Factoring integers, Producing primes and the RSA cryptosystem

KATHMANDU UNIVERSITY

December 14, 2005

RSA_{2048} = 2519590847565789349402718324004839857142928212620403202777713783604366202070759555626401852588078440691829064124951508218929855914917618450280848912007284499268739280728777673597141834727026189637501497182469116507761337985909570009733045974880842840179742910064245869181719511874612151517265463228221686998754918242243363725908514186546204357679842338718477444792073993423658482382428119816381501067481045166037730605620161967625613384414360383390441495263443219011465754445417842402092461651572335077870774981712577246796292638635637328991215483143816789988504044536402352738195137863656439121201039712282212072035 7

RSA_{2048} is a 617 (decimal) digit number

http://www.rsasecurity.com/rsalabs/node.asp?id=2093

$$RSA_{2048}=p \cdot q, \quad p,q \approx 10^{308}$$

PROBLEM: *Compute p and q*

PRICE: 200.000 US$ ($\sim 13,948,300.17$ NPR)!!

Theorem. If $a \in \mathbb{N}$ $\exists!\ p_1 < p_2 < \cdots < p_k$ *primes* s.t. $a = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$

Regrettably: RSAlabs believes that factoring in one year requires:

number	computers	memory
RSA_{1620}	1.6×10^{15}	120 Tb
RSA_{1024}	$342,000,000$	170 Gb
RSA_{760}	215,000	4Gb.

http://www.rsasecurity.com/rsalabs/node.asp?id=2093

Challenge Number	Prize ($US)
RSA_{576}	$10,000
RSA_{640}	$20,000
RSA_{704}	$30,000
RSA_{768}	$50,000
RSA_{896}	$75,000
RSA_{1024}	$100,000
RSA_{1536}	$150,000
RSA_{2048}	$200,000

http://www.rsasecurity.com/rsalabs/node.asp?id=2093

Challenge Number	Prize ($US)	Status
RSA_{576}	$10,000	Factored December 2003
RSA_{640}	$20,000	Not Factored
RSA_{704}	$30,000	Not Factored
RSA_{768}	$50,000	Not Factored
RSA_{896}	$75,000	Not Factored
RSA_{1024}	$100,000	Not Factored
RSA_{1536}	$150,000	Not Factored
RSA_{2048}	$200,000	Not Factored

History of the "Art of Factoring"

- 220 BC Greeks (Eratosthenes of Cyrene)
- 1730 Euler $2^{2^5}+1=641\cdot 6700417$
- 1750–1800 Fermat, Gauss (Sieves - Tables)
- 1880 Landry & Le Lasseur: $2^{2^6}+1=274177\times 67280421310721$
- 1919 Pierre and Eugène Carissan (Factoring Machine)
- 1970 Morrison & Brillhart
  $2^{2^7}+1=59649589127497217\times 5704689200685129054721$
- 1980, Richard Brent and John Pollard $2^{2^8}+1=1238926361552897\times 93461639715357977769163558199606896584051237541638188580280321$
- 1982 Quadratic Sieve **QS** (Pomerance) $\rightsquigarrow$ Number Fields Sieve **NFS**
- 1987 Elliptic curves factoring **ECF** (Lenstra)

