CARDIFF UNIVERSITY
EXAMINATION PAPER

Academic Year: 2000/2001
Examination Period: SPRING
Examination Paper Number: CM0321
Examination Paper Title: Data Security
Duration: Three hours

Do not turn this page over until instructed to do so by the Senior Invigilator.

Structure of Examination Paper:

There are TWO pages.
There are FOUR questions in total.
There are no appendices.

The maximum mark for the examination paper is 100% and the mark obtainable for a question or part of a question is shown in brackets alongside the question.

Students to be provided with:

The following items of stationery are to be provided:
ONE answer book.

Instructions to Students:
Answer THREE questions.
1. (i) Prove that the number of \( n \)-long \( q \)-ary vectors at Hamming distance \( \leq d \) from a given vector \( x \) is

\[
\sum_{k=0}^{d} \binom{n}{k} (q - 1)^k
\]

[5]

(ii) Let \( d \) be a positive integer. Define \( A_q(n, d) \) to be the maximum value of \( M \) such that there exists a \( q \)-ary \((n, M, d)\) code, i.e. a general code over \( q \) symbols, consisting of vectors of length \( n \) and having \( M \) points with minimum Hamming distance \( d \) apart. Using the result of (i) prove the Gilbert-Varshamov bound, i.e. that whether \( d \) is odd or even

\[
A_q(n, d) \sum_{k=0}^{d-1} \binom{n}{k} (q - 1)^k \geq q^n
\]

[10]

(iii) We are given a non-symmetric binary channel for transmitting binary strings. The probability of a zero or one being received correctly is \( p_0 \) or \( p_1 \) respectively. The only thing we can do to maximise the probability of a string of \( n \)-bits being received correctly is to count the number of ones and zeros in the string and then interchange ones and zeros before transmission if this will give a larger probability of correct reception. Let \( r \) be the ratio of number of zeros over total number \( n \) of bits in the bit string. Prove that (as intuition might suggest) regardless of the values of \( n \), \( p_0 \) and \( p_1 \) the correct rule for when to interchange ones and zeros is:

"Interchange if \( r < 1/2 \) and \( p_0 > p_1 \) or if \( r > 1/2 \) and \( p_0 < p_1 " \)

[5]

2. (a) A linear code over a finite field (mod \( q \)) has a \( k \times n \) generator matrix \( G \) in standard form. Explain how to construct the parity check matrix. [2]

(b) Given the parity check matrix for a linear code explain the decoding algorithm of a well designed code (define all terms that you use). [8]

(c) The generator matrix in standard form of a binary code \( C \) is given by

\[
G = \begin{pmatrix}
1 & 0 & 0 & 1 & 1 \\
0 & 1 & 0 & 0 & 1 \\
0 & 0 & 1 & 1 & 0
\end{pmatrix}
\]

(i) Give the parity check matrix of \( C \). [5]

(ii) The code \( C \) is not well designed, and the received vector is \{1, 0, 1, 1, 1\}. Find the two coset leaders associated with this vector. What are the two most likely possibilities for the original \( k = 3 \) message bits? [5]
3. (a) Describe Shamir’s method for sharing a secret key among \( n \) persons in such a way that any subset of \( t \) or more can recover the key but no subset of less than \( t \) can do so. \[5\]

(b) In the following case \( n = 4 \) and \( t = 3 \) and Shamir’s method has been used to provide information about the secret key. Persons 1, 2, and 4 decide to recover the key. Their numbers are

\[ f(1) = 3, f(2) = 5, f(4) = 1 \text{ and all arithmetic is } (\text{mod } 31). \]

Find the polynomial and hence the key and also the value of \( f(3) \). \[5\]

(c) Now suppose that in Shamir’s method \( n = 4, t = 3 \) and \( p = 24 \) and the polynomial is

\[ f(x) = K + bx + cx^2 \equiv 11 + x + 2x^2 \pmod{24} \]

Recipients 1, 2 and 4 are issued \( f(1) = 14, f(2) = 21 \) and \( f(4) = 23 \pmod{24} \) accordingly.

