d(y) \equiv y^L \ mod \ N$,

where $x^{KL} \equiv x \ mod \ N$ for all x that have no factors in common with N.

Caution: We must choose N, K, and L with great care ...

RSA helps Alice and Bob to exchange keys

1. Alice generates two large prime numbers, p and q and multiplies them together to get $N = pq$.

2. She carefully chooses a special number K and defines $e(x) = x^K \ (mod\ N)$.

3. She does some math to find a number L so that $x^{KL} \equiv x \ (mod\ N)$ for all x relatively prime to p and q.

4. She defines $d(x) = x^L \ (mod\ N)$.

5. She tells Bob about K and N. Bob will use these numbers to send secret messages to Alice. Alice will unscramble them using L and N.

Eve appears on the scene ...

Bob sends a message to Alice using K and N. Alice reads it using L and N.

Unfortunately, Eve intercepts the message, and had previously intercepted K and N using a sniffer attached to Bob's ISP.

Eve also knows the mathematics of RSA, and she is a whiz at computing, so she tries to find L. She sets her computer running, ...

... years have passed, and Bob no longer cares if Eve decodes his message, nor does Eve care. Still, her computer continues to chug away. In the lonely silence of her study, it tries increasinlgy large factors of N ... *sniff* ...

Why can't Eve break Alice's code?

1. Alice found L by solving the congruence

   $$KL \equiv 1 \ mod \ (p-1)(q-1)$$

   She needs p and q for this.

2. Eve knows K, and she knows N, but she doesn't know the factorization $N = pq$. So she doesn't know which congruence to solve.

3. It is "easy" to manufacture large prime numbers, but it is "hard" to factor large integers into primes.

Drats!!

Appendix:The Mathematical Core of RSA

... the key idea is due to Fermat (ca. 1650)

Theorem:

$$x^{p-1} \equiv 1 \ mod \ p$$

if p does not divide x.

Example

$$3^{1008} \equiv 1 \ mod \ 1009$$

Generalizations

$$x^{(p-1)(q-1)} \equiv 1 \ mod \ pq$$

if p and q do not divide x.

A computation: $7^{100}\ mod\ 1009$

Since $100 = 64 + 32 + 4$,

$$7^{100} = 7^{64} \times 7^{32} \times 7^{4}$$

Table of numbers 7^{2^n}:
$7^2 = 49$
$7^4 = 49^2 = 2401 \equiv 383$
$7^8 = 383^2 = 146689 \equiv 384$, etc.

Then

$$7^{100} \equiv 256 \times 993 \times 383 = 973,616,664 \equiv 227$$

These computations are *fast*.

Appendix: The Factoring Problem

How hard is it?

Martin Gardner's Challenge

A message encoded with RSA using a 129 digit number:

$N = 114,381,625,757,888,867,669,235,779,976,$
$146,612,010,218,296,721,242,362,562,561,842,$
$935,706,935,245,733,897,830,597,123,563,958,$
$705,058,989,075,147,599,290,026,879,543,541$

Challenge announced August 1977 in the *Scientific American.*

$\Rightarrow$ 17 years $\Rightarrow$

Challenge solved April 26, 1994 by a team of 600 volunteers ... 5000 MIPS-years.

Plaintext = "the magic words are squeamish ossifrage".

“Latest” RSA challenge

155-digit number factored, August 22, 1999

35.7 CPU-years on

160 SGI and Sun workstations (175-400 MHz)

8 SGI Origin 2000 processors (250 MHz)

120 Pentium II PC’s (300-450 MHz)

4 Digital/Compaq boxes (500 MHz)

7.4 calendar months

www.rsasecurity.com/rsalabs/challenges/factoring/rsa155.html

Another computation:

Factor Martin Gardner's N

Rough estimate of time needed — simplest method for factoring (trial division).

Divide by 2, 3, 4, 5, 6, ..., $[\sqrt{N}]$ to find the factors.

