CS 290
Host-based Security and Malware

Christopher Kruegel
chris@cs.ucsb.edu

Cryptography

Cryptography

- (One) definition of cryptography

  Mathematical techniques related to aspects of information security such as

  - confidentiality
    - keep content of information from all but authorized entities
  - integrity
    - protect information from unauthorized alteration
  - authentication
    - identification of data or communicating entities
  - non-repudiation
    - prevent entity from denying previous commitments or actions

Taxonomy

- Unkeyed primitives
  - hash functions
  - random sequences
- Symmetric-key primitives
  - block ciphers
  - stream ciphers
  - signatures
  - pseudorandom sequences
- Public-key primitives
  - public-key ciphers
  - signatures

Symmetric-key Cryptography

- Consider an encryption scheme with key pair (e,d)
  - scheme is called a symmetric-key scheme
    if it is "relatively" easy to obtain d when e is know
  - often e = d
- Block cipher
  - break up plaintext into strings (blocks) of fixed length t
  - encrypt one block at a time
  - uses *substitution* and *transposition (permutation)* techniques
- Stream Cipher
  - special case of block cipher with block length t = 1
  - however, substitution technique can change for every block
  - key stream (e_1, e_2, e_3, …)

Public-key Cryptography

- Consider an encryption scheme with key pair (e,d)
  - scheme is called a public-key scheme
    if it is computationally infeasible to determine d when e is known

- In public-key schemes, E_e is usually a *trapdoor one-way function* and d is the trapdoor

- One-way function
  - A function f: X → Y is called a trapdoor function, if f(x) is “easy” to compute for all x ∈ X, but for most y ∈ Y, it is infeasible to find a x such that f(x) = y.
  - calculating the exponentiation of an element a in a finite field [a^p (mod n)]
  - multiplication of two large prime numbers [n = p*q]

Public-key Cryptography

- Trapdoor one-way function
  - A trapdoor function $f: X \rightarrow Y$ with the additional property that given some additional information (called the trapdoor information) it becomes feasible for all $y \in Y$ to find a x such that $f(x) = y$.
- No longer necessary to transfer a secret key over a secure channel
- Significant problem is binding of public key to a certain person (authentication)
  - otherwise, an attacker can substitute his own public key for the victim's key
- Key certificates are needed
  - public key infrastructure (PKI)
  - idea is to cryptographically bind a public key to a certain entity via certificates
  - certificates commonly issued by certification authorities (CAs)
  - chain of trust is traced to a root CA (whose public key must be known by all participants)

Cryptanalysis

- Fundamental
  - all alphabets and the encryption/decryption functions are public knowledge
  - only the selection of the key pair remains secret
- System is breakable
  - if a third party can (without the knowledge of the key pair) systematically recover plaintext from corresponding ciphertext *within some appropriate time frame*
  - exhaustive key search must be made impossible
- Cryptanalysis
  - study of techniques to defeat cryptographic techniques

Cryptanalysis

- Different model (power) of adversary assumed
  - Known-Ciphertext Attack (KCA)
    - you only know the ciphertext
    - requires you know something about the plaintext (e.g., it's English text, an MP3, C source code, …)
    - this is the model for the Sunday cryptograms which use substitution
  - Known-Plaintext Attack (KPA)
    - you have some number of plaintext-ciphertext pairs, but you cannot choose which plaintexts you would like to see
  - Chosen-Plaintext Attack (CPA)
    - you get to submit plaintexts of your choice to an encryption oracle (black box) and receive the ciphertexts in return

Cryptanalysis

- Known-Ciphertext Attack (KCA)
  - weak attack model
  - works only when weak ciphers are used (simple substitution algorithms)

- Attacker can use frequency analysis
  - assumption is that symbols (letters) do not appear with the same frequency in the plaintext
  - this assumption holds with high probability if natural language texts are encrypted
  - in the English language, most frequent letters are E T N R O A S (in this order)

- Attack
  - analyze frequency of symbols in ciphertext
  - assume that symbols with high frequency correspond to frequent letters
  - try to reconstruct plaintext

