

# ENHANCING THE PERFORMANCE OF FACTORING ALGORITHMS

GIVEN  $n$  FIND  $p_1, p_2, \dots, p_k$  SUCH THAT

$$n = p_1^{d_1} \times p_2^{d_2} \dots \times p_k^{d_k}$$

WHERE  $p_i$  ARE PRIMES

- FACTORING IS CONSIDERED TO BE A VERY HARD.
- THE BEST KNOWN ALGORITHM HAS A RUNNING TIME OF  $O(e^{c \times (\ln(n))^{1/3} \times \ln(\ln(n))^{2/3}})$
- FACTORING LARGE COMPOSITE NUMBERS IS INFEASIBLE FOR LARGE  $n$  WITH CURRENT ALGORITHMS AND COMPUTING TECHNOLOGY.

# MAJOR APPLICATIONS

- PUBLIC KEY CRYPTOGRAPHY
  - RSA PUBLIC KEY CRYPTOSYSTEM'S SECURITY DEPENDS ON THE ASSUMPTION THAT FACTORIZATION IS A VERY HARD PROBLEM.
- THEOREM PROVING
  - PROOFS OF CERTAIN THEOREMS IN NUMBER THEORY HAVE RELIED ON THE FACTORIZATION OF LARGE COMPOSITE NUMBERS.

# FACTORING PROGRESS IN THE LAST 20 YEARS

- RSA-129 A 17-YEAR OLD CHALLENGE PROBLEM WAS SOLVED IN APRIL 1994.
- RSA-129 =  
11438162575788886766923577997614661201021829  
67212423625625618429357069352457338978305971  
23563958705058989075147599290026879543541
- =  
34905295108476509491478496199038981334177646  
38493387843990820577  
×  
32769132993266709549961988190834  
461413177642967992942539798288533
- USING FACTORING METHODS AND COMPUTER TECHNOLOGY AVAILABLE IN 1977, FACTORING RSA-129 WOULD HAVE REQUIRED **40 QUADRILLION YEARS**
- RSA-129 WAS FACTORED IN ONLY EIGHT MONTHS.
- MUCH OF THE SPEEDUP COMES FROM FASTER ALGORITHMS
- RSA-129 WAS FACTORED USING A MULTIPLE POLYNOMIAL QUADRATIC SIEVE (MPQS)

# KNOWN FACTORING ALGORITHMS

- PRIMITIVE DIVISION ALGORITHM -- EXPONENTIAL RUNNING TIME
- EULER'S ALGORITHM
- SHOR'S ALGORITHM --  $O(e^{\# \text{ OF DIGITS IN } n})$
- SHANK'S SQUARE FORM ALGORITHM
- WILLIAM'S P+1 ALGORITHM
- POLLARD'S P-1 ALGORITHM
- POLLARD'S RHO ALGORITHM --  $O(\text{SQRT}(p))$   
WHERE P IS THE LARGEST FACTOR OF n
- CONTINUED FRACTION ALGORITHM --  
 $O(e^{\text{SQRT}(2 \times \text{LN}(n) \times \text{LN}(\text{LN}(n)))})$
- CLASS GROUP METHOD -- SUB EXPONENTIAL
- FERMAT'S ALGORITHM --  $O(e^{\# \text{ OF DIGITS IN } n})$
- DIXON'S RANDOM SQUARE ALGORITHM  
-- SUB EXPONENTIAL
- MULTIPLE POLYNOMIAL QUADRATIC SIEVE  
--  $O(e^{\text{SQRT}(\text{LN}(n) \times \text{LN}(\text{LN}(n)))})$
- NUMBER FIELD SIEVE --  $O(e^{c \times (\text{LN}(n)^{1/3} \times \text{LN}(\text{LN}(n)^{2/3}))})$   
FASTEST KNOWN FACTORING ALGORITHM FOR  
LARGE INTEGERS
- ELLIPTIC CURVE ALGORITHM  
--  $O(e^{\text{SQRT}(\text{LN}(n) \times \text{LN}(\text{LN}(n)))})$
- VALLES TWO-THIRDS ALGORITHM  
-- sub exponential

