The RSA* encryption algorithm

* Rivest-Shamir-Adleman, 1977
(Also developed in 1973 by British intelligence, declassified 1997.)

Basic premise:
- Multiplying is __easy__, but
- Factoring is __hard__.

Overview of RSA:

(A) Encoding.

(1) "Numerize:" convert letters to numbers in some simple way. E.g. $A \to 11$, $B \to 12, \ldots,$ $Z \to 36$. (Can do punctuation etc. similarly.)
E.g. MATH $\to$ 23113018.

(2) Call the resulting message n (e.g. $n = 23113018$). Raise n to a large positive integer k.

(3) Take the result and compute its __remainder__ r after division by another large positive integer m. That is, write

$$n^k = m \cdot q + r \qquad (0 \le r < m).$$

(Always possible by the "Division Algorithm:" more soon.)

Then r __is__ the coded message to be sent. Example:

If the message is "MATH," then, again, $n = 23113018$. Choose $k = 137$, $m = 555$. We compute that

$$n^k = \overbrace{7{,}052{,}\ldots{,}811{,}968}^{1009 \text{ digits}}$$
$$= 555 \cdot \underbrace{12{,}707{,}\ldots 273{,}535}_{1007 \text{ digits}} + 43. \quad (*)$$

So $r = 43$.

(B) Decoding.

(1) Under the right conditions, if you know k and m, and how m factors, you can recover the original message n from the coded message r.

(2) If you can't factor m, deducing n from r can be essentially impossible. And remember: factoring is hard.

(3) RSA is "public key:" k and m can be shared publicly, so anyone can encode, but knowing k and m is not enough to decode.

(D) Some number theory.

Notation:

(i) $\mathbb{Z}$ denotes the integers $\ldots, -2, -1, 0, 1, 2, \ldots$

(ii) $\mathbb{N}$ denotes the natural numbers $1, 2, 3, \ldots$

(iii) The symbol "$\in$" means "belongs to". E.g. $m \in \mathbb{Z}$ means m is an integer.

(iv) The symbol "$|$" means "divides" or "goes into evenly." E.g. $4|24$. Note that $24 = 4 \cdot 6$: in general, $a|b$ means $b = ac$ for some $c \in \mathbb{Z}$.

Now recall: to encode a message n in RSA, we write

$$n^k = m \cdot q + r \quad (0 \leq r < m)$$

for given $k, m \in \mathbb{N}$. We rewrite this as

$$n^k - r = mq, \quad \text{so} \quad m | (n^k - r).$$

MORAL: to study RSA, we should study phenomena like

$$m | (a-b)$$

for $a, b, m \in \mathbb{Z}$.

Definition.
Let $a, b, m \in \mathbb{Z}$. We say "a is congruent to b mod m," and write

$$a \equiv b \pmod{m},$$

if $m | (a-b)$.

Examples:

$122 \equiv 87 \pmod{5}$, (since $122 - 87 = 35 = 5 \cdot 7$, so $5 | (122-87)$),

$-13 \equiv 8 \pmod{7}$,

$241137 \equiv 137 \pmod{1000}$,

$k \equiv 0 \pmod{2}$ for any even $k \in \mathbb{Z}$,

$3^5 \equiv 3 \pmod{24}$ (since $3^5 - 3 = 240 = 24 \cdot 10$),

$23113018^{137} \equiv 43 \pmod{555}$ (by (*) above),

etc.

Properties of "mod m:"

__Proposition.__ Let $a, b, c, d, m \in \mathbb{Z}$.
If $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, then:

(i) $a + c \equiv b + d \pmod{m}$.

(ii) $a - c \equiv b - d \pmod{m}$.

(iii) $ac \equiv bd \pmod{m}$.

(Proof to follow.)

Example using Part (i) of the proposition: we have

$$25 \equiv 4 \pmod{7}$$

and $75 \equiv -2 \pmod{7}$,

so, adding, we find that

$$25 + 75 \equiv 4 + (-2) \pmod{7};$$

that is

$$100 \equiv 2 \pmod{7}.$$

(You should check this.)