Cryptanalysis

- Frequency analysis has to be adapted when poly-alphabetic substitution is used
  - in this case, the number of different permutations is most difficult part to find out
  - once the number N of different permutations is known, the ciphertext can be divided into N groups
  - apply frequency analysis individually for each group

- Example with 3 permutations (from the Vigenere cipher)

plaintext : `THISC IPHER ISCER TAINL YNOTS ECURE`

ciphertext: `WOSVJ SSOOU PCFLB WHSQS IQVDV LMXYO`

Group 1: `W, V, S, U, F, W, Q, Q, V, X | V(S), W(T), Q(N)`

Group 2: `O, J, O, P, L, H, S, V, L, Y | O(H)`

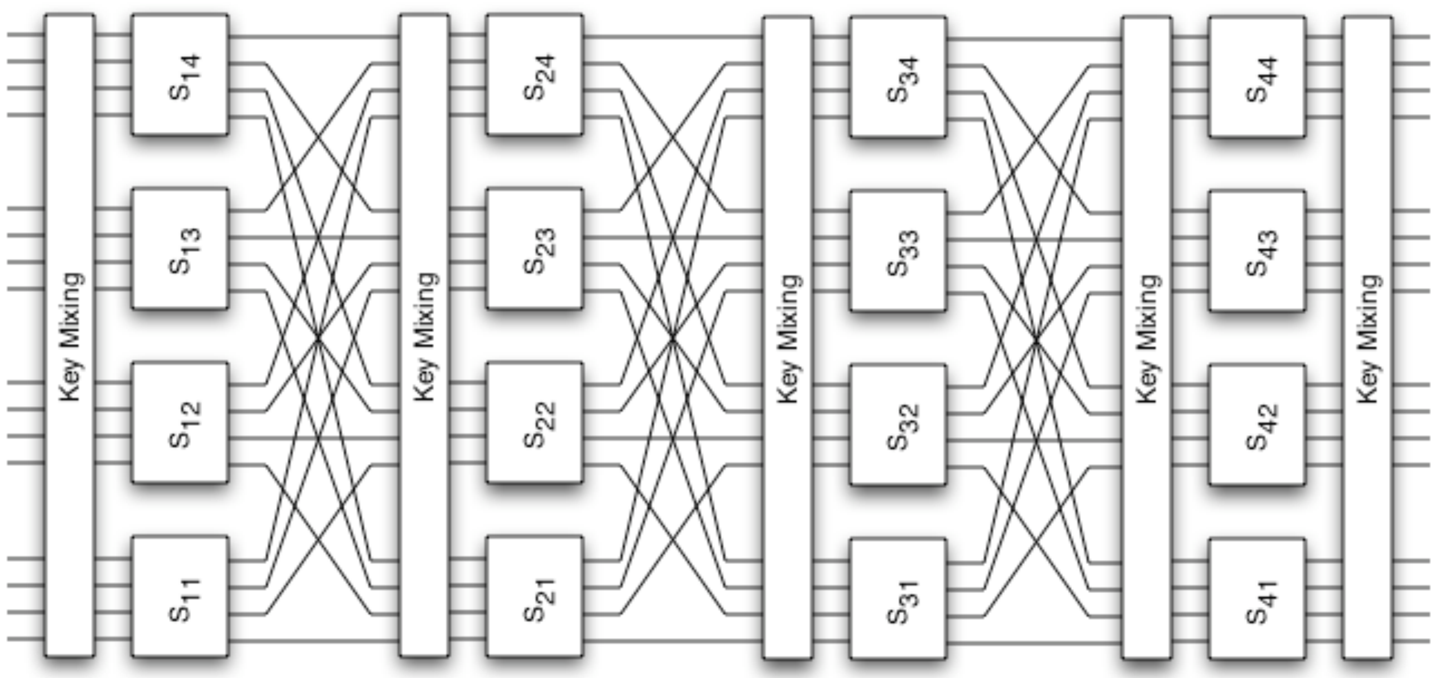
Group 3: `S, J, O, C, B, S, I, D, M, O | S(I), O(E)`

