Estatística

GERADOR DE NÚMEROS PSEUDO-ALEATÓRIOS

Números aleatórios

- Um número aleatório é aquele que é retirado de um conjunto de valores possíveis onde cada um é igualmente provável de ocorrer, ou seja, uma *distribuição uniforme.*

- Em uma sequência de números aleatórios, cada número sorteado deve ser *estatisticamente independente* dos outros.

distribuição uniforme: distribuição de probabilidade simétrica em que um número finito de valores é igualmente provável de ser observado.

estatisticamente independentes: ocorrência de um não afeta a probabilidade de ocorrência do outro, portanto não afeta as respectivas probabilidades.

Números aleatórios

Com o advento dos computadores, veio a necessidade de introduzir aleatoriedade em programas.

No entanto, é muito difícil fazer com que um computador faça algo ao acaso, já que ele segue instruções e é totalmente previsível.

Então, surgiram duas soluções:

- Geradores de Números Aleatórios Verdadeiros (TRNG)
- Geradores de Números Pseudo-Aleatórios (PRNG)

GERADORES DE NÚMEROS ALEATÓRIOS VERDADEIROS (TRNG)

Hardwares que extraem aleatoriedade de um fenômeno físico e introduzem no computador

decaimento de fontes radioativas, ruídos atmosféricos etc.

- ~~X~~ produzem apenas um número limitado de bits aleatórios por segundo, por isso são bastante ineficientes em comparação com os PRNGs, levando muito mais tempo para produzir a mesma quantidade de números.
- ~~X~~ não são determinísticos, ou seja, uma dada sequência de números não pode ser reproduzida intencionalmente.
- ~~X~~ não têm período, o que significa que não há uma repetição após n valores.
    - Site: random.org, ruídos atmosféricos.

GERADORES DE NÚMEROS PSEUDO-ALEATÓRIOS (PRNG)

Algoritmos que usam fórmulas matemáticas ou tabelas pré-calculadas

os números parecem aleatórios, mas são pré-determinados.
se a imprevisibilidade for necessária, podem obter as sementes dos resultados de um TRNG.

- ~~X~~ são muito rápidos e eficientes, podendo produzir vários números em um curto espaço de tempo.
- ~~X~~ são determinísticos, ou seja, uma dada sequência pode ser reproduzida se o valor inicial (semente) for conhecido.
- ~~X~~ geralmente são periódicos, portanto a sequência acabará se repetindo, mesmo que demore anos como é o caso dos PRNGs modernos.

  ex.: aplicativos de simulação e modelagem.

  - não são adequados para criptografia nem jogos de azar.

GERADORES DE NÚMEROS PSEUDO-ALEATÓRIOS (PRNG)

Definição matemática

P – uma distribuição de probabilidades em R e RB (onde B é o conjunto Borel padrão na linha real)

II – uma coleção não vazia de conjuntos de Borel $I \subseteq B$, por exemplo: $I = \{(-\infty, t] : t \in R\}$. Se I não é especificado, pode ser qualquer B ou $= \{(-\infty, t] : t \in R\}$, dependendo do contexto.

A$\subseteq R$ – um conjunto não vazio (não necessariamente um conjunto de Borel). amiúde A é um conjunto entre P suporte e seu interior; por exemplo, se P é a distribuição uniforme no intervalo (0,1], A pode ser (0,1] . Se A não é especificado, presume-se que seja algum conjunto contido no suporte de P e contendo seu interior, dependendo do contexto.

Chamamos de função $ff : N_1 \to R$ (onde $N_1 = \{1, 2, 3,...\}$ é o conjunto de inteiros positivos) um gerador de números pseudoaleatórios para P dado I tendo valores em A se e somente se $f(N_1) \subseteq$ A

$$\forall \mathrm{E} \in I \quad \forall 0 < \epsilon R \quad \exists N \in N_1 \quad \forall N \leq n \in N_1, \left| \frac{\#\{i \in \{1,2,...,n\} : f(i) \in \mathrm{E}\}}{n} - P(\mathrm{E}) \right| < \varepsilon$$

($\#S$ denota o número de elementos no conjunto finito S)

Pode-se mostrar que se f é um gerador de números pseudoaleatórios para a distribuição uniforme em (0,1) e se F é a CDF (função de distribuição cumulativa) de alguma distribuição de probabilidade dada P então $F^* \circ f$ é um gerador de números pseudoaleatórios para P onde $F^* : (0,1) \to R$ é o percentil de P, ou seja, $F^*(x) := inf\{t \in R : x \leq F(t)\}$. Intuitivamente, uma distribuição arbitrária pode ser simulada a partir de uma simulação da distribuição uniforme padrão.

GERADORES DE NÚMEROS PSEUDO-ALEATÓRIOS CRIPTOGRAFICAMENTE SEGUROS (CSPRNG)

São PRNGs com propriedades que os tornam adequados para o uso na criptografia

fazem parte do Sistema Operacional ou vêm de alguma outra fonte segura.

Requisitos para ser CSPRNG:

- ~~✗~~ deve satisfazer o teste do próximo bit, ou seja, dados os primeiros k bits de uma sequência aleatória, não há algoritmo de tempo polinomial capaz de predizer o bit k+1 com probabilidade de sucesso maior que 50%
  - Andrew Yao (informático) provou em 1982 que um gerador que passe nesse teste, passará em qualquer outro teste estatístico de tempo polinomial para aleatoriedade.
- ~~✗~~ deve resistir a extensões de compromisso de estado, isto é, caso parte ou o todo de seus estados seja revelado, deve ser impossível reconstruir o fluxo de números anterior.

GERADORES DE NÚMEROS PSEUDO-ALEATÓRIOS CRIPTOGRAFICAMENTE SEGUROS (CSPRNG)

Como funciona:

Há diversas formas de criar números aleatórios por meio de CSPRNGs.

Uma delas baseia-se em reunir uma sequência de n bits verdadeiramente aleatórios, onde n é um número grande o suficiente para impedir ataques por força bruta.

Com a tecnologia atual, um valor ideal seria n=128, pois 2^128 torna inviável testar todas as combinações possíveis.

Após isso, o sistema codifica esses valores e os comprime, aplicando uma função Hash segura, como a SHA-256, o que torna o valor inicial indistinguível.

ex.: podem ser usados em jogos de azar, ou qualquer outro ambiente que necessite de criptografia, como sistemas bancários.

CIFRA DE USO ÚNICO (ONE-TIME PAD)

Primeiramente descrita pelo banqueiro e criptografista Frank Miller em 1882, consiste num algoritmo em que o plain-text (texto puro) é combinado, caractere por caractere, a uma chave secreta aleatória que deve ter, no mínimo, a mesma quantidade de caracteres do texto.

Para garantir a segurança completa, a chave só deve ser usada uma vez, além de ser imediatamente destruída após o uso.

As mensagens cifradas não oferecem nenhuma informação além do tamanho máximo possível, portanto, se usadas da forma correta, as cifras de uso único são consideradas seguras para a tecnologia atual.

OBRIGADA!

Universidade de São Paulo (USP)

Sistemas de Informação (2021)