(i) Show that multiple solutions exist and explain why. \[5\]

(ii) Verify that the solution corresponding to \( c = 2 \) is totally correct and find the value of \( K \) given by the solution with \( c = 10 \). \[5\]

4. (a) Define Euler’s function \( \phi(n) \) and find the value of \( \phi(4725) \). \[2\]

(b) State the Euler-Fermat theorem. \[2\]

(c) Explain the RSA algorithm for public key encryption/decryption. \[6\]

(d) A public key system which uses the RSA algorithm is based on the modulus \( n = 187 \) and encryption key \( e = 107 \).

(i) What is the decryption key? \[5\]

(ii) If an encrypted number is 7 what is the original number? \[5\]
Solutions

1. (i) **Lemma.** The number of $n$-long $q$-ary vectors at Hamming distance $\leq d$ from a given vector $x$ is

$$\sum_{k=0}^{d} \binom{n}{k} (q - 1)^k$$

**Proof.** If $y$ is such a vector then $y$ differs from $x$ in exactly $k \leq d$ places and there are

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

different ways of selecting these places. At each of the $k$ places we can choose the corresponding element of $y$ to have any one of $q - 1$ values, so for each vector $y$ we have $(q - 1)^k$ possibilities. Hence there are

$$\sum_{k=0}^{d} \binom{n}{k} (q - 1)^k$$

such vectors in all.

(ii) **Prove the Gilbert-Varshamov bound,** i.e. that whether $d$ is odd or even

$$A_q(n, d) \sum_{k=0}^{d-1} \binom{n}{k} (q - 1)^k \geq q^n$$

**Proof.** If $C$ is an $(n, M, d)$ code with the maximum number of codewords then there can be no vector which is not a codeword, and which is at least $d$ distant from all the codewords of $C$ (otherwise $M$ could be increased to at least $M + 1$). Therefore the vectors at distance $d - 1$ from all the codewords of $C$ must account for all possible vectors, some of these may be counted two or more times, so using the lemma their number is

$$A_q(n, d) \left( 1 + \binom{n}{1}(q-1) + \binom{n}{2}(q-1)^2 + ... + \binom{n}{d-1}(q-1)^{d-1} \right) \geq q^n$$

i.e.

$$A_q(n, d) \sum_{k=0}^{d-1} \binom{n}{k} (q - 1)^k \geq q^n$$

which proves the result.

(ii) The probability of $n$ bits being received correctly is $p_0^n p_1^{(1-p_0)n}$. If we interchange ones and zeros this becomes $p_0^{(1-p_0)n} p_1^n$. The question is when to swap, i.e. when is

$$p_0^n p_1^{(1-p_0)n} < p_0^{(1-p_0)n} p_1^n$$

Taking logs we have

$$n \log p_0 + (1 - r)n \log p_1 < (1 - r)n \log p_0 + rn \log p_1$$

Cancelling through by $n \geq 1$ and rearranging (noting $\log p_0 < 0$ since $p_0 < 1$) we have
\[ \frac{r}{1 - r} + \frac{\log p_1}{\log p_0} > 1 + \frac{r \cdot \log p_1}{1 - r \cdot \log p_0} \]
i.e.

\[ \frac{r}{1 - r} - 1 > \left( \frac{r}{1 - r} - 1 \right) \frac{\log p_1}{\log p_0} \]

We note that \( r/(1 - r) - 1 = (2r - 1)/(1 - r) > 0 \) if \( r > 1/2 \), in which case the above inequality yields \( \log p_0 < \log p_1 \) (note multiplying by \( \log p_0 \) again reverses the inequality) i.e. \( p_0 < p_1 \). If \( r < 1/2 \) we obtain \( \log p_0 > \log p_1 \), i.e. \( p_0 > p_1 \). This gives the required rule.
2. (a) A linear code over a finite field (mod q) has a \( k \times n \) generator matrix \( G \) in standard form. Explain how to construct the parity check matrix. \[3\]

Write the \( k \times n \) generator matrix in the standard form \( G = [I_k, A] \), where \( I_k \) is the \( k \times k \) identity matrix and \( A \) is some \( k \times (n - k) \) matrix. Then the parity check matrix is given by \( H = [-A^T, I_{n-k}] \), where all matrix components are reduced (mod q).