Carissan's ancient Factoring Machine

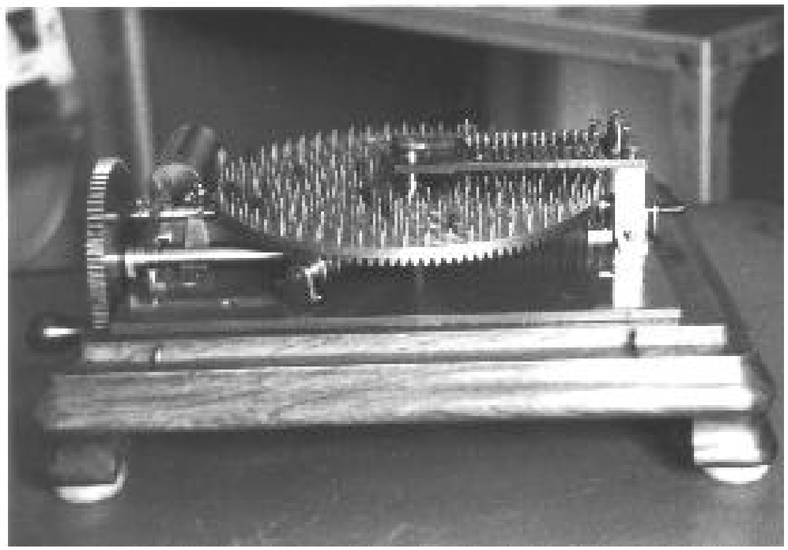


Figure 1: Conservatoire Nationale des Arts et Métiers in Paris

`http://www.math.uwaterloo.ca/ shallit/Papers/carissan.html`

Figure 2: Lieutenant Eugène Carissan

$225058681 = 229 \times 982789$	2 minutes
$3450315521 = 1409 \times 2418769$	3 minutes
$3570537526921 = 841249 \times 4244329$	18 minutes

Contemporary Factoring 1/2

1. 1994, Quadratic Sieve (QS): (8 months, 600 voluntaries, 20 countries)
   D.Atkins, M. Graff, A. Lenstra, P. Leyland

$$RSA_{129} = 114381625757888867669235779976146612010218296721242362562561842935706$$
$$935245733897830597123563958705058989075147599290026879543541 =$$
$$= 3490529510847650949147849619903898133417764638493387843990820577 \times$$
$$32769132993266709549961988190834461413177642967992942539798288533$$

2. (February 2 1999), Number Fields Sieve (NFS): (160 Sun, 4 months)

$$RSA_{155} = 109417386415705274218097073220403576120037329454492059909138421314763499842$$
$$88934784717997257891267332497625752899781833797076537244027146743531593354333897 =$$
$$= 102639592829741105772054196573991675900716567808038066803341933521790711307779 \times$$
$$106603488380168454820927220360012878679207958575989291522270608237193062808643$$

3. (December 3, 2003) (NFS): J. Franke et al. (174 decimal digits)

$$RSA_{576} = 188198812920607963838697239461650439807163563379417382700763356422988859715234$$
$$6654853190606065047430453173880113033967161996923212057340318795506569962213051687593076 50257059 =$$
$$= 398075086424064937397125500550386491199064362342526708406385189575946388957261768583317 \times$$
$$472772146107435302536223071973048224632914695302097116459852171130520711256363590397527$$

4. (May 9,2005) (NFS): F. Bahr, et al (663 binary digits)

$$RSA_{200} = 27997833911221327870829467638722601621070446786955428537560009929326128400107609345671052955360856061822351910951365788637105954482006576775098580557613579098734950144178863178946295187237869221823983 =$$
$$3532461934402770121272604978198464368671197400197625023649303468776121253679423200058547956528088349 \times$$
$$7925869954478333033347085841480059687737975857364219960734330341455767872818152135381409304740185467$$

Contemporary Factoring 2/2

Elliptic curves factoring (ECM) H. Lenstra (1985) - small factors (50 digits)

- ❻ (1993) A. Lenstra, H. Lenstra, Jr., M. Manasse, and J. Pollard $2^{2^9}+1 = 2424833 \times 7455602825647884208337395736200454918783366342657 \times p99$
- ❻ (April 6, 2005) (ECM) B. Dodson $3^{466}+1$ is divisible by 709601635082267320966424084955776789770864725643996885415676682297;
- ❼ (Sept. 5, 2005) (ECM) K. Aoki & T. Shimoyama $10^{311}-1$ is divisible by 4344673058714954477761314793437392900672885445361103905548950933

For updates see Paul Zimmerman's "*Integer Factoring Records*":

`http://www.loria.fr/ zimmerma/records/factor.html`

More infoes about fatroring in

`http://www.crypto-world.com/FactorWorld.html`

Update on "factorization of Fermat Numbers":

`http://www.prothsearch.net/fermat.html`

Last Minute News

Date: Thu, 10 Nov 2005 22:07:26 -0500
From: Jens Franke <franke@math.uni-bonn.de>
To: NMBRTHRY@LISTSERV.NODAK.EDU

We have factored RSA640 by GNFS. The factors are

1634733645809253848443133883865090859841783670033092312181110852389333100104508151212118167511579

and

1900871281664822113126851573935413975471896789968515493666638539088027103802104498957191261465571

We did lattice sieving for most special q between 28e7 and 77e7 using factor base bounds of 28e7 on the algebraic side and 15e7 on the rational side. The bounds for large primes were 2^ 34. This produced 166e7 relations. After removing duplicates 143e7 relations remained. A filter job produced a matrix with 36e6 rows and columns, having 74e8 non-zero entries. This was solved by Block-Lanczos.

