CHƯƠNG 5

CHUỖI BIT GIẢ NGẪU NHIÊN

PSEUDORANDOM BITS AND SEQUENCES

Nội dung

- Sinh bit ngẫu nhiên và giả ngẫu nhiên

Nội dung

- Sinh bit ngẫu nhiên và giả ngẫu nhiên
- Kiểm định tính ngẫu nhiên

Nội dung

- Sinh bit ngẫu nhiên và giả ngẫu nhiên
- Kiểm định tính ngẫu nhiên
- Thuật toán sinh bit giả ngẫu nhiên an toàn trong mã hóa

Sinh bit ngẫu nhiên, giả ngẫu nhiên

- **Sinh bit ngẫu nhiên** là thiết bị hay thuật toán cho dãy các bit độc lập thống kê và không chệch theo nghĩa xác suất cho bit 0 và bit 1 là như nhau.

Sinh bit ngẫu nhiên, giả ngẫu nhiên

- **Sinh bit ngẫu nhiên** là thiết bị hay thuật toán cho dãy các bit độc lập thống kê và không chệch theo nghĩa xác suất cho bit 0 và bit 1 là như nhau.
- Một dãy bit ngẫu nhiên có thể được sử dụng để sinh ra số ngẫu nhiên phân phối đều.

Sinh bit ngẫu nhiên, giả ngẫu nhiên

- **Sinh bit ngẫu nhiên** là thiết bị hay thuật toán cho dãy các bit độc lập thống kê và không chệch theo nghĩa xác suất cho bit 0 và bit 1 là như nhau.

- Một dãy bit ngẫu nhiên có thể được sử dụng để sinh ra số ngẫu nhiên phân phối đều.

- **Sinh bit giả ngẫu nhiên** là thuật toán tất định mà với một dãy bit độ dài ngẫu nhiên k cho trước thì cho dãy bit (độ dài $\geq k$) có "tính ngẫu nhiên". Dãy bit đầu vào gọi là hạt gieo (seed), kết quả cho ra là dãy bit giả ngẫu nhiên.

Định nghĩa

- *Một sinh bit giả ngẫu nhiên được gọi là vượt qua kiểm định thống kê thời gian đa thức nếu không có thuật toán với thời gian đa thức nào có thể phân biệt chính xác giữa một dãy kết quả đó và dãy ngẫu nhiên có cùng độ dài với xác suất* $\geq \frac{1}{2}$.

Định nghĩa

- *Một sinh bit giả ngẫu nhiên được gọi là vượt qua kiểm định thống kê thời gian đa thức nếu không có thuật toán với thời gian đa thức nào có thể phân biệt chính xác giữa một dãy kết quả đó và dãy ngẫu nhiên có cùng độ dài với xác suất* $\geq \frac{1}{2}$.
- *Một sinh bit giả ngẫu nhiên được gọi là vượt qua kiểm định bit kế tiếp nếu không có thuật toán với thời gian đa thức nào mà với l bit đầu tiên biết trước mà có thể dự đoán được bit thứ l+1 với xác suất* $\geq \frac{1}{2}$.

Định nghĩa

- *Một sinh bit giả ngẫu nhiên được gọi là vượt qua kiểm định thống kê thời gian đa thức nếu không có thuật toán với thời gian đa thức nào có thể phân biệt chính xác giữa một dãy kết quả đó và dãy ngẫu nhiên có cùng độ dài với xác suất* $\geq \frac{1}{2}$.
- *Một sinh bit giả ngẫu nhiên được gọi là vượt qua kiểm định bit kế tiếp nếu không có thuật toán với thời gian đa thức nào mà với l bit đầu tiên biết trước mà có thể dự đoán được bit thứ l+1 với xác suất* $\geq \frac{1}{2}$.
- *Một sinh bit giả ngẫu nhiên mà vượt qua kiểm định bit kế tiếp thì được gọi là sinh bit giả ngẫu nhiên an toàn trong mã hóa.*

Kỹ thuật sinh dãy bit ngẫu nhiên

- Sinh bit ngẫu nhiên dựa theo phần cứng

Kỹ thuật sinh dãy bit ngẫu nhiên

- Sinh bit ngẫu nhiên dựa theo phần cứng
- **Sinh bit ngẫu nhiên dựa theo phần mềm**

Kỹ thuật sinh dãy bit ngẫu nhiên

- Sinh bit ngẫu nhiên dựa theo phần cứng
- **Sinh bit ngẫu nhiên dựa theo phần mềm**
- **Sử dụng kỹ thuật de-skewing tránh biases**

