Written exam in TSIT03 Cryptology

14:00–18:00, 28 October 2015

Jan-Åke Larsson
Institutionen för Systemteknik,
Linköpings Universitet

Permitted equipment: General dictionaries between English and another language without personal notes. (Thus not specific scientific dictionaries with or without formulæ.)

Solutions: Solutions will be posted on the course home page after the exam.

Grading: Grade n requires at least $6n$ points,
this translates to ECTS grade $20 - 2n$ (converted to hexadecimal)

Other information: Answers can be in English or Swedish

1. Information theory
   (a) Which are the most common and second-most common letter in English? (1p)
   (b) Why is the probability of the digram "TH" not equal to the probability of "T" multiplied with the probability of "H"? What is the probability-theory notion called that captures this property? (1p)
   (c) Give the formula for Shannon entropy in terms of a probability distribution $P(X = x)$. (1p)
   (d) The Shannon entropy connects with a particular kind of code that is adapted to the source distribution $P(X = x)$. What is the code called, and what is the key property of the code (in words, no mathematics is needed)? (2p)
2. Block ciphers
   (a) What block size does AES have? What is the recommended key length for AES? (2p)
   (b) AES uses calculation in the finite field GF(256) with the primitive polynomial $X^8 + X^4 + X^3 + X + 1$. Calculate $X^2 * (X^7 + X^6 + X^3 + X + 1)$ in the field. (1p)
   (c) Draw a diagram of Cipher Feedback mode. Give two good properties of the mode. (2p)
   (d) What is the Horton principle (according to Bruce Schneier)? (1p)

3. Asymmetric ciphers

   (a) How do you choose the RSA parameters p, q, n, e, and d? (2p)

   (b) How is encryption and decryption done in RSA? (1p)

   (c) How is signing and verification done in RSA? (1p)

   (d) In RSA, what is a suitable number of bits for the public key n to give reasonable security? (1p)

   (e) What attack is possible if two public keys n_1 and n_2 happen to share one of the primes p? Is this a relevant question for existing implementations? (2p)

4. Message authentication, and digital signatures

   (a) What is the technical difference between a Message Authentication Code and a digital signature? (2p)

   (b) What are the effects of this, in terms of who can create a MAC, and who can create a signature? Who can verify a MAC, and who can verify a signature? (2p)

   (c) What is a "blind signature"? (1p)

   (d) Describe how a blind signature can be created using RSA. (2p)

5. Key distribution

   (a) Describe Diffie-Hellman key exchange in modular arithmetic, and list public and secret parameters. How are the personal public parameters calculated from the secret and the general parameters? How is the shared key calculated? (3p)

   (b) Draw a diagram of the Station-To-Station protocol, and describe the steps. What are the differences to ordinary Diffie-Hellman key exchange? (3p)

6. Discrete log one-way functions

   (a) You should have listed a prime number p as one of the public parameters in 5(a). What influences the length (in bits) of p, and what is the currently recommended length? (1p)

   (b) Under the assumption that one particular length of p is appropriate for several users, give two arguments why it is a good idea to use the same p. (1p)

   (c) Under the assumption that one particular length of p is appropriate for several users, give one argument why it is NOT a good idea to use the same p. (1p)

   (d) Do current implementations follow c) or d), and is this a good idea? Why? (2p)

7. Zero knowledge

   Peggy claims to know a square root s of $t \bmod n = pq$, where p and q are large primes. How would Peggy convince Victor that the claim is true without revealing the square root s? (3p)

Solutions

1. Information theory

   (a) Which are the most common and second-most common letter in English? (1p)

   - "E" and "T", in that order

   (b) Why is the probability of the digram "TH" not equal to the probability of "T" multiplied with the probability of "H"? What is the probability-theory notion called that captures this property? (1p)

   - The combination "TH" is much more common than one would think: on average, every fourth "H" is in "TH". The letters in a text depend on each other, the distribution of a letter that follows another letter depends on which letter is first.

