Internet security and prime numbers

Stanley Chang
Department of Mathematics
Wellesley College

January 11, 2022

Credit Card Usage

There were 368.92 billion purchase transactions in 2018.

The 2023 forecast is that 347 billion email messages will be sent per day by 4.3 billion users.

Encrypting and Decrypting

Unscramble this word:

istsouphonsa

The reader might try subdivide: ist||sou||phon||sa

sousaphonist 

Here is a random encryption: tponssuaoshi

Here is another, but not so random: passoutohsin

cushyarsonon: asynchronous
celeryspabox: Brexocalypse
cyanidetoast: staycationed

Encrypting and Decrypting

Try to decrypt the following!

stinkfanneer
roachmodstew
luluscornfad
defusehamsis
piemansmarts
blendourmaid
exgoofipauls

Rearrangement

Rearrange these letters to find the first four lines of an English poem. Can a human or a computer do it?

eeithnsehauamwomneltmeeihllernehsatechswneasoorbey
eaoyronwepotdeetifshyaeoofwnuwrrdoggaurtidkhhanthee
tthholenvogubosowlshdrealbaetsehcinsth

That time of year thou mayst in me behold,
When yellow leaves, or none, or few, do hang,
Upon the boughs that shake against the cold,
Bare ruin'd choirs where late the sweet birds sang.

Shakespeare, Sonnet 73

Hybrids from Interweaving

Interweave the titles


Pirates of the Carribean


The Empire Strikes Back

pithreematpiesorestrikfetshecabraribecank

But then can you unravel it?

pithreematpiesorestrikfetshecabraribecank

Interweaving Text

Here are three names that are interwoven together. Identify them.

fhiramllincarcisky feryorthoddeochdamoop crelipomontounlase
fhiramllincarcisky feryorthoddeochdamoop crelipomontounlase

 Hillary Rodham Clinton

framinccisk feyorthdeocdoop repomoulase
framinccisk feyorthdeocdoop repomoulase

 Francis Ford Coppola

mick eytheodo remouse

 Mickey Theodore Mouse

Interweaving Text

rlogutehuiorsabag emwaadergishyin nagsltonccobearverrtgt
rlogutehuiorsabag emwaadergishyin nagsltonccobearverrtgt

Louisa May Alcott

rgutehorbag ewadergishin ngstoncbearverrg

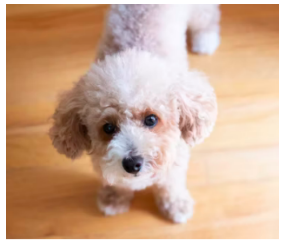
rgutehorbag ewadergishin ngstoncbearverrg

George Washington Carver

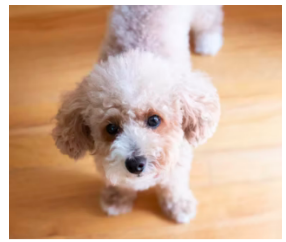
ruthba dergi nsberg

Ruth Bader Ginsberg

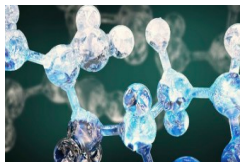
Given a hybrid, can you identify its parentage?



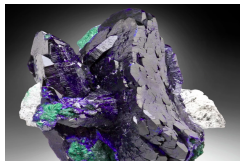
Can you recover purebreds from two hybrids?