Cryptanalysis

- Better ciphers require more advanced attack techniques
- Two well-known techniques against secret-key block ciphers are
  - *linear cryptanalysis*
    - developed 1993 by Matsui
  - *differential cryptanalysis*
    - discovered three times by NSA, IBM, and Biham and Shamir
- We use a simple four round SPN as example
  - 16 bit key, 16 bit block size
  - S-Box with the following mapping (4 bit input → 4 bit output)

0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
E	4	D	1	2	F	B	8	3	A	6	C	5	9	0	7

Cryptanalysis



Cryptanalysis

- Linear cryptanalysis
  - known plaintext attack
  - exploits high probability occurrences of linear relationships between plaintext, ciphertext, and key bits
  - linear with regards to bitwise operation modulo 2 (i.e., XOR)
  - expressions of form $X_{i1} \oplus X_{i2} \oplus X_{i3} \oplus \ldots \oplus X_{iu} \oplus Y_{j1} \oplus Y_{j2} \oplus \ldots \oplus Y_{jv} = 0$

    X_i = i-th bit of input plaintext [X_1, X_2 , ...]

    Y_j = j-th bit of output ciphertext [Y_1, Y_2, …]

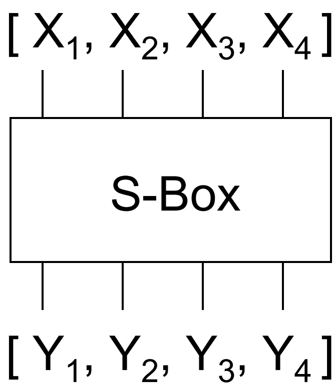
  - for a perfect cipher, such relationships hold with probability 1/2
  - for vulnerable cipher, the probability p might be different from 1/2

  → a bias |p - 1/2| is introduced

Linear Cryptanalysis

- 2 steps
  - analyze the linear vulnerability of a single S-Box
  - connect the output of an S-Box to the input of the S-Box in the next round and “pile up” probability bias

- To analyze a single S-Box, check all possible linear approximations



0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
E	4	D	1	2	F	B	8	3	A	6	C	5	9	0	7

Linear Cryptanalysis

X1	X2	X3	X4	Y1	Y2	Y3	Y4	X1 ⊕ X3 ⊕ X4 = Y2	X2 = Y2 ⊕ Y4
0	0	0	0	1	1	1	0	F	F
0	0	0	1	0	1	0	0	T	F
0	0	1	0	1	1	0	1	T	T
0	0	1	1	0	0	0	1	T	F
0	1	0	0	0	0	1	0	T	F
0	1	0	1	1	1	1	1	T	F
0	1	1	0	1	0	1	1	F	T
0	1	1	1	1	0	0	0	T	F
1	0	0	0	0	0	1	1	F	F
1	0	0	1	1	0	1	0	T	T
1	0	1	0	0	1	1	0	F	F
1	0	1	1	1	1	0	0	T	F
1	1	0	0	0	1	0	1	T	F
1	1	0	1	1	0	0	1	T	T
1	1	1	0	0	0	0	0	T	F
1	1	1	1	0	1	1	1	T	F

Linear Cryptanalysis

- Linear approximations with many true or many false entries are interesting

  $p(X1 \oplus X3 \oplus X4 = Y2) = 12/16 = 0.75$ [bias = 0.25]

  $p(X2 = Y2 \oplus Y4) = 4/16 = 0.25$ [bias = -0.25]

- How to connect probabilities between different rounds?

  consider the following equations, when bias of X1 is b1, and bias of X2 is b2

  $$p(X1 \oplus X2 = 0) = p(X1)*p(X2) + (1-p(X1))*(1-p(X2))$$
  $$= (1/2+b1)*(1/2+b2) + (1/2-b1)*(1/2-b2)$$
  $$= 1/2 + 2*b1*b2$$