Sieving has been done on 80 2.2 GHz Opteron CPUs and took 3 months. The matrix step was performed on a cluster of 80 2.2 GHz Opterons connected via a Gigabit network and took about 1.5 months.

Calendar time for the factorization (without polynomial selection) was 5 months.

More details will be given later.

F. Bahr, M. Boehm, J. Franke, T. Kleinjung

RSA

Adi Shamir, Ron L. Rivest, Leonard Adleman (1978)

The RSA cryptosystem

1978 R. L. Rivest, A. Shamir, L. Adleman (Patent expired in 1998)

Problem: Alice wants to send the message $\mathcal{P}$ to Bob so that Charles cannot read it

$$\begin{array}{ccc} A\ (Alice) & \longrightarrow & B\ (Bob) \\ & \uparrow & \\ & C\ (Charles) & \end{array}$$

1. KEY GENERATION — Bob has to do it
2. ENCRYPTION — Alice has to do it
3. DECRYPTION — Bob has to do it
4. ATTACK — Charles would like to do it

Bob: Key generation

- He chooses randomly p and q primes $(p, q \approx 10^{100})$
- He computes $M = p \times q$, $\varphi(M) = (p-1) \times (q-1)$
- He chooses an integer e s.t.

$$0 \leq e \leq \varphi(M) \quad \text{and} \quad \gcd(e, \varphi(M)) = 1$$

NOTE. One could take $e = 3$ and $p \equiv q \equiv 2 \bmod 3$

Experts recommend $e = 2^{16} + 1$

- He computes arithmetic inverse d of e modulo $\varphi(M)$

  (i.e. $d \in \mathbb{N}$ (unique $\leq \varphi(M)$) s.t. $e \times d \equiv 1 \pmod{\varphi(M)}$)
- Publishes (M, e) **public key** and hides **secret key** d

Problem: How does Bob do all this?- We will go came back to it!

Alice: Encryption

Represent the message $\mathcal{P}$ as an element of $\mathbb{Z}/M\mathbb{Z}$

(for example) $A \leftrightarrow 1 \quad B \leftrightarrow 2 \quad C \leftrightarrow 3 \quad \dots \quad Z \leftrightarrow 26 \quad AA \leftrightarrow 27 \dots$

$$\texttt{NEPAL} \leftrightarrow 14 \cdot 26^4 + 5 \cdot 26^3 + 16 \cdot 26^2 + 26 + 12 = 6496398$$

Note. Better if texts are not too short. Otherwise one performs some *padding*

$$\mathcal{C} = E(\mathcal{P}) = \mathcal{P}^e \pmod M$$

Example: $p = 9049465727$, $q = 8789181607$, $M = 79537397720925283289$, $e = 2^{16} + 1 = 65537$, $\mathcal{P} = \texttt{NEPAL}$:

$$E(\texttt{NEPAL}) = 6496398^{65537} \pmod{79537397720925283289}$$

$$= 68059003759328352940 = \mathcal{C} = \texttt{ZKUFANERFPXDKAA}$$

Bob: Decryption

$$\mathcal{P} = D(\mathcal{C}) = \mathcal{C}^d \pmod{M}$$

Note. Bob decrypts because he is the only one that knows d.

Theorem. (Euler) If $a, m \in \mathbb{N}$, $\gcd(a, m) = 1$,

$$a^{\varphi(m)} \equiv 1 \pmod{m}.$$

If $n_1 \equiv n_2 \bmod \varphi(m)$ then $a^{n_1} \equiv a^{n_2} \bmod m$.