The Enterprise of Academia



Chemistry

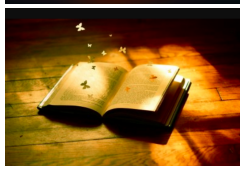


Metallurgy

Physics

Biology

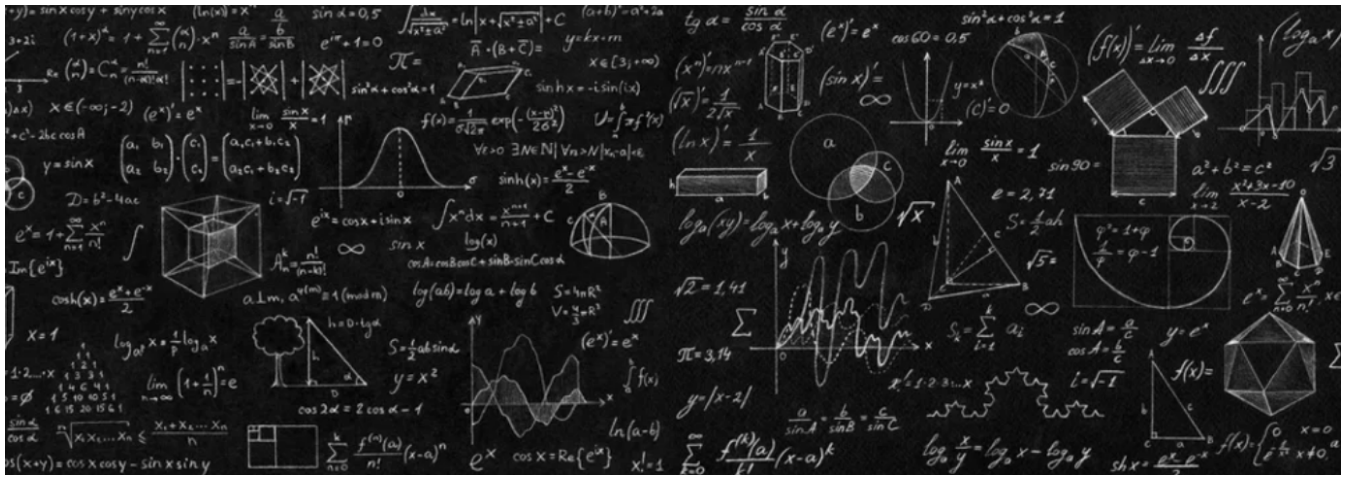
Acoustics



Social Sciences and Humanities

Mathematical Tools

How is mathematics involved in the practice of encryption/decryption, coupling/decoupling, combining/separating, especially for computer security?



Answer: Multiplication

Multiplication

Calculate 128×412.

$$
\begin{array}{r}
128 \\
412 \\
\hline
256 \\
128 \\
512 \\
\hline
52736
\end{array}
$$

Multiplication

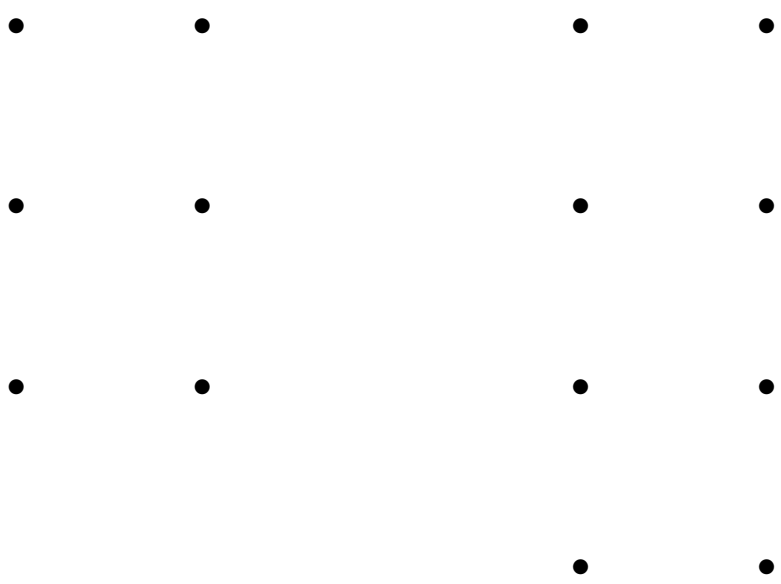
Calculate $5915587277 \times 1500450271$.

$$
\begin{array}{r}
5915587277 \\
1500450271 \\
\hline
5915587277 \\
41409110939 \\
11831174554 \\
\cdots\cdots\cdots \\
\hline
8876044532898802067
\end{array}
$$

Composite numbers

The opposite of multiplication is **factorization**, which lies at the center of computer security.