Linear Cryptanalysis

- Now, we show how we can eliminate intermediate variables

$p(X1 \oplus X2 = 0) = 1/2 + b1,2$

$p(X2 \oplus X3 = 0) = 1/2 + b2,3$

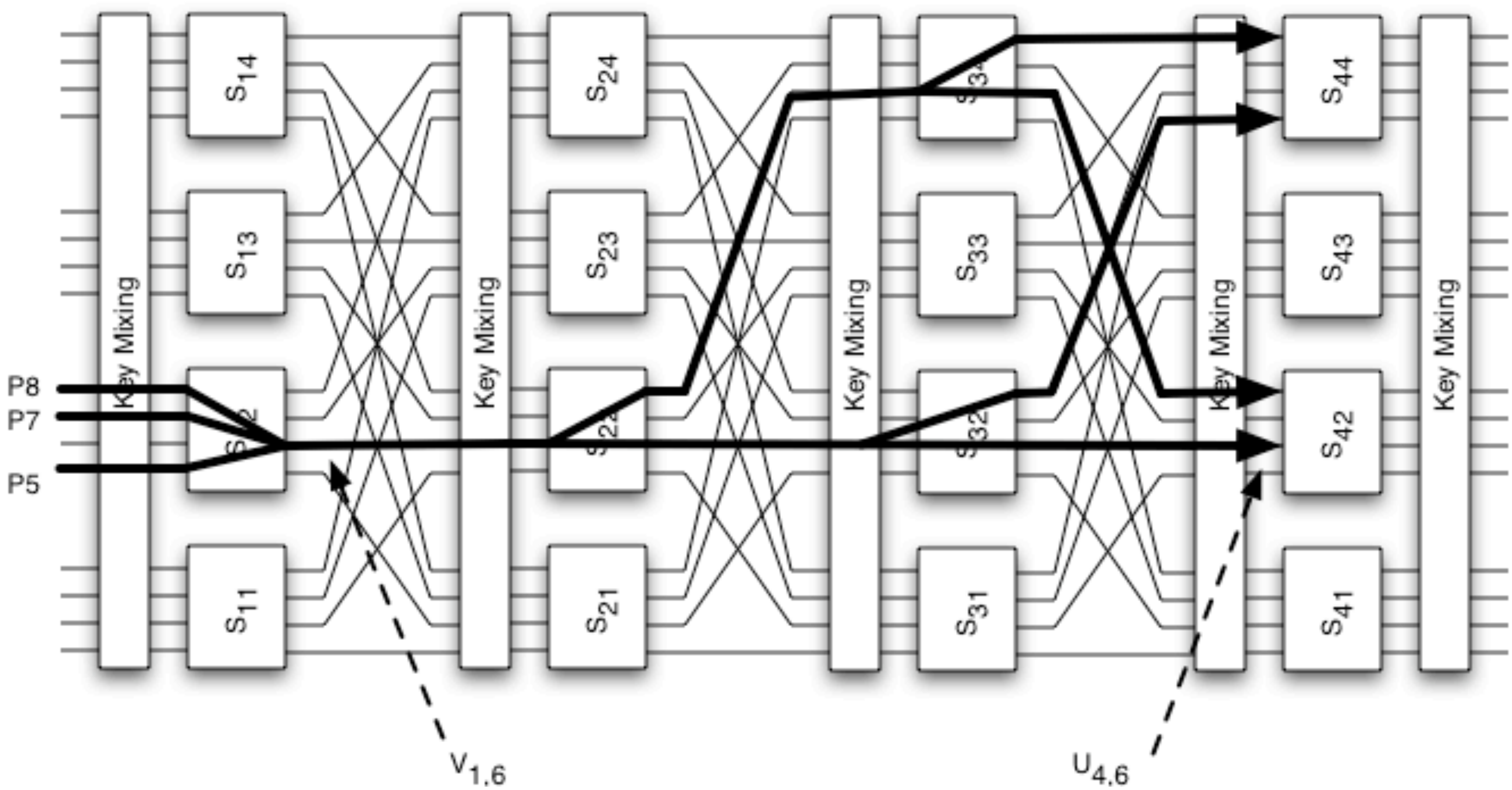
$p(X1 \oplus X3 = 0) = p([X1 \oplus X2] \oplus [X2 \oplus X3] = 0)$

$= p(X1 \oplus X3 = 0)$

$= 1/2 + 2*b1,2 *b2,3$

- Let $U_i(V_i)$ represent the 16-bit block of bits at the input (output) of the S-Box of round i. Then, let $U_{i,k}$ denote the k-th bit of the i-th round of the cipher. Similarly, let K_i represent the key of round i.