Therefore $(ed \equiv 1 \bmod \varphi(M))$

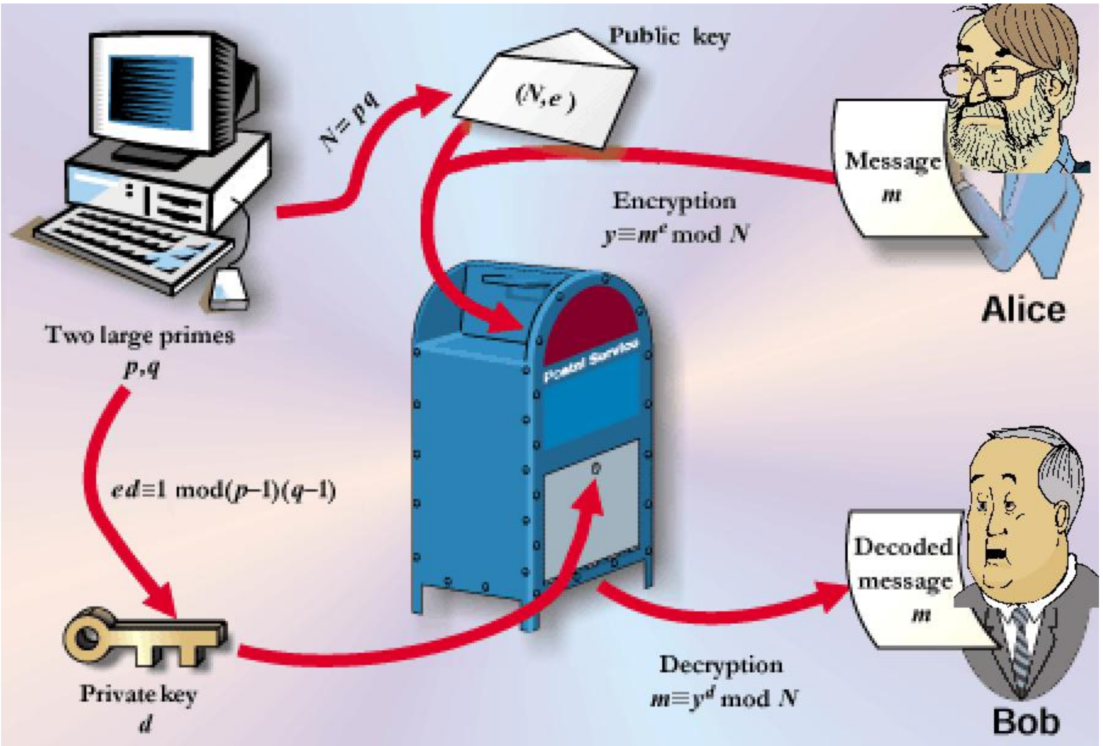
$$D(E(\mathcal{P})) = \mathcal{P}^{ed} \equiv \mathcal{P} \bmod M$$

Example(cont.):$d = 65537^{-1} \bmod \varphi(9049465727 \cdot 8789181607) = 5717391406064378015$3

$D($`ZKUFANERFPXDKAA`$) =$

$680590037593283529 40^{5717391406064378015 3} (\mathrm{mod}\, 79537397720925283289) =$ `NEPAL`

RSA at work


Public key
(N,e)
N=pq
Message
m
Encryption
y≡m^e mod N
Alice
Two large primes
p,q
ed≡1 mod(p−1)(q−1)
Decoded
message
m
Private key
d
Decryption
m≡y^d mod N
Bob

Repeated squaring algorithm

Problem: How does one compute $a^b \bmod c$?

$$68059003759328352940^{57173914060643780153} \pmod{79537397720925283289}$$

- Compute the binary expansion $b = \sum_{j=0}^{[\log_2 b]} \epsilon_j 2^j$

  $57173914060643780153 = 110001100101110010100010111110101011110011011000100100011000111001$

- Compute recursively $a^{2^j} \bmod c$, $j = 1, \dots, [\log_2 b]$:

$$a^{2^j} \bmod c = \left(a^{2^{j-1}} \bmod c\right)^2 \bmod c$$

- Multiply the $a^{2^j} \bmod c$ with $\epsilon_j = 1$

$$a^b \bmod c = \left(\prod_{j=0, \epsilon_j=1}^{[\log_2 b]} a^{2^j} \bmod c\right) \bmod c$$

$$\#\{\textbf{oper. in } \mathbb{Z}/c\mathbb{Z} \textbf{ to compute } a^b \bmod c\} \leq 2\log_2 b$$

`ZKUFANERFPXDKAA` is decrypted with 131 operations in

$$\mathbb{Z}/795373977209252832 89\mathbb{Z}$$

PSEUDO CODE: $e_c(a,b) = a^b \bmod c$

$e_c(a,b)$	=	if	$b=1$	then	$a \bmod c$
		if	$2\|b$	then	$e_c(a,\frac{b}{2})^2 \bmod c$
		else			$a * e_c(a,\frac{b-1}{2})^2 \bmod c$

To encrypt with $e = 2^{16}+1$, only 17 operations in $\mathbb{Z}/M\mathbb{Z}$ are enough

Key generation

Problem. Produce a random prime $p \approx 10^{100}$

Probabilistic algorithm (type Las Vegas)	
1.	Let $p = \textsc{Random}(10^{100})$
2.	If ISPRIME(p)=1 then OUTPUT=p else goto 1

subproblems:

A. How many iterations are necessary?

(i.e. how are primes distributes?)