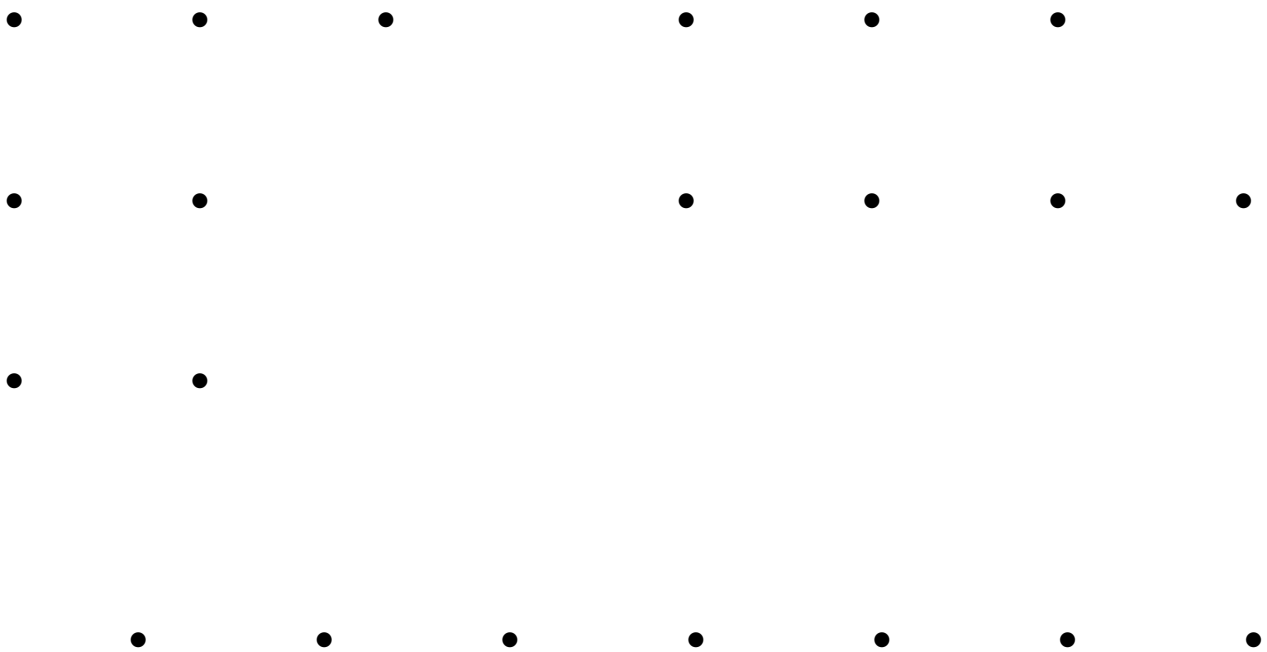
Consider $n = 6$ and $n = 8$. Put them into rectangles.



The numbers 6 and 8 are **composite** (decomposable).

Prime numbers

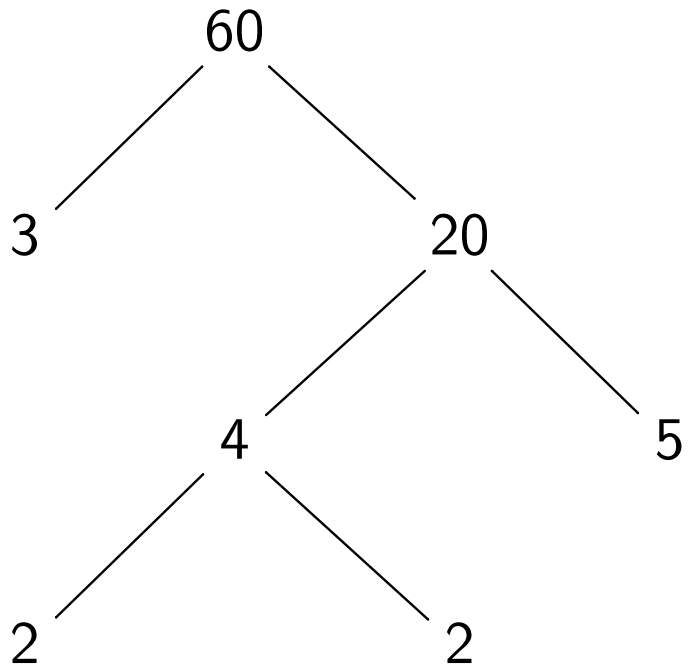
Consider $n = 7$.