Sinh bit dãy ngẫu nhiên dựa theo phần mềm

- Dựa vào thời gian của hệ thống

Sinh bit dãy ngẫu nhiên dựa theo phần mềm

- Dựa vào thời gian của hệ thống
- Thời điểm ấn phím hoặc di chuyển chuột

Sinh bit dãy ngẫu nhiên dựa theo phần mềm

- Dựa vào thời gian của hệ thống
- Thời điểm ấn phím hoặc di chuyển chuột
- Số các buffer được sử dụng để thao tác dữ liệu

Sinh bit dãy ngẫu nhiên dựa theo phần mềm

- Dựa vào thời gian của hệ thống
- Thời điểm ấn phím hoặc di chuyển chuột
- Số các buffer được sử dụng để thao tác dữ liệu
- Do người dùng chọn

Sinh bit dãy ngẫu nhiên dựa theo phần mềm

- Dựa vào thời gian của hệ thống
- Thời điểm ấn phím hoặc di chuyển chuột
- Số các buffer được sử dụng để thao tác dữ liệu
- Do người dùng chọn
- Các giá trị thống kê về mạng hoặc hệ thống đang vận hành

Kỹ thuật de-skewing tránh biases

- **De-skewing** là kỹ thuật sử dụng khi nguồn sinh bit ngẫu nhiên bị tác động dẫn đến kết quả thiên lệch, chẳng hạn xác suất cho bit 1 không bằng 0.5, hoặc bit sinh sau có tương quan với bit trước đó.

Kỹ thuật de-skewing tránh biases

- **De-skewing** là kỹ thuật sử dụng khi nguồn sinh bit ngẫu nhiên bị tác động dẫn đến kết quả thiên lệch, chẳng hạn xác suất cho bit 1 không bằng 0.5, hoặc bit sinh sau có tương quan với bit trước đó.

- **Ví dụ**: Giả sử một sinh bit cho bit 1 với xác suất $p, p \neq 0.5$ và bit 0 với xác suất $1 - p$. Khi đó xác suất cho cặp 01 và 10 là như nhau và bằng $p(1 - p)$, xem chúng như kết quả trong khi bỏ qua trường hợp cặp 00, 11.

Thuật toán sinh dãy bit giả ngẫu nhiên

- ANSI X9.17

Thuật toán sinh dãy bit giả ngẫu nhiên

- ANSI X9.17
- FIPS 186

Thuật toán sinh dãy bit giả ngẫu nhiên

- ANSI X9.17
- FIPS 186
- Các thuật toán trên có sử dụng các hàm băm SHA và chuẩn mật mã DES

Thuật toán sinh bit ngẫu nhiên ANSI X9.17

Thuật toán sinh bit ngẫu nhiên ANSI X9.17

Input: Hạt gieo ngẫu nhiên 64 bit s, số m và khóa k của DES

Output: Sinh ra một dãy các số giả ngẫu nhiên 64-bit m

1. Tính $I = E_k(D)$, với D là biểu diễn nhị phân 64-bit lấy được từ giá trị ngày giờ tại thời điểm đó
2. Thực hiện vòng lặp:

   For i from 1 to m do:

   $x_i \longleftarrow E_k(I \oplus s)$

   $s \longleftarrow E_k(x_i \oplus I)$
3. Kết quả là dãy $(x_1, x_2, \ldots x_m)$

Định nghĩa

- *Gọi* $s = s_0, s_1, \ldots$ *là dãy bit vô hạn,* $s^n = s_0, s_1, \ldots, s_{n-1}$ *là dãy n bit đầu tiên của s.*

Định nghĩa

- *Gọi* $s = s_0, s_1, \ldots$ *là dãy bit vô hạn,* $s^n = s_0, s_1, \ldots, s_{n-1}$ *là dãy n bit đầu tiên của s.*
- *Dãy s được gọi là N-tuần hoàn nếu* $s_i = s_{i+N}, \forall i \geq 0$, *khi đó N gọi là chu kỳ và* s^N *gọi là một chu trình.*

Định nghĩa

- *Gọi* $s = s_0, s_1, \ldots$ *là dãy bit vô hạn,* $s^n = s_0, s_1, \ldots, s_{n-1}$ *là dãy n bit đầu tiên của s.*
- *Dãy s được gọi là N-tuần hoàn nếu* $s_i = s_{i+N}, \forall i \geq 0$, *khi đó N gọi là chu kỳ và* s^N *gọi là một chu trình.*
- *Ta gọi một quãng của s là các dãy bit con liên tiếp toàn là bit 1 hoặc 0 mà phía trước và sau dãy là bit khác. Nếu 1 quãng toàn là 1 thì gọi là 1 khối còn toàn 0 thì gọi là một khoảng trống.*