(b) Given the parity check matrix for a linear code explain the decoding algorithm of a well designed code (define all terms that you use).

For each received word \( x \) (row vector) compute \( y = Hx^T \) (mod q) (where T denotes transpose). Then \( y \) is the \textit{syndrome} of \( x \). If \( y = \mathbf{0} \) then no error correction is required. Otherwise, from the (syndrome, coset-leader) lookup table find the coset leader. The coset leader is the maximum likelihood error vector. To decode subtract the coset leader from the received word. The first \( k \) components of the result are the corrected version of the message bits. \[4\]

\textbf{Definition.} In the code \( C \) the \textit{coset} \( \text{Coset}(x) \) of any vector \( x \) is defined as

\[
\text{Coset}(x) = \{ z | z = x + w, w \in C \}
\] \[2\]

\textbf{Definition.} The \textit{coset leader} is the vector in the coset which has the least number of non-zero (mod q) components. \[2\]

In a well designed code the coset leader is unique for every coset. Note that every element of a coset has the same syndrome and before the code is used we construct a (syndrome, coset-leader) lookup table.

(c) The generator matrix in standard form of a binary code is given by

\[
G = \begin{pmatrix}
1 & 0 & 0 & 1 & 1 \\
0 & 1 & 0 & 0 & 1 \\
0 & 0 & 1 & 1 & 0
\end{pmatrix}
\]

(i) Give the parity check matrix of \( C \). \[5\]

\[
H = \begin{pmatrix}
1 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 0 & 1
\end{pmatrix}
\]

(ii) The code is not well designed, and the received vector is \( \{1, 0, 1, 1, 1\} \). Find the two coset leaders associated with this vector. What are the two most likely possibilities for the original \( k = 3 \) message bits? \[5\]

The coset associated with \( \{1, 0, 1, 1, 1\} \) is

\[
\{ \{1, 0, 1, 1, 1\}, \{1, 0, 0, 0, 1\}, \{1, 1, 1, 1, 0\}, \{1, 1, 0, 0, 0\}, \{0, 0, 1, 0, 0\}, \{0, 0, 0, 1, 0\}, \{0, 1, 1, 0, 1\}, \{0, 1, 0, 1, 1\} \}
\]

The coset leader is not unique and is either \( \{0, 0, 1, 0, 0\} \) or \( \{0, 0, 0, 1, 0\} \) (so the code is not well designed) and the transmitted \( k = 3 \) bits are equally likely to be \( \{1, 0, 0\} \) or \( \{1, 0, 1\} \).
3. (a) **Shamir’s method.** The trusted administrator chooses a polynomial of degree \( t - 1 \) with integer coefficients \((\text{mod } p)\), where \( p \) is a large prime \( p \gg n \). The constant term in this polynomial is \( K \) (the secret number), the other coefficients are chosen at random \((\text{mod } p)\).

Let this polynomial be

\[
f(x) = K + a_1x + a_2x^2 + \ldots + a_{t-1}x^{t-1}
\]

The \( n \) recipients are allocated the identifying labels 1, 2, ..., \( n \) and the \( i \)th recipient is told the value of \( f(i) \) \((\text{mod } p)\). In order to recover the secret number \( K \) it is sufficient to find the polynomial \( f(x) \), because then \( K = f(0) \).

- This can be done if and only if \( t \) pairs of values \((i, f(i))\) are available.

(b) From the question

\[
\begin{align*}
\text{(1)} & \quad f(1) = K + a_1 + a_2 = 3 \pmod{31} \\
\text{(2)} & \quad f(2) = K + 2a_1 + 4a_2 = 5 \pmod{31} \\
\text{(4)} & \quad f(4) = K + 4a_1 + 16a_2 = 1 \pmod{31}
\end{align*}
\]

Solving these congruences by Gaussian elimination gives \( \{K, a_1, a_2\} = \{19, 6, 9\} \). Therefore secret key is \( K = 19 \) and \( f(3) = 25 \pmod{31} \).