   (c) Give the formula for Shannon entropy in terms of a probability distribution $P(X = x)$. (1p)

   - $H(X) = -\sum_x P(X = x) \log_2 P(X = x)$

   (d) The Shannon entropy connects with a particular kind of code that is adapted to the source distribution $P(X = x)$. What is the code called, and what is the key property of the code (in words, no mathematics is needed)? (2p)

   - Huffman code
   - Letters that have high probability are given short code words, and letters that have low probability are given long code words (the length is chosen to give as short average as possible)

2. Block ciphers

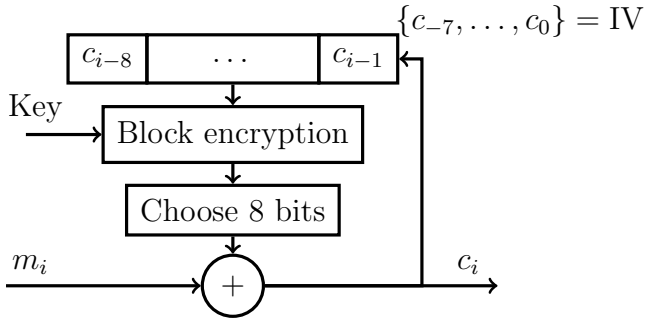
   (a) What block size does AES have? What is the recommended key length for AES? (2p)

   - Block size is 128 bits, and the possible (and recommended) key sizes are 128, 192 and/or 256 bits

   (b) AES uses calculation in the finite field GF(256) with the primitive polynomial $X^8 + X^4 + X^3 + X + 1$. Calculate $X^2 * (X^7 + X^6 + X^3 + X + 1)$ in the field. (1p)

   - $X^2 * (X^7 + X^6 + X^3 + X^+1) = X^9 + X^8 + X^5 + X^3 + X^2$
     $= X * (X^8 + X^7 + X^4 + X^2 + X)$
     $= X * (X^8 + X^7 + X^4 + X^2 + X + X^8 + X^4 + X^3 + X + 1)$
     $= X * (X^7 + X^3 + X^2 + 1) = X^8 + X^4 + X^3 + X = 1$

   (c) Draw a diagram of Cipher Feedback mode. Give two good properties of the mode. (2p)



- CFB can encrypt smaller message pieces than whole blocks
- CFB can be used to test message integrity

(d) What is the Horton principle (according to Bruce Schneier)? (1p)

- Authenticate what is meant, not what is being said

3. Asymmetric ciphers

(a) How do you choose the RSA parameters p, q, n, e, and d? (2p)

- The parameters p and q are random very large primes, and $n = pq$. An important number here is $\varphi(n) = (p-1)(q-1)$, and e is chosen so that $\gcd(e, (p-1)(q-1)) = 1$. The final parameter d is chosen so that $ed = 1 \mod (p-1)(q-1)$.

(b) How is encryption and decryption done in RSA? (1p)

- With message m and cryptotext c, encryption and decryption are: $c = m^e \mod n$; $m = c^d \mod n$

(c) How is signing and verification done in RSA? (1p)

- With message m and signature s, signing and verification are: $s = m^d \mod n$, compare m and $s^e \mod n$.

(d) In RSA, what is a suitable number of bits for the public key n to give reasonable security? (1p)

- 2-3kbit of key is currently estimated to give 20-30 years protection

(e) What attack is possible if two public keys n_1 and n_2 happen to share one of the primes p? Is this a relevant question for existing implementations? (2p)

- If two public keys n_1 and n_2 happen to share one of the primes p, the extended Euclidean algorithm gives you $\gcd(n_1, n_2) = p$ immediately. This is relevant, because it seems some implementations choose between very few primes. Lenstra, Hughes et al. were able to factor 0.2% of a sample of public keys gathered from the Internet in early 2012.

4. Message authentication, and digital signatures

(a) What is the technical difference between a Message Authentication Code and a digital signature? (2p)

- A digital signature is created in an asymmetric-key system while a MAC is created in a symmetric-key system

(b) What are the effects of this, in terms of who can create a MAC, and who can create a signature? Who can verify a MAC, and who can verify a signature? (2p)

- A digital signature can only be created by the one who knows the secret key, but can be verified by anyone that has the public key. A MAC can be created by anyone who can verify its correctness.