B. How does one check if p is prime?

(i.e. how does one compute ISPRIME(p)?) $\rightsquigarrow$ Primality test

False Metropolitan Legend: Check primality is equivalent to factoring

A. Distribution of prime numbers

$$\pi(x) = \#\{p \leq x \text{ t. c. } p \text{ is prime}\}$$

Theorem. (Hadamard - de la vallee Pussen - 1897)

$$\pi(x) \sim \frac{x}{\log x}$$

Quantitative version:

Theorem. (Rosser - Schoenfeld) if $x \geq 67$

$$\frac{x}{\log x - 1/2} < \pi(x) < \frac{x}{\log x - 3/2}$$

Therefore

$$0.0043523959267 < Prob\left((\textsc{Random}(10^{100}) = \texttt{prime}\right) < 0.004371422086$$

If P_k is the probability that among k random numbers$\leq 10^{100}$ there is a prime one, then

$$P_k = 1 - \left(1 - \frac{\pi(10^{100})}{10^{100}}\right)^k$$

Therefore

$$0.663942 < P_{250} < 0.66554440$$

To speed up the process: One can consider only odd random numbers not divisible by 3 nor by 5.

Let

$$\Psi(x, 30) = \#\left\{n \leq x \text{ s.t. } \gcd(n, 30) = 1\right\}$$

To speed up the process: One can consider only odd random numbers not divisible by 3 nor by 5.

Let

$$\Psi(x,30) = \#\{n \leq x \text{ s.t.} \gcd(n,30) = 1\}$$

then

$$\frac{4}{15}x - 4 < \Psi(x,30) < \frac{4}{15}x + 4$$

Hence, if P'_k is the probability that among k random numbers $\leq 10^{100}$ coprime with 30, there is a prime one, then

$$P'_k = 1 - \left(1 - \frac{\pi(10^{100})}{\Psi(10^{100},30)}\right)^k$$

and

$$0.98365832 < P'_{250} < 0.98395199$$

B. Primality test

Fermat Little Theorem. If p is prime, $p \nmid a \in \mathbb{N}$

$$a^{p-1} \equiv 1 \bmod p$$

NON-primality test

$$M \in \mathbb{Z}, \quad 2^{M-1} \not\equiv 1 \bmod M \Longrightarrow \quad M \text{composite!}$$

EXAMPLE: $2^{RSA_{2048}-1} \not\equiv 1 \bmod RSA_{2048}$

Therefore RSA_{2048} is composite!

Fermat little Theorem does not invert. Infact

$$2^{93960} \equiv 1 \pmod{93961} \quad \text{but} \quad 93961 = 7 \times 31 \times 433$$

Strong pseudo primes

From now on $m \equiv 3 \bmod 4$ (just to simplify the notation)

Definition. $m \in \mathbb{N}$, $m \equiv 3 \bmod 4$, composite is said strong pseudo prime (SPSP) in base a if

$$a^{(m-1)/2} \equiv \pm 1 \pmod{m}.$$

Note. If $p > 2$ prime $\Longrightarrow$ $a^{(p-1)/2} \equiv \pm 1 \pmod{p}$

Let $\boxed{\mathcal{S} = \{a \in \mathbb{Z}/m\mathbb{Z} \text{ s.t. } \gcd(m, a) = 1, a^{(m-1)/2} \equiv \pm 1 \pmod{m}\}}$

① $\mathcal{S} \subseteq (\mathbb{Z}/m\mathbb{Z})^*$ subgroup

② If m is composite $\Longrightarrow$ proper subgroup

③ If m is composite $\Longrightarrow$ $\#\mathcal{S} \leq \frac{\varphi(m)}{4}$

④ If m is composite $\Longrightarrow$ $Prob(m \text{ SPSP in base } a) \leq 0,25$

Miller–Rabin primality test

Let $m \equiv 3 \bmod 4$

MILLER RABIN ALGORITHM WITH k ITERATIONS
$N = (m-1)/2$
`for` $j = 0$ `to` k `do` $a =$`Random`(m)
`if` $a^N \not\equiv \pm 1 \bmod m$ `then OUPUT=(`m `composite): END`
`endfor OUTPUT=(`m `prime)`