Định nghĩa

- *Gọi* $s = s_0, s_1, \ldots$ *là dãy bit vô hạn,* $s^n = s_0, s_1, \ldots, s_{n-1}$ *là dãy n bit đầu tiên của s.*
- *Dãy s được gọi là N-tuần hoàn nếu* $s_i = s_{i+N}, \forall i \geq 0$, *khi đó N gọi là chu kỳ và* s^N *gọi là một chu trình.*
- *Ta gọi một quãng của s là các dãy bit con liên tiếp toàn là bit 1 hoặc 0 mà phía trước và sau dãy là bit khác. Nếu 1 quãng toàn là 1 thì gọi là 1 khối còn toàn 0 thì gọi là một khoảng trống.*
- *Hàm tự tương quan của s, ký hiệu là* $C(t)$ *xác định bởi:*

$$C(t) = \frac{1}{N}\sum_{i=0}^{N-1}(2s_i - 1)(2s_{i+t} - 1), \quad 0 \leq t \leq N-1.$$

Định đề ngẫu nhiên của Golomb

Giả sử s là dãy tuần hoàn chu kỳ N và s^N là chu trình. Các định đề về tính ngẫu nhiên của dãy s gồm:

R1 Trong chu trình s^N, số các phần tử 0 khác số các phần tử 1 không quá 1.

Định đề ngẫu nhiên của Golomb

Giả sử s là dãy tuần hoàn chu kỳ N và s^N là chu trình. Các định đề về tính ngẫu nhiên của dãy s gồm:

R1 Trong chu trình s^N, số các phần tử 0 khác số các phần tử 1 không quá 1.

R2 Trong chu trình s^N, có ít nhất $1/2^k$ số quãng có độ dài bằng k. Hơn nữa với mỗi độ dài thì số các khối và số các khoảng trống gần như bằng nhau.

Định đề ngẫu nhiên của Golomb

Giả sử s là dãy tuần hoàn chu kỳ N và s^N là chu trình. Các định đề về tính ngẫu nhiên của dãy s gồm:

R1 Trong chu trình s^N, số các phần tử 0 khác số các phần tử 1 không quá 1.

R2 Trong chu trình s^N, có ít nhất $1/2^k$ số quãng có độ dài bằng k. Hơn nữa với mỗi độ dài thì số các khối và số các khoảng trống gần như bằng nhau.

R3 Hàm tự tương quan nhận giá trị xác định bởi:

$$N.C(t) = \sum_{i=0}^{N-1}(2s_i - 1)(2s_{i+t} - 1) = \begin{cases} N, \text{ nếu } t = 0 \\ K, \text{ nếu } \quad 1 \leq t \leq N-1. \end{cases}$$

Định đề ngẫu nhiên của Golomb

Giả sử s là dãy tuần hoàn chu kỳ N và s^N là chu trình. Các định đề về tính ngẫu nhiên của dãy s gồm:

R1 Trong chu trình s^N, số các phần tử 0 khác số các phần tử 1 không quá 1.

R2 Trong chu trình s^N, có ít nhất $1/2^k$ số quãng có độ dài bằng k. Hơn nữa với mỗi độ dài thì số các khối và số các khoảng trống gần như bằng nhau.

R3 Hàm tự tương quan nhận giá trị xác định bởi:

$$N.C(t) = \sum_{i=0}^{N-1}(2s_i - 1)(2s_{i+t} - 1) = \begin{cases} N, \text{ nếu } t = 0 \\ K, \text{ nếu } \quad 1 \leq t \leq N-1. \end{cases}$$

- **Một dãy các bit nhị phân thỏa mãn các định đề Golomb trên gọi là dãy giả nhiễu hay dãy pn.**

- **Kiểm định tần số**:

$$X_1 = \frac{(n_0 - n_1)^2}{n} \sim \chi_1^2,$$

với n_0, n_1 là tấn số của bit 0 và 1 tương ứng.

- **Kiểm định tần số**:

$$X_1 = \frac{(n_0 - n_1)^2}{n} \sim \chi_1^2,$$

với n_0, n_1 là tần số của bit 0 và 1 tương ứng.

- **Kiểm định hai bit**:

$$X_2 = \frac{4}{n-1}(n_{00}^2 + n_{01}^2 + n_{10}^2 + n_{11}^2) - \frac{2}{n}(n_0^2 + n_1^2) + 1 \sim \chi_2^2,$$

với n_{00}, n_{01}, n_{10}, n_{11}, n_0, n_1 là tần số của 00, 01, 10, 11, 0 và 1 tương ứng.

- **Kiểm định tần số**:

$$X_1 = \frac{(n_0 - n_1)^2}{n} \sim \chi_1^2,$$

với n_0, n_1 là tần số của bit 0 và 1 tương ứng.

- **Kiểm định hai bit**:

$$X_2 = \frac{4}{n-1}(n_{00}^2 + n_{01}^2 + n_{10}^2 + n_{11}^2) - \frac{2}{n}(n_0^2 + n_1^2) + 1 \sim \chi_2^2,$$

với n_{00}, n_{01}, n_{10}, n_{11}, n_0, n_1 là tần số của 00, 01, 10, 11, 0 và 1 tương ứng.

