tween data, based on concepts similar to those described above.

C. B. FINKELSTEIN

DATA ENCRYPTION

For articles on related subjects *see* CRIME AND COMPUTER SECURITY; DATA COMMUNICATIONS; DATA SECURITY; and KEY.

Cryptography is the science of transforming messages for the purpose of making the message unintelligible to all but the intended receiver of the message. The term *data encryption* refers to the use of cryptographic methods in computer communications for the same reason but also implies the additional goals of providing assurance to the receiver that the message is not a forgery, and/or allowing the receiver to prove to a third party that the message is not a forgery. These various aims are called, respectively, the goals of *communication security, authentication,* and *digital signatures.*

The transformation used to encipher a message typically involves both a general method, or algorithm, and a key. While the general method used by a pair of correspondents may be public knowledge, some or all of the key information must be kept secret. The process of transforming (enciphering) a message is to apply the enciphering algorithm to the message, where the key is used as an auxiliary input to control the enciphering. The reverse operation (deciphering) is performed similarly.

Classical encryption techniques involve such operations as substituting for each message letter a substitute letter; in this case, the key is the correspondence between message (plaintext) letters and the enciphered message (ciphertext) letters. Such *substitution ciphers* can also be based on substituting for two or more letters at a time. Another common technique is to permute the order of the message letters using an algorithm whose steps are determined by a *key.* Many complicated hand or mechanical ciphers have been developed in the last few centuries; the reader should consult Kahn (1967) for details. These techniques are insecure in general; the breaking of the German Enigma cipher during World War II attests to the vulnerability of even complicated rotor-machine ciphers.

The *one-time pad* is a technique which provides the ultimate in security: It is provably unbreakable. To en-

and distributing the large amount of key information required.

Cryptosystems which, unlike the one-time pad, depend upon an amount of key information which is independent of message length are breakable in theory. What makes them usable in practice is that the person trying to break the cipher (the *cryptanalyst*) must use an impractical or infeasible amount of computational resources in order to break the cipher. These ciphers are constructed so that the "work-factor" in breaking them is high enough to prevent a successful attack.

The major application of cryptography today is for data transmitted between computers in computer communication networks and for computer data encrypted for storage.

The most widely used cipher in the U.S. for the encryption of stored or transmitted computer data is undoubtedly the Data Encryption Standard (DES), which was designed at IBM and approved as a standard by the National Bureau of Standards in 1976. The DES enciphers a 64-bit message block under control of a 56-bit key to produce a 64-bit ciphertext. The enciphering operation consists of roughly 16 iterations of the following two steps.

1. Exchange the left half of the 64-bit message with the right half.
2. Replace the right half of the message with the bit-wise exclusive-or of the right half and a 32-bit word which is a complicated function f of the left half, the key, and the iteration number. The function f involves in part a number of substitutions of short sub-blocks using specially constructed substitution tables (S-boxes) and permutations of the individual bit positions. The basic DES function has been implemented by a large number of manufacturers on special-purpose LSI chips, which can encipher at megabit per second rates.

Some applications (e.g., enciphering a line to a user's terminal) require that blocks shorter than 64 bits (e.g., a byte) be individually enciphered. The basic DES block can be used for this application in *cipher feedback mode:* each message byte is enciphered by an exclusive-or with the left-most byte of the result of taking the last eight ciphertext bytes and using them as input to the DES to obtain another 64-bit block of ciphertext.

Conventional cryptosystems (including DES) use

As a small example of the RSA method, the word "IT" can be encrypted as follows. Using the representation $A = 01$, $B = 02$, . . ., $Z = 26$, we obtain the number 0920 for IT. Then with $n = 2773 = 47 \cdot 59$ and $e = 17$, we obtain the ciphertext:

$$C = 920^{17}(\text{modulo } 2773) = 948.$$

Using $p = 47$ and $q = 59$, a value of $d = 157$ can be derived, from which we can calculate 948^{157} (modulo 2773) = 920, the original message.

Fig. 1. Data encryption using the RSA method.

The friends can send to the creator of the enciphering key enciphered mail that only the creator can read. (Even if a cryptanalyst obtains a copy of the enciphering key, it does no good.) This demonstrates the flexibility of a public-key cryptosystem for *key distribution,* an area where conventional cryptosystems are awkward because all keys must be kept secret. Public-key cryptosystems can also be used to provide *digital signatures:* A user can create a signature for a message by enciphering it with a private key. (Here the enciphering/deciphering roles of the public/private keys are reversed.) Someone else can check the validity of the signature by checking that it deciphers to the message using the signer's public key. This capability of public-key cryptosystems promises to have important applications in electronic funds transfer systems (*q.v.*).

The first proposal for a function to implement public-key cryptosystems was by Rivest, Shamir and Adleman (1978). Their cryptosystem (the so-called *RSA cipher*) enciphers a message M (first coded into numeric form by, for example, setting $A = 01$, $B = 02$, etc.) using a public key *(e,n)* to obtain a ciphertext C as follows.

$$C = M^e(\text{mod } n).$$

That is, C is the remainder of M^e when divided by n. Here all quantities are large numbers (several hundred bits long), and n is the product of two very large prime numbers p and q. The security of the cipher rests mainly on the practical impossibility of factoring the number n into its parts p and q. The deciphering operation is similar, except that the exponent is different:

$$M = C^d(\text{mod } n).$$

Since d depends on p and q (in a way too complicated to

Fig. 1. Another public-key cryptosystem has been proposed by Merkle and Hellman (1978). This technique is based on the difficulty of determining which subset of a given list of numbers adds up to a given target number (this is the NP-complete "knapsack" problem). A public key consists of a list, $a_1, a_2 \ldots a_{200}$, of two hundred 200-bit numbers. A 200-bit message, $m_1, \ldots, m_{200}$, is enciphered as follows.

$$c = \sum_{i=1}^{200} m_i \cdot a_i.$$

Deciphering is more complicated and depends on the special manner in which a_i is constructed. This technique does not easily lend itself to the creation of digital signatures (as does RSA) but it does permit substantially higher enciphering rates.

References

1967. Kahn, D. *The Codebreakers.* New York: Macmillan.

1976. Diffie, W. and Hellman, M. "New Directions in Cryptography," *IEEE Trans. Information Theory* **IT-22,** pp. 644–654 (November).

1977. FIPS Publication 46. *Specifications for the Data Encryption Standard.*

1978. Rivest, R., Shamir, A., and Adleman, L. "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems," *Comm. ACM,* pp. 120–126 (February).

1978. Merkle, R. and Hellman, M. "Hiding Information and Signatures in Trapdoor Knapsacks," *IEEE Trans. Information Theory* **IT-24,** pp. 535–530 (September).

1979. Diffie, W. and Hellman, M. "Privacy and Authentication: An Introduction to Cryptography," *Proc. IEEE* **67,** pp. 397–427 (March).

R. L. Rivest