# THE QUADRATIC SIEVE

## (our research interest)

- THE QUADRATIC SIEVE WORKS BY FINDING INTEGERS  $x, y$  SO THAT  $x^2 = y^2 \pmod n$  BUT  $x \not\equiv \pm y \pmod n$ .
- THE FIRST CONGRUENCE IMPLIES THAT  $n$  DIVIDES  $(x-y)(x+y)$ , WHILE THE SECOND CONGRUENCE IMPLIES THAT  $n$  DOES NOT DIVIDE  $x - y$  OR  $x + y$ .
- IT FOLLOWS THAT AT LEAST ONE PRIME FACTOR OF  $n$  DIVIDES  $x - y$  AND AT LEAST ONE PRIME FACTOR OF  $n$  DOES NOT DIVIDE  $x - y$ .
- THEREFORE,  $\gcd(n, x-y)$  IS A PROPER FACTOR OF  $n$ .
- THE QUADRATIC SIEVE ALGORITHM IGNORES THE CONDITION  $x \not\equiv \pm y \pmod n$  AND SEEKS MANY RANDOM SOLUTIONS TO  $x^2 = y^2 \pmod n$ .
- IF  $n$  IS ODD AND NOT A PRIME POWER, AT LEAST HALF OF ALL SOLUTIONS WITH  $y \not\equiv 0 \pmod n$  SATISFY  $x \not\equiv \pm y \pmod n$  AND FACTOR  $n$ .
- MANY CONGRUENCES (CALLED RELATIONS) OF THE FORM  $z^2 = q \pmod n$  ARE PRODUCED WITH  $q$  FACTORED COMPLETELY.
- LINEAR ALGEBRA IS USED OVER THE FIELD  $\text{GF}(2)$  WITH TWO ELEMENTS TO PAIR THESE PRIMES AND FIND A SUBSET OF THE RELATIONS IN WHICH THE PRODUCT OF THE  $q$ 's IS A SQUARE,  $y^2$ , SAY.
- LET  $x$  BE THE PRODUCT OF THE  $z$ 's IN THESE RELATIONS. THEN  $x^2 = y^2 \pmod n$ , AS DESIRED.

# VARIATIONS OF THE MULTIPLE POLYNOMIAL QUADRATIC SIEVE (MPQS)

## *HYPERCUBE VARIATION OF THE MULTIPLE POLYNOMIAL QUADRATIC SIEVE (HMPQS)*

- HMPQS IS A VARIATION ON THE QUADRATIC SIEVE (QS) ALGORITHM, WHICH INSPECTS MANY QUADRATIC POLYNOMIALS LOOKING FOR QUADRATIC RESIDUES WITH SMALL PRIME FACTORS.
- THE POLYNOMIALS ARE ORGANIZED AS THE NODES OF AN N-DIMENSIONAL CUBE.
- SINCE CHANGING POLYNOMIALS ON THE HYPERCUBE IS CHEAP, THE OPTIMAL VALUE FOR THE SIZE OF THE SIEVING INTERVAL IS MUCH SMALLER THAN IN OTHER IMPLEMENTATIONS OF THE MULTIPLE POLYNOMIAL QUADRATIC SIEVE (MPQS).

- HMPQS IS SUBSTANTIALLY FASTER THAN MPQS.

*DOUBLE LARGE PRIME VARIATION OF  
MULTIPLE POLYNOMIAL QUADRATIC SIEVE*

- THIS VARIATION OF THE QUADRATIC SIEVE IS KNOWN TO BE FASTER THAN MPQS BY APPROXIMATELY A FACTOR OF 2:5 FOR SUFFICIENTLY LARGE  $n$ .

*TRIPLE LARGE PRIME VARIATION OF MPQS  
(TPMPQS)*

- WE COMBINED HMPQS WITH THIS VARIATION OF THE MPQS
- WE ARE CURRENTLY TESTING THIS VARIATION ON AN IBM SP 500 SUPER COMPUTER.
- PRELIMINARY RESULTS INDICATE THAT THIS IS THE FASTEST VERSION OF THE MPQS TO DATE.

# OUR WORK

- WE COMBINED THE HMPQS AND TPMPQS ALGORITHMS.
- NEW ALGORITHM WAS PARALLELIZED TO TAKE ADVANTAGE OF AN IBM SP 500 SUPERCOMPUTER.
- THE ALGORITHM IS USING MPI LIBRARY ON THE IBM SP 500
- THE IBM SP 500 IS A MESSAGE PASSING ARCHITECTURE THAT HAS 272 PROCESSORS.
- THE MACHINE IS NUMBER 72 ON THE TOP500 NOVEMBER 3RD, 2000 LIST. THIS LIST RANKS THE MOST POWERFUL SUPERCOMPUTERS IN THE WORLD.
- FUTURE WORK MIGHT INVOLVE WRITING A DISTRIBUTED VERSION OF THE ALGORITHM THAT USES IDLE COMPUTING RESOURCES ON THE CAMPUS