- **Kiểm định Poker**

- **Kiểm định tần số**:

$$X_1 = \frac{(n_0 - n_1)^2}{n} \sim \chi_1^2,$$

với n_0, n_1 là tần số của bit 0 và 1 tương ứng.

- **Kiểm định hai bit**:

$$X_2 = \frac{4}{n-1}(n_{00}^2 + n_{01}^2 + n_{10}^2 + n_{11}^2) - \frac{2}{n}(n_0^2 + n_1^2) + 1 \sim \chi_2^2,$$

với n_{00}, n_{01}, n_{10}, n_{11}, n_0, n_1 là tần số của 00, 01, 10, 11, 0 và 1 tương ứng.

- **Kiểm định Poker**
- **Kiểm định các quãng (run)**

- **Kiểm định tần số**:

$$X_1 = \frac{(n_0 - n_1)^2}{n} \sim \chi_1^2,$$

với n_0, n_1 là tần số của bit 0 và 1 tương ứng.

- **Kiểm định hai bit**:

$$X_2 = \frac{4}{n-1}(n_{00}^2 + n_{01}^2 + n_{10}^2 + n_{11}^2) - \frac{2}{n}(n_0^2 + n_1^2) + 1 \sim \chi_2^2,$$

với n_{00}, n_{01}, n_{10}, n_{11}, n_0, n_1 là tần số của 00, 01, 10, 11, 0 và 1 tương ứng.

- **Kiểm định Poker**
- **Kiểm định các quãng (run)**
- **Kiểm định tự tương quan**

Thuật toán sinh bit giả ngẫu nhiên an toàn trong mã hóa

- Thuật toán sinh bit ngẫu nhiên RSA

Thuật toán sinh bit giả ngẫu nhiên an toàn trong mã hóa

- Thuật toán sinh bit ngẫu nhiên RSA
- Thuật toán sinh bit ngẫu nhiên Micali-Schnorr

Thuật toán sinh bit giả ngẫu nhiên an toàn trong mã hóa

- Thuật toán sinh bit ngẫu nhiên RSA
- Thuật toán sinh bit ngẫu nhiên Micali-Schnorr
- Thuật toán sinh bit ngẫu nhiên Blum-Blum-Shub

Thuật toán sinh bit ngẫu nhiên RSA

Thuật toán sinh bit ngẫu nhiên RSA

Output: sinh ra một dãy các bit $z_1, z_2, \ldots, z_l$ có độ dài l

1. Thiết lập tham số. Sinh ra hai số nguyên tố kiểu RSA bí mật p, q, tính $n = pq$ và $\phi = (p-1)(q-1)$. Chọn ngẫu nhiên một số tự nhiên e, $1 < e < \phi, gcd(e, \phi) = 1$

2. Chọn ngẫu nhiên một số tự nhiên x_0 trong khoảng $[1, n-1]$ làm hạt gieo ngẫu nhiên.

3. Thực hiện vòng lặp:

   For i from 1 to l do:

   $x_i \longleftarrow x_{i-1}^e \mod n$

   $z_i \longleftarrow$ bit ít ý nghĩa nhất trong biểu diễn nhị phân của x_i

4. Kết quả là dãy bit $z_1, z_2, \ldots z_l$

Thuật toán sinh bit ngẫu nhiên Micali-Schnorr

Thuật toán sinh bit ngẫu nhiên Micali-Schnorr

Output: sinh ra một dãy các bit ngẫu nhiên

1. Thiết lập tham số. Sinh ra hai số nguyên tố kiểu RSA bí mật p, q, tính $n = pq$ và $\phi = (p-1)(q-1)$. Đặt $N = [\log_2 n] + 1$. Chọn một số tự nhiên e, $1 < e < \phi$, với $gcd(e, \phi) = 1$ và $80e \leq N$. Đặt $k = [N(1 - 1/e)]$ và $r = N - k$

2. Chọn ngẫu nhiên một số tự nhiên x_0 trong khoảng $[1, n-1]$ làm hạt gieo ngẫu nhiên

3. Thực hiện vòng lặp:

   For i from 1 to l do:

   $y_i \longleftarrow x_{i-1}^e \mod n$

   $x_i \longleftarrow$ lấy r bit ý nghĩa nhất của y_i

   $z_i \longleftarrow$ lấy k bit ít ý nghĩa nhất của y_i

4. Kết quả là dãy gồm $k.l$ bit $z_1 \parallel z_2 \parallel \ldots \parallel z_l$, ở đó $\parallel$ là hàm ghép.

Tài liệu tham khảo

Handbook of Applied Crytography, chapter 5

http://cacr.uwaterloo.ca/hac/