Linear Cryptanalysis



Linear Cryptanalysis

- With probability 0.75 (and bias = 0.25), we have

  $V_{1,6} = U_{1,5} \oplus U_{1,7} \oplus U_{1,8}$

  $= (P_5 \oplus K_{1,5}) \oplus (P_7 \oplus K_{1,7}) \oplus (P_8 \oplus K_{1,8})$

- For the second round, we obtain with probability 0.25 (bias = -0.25)

  $V_{2,6} \oplus V_{2,8} = U_{2,6} \oplus K_{2,6}$

- Because $U_{2,6} = V_{1,6}$, we can connect these two equations and get

  $V_{2,6} \oplus V_{2,8} = (P_5 \oplus K_{1,5}) \oplus (P_7 \oplus K_{1,7}) \oplus (P_8 \oplus K_{1,8}) \oplus K_{2,6}$

  which can be rewritten as

  $V_{2,6} \oplus V_{2,8} \oplus P_5 \oplus P_7 \oplus P_8 \oplus K_{1,7} \oplus K_{1,8} \oplus K_{2,6} = 0$

  This holds with a probability (see before) of 1/2 + 2*0.25*(-0.25) = 0.375

Linear Cryptanalysis

- We continue to eliminate intermediate variables in intermediate rounds to obtain

  $U4,6 \oplus U4,8 \oplus U4,14 \oplus U4,16 \oplus P5 \oplus P7 \oplus P8 \oplus \sum = 0$

  where $\sum$ is a constant factor (either 0 or 1 that depends on a number of key bits)

  This equation holds with a probability of 15/32 (with a bias of -1/32).

  Because $\sum$ is fixed, we know the following linear approximation of the cipher that holds with probability 15/32 or 17/32 (depending on whether $\sum$ is 0 or 1):

  $U4,6 \oplus U4,8 \oplus U4,14 \oplus U4,16 \oplus P5 \oplus P7 \oplus P8 = 0$

Linear Cryptanalysis

- Given an equation that relates the input to the last round of S-Boxes to the plaintext, how can we get the key?
- We attack parts of the key (called target subkey) of the last round, in particular those bits of the key that connect the output of our S-Boxes of interest with the ciphertext

  Given the equation $U_{4,6} \oplus U_{4,8} \oplus U_{4,14} \oplus U_{4,16} \oplus P_5 \oplus P_7 \oplus P_8 = 0$, we look at the 8 bits $K_{5,5}$ - $K_{5,8}$ and $K_{5,13}$-$K_{5,16}$

Linear Cryptanalysis

- Idea
  - for a large number of ciphertext and plaintext pairs, we first feed the ciphertext back into the active S-Boxes S_{42} and S_{44}
  - because we do not know the target subkey, we have to repeat this feedback procedure for all possible 256 keys
  - for each subkey, we keep a count on how often the linear equation holds
  - when the wrong subkey is used
    - the equation will hold with probability 1/2 (similar to using random values)
  - when the correct subkey is used
    - the equation will hold with more or less often than 1/2 (depending on the bias)

→ after all pairs of plaintext and ciphertext are checked, we take the subkey with the count that differs most from 1/2

Differential Cryptanalysis

- Similar in spirit to linear cryptanalysis
- Chosen plaintext attack
- Instead of linear relationships, sensitivity to modifications of the input are analyzed
  - when certain bits of the input are changed, how does the output change
  - for an ideal cipher, a single bit flip in the input makes all output bits change with a probability of 1/2
  - not always the case
  - probabilistic attack that targets the key of the last round

Conclusion

- Cryptographic schemes
  - symmetric-key cryptography
    - block ciphers
    - DES, SPN, Feistel networks
    - stream ciphers
  - public-key cryptography
    - RSA

- Cryptanalysis
  - frequency analysis
  - linear and differential cryptanalysis

    tutorial on this topic available under http://www.engr.mun.ca/~howard/