NEWS RELEASE

NTT

January 8, 2010

New World Record for "Integer Factorization Problem" the Basis for the Security of Public Key Cryptography

- Successful Full Factorization of a 768-Bit Composite Number Using General Number Field Sieve -

Nippon Telegraph and Telephone Corporation (hereafter NTT, headquarters: Chiyoda-ku, Tokyo, President and CEO: Satoshi Miura) through collaborative research with Ecole Polytechnique Fédérale de Lausanne (hereafter EPFL), Bonn University, Institut National de Recherche en Informatique et en Automatique (hereafter INRIA), and Centrum Wiskunde & Informatica (hereafter CWI) set a new world record for a 768-bit (232-digit decimal) composite number that far exceeds the previous record (663-bits 200-digit decimal). This was achieved through integer factorization using the general number field sieve[*1] for the integer factorization problem[*2], which for only a few bits requires a significantly large number of calculations.

<Background and Significance of Research>

Accompanying the fundamental diffusion of the Internet, practical convenient service such as net settlement and Internet banking have become commonplace and the number of transactions involving secret information over the Internet has increased. Thus, guaranteeing a high level of information security is indispensable to these socio-economic activities that use networks.

NTT Information Sharing Platform Laboratories (hereafter NTT Laboratories) studies new encryption technologies and examines the security of existing encryption schemes to provide information security.

Examination of the number of bits that can be factored is important to estimate more precisely the security and strength of the RSA scheme[*3] because the RSA scheme, which is widely used as a public key encryption[*4] scheme, is based on the difficulty of the integer factorization problem.

This release reports on the achievement of a factorization that overwhelms 700 bits, and this suggests that 1024-bit integer factorization can be achieved in the near future with a size that can be readily used in the current RSA scheme implementation. There has been growing interest in the use of stronger and more efficient encryption technology.

At the current level of integer factorization technology and computing power, by estimating how large of a composite number to which integer factorization can be applied, estimating the possibility of cryptanalysis becomes possible, and from those results we can establish the pertinent RSA encryption key length and contribute to the construction of a future secure and stable encryption system.

<Contents of Research>

This factorization is achieved by the number field sieve, which is currently known as

the fastest algorithm to factor large composite numbers. The general number field sieve comprises five steps: polynomial selection, sieving, filtering, linear algebra, and square root. There are many parameters to be chosen in each step. Currently, nobody knows of an efficient method to choose these parameters appropriately, though the choice of these parameters heavily depends on the subsequent computational load. We succeeded in factoring the number very quickly by properly choosing the parameters. Among the five steps, the sieving and linear algebra steps require the highest calculation loads. In this analysis, a description of how each step is performed is given below.

(1) Polynomial Selection
This very important step determines the remaining calculation load, but at this point nobody knows an effective method of determining how much time is required and how to find optimal polynomials. This time, from search results we selected a polynomial from Bonn University in the summer of 2005 that has a calculation load equal to that for 20 years of operation on a 2.2 GHz Opteron[*5] processor. Subsequently, early in 2007 at EPFL, we spent a calculation load equal to that for 20 years of operation on a 2.2 GHz Opteron processor, but we did not find anything better.

(2) Sieving
Although this step accounts for the majority of the entire calculation load, since the calculations can be comparatively simply distributed we performed parallel calculations among multiple participating organizations. For the calculations this time, to conform to the memory capacity of the computers we prepared several parameters. We began in the summer of 2007 and ended in April 2009. Almost all of the processing was done between the spring of 2008 and March 2009. Sieving was mainly performed at NTT Laboratories, EPFL, Bonn University, INRIA, and CWI using a variety of PCs and clusters. In all, a calculation load equal to that for 1500 years of operation on a 2.2 GHz Opteron processor was required.

(3) Filtering
By performing this step, the subsequent linear algebra step becomes significantly faster. We used a cluster and 8 core computers equipped with 10 TB hard disks at the EPFL. The calculation load was less than or equivalent to 6 months of operation on a 2.66 GHz Core2[*6] processor, including unnecessary computations that were done with many parameter trials.

(4) Linear Algebra (Solving a System of Simultaneous Equations)
This step theoretically requires an extremely high calculation load and distributed computing[*7] is difficult. This time, we used a few clusters, and we developed and used a technique where even if the speed or availability were different for each cluster we could still efficiently perform the calculations. Using the clusters at NTT Laboratories and EPFL, and ALADDIN-G5K[*8] in France, which is operated efficiently by INRIA, we solved simultaneous equations from a sparse matrix generated by filtering. The calculation load was equal to that for 155 years of operation on a 2.2 GHz Opteron processor. From these results we obtained solutions that can be analyzed.