Monte Carlo primality test

$Prob$(Miller Rabin says m `prime` and m is composite) $\lesssim \frac{1}{4^k}$

In the real world, software uses Miller Rabin with $k = 10$

Deterministic primality tests

Theorem. (Miller, Bach) If m is composite, then

$$\textbf{GRH} \Rightarrow \ \exists a \leq 2\log^2 m \text{ s.t. } \ a^{(m-1)/2} \not\equiv \pm 1 \pmod m.$$

(i.e. *m is not SPSP in base a.*)

Consequence: "Miller–Rabin de–randomizes on GRH" ($m \equiv 3 \bmod 4$)

for	$a = 2$ to $2\log^2 m$	do
	if $a^{(m-1)/2} \not\equiv \pm 1 \bmod m$	then
		OUPUT=(m composite): END
endfor		OUTPUT=(m prime)

Deterministic Polynomial time algorithm

It runs in $O(\log^5 m)$ operations in $\mathbb{Z}/m\mathbb{Z}$.

Certified prime records

Top 10 Largest primes:

1	$2^{25964951} - 1$	7816230	Nowak	2005	Mersenne	42?
2	$2^{24036583} - 1$	7235733	Findley	2004	Mersenne	41?
3	$2^{20996011} - 1$	6320430	Shafer	2003	Mersenne	40?
4	$2^{13466917} - 1$	4053946	Cameron	2001	Mersenne	39
5	$27653 \times 2^{9167433} + 1$	2759677	Gordon	2005		
6	$28433 \times 2^{7830457} + 1$	2357207	SB7	2004		
7	$2^{6972593} - 1$	2098960	Hajratwala	1999	Mersenne	38
8	$5359 \times 2^{5054502} + 1$	1521561	Sundquist	2003		
9	$4847 \times 2^{3321063} + 1$	999744	Hassler	2005		
10	$2^{3021377} - 1$	909526	Clarkson	1998	Mersenne	37

- Mersenne's Numbers: $M_p = 2^p - 1$
- For more see

`http://primes.utm.edu/primes/`

The AKS deterministic primality test

Department of Computer Science & Engineering,
I.I.T. Kanpur, Agost 8, 2002.

Nitin Saxena, Neeraj Kayal and Manindra Agarwal

New deterministic, polynomial–time, primality test.

Solves #1 open question in computational number theory

http://www.cse.iitk.ac.in/news/primality.html

How does the AKS work?

Theorem. (AKS) Let $n \in \mathbb{N}$. Assume q, r primes, $S \subseteq \mathbb{N}$ finite:

- $q | r - 1$;
- $n^{(r-1)/q} \bmod r \notin \{0, 1\}$;
- $\gcd(n, b - b') = 1, \quad \forall b, b' \in S$ (distinct);
- $\binom{q+\#S-1}{\#S} \geq n^{2\lfloor\sqrt{r}\rfloor}$;
- $(x + b)^n = x^n + b$ in $\mathbb{Z}/n\mathbb{Z}[x]/(x^r - 1), \quad \forall b \in S$;

Then n is a power of a prime

Bernstein formulation

Fouvry Theorem (1985) $\Longrightarrow$ $\exists r \approx \log^6 n, s \approx \log^4 n$

$\Longrightarrow$ AKS runs in $O(\log^{17} n)$ operations in $\mathbb{Z}/n\mathbb{Z}$.

Many simplifications and improvements: Bernstein, Lenstra, Pomerance.....

Why is RSA safe?

- It is clear that if Charles can factor M,
  then he can also compute $\varphi(M)$ and then also d so to decrypt messages
- Computing $\varphi(M)$ is equivalent to completely factor M. In fact

$$p, q = \frac{M - \varphi(M) + 1 \pm \sqrt{(M - \varphi(M) + 1)^2 - 4M}}{2}$$

- **RSA Hypothesis.** The only way to compute efficiently

  $$x^{1/e} \bmod M, \quad \forall x \in \mathbb{Z}/M\mathbb{Z}$$

  (i.e. decrypt messages) is to factor M

  In other words

  The two problems are polynomially equivalent

Two kinds of Cryptography

- **Private key (or symmetric)**
  - Lucifer
  - DES
  - AES
- **Public key**
  - RSA
  - Diffie–Hellmann
  - Knapsack
  - NTRU