It is **prime** (or indecomposable).

Factor trees

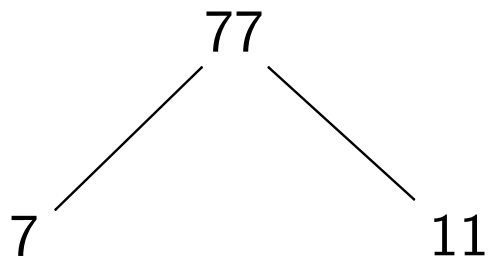
We use factor trees to break down a number to its prime factors.



$$60 = 2 \times 2 \times 3 \times 5$$

Factor trees

Some numbers have very small trees.



$$77 = 7 \times 11$$

Such a number is called a **semiprime**.

Some terminology

A number ≥ 2 is **composite** if it can be written as a nontrivial product $n = ab$. Example: $60 = 4 \times 15$.

The numbers a and b are called (nontrivial) **factors** of n.

If $n \geq 2$ cannot be written in this way, then it is called **prime**. Example: 17 and 29 are primes.

If n can be written as a product of exactly two primes, then it is called a **semiprime**. Example: $77 = 7 \times 11$.

A list of prime numbers

2	3	5	7	11	13	17	19	23	29
31	37	41	43	47	53	59	61	67	71
73	79	83	89	97	101	103	107	109	113
127	131	137	139	149	151	157	163	167	173
179	181	191	193	197	199	211	223	227	229
233	239	241	251	257	263	269	271	277	281
283	293	307	311	313	317	331	337	347	349
353	359	367	373	379	383	389	397	401	409
419	421	431	433	439	443	449	457	461	463
467	479	487	491	499	503	509	521	523	541
547	557	563	569	571	577	587	593	599	601
607	613	617	619	631	641	643	647	653	659
661	673	677	683	691	701	709	719	727	733
739	743	751	757	761	769	773	787	797	809
811	821	823	827	829	839	853	857	859	863
877	881	883	887	907	911	919	929	937	941
947	953	967	971	977	983	991	997	1009	1013
1019	1021	1031	1033	1039	1049	1051	1061	1063	1069
1087	1091	1093	1097	1103	1109	1117	1123	1129	1151
1153	1163	1171	1181	1187	1193	1201	1213	1217	1223
1229	1231	1237	1249	1259	1277	1279	1283	1289	1291

How many primes are there?

Ten-digit primes

5915587277

1500450271

We multiplied them together earlier.

Twenty-digit primes

48112959837082048697

54673257461630679457

Can we write down some big primes?

Thirty-digit primes

$$671998030559713968361666935769$$

$$282174488599599500573849980909$$

Bigger is better.

100-digit primes

$$2074722246773485207821695222107608587480996474721117$$
$$292752992589912196684750549658310084416732550077$$

$$2367495770217142995264827948666809233066409497699870$$
$$1120031493523803751248552300684871093732262519 83$$

