In a JUMBLE

Next time you need to send a confidential message, do not mark it 'For your eyes only', encrypt it

The year was 1941. Japan was already part of World War II and all correspondence addressed to the Japanese ambassador in the US was under scrutiny by American agents. Yet, when in December, the Japanese attacked Pearl Harbour, America was taken by surprise. The Americans had no clue about the impending attack because they could not crack the code used by the Japanese in their communication.

However, encryption or cryptography—the scrambling of data in order to make it unintelligible to all but the intended recipient—is a science that goes beyond World War II.

Encryption methods, right from the simplistic translation table, have been used through the ages, even as newer and more secure encryption methods (like the super-secure public key cryptography) evolve.

Translation Table

Probably the simplest and the oldest method of encryption is the translation table. A table containing (random) characters is used to map the original data

into the encrypted data. The same table is used to decrypt the data. A very simple example of this is the computer called HAL 9001 in the famous movie of the 1970s, *2001: A Space Odyssey*. The letters 'H', 'A' and 'L' precede the letters 'I', 'B' and 'M' in the English alphabet.

Though simple and fast, translation table codes are easy to break. This method can be strengthened by the use of two or more tables. To make decryption tougher, the second translation table can be based on output generated by the first. So unless a cracker knows the number of translation tables involved, he will find it very difficult to crack the code.

CHECKING THE SEAL

You have just encrypted some data. How do you ensure that the recipient reads the data correctly and that the data does not get corrupted on the way? You attach a signature to the encrypted data that can be used to check if everything is fine.

One way to do this is via a checksum: add up all the ASCII values of the original data and include the result in the encrypted data. The recipient upon decryption, performs the same operation to check if the two values match.

This method is not very reliable because it depends upon simple addition both ABCD and BDCA will give the same checksum value.

Another algorithm called CRC (Cyclic Redundancy Check) fixes this flaw. CRC passes the data through a complex polynomial function, then stores the remainder as the CRC value. CRC is quite good at detecting corrupted data. 16-bit CRC, for instance, has an error possibility of 1 in 65,536 (2^{16} = 65,536). 32-bit CRC, which is the normal minimum for CRC

Data Repositioning

Another way to encrypt data is by data repositioning. Letters of the alphabet are arranged in an apparently random pattern to form anagrams. The receiver uses the same pattern to decrypt the text.

T	x		r
h	t	b	y
i		e	p
s	w		t
	i	e	e
t	l	n	d
e	l	c	.

Consider this text for example: "This text will be encrypted." This sentence has 28 characters inclusive of spaces and the fullstop. 28 is 7 times 4, so by making a 7x4 table and arranging the characters vertically in columns, you get something like the table at left

Reading the table horizontally gives you "Tx rhtbyi epsw t ieetlndelc." There, your text is encrypted! Unfortunately, jumbling letters is one of the most primitive methods of encryption, which even an amateur can decipher in a few minutes.

Like translation tables, computers can handle data repositioning rather well. This method can be made difficult to crack by making anagrams at the 'bit' level. For instance, instead of jumbling the letters of the alphabet, a computer can change the bit values of characters, making the encryption a little more difficult to crack.

Key-based encryption

Present day data encryption schemes are key based—the user specifies a 'key' or a password and then uses it to

Illustration: FARZANA

encrypt the text. The receiver again uses a key to decrypt the data. He could either use the symmetric key method, where the same key is used for both encryption and decryption, or the asymmetric key method, where the two keys are different but form a unique pair. In the latter case, the encrypting key is called the 'public' key and the decrypting key is called the 'private' key. Using a good algorithm (a mathematical equation), it will not be possible to generate one key from the other.

PGP for Windows, available from www.pgpi.com

Very few operations in mathematics are truly irreversible. In almost all cases, if 'a' can be transformed to 'b', the reverse is also possible. The trick, therefore, is to use an operation, reversing which will give an undefined value (an extreme example could be 'divide by zero'), thereby forcing a cracker to try all possible combinations to find the one that fits.

One of the most popular algorithms, RSA, was invented in 1977 by three MIT scientists, Ronald Rivest, Adi Shamir, and Leonard Adelman (their initials form the term RSA). This algorithm uses very large prime numbers, sometimes as large as 128 digits, to generate the public and the private keys (*See Tech-Talk*).

The most popular privacy software, PGP (Pretty Good Privacy) uses public key cryptography, particularly the RSA algorithm. However, RSA is slow so PGP first uses a randomly generated symmetric key to encode data, and then encodes the symmetric key using RSA. The only way to get to the data is by decrypting the symmetric key.

Steganography

Encryption is a safe way to pass confidential information. However, if you want to hide the fact that you are even attempting to pass encrypted information to a certain client, you can mask the transfer of information by 'steganographing' it.

Prior to the personal computing revolution of the 1980s, various methods of steganography have been in use. The most famous being the German microdot, developed during World War II, where a secret message was photographically reduced to the size of a dot and affixed above the letter 'i', or any other punctuation symbol in a dummy message.

Today steganography involves hiding data within images or audio clips. It works on the principle that files containing digital images or audio can be modified to a certain extent without affecting the image (or audio). A certain feature of the Graphics Interchange Format (GIF) allows you to add comments to a GIF file. Any comments added this way do not appear in the image, but this is not steganography because anyone looking for such 'hidden' comments in a GIF file will easily find them.

Social Engineering

Progress in data encryption points in the direction of greater privacy. However, encryption alone is not enough. Alertness and presence of mind are equally important. Crackers and samurai on the Net use 'social engineering'–the art of convincing people into disclosing their passwords or other such important information. The standard method is to pose as an engineer from the victim's ISP and ask for the password for 'maintenance reasons'.

Remember: Your ISP, or any other Web-based service, does not need your password for any reason. So, never give it away; no matter how convincing the person seems.

KIRAN JONNALAGADDA

TECH-TALK

DATA ENCRYPTION ALGORITHMS

RSA is based on the idea of factoring a function that is very easy to calculate, but extremely difficult to reverse. Two prime numbers p and q when multiplied give N. However if you know only the value of N, p and q are hard to calculate, especially if N is a large number. RSA uses factors of over 500 bits (150 digits) making even trial-and-error pretty tough.

Encryption is carried out using blocks of data each block treated as a sequence of bits, not bytes with a slightly lesser number of digits than N. The block is considered as a single number, and multiplied e times by itself (for PGP, e is usually 17). The result thus obtained is divided by N and the remainder thus obtained is stored as the encrypted message. To decrypt, the recipient uses another special number k, where (ke-1) is divisible by (p-1) (q-1). The value k is chosen such that multiplying the encrypted message k times by itself and then dividing by N, gives the original message as the remainder.

To find k, you have to know p and q. e and N from the public key and can be freely distributed, while k is the private key and must be kept secret. e and k are symmetric in the sense that if you use either one as an encryption key, the other can be used as a decryption key. This is different from the symmetric key encryption method where e and k would both be the same number.

RSA is patented by MIT, who granted exclusive rights to license the product to RSA Data Security, Inc (RSADSI). PGP has rights to use RSA for non-commercial use.

Various other algorithms have evolved to supersede RSA, the most notable being Blowfish. Bruce Schneier s Blowfish is a new, unpatented, royalty- and licence-free encryption algorithm that is much faster than other popular algorithms like DES and IDEA. Source code implementations of this algorithm are also freely available.