(c) What is a "blind signature"? (1p)

- Bob wants to prove that he has created a document at a certain time, but keep it secret, and Alice agrees to help him. She signs the document while the contents is hidden to her. Bob still gets a valid signature for his document.

(d) Describe how a blind signature can be created using RSA. (2p)

- Alice sets up standard RSA, keeping d for herself.
- Bob chooses a random integer k, and gives Alice the message

$$t = k^e m \bmod n$$

- The number t is random to Alice, but she signs the message and gives the signature to Bob

$$s = t^d = k^{ed} m^d = k m^d \bmod n$$

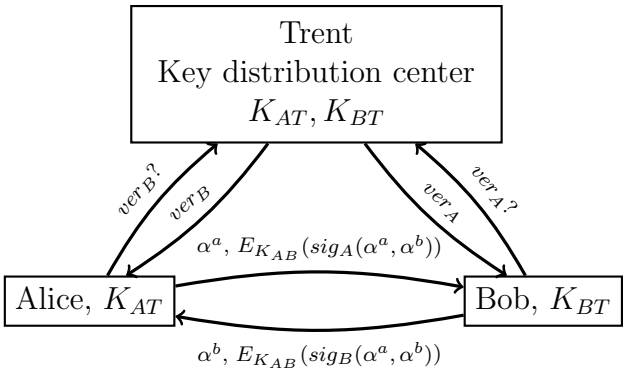
- Bob can now divide by k and retrieve m^d, Alice's signature for m.

5. Key distribution

(a) Describe Diffie-Hellman key exchange in modular arithmetic, and list public and secret parameters. How are the personal public parameters calculated from the secret and the general parameters? How is the shared key calculated? (3p)

- The public parameters are a prime p and a primitive root α mod p. Alice's secret parameter is a random integer a, while Bob's secret parameter is a random integer b. Alice makes α^a mod p public, while Bob makes α^b mod p public. Both can now create the shared key $k = (\alpha^a)^b = (\alpha^b)^a$ mod p

(b) Draw a diagram of the Station-To-Station protocol, and describe the steps. What are the differences to ordinary Diffie-Hellman key exchange? (3p)



- Do Diffie-Hellman, but also exchange encrypted versions (1p) of a signature of both public keys
- Retrieve the verification procedure for the counterpart from a third trusted party (should be in the figure)

6. Discrete log one-way functions

(a) You should have listed a prime number p as one of the public parameters in 5(a). What influences the length (in bits) of p, and what is the currently recommended length? (1p)

- The length determines the security, 3kbit is the current recommendation for $\sim$30 years security.

(b) Under the assumption that one particular length of p is appropriate for several users, give two arguments why it is a good idea to use the same p. (1p)

- They can use the same secret key in several connections, and
- They do not need to negotiate which prime number p to use

(c) Under the assumption that one particular length of p is appropriate for several users, give one argument why it is NOT a good idea to use the same p. (1p)

- If many users use the same p it may suddenly be worth the effort to compute the discrete log for that p.

(d) Do current implementations follow c) or d), and is this a good idea? Why? (2p)

- They use the same p (at a given security), and no, this is not a good idea, the rumor is that NSA has the discrete log function for one of the supposedly secure ps.

7. Zero knowledge

Peggy claims to know a square root s of t mod $n = pq$, where p and q are large primes. How would Peggy convince Victor that the claim is true without revealing the square root s? (3p)

- (a) Peggy uses a random number r, computes $x = r^2$ and sends x to Victor
  (b) Victor asks either for the square root of x, or the square root of xt
  (c) If Victor asked for the square root of x, Peggy returns $y = r$. If he asked for the square root of xt, Peggy returns $y = rs$.
  (d) Victor checks that $y^2 = x$ or $y^2 = xt$ depending on what he asked for
  (e) If Peggy can do this enough times (she must use different random numbers r each time), Victor will conclude that Peggy knows s, the square root of t