A list of semiprime numbers

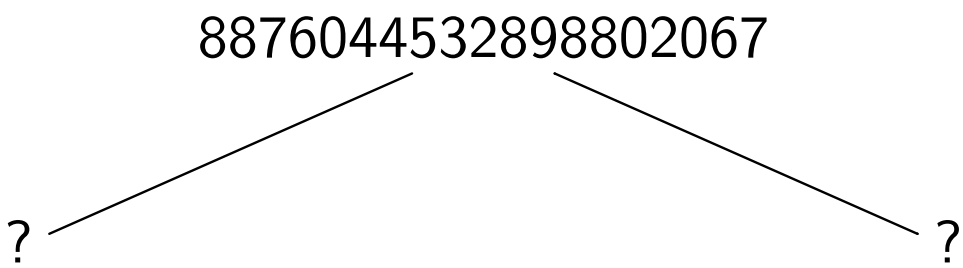
4, 6, 9, 10, 14, 15, 21, 22, 25, 26, 33, 34, 35, 38, 39, 46, 49, 51, 55, 57, 58, 62, 65, 69, 74, 77, 82, 85, 86, 87, 91, 93, 94, 95, 106, 111, 115, 118, 119, 121, 122, 123, 129, 133, 134, 141, 142, 143, 145, 146, 155, 158, 159, 161, 166, 169, 177, 178, 183, 185, 187, 194, 201, 202, 203, 205, 206, 209, 213, 214, 215, 217, 218, 219, 221, 226, 235, 237, 247, 249, 253, 254, 259, 262, 265, 267, 274, 278, 287, 289, 291, 295, 298, 299, 301, 302, 303, 305, 309, 314, 319, 321, 323, 326, 327, 329, 334, 335, 339, 341, 346, 355, 358, 361, 362, 365, 371

In particular,

$$5915587277 \times 1500450271 = 8876044532898802067$$

is a semiprime.

But can you easily factor large semiprimes?



Frank Nelson Cole

In 1903, Frank Nelson Cole made a presentation to the meeting of the American Mathematical Society where he identified the factors of the Mersenne number $M_{67} = 2^{67} - 1$.

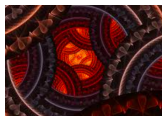
$$147,573,952,589,676,412,927$$

$$=$$

$$193,707,721 \quad \times \quad 761,838,257,287$$

This number is a semiprime.

The RSA Factoring Challenge


created by RSA Laboratories on March 18, 1991;


encouraged research into computational number theory;


focused on factoring large integers;


involved a cash prize;


concluded but people are still trying.

RSA = Rivest, Shamir, Adleman

A 300-bit key

RSA-100 (300 bits) =

1522605027922533360535618378132637429718068114961380688657908494580122963258952897654000350692006139

is the product of

37975227936943673922808872755445627854565536638199

and

40094690950920881030683735292761468389214899724061

Lenstra, 1991, a few days, $1000

A 640-bit key

RSA-193 (640 bits) =

3107418240490043721350750035888567930037346022
8427275457201619488232064405180815045563468296
7172328678243791627283803341547107310850191954
8529007337724822783525742386454014691736602477
652346609

is the product of

1634733645809253848443133883865090859841783670033
092312181110852389333100104508151212118167511579

and

1900871281664822113126851573935413975471896789968
515493666638539088027103802104498957191261465571

Böhm, 2005, five months, $20,000

RSA Factoring Challenge

RSA-100	RSA-110	RSA-120	RSA-129	RSA-130	RSA-140
RSA-150	RSA-155	RSA-160	RSA-170	RSA-180	RSA-190
RSA-200	RSA-210	RSA-220	RSA-230	RSA-232	RSA-240
RSA-250	RSA-260	RSA-270	RSA-280	RSA-290	RSA-300
RSA-309	RSA-310	RSA-320	RSA-330	RSA-340	RSA-350
RSA-360	RSA-370	RSA-380	RSA-390	RSA-400	RSA-410
RSA-420	RSA-430	RSA-440	RSA-450	RSA-460	RSA-470
RSA-480	RSA-490	RSA-500	RSA-576	RSA-617	RSA-640
RSA-704	RSA-768	RSA-896	RSA-1024	RSA-1536	RSA-2048

RSA Factoring Challenge

RSA-100	RSA-110	RSA-120	RSA-129	RSA-130	RSA-140
RSA-150	RSA-155	RSA-160	RSA-170	RSA-180	RSA-190
RSA-200	RSA-210	RSA-220	RSA-230	RSA-232	RSA-240
RSA-250	RSA-260	RSA-270	RSA-280	RSA-290	RSA-300
RSA-309	RSA-310	RSA-320	RSA-330	RSA-340	RSA-350
RSA-360	RSA-370	RSA-380	RSA-390	RSA-400	RSA-410
RSA-420	RSA-430	RSA-440	RSA-450	RSA-460	RSA-470
RSA-480	RSA-490	RSA-500	RSA-576	RSA-617	RSA-640
RSA-704	RSA-768	RSA-896	RSA-1024	RSA-1536	RSA-2048