(c) Now suppose that in Shamir’s method \( n = 4, t = 3 \) and \( p = 24 \) and the polynomial is

\[
f(x) = K + bx + cx^2 = 11 + x + 2x^2 \pmod{24}
\]

Recipients 1, 2 and 4 are issued \( f(1) = 14, f(2) = 21 \) and \( f(4) = 23 \pmod{24} \) accordingly.

(i) **Show that multiple solutions exist and explain why.**

Each of these 6 possible values for \( c \) leads to values of \( b \) and \( K \). But only one solution, the one corresponding to \( c = 2 \), i.e. \( \{K, b, c\} = \{11, 1, 2\} \) is totally correct. The other solutions satisfy all the conditions but may lead to the wrong value of \( K \). Thus if we take \( c = 10 \) we find that \( b = 1 \) and

\[
K = 3, \text{ i.e. } f(x) = 3 + x + 10x^2 \pmod{24}
\]

which provides the right values at \( x = 1, 2 \) and 4 but which is nevertheless incorrect.
4. (a) $\varphi(n)$ is the number of integers $k$ ($1 \leq k \leq n$) such that $k$ and $n$ are coprime (have no non-trivial factors in common).

$4725 = 25 \times 189 = 25 \times 27 \times 7 = 3^3 \times 5^2 \times 7$. So that using the formula given in the notes

$$\varphi(4725) = 3^3 \times 5^2 \times 7 \left(1 - \frac{1}{3}\right) \left(1 - \frac{1}{5}\right) \left(1 - \frac{1}{7}\right)$$

$$= 3^2 \times 5 \times 2 \times 4 \times 6 = 2160$$

(b) Euler-Fermat theorem. Given coprime integers $a, m$ then

$$a^{\varphi(m)} \equiv 1 \pmod{m}$$

(c) The RSA public key encryption system The system is based on choosing three integers ($e, p, q$) where $p$ and $q$ are large primes ($= 10^{200}$) and $e$ is coprime to $(p - 1)(q - 1)$.

Let $n = pq$. Then $e$ and $n$ are made public and only the owner of the system knows $p$ and $q$.

To encrypt the message:

1. Convert the message into a stream of decimal digits (the method for doing this is not secret, indeed it must be known).
2. Cut the stream into blocks of equal length and convert these to decimal numbers $m_1, m_2, \ldots$, where $m_i < n$.
3. Raise each number $m_i$ to the power $e$ and replace $m_i^e$ by its remainder $r_i \pmod{n}$ (so $m_i^e \equiv r_i \pmod{n}$ in mathematical notation).
4. The encrypted message is then the stream $r_1, r_2, r_3, \ldots$

To decrypt the message:

1. Let $d$ be the number that has the property

$$ed \equiv 1 \pmod{(p-1)(q-1)}$$

(this is not too hard to compute, it can be done by using Euclid’s algorithm, and can be worked out once and for all and saved).
2. Then

$$r_i^d \equiv (m_i^e)^d \equiv m_i^{ed} \equiv m_i^{\varphi(n)+1} \equiv m_i \pmod{n}$$

by the Euler-Fermat theorem, and the decrypted message is $m_1, m_2, m_3, \ldots$

(d) Factorising 187 gives $n = 11 \times 17 = 187$. So $p = 11, q = 17$. So $\varphi(n) = 10 \times 16 = 160$. The public key exponent is $e = 107$. Hence the decryption key $d$ is the solution to $107d \equiv 1 \pmod{160}$.

(i) Using Euclid’s algorithm (for example)

$$160 = 1 \times 107 + 53$$

$$107 = 2 \times 53 + 1$$

$$53 = 53 \times 1 + 0$$

So $1 = 107 \times 2 \times 53 = 107 \times 2 \times (160 - 107) = 3 \times 107 - 2 \times 160 \equiv 3 \times 107 \pmod{160}$ and the required
decryption key is $d = 3$.

[Or can use modular arithmetic and Euler-Fermat theorem to work out $139^{e-1}$ in a similar way to below or can just do it by trying small values for $d$]

(ii) Since we have found $d = 3$ when $e = 107$ the decrypted value of 7 is $7^3 \text{ (mod 187)}$.

Now $7^3 = 343 \equiv 156 \text{ (mod 187)}$. So the decrypted value is 156.