— One million divisions per second (10^6).

— $N \sim 10^{129}$, so $\sqrt{N} \sim 10^{64}$ divisions.

— 10^{58} seconds.

— 3×10^7 seconds in a year.

These computations are slow: 10^{50} years!

How to crack the code? — $N \sim 10^{300}$

Need a breakthrough:

— Technical (machines)

— Theoretical (mathematics)

Unsolved problem: Is factoring "hard?"

— Evidence: 3,000 years of experience

— Need: *lower* bounds on how hard it is to factor.

Complexity.

Good algorithms

Running time $\sim (\log N)^k$.

Bad algorithms

Running time $\sim e^{(\log N)^k}$.

Good: Finding primes, computing powers mod N, solving congruences

Bad (as far as we know): Factoring

Complexity of factoring algorithms

Trial division $= \sqrt{N} = e^{0.5 \log N}$

Quadratic sieve $\sim e^{(1+o(1))(\log n \log\log n)^{1/2}}$

Elliptic curve $\sim e^{(1+o(1))(2 \log p \log\log p)^{1/2}}$

Number field sieve $\sim e^{(1.92+o(1))(\log n)^{1/3}(\log\log n)^{2/3}}$

$N = 10^{129}$
— Trial division: 10^{64}
— Number field sieve: 10^{17}

$N = 10^{300}$
— Trial division: 10^{150}
— Number field sieve: 10^{25}

CRYPTOGRAPHY PROBLEMS

The material in these notes, plus that in [Singh] and [Davis] should be enough to decrypt the following messages.

1. Decode: KYRFGQDSL

The remaining problems will be harder.

2. Decode:

LEDFKPKLEX DP VODQQHJ DJ QEDP MOZJC AKKB, QEH TJDUHOPH, VEDRE PQZJCP RKJQDJTZFFX KLHJ QK KTO MZYH. ATQ QEH AKKB RZJJKQ AH TJCHOPQKKC TJFHPP KJH IDOPQ FHZOJP QK RKGLOHEHJC QEH FZJMTZMH ZJC OHZC QEH FHQQHOP DJ VEDRE DQ DP RKGLKPHC. DQ DP VODQQHJ DJ QEH FZJMTZMH KI GZQEHGZQDRP, ZJC DQP REZOZRQHOP ZOH QODZJMFHP, RDORFHP, ZJC KQEHO MHKGHQODR IDM-TOHP VDQEKTQ VEDRE DQ DP ETGZJFX DGLKPPDAFH QK TJCHOPQZJC Z PDJMFH VKOC KI DQ.

— MZFDFHK

3. Decode:

HFSGLQUIE PUB UVTTG MKRRH HEQ

Vigenere, keyword CRYPTOGRAPHY. From Davis, p. 8

4. Decode:

23, 52, 85, 91, 15, 06, 53, 61, 30, 72, 23

"Numerical Vigenere" — pseudorandom sequence based on a seed (the "keyword"). See section 6 of Davis. Use the character encoding on page 9.

5. Decode:

14756

RSA with N = 16781, e = 5. The result is a nice four-digit number. See [Davis], section 9.

References:

Childs, Lindsay: A Concrete Introduction to Higher Algebra, 2nd edition (Springer, 1995)

Koblitz, Neal: A Course in Number Theory and Cryptography (Springer, 1994)

Silverman, Joseph: A Friendly Introduction to Number Theory (Prentice Hall, 1996).

Singh, Simon: The Code Book (Anchor Books, 1999).

Stinson, Douglas R.: Cryptography, Theory and Practice (CRC 1995).

References on the Web:

Tom Davis' article:

http://mathcircle.berkeley.edu/BMC3/crypto.pdf

www.rsasecurity.com

www.rsasecurity.com/rsalabs

www.4thestate.co.uk/cipherchallenge

This article can be found at:

www.math.utah.edu/~carlson/ugc/crypt/