RSA Factoring Challenge

RSA Number	Decimal Digits	Binary Digits	Prize	When
RSA-100	100	330	$1000	1991
RSA-120	120	397	$4429	1992
RSA-130	130	430	$14527	1996
RSA-155	155	512	$9383	1999
RSA-174	174	576	$10000	2003
RSA-193	193	640	$20000	2005
RSA-212	212	704	$30000	2012
RSA-232	232	768	$50000	2009
RSA-250	250	829	☹	2020
RSA-270	270	896		
RSA-300	300	1024		
RSA-463	463	1536		
RSA-617	617	2048		

RSA Factoring Challenge

During the RSA-155 cracking in 1999, it took 290 computers on the Internet and a supercomputer 4 months to factor a 512 bits (155 decimal digits) integer with two large prime factors. The computing power needed was estimated at 8000 Mips-years (a Mips is a million of processor instructions per second).

To see what this means for bigger values of $n = pq$, it is possible to use the estimated factoring time as the number d of digits of n goes to infinity, given by $K \exp(d^{1/3} \log^{2/3}(d))$ for some constant K.

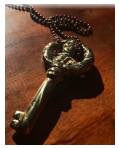
Number of digits	Computing power
155	8000 Mips-years
1024	800,000 Mips-centuries
2048	20,000,000,000 Mips-centuries

Public key cryptography

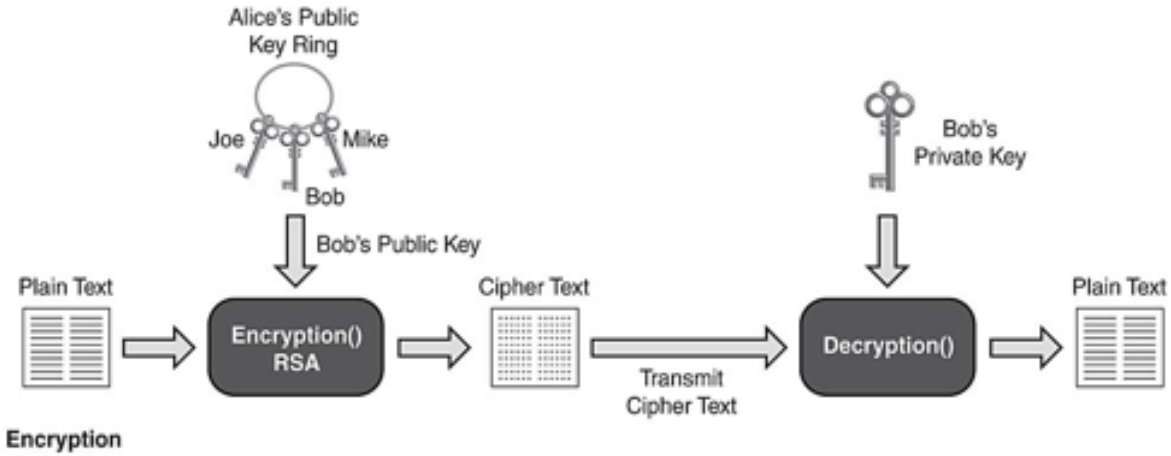
Alice wants to send a secret message to Bob.

Bob sends an open padlock to Alice in the mail. He has the key.

Alice puts her message in a box and locks it with Bob's padlock.



She sends it in the mail to Bob, and Bob opens it with the key.