(5) Square Root (Computing a Square Root for Algebraic Numbers and the Greatest Common Divisor)
Although this step uses a mathematically high level of logic, not a particularly high

calculation load is required. Using the computers established at EPFL, we obtained the solution below in less than a few hours.

123018668453011775513049495838496272077285356959533479219732245
215172640050726365751874520219978646938995647494277406384592519
255732630345373154826850791702612214291346167042921431160222124
047927473779408066535141959745985690214341 3

=

334780716989568987860441698482126908177047949837137685689124313
889828837938780022876147116525317430877378144679994 89

X

367460436667995904282446337996279526322791581643430876426760322
838157396665112792333734171433968102700927987363089 17

<Future Outlook>
Encryption technology will become increasingly important to the industry for protecting information as the info-communication society develops. NTT Laboratories are continuously evaluating the security of encryption technology as a whole and diligently spreading the new public key encryption scheme "Elliptic Curve Cryptography*9," which uses elliptical curve arithmetic rules as the next generation encryption.
In the future, we will promote a wide range of security research from cryptography to social influences, and investigate a safe and secure network society.

<Glossary>
*1 General Number Field Sieve (GNFS)
GNFS is an integer factorization algorithm that was originally proposed by J. Pollard and others. A. K. Lenstra et al. refined the algorithm in early 90s. The well-known GNFS is asymptotically the fastest algorithm for a non-special-type composite number factorization algorithm such as that used in the RSA scheme. While the run time for the so-called trial division, which employs division by 2, 3, 5, 7, etc., is the exponential time, the GNFS is evaluated to be completed in subexponential time. However, since the evaluation of the currently known run time is the upper bound of the average, we must compile calculation experiment data in order to estimate accurately a definitive run time for practical numbers.

*2 Integer Factorization Problem
This problem involves resolving composite numbers into a multiplication of prime numbers. While a small composite number can be factored into prime numbers in a short time, it is difficult to complete the calculations for a large number in a realistic amount of time. However, for a case with a prime factor that is not so large, we can use the elliptic curve method to obtain the prime factor.
For the factorization of the product of two large prime factors that comprise a composite number used in the RSA scheme, we utilize the number field sieve. Currently, the general number field sieve is the fastest method for the composite numbers used in the RSA scheme.

*3 RSA Scheme

RSA, which comes from the first letter of each developer Rivest, Shamir, and Adleman, is a public key encryption and digital signature scheme that was released in 1978. The RSA scheme is currently the most widely used digital signature scheme. Until now, various improvements have been investigated, and several of these improvements are included in digital signature method guidelines and the E-government recommended encryption list in Japan. The security of the RSA scheme is dependent on the parameter called "modulus," and the larger it is the more secure but the processing capability decreases. Currently, the modulus of 1024 bits is widely used.

*4 Public Key Encryption
This concept was proposed by Diffie and Hellman in 1976. The RSA encryption scheme is the most famous application instance of this concept. While the conventional encryption scheme keeps the key used for both encryption and decryption secret, the key for encryption can be disclosed for public key cryptography, and only the decryption key needs to be kept secret.

*5 Opteron
AMD developed AMD64 the so-called "64-bit CPU" architecture, which is an extension of IA-32 the 32-bit architecture developed by Intel. The Opteron CPU is based on the AMD64 architecture.

*6 Core2
Intel developed EM64T as a 64-bit architecture that is compatible with AMD64. The Core2 CPU is based on the EM64T architecture.

*7 Distributed Computing
This technology distributes a large number of computations and computes them using many computers. It is difficult to achieve distributed computing when a part of the computation depends on another part of the computation.

*8 ALADDIN-G5K
An infrastructure distributed in 9 sites around France for research in large-scale parallel and distributed systems.

*9 Elliptic Curve Cryptography
Encryption schemes that encrypt and decrypt based on the special addition defined by the formulae for the points of an elliptic curve. The difficulty of the cryptanalysis is believed to be similar to solving the elliptic curve discrete logarithm problem, and currently nobody knows of an efficient algorithm to solve it.
NTT Laboratories developed an elliptic curve key agreement scheme called PSEC-KEM.

For more information, contact:

NTT Information Sharing Laboratory Group
(Information Sharing Platform Laboratories)
Planning Dept. Public Relations
TEL:0422-59-3663
E-mail:islg-pr@lab.ntt.co.jp