고급소프트웨어실습I
1주차 보고서

학번 : 20171663
이름 : 이도훈

선형 합동 생성기(linear congruential generator)는 널리 알려진 유사난수 생성기로 간단한 난수를 생성하기 위해 이용할 수 있다. $X_{n+1} = (a \cdot X_n + c)\%m$의 점화식으로 표현할 수 있으며 이때 각각의 변수는 $m > 0, 0 < a < m, 0 < c < m, 0 \le X_0 < m$의 조건을 가진다. 선형 합동 생성기의 state vector는 바로 이전에 생성된 난수이고 최대 m가지 경우가 있으므로 난수의 주기는 최대 m이다. 최대 주기 m을 갖기 위한 필요충분조건은 다음과 같다.

1. c와 m이 서로소여야 한다.
2. a - 1이 m의 모든 소인수로 나눠져야 한다.
3. m이 4의 배수일 경우, a - 1도 4의 배수여야 한다.

선형 합동 생성기가 생성해 내는 난수의 질은 그 인자에 따라 큰 폭으로 달라지고 인자에 따라서는 적절치 못한 초깃값 때문에 문제가 생기기도 한다.

선형 합동 생성기의 문제점은 최대 주기 m을 가지도록 인자를 선택했더라도 아주 좋은 품질의 난수열을 생성해 내지 못한다. 선형 합동 생성기가 만드는 연속된 난수들 사이에 상당한 상관 관계가 존재하기 때문에 몬테 카를로 시뮬레이션에 적합하지 않으며, 마지막으로 생성된 난수를 알면 그 뒤에 만들어질 난수를 모두 예측할 수 있기 때문에 암호학에 사용하기는 적합하지 않다.

선형 합동 생성기를 쓸 수 있는 대부분의 환경에서는 메르센 트위스터와 같은 생성기를 쓰는 것이 난수의 질이나 속도 면에서 더 좋다. 반면 선형 합동 생성기는 이들 생성기가 적용되기 힘든 환경에서 유리한데, 예를 들어 메르센 트위스터는 2킬로바이트 정도의 메모리를 요구하지만 많은 임베디드 환경에서는 이보다 적은 메모리만을 가지고 있는 경우가 많다. 또한 상관 관계에 대한 고려가 필요하지 않은 경우에도 선형 합동 생성기가 사용되는데, 한 예로 대부분의 메르센 트위스터 구현에서는 의사 난수 생성기를 사용해서 초기값으로부터 더 큰 초기화 벡터를 만들어 낸다.

메르센 트위스터는 1997년에 마츠모토 마코토와 니시무라 다쿠지가 개발한 유사난수 생성기이다. 메르센 트위스터는 동일한 사람들이 개발한 TT800 생성기의 개선판으로, 기존 생성기들의 문제점들을 개선하면서 매우 질이 좋은 난수를 빠르게 생성할 수 있도록 설계되었다. 메르센 트위스터 중 가장 흔히 사용되는 생성기인 MT19937은 유사난수 생성기로서 다양한 장점을 가지고 있다.

첫 번째로는 생성해 내는 난수의 주기가 $2^{19937}-1$로 매우 크다.

두 번째 장점은 생성된 난수는 623차원까지 동일분포되어 있다는 것이다. 난수를 623개까지 짝지어서 623차원 하이퍼큐브에 해당하는 좌표에 점을 찍어도 일관성을 발견할 수 없으며, 따라서 연속된 숫자들 사이의 관계가 매우 낮다. 이러한 점 때문에 MT19937은 시뮬레이션에서 자주 사용된다.

세 번째는 다이하드 테스트를 비롯한 다양한 확률적 시험을 통과한다.

네 번째로는 비트 연산만으로 알고리즘의 구현이 가능하기 때문에 매우 빠르다.

하지만 MT19937은 여러 단점도 가지고 있다.

첫 번째는 생성기의 적어도 624개의 숫자를 담을 수 있는 공간이 필요해 state

vector가 비교적 큰 편이다. 이는 대부분의 환경에서는 큰 문제가 되지 않지만 임베디드 환경과 같은 적은 메모리만 사용이 가능한 환경에서는 문제가 된다.
두 번째로는 선형 합동 생성기와 마찬가지로 암호학적으로 안전하게 설계되어 있지 않다. 주기와 난수 범위 등의 난수의 특징을 알고 있다면 유한한 수의 난수로 현재 상태를 확인하여 추후에 나올 난수를 예측 가능하다
마지막으로는 MT19937의 초기 구현의 상태 초기화 루틴은 주어진 초기화 값의 최상위 비트를 상태에 잘 반영하지 않는 문제가 있었다.
추가적으로 2006년에 개발된 SIMD 기반 메르센 트위스터는 MT19937에 비해 약 2배 정도 더 빠르다.