Public key cryptography

1. You want to buy something from Amazon and you want to send your credit card number $T = 1424124153661414$ (plaintext).
2. Amazon sends you a huge semiprime n, but only it knows the factorization $n = pq$.
3. The semiprime n transforms the plaintext $T = 1424124153661414$ into some alternate text: $C = 9987140172310982$ (ciphertext). This number is sent to Amazon.
4. Amazon uses the factorization $n = pq$ to turn the ciphertext C back into plaintext T.
5. If n and C are intercepted, then C cannot be decoded without knowledge of p and q.

A short discursion

Fermat's Little Theorem: If p is prime and m is not divisible by p, then $m^{p-1} \equiv 1 \bmod p$.

Euler's Theorem: If $n \geq 2$ is an integer and m is relatively prime to n, then $m^{\phi(n)} \equiv 1 \bmod n$.

Here $\phi(n)$ is the totient function and equals the number of values in $\{1, 2, \ldots, n-1\}$ which are relatively prime to n.

If $de \equiv 1 \bmod \phi(n)$, then $de = 1 + k\phi(n)$ for some k. If m is relatively prime to n, then

$$(m^e)^d = m^{de} = m^{1+k\phi(n)} = m \cdot (m^{\phi(n)})^k \equiv m \cdot 1^k \bmod n \equiv m \bmod n.$$

An example

1. Generate two large primes p and q.
   Example $p = 23$ and $q = 37$.
2. Compute $n = pq$ and $\phi = (p-1)(q-1)$.
   Then $n = 851$ and $\phi = 792$.
3. Select random e between 1 and ϕ such that $\gcd(e, \phi) = 1$.
   For example $e = 91$.
4. Compute d between 1 and ϕ such that $de \equiv 1 \bmod \phi$.
   By Euclidean algorithm $d = 235$.
5. The message is an integer m between 0 and $n-1$.
   For example $m = 118$.

An example

1. Recall $p = 23$, $q = 37$, $n = 851$, $\phi = 792$, $e = 91$, $d = 235$, $m = 118$.
   Here $pq = n$ and $\phi = (p-1)(q-1)$ and $ed \equiv 1 \bmod \phi$.
2. Wellesley knows p, q, n, d, e, while Alice knows m.
3. Wellesley sends e and n to Alice. These numbers are the public key.
4. Alice computes $c \equiv m^e \bmod n$. Send c to Wellesley.
   Here $c = 118^{91} \bmod 851 \equiv 303 \bmod 851$.
5. Wellesley computes c^d modulo 851 to retrieve original m.
   Here $c^d \equiv 303^{235} \bmod 851 \equiv 118 \bmod 851$.
6. The numbers e, n, c may be intercepted. But without d ....

Hacking

Internet security is based on the level of difficulty required to factor a large number into two constituent primes.

A hacker can break into your private information by finding the key:

1. by obtaining the information somehow,
2. by factoring it with clever algorithms and computing power.

Some semiprimes are easier to crack than others!

(Think ist||sou||phon||sa versus passoutohsin.)

The future of security

The security of the key depends on the keylength.

1. 32-bit and 64-bit keys are useless for security;
2. 80-bit and 128-bit keys can be broken if one has sufficient storage space;
3. 512-bit keys can be broken by a dedicated amateur who has access a few dozen big PCs;
4. 1024-bit keys cannot be broken by an amateur for at least 10 years, but it is at the verge of being compromised with a sizeable budget and computational power;
5. 2048-bit keys are safe right now. ☺

Public key cryptography

Why not just choose a key with a million digits?

Answer: It takes too long to encrypt and decrypt.

The Game: To pick a keylength that is difficult to break but which easy to decrypt.

Thanks for your attention.

Thanks go to the Wellesley College Alumni Association, Hillary Rodham Clinton, George Washington Carver, Francis Ford Coppola, Mickey Theodore Mouse, Louisa May Alcott, Ruth Bader Ginsberg, William Shakespeare, Alice in Wonderland, Bob Marley, Captain Jack Sparrow, Luke Skywalker, Darth Vader, Maltipoos, Sousaphonists, Frank Nelson Cole, Ron Rivest, Adi Shamir, Leonard Adleman, Henrik Lenstra, M. Böhm, Pierre de Fermat, and Leonhard Euler.