DOI: 10.3745/KIPSTC.2010.17C.1.047

키유도함수의 통계적 난수성 평가 방법

강 주 성[†] · 이 옥 연[†] · 염 지 선[††] · 조 진 웅[†††]

요 약

암호시스템에 사용되는 알고리즘의 기본적인 안전성 평가 항목은 난수성이다. 미국의 표준기술원 NIST는 차세대 암호알고리즘 AES를 선정하는 과정에서 블록암호의 난수성을 통계적으로 평가할 수 있는 패키지를 제안하였다. 이 패키지는 입출력 길이가 동일한 함수인 블록암호에 적합하도록 구성되어 있으므로 대부분 확장된 출력 길이를 갖는 키유도함수의 난수성 평가에 그대로 적용하는 것은 무리가 있다. 본 논문에서는 입력 길이보다 확장된 다중 블록을 출력하는 대표적인 암호 구성 요소인 키유도함수에 적합한 통계적 난수성 평가 방법으로 NIST의 방식을 개선한 것을 제안한다. 그리고 제안된 방법에 의하여 3GSM과 NIST에서 표준으로 권고하고 있는 키유도함수에 대한 통계적 난수성 평가 결과를 제시한다.

키워드 : 통계적 난수성 검정, 키유도함수, 다중 블록 출력 함수

A Method of Statistical Randomness Test for Key Derivation Functions

Ju-Sung Kang[†] · Ok-Yeon Yi[†] · Ji-Sun Youm[††] · Jin Woong Cho[†††]

ABSTRACT

Randomness is a basic security evaluation item for the most cryptographic algorithms. NIST has proposed a statistical test suit for random number generators for cryptographic applications in the process of AES project. However the test suit of NIST is customized to block ciphers which have the same input and output lengths. It needs to revise NIST′s test suit for key derivation functions which have multiple output blocks. In this paper we propose a revised method of NIST′s statistical randomness test adequate to the most key derivation functions and some experimental results for key derivation functions of 3GSM and NIST.

Keywords : Statistical Randomness Test, Key Derivation Function, Multiple Block Output Function

1. 서 론

정보보호시스템에서 암호 알고리즘이 올바르게 구동되기 위해서는 안전성 높은 대칭키나 공개키 암호알고리즘에 적합한 키관리시스템이 반드시 필요하다. 키관리시스템은 마스터키, 세션키, 암호화키, 인증키 등 키의 중요성 수준과 용도에 따라 다양한 키들을 안전하게 사용하기 위한 것이다. 다양한 용도의 키들은 하나의 마스터키로부터 유도되는 것이 보통이며, 이 때 사용되는 알고리즘이 키유도함수(key derivation function)이다. 그러므로 키관리시스템에서 핵심적 위치를 점하고 있는 키유도함수의 안전성은 정보보호시스템 전체의 안전성 차원에서 매우 중요한 요소이다. 하지만 키유도함수의 안전성에 대한 연구 결과는 그동안 상대적으로 빈약한 실정이다. 본 논문에서 우리는 키유도함수의 안전성 평가에서 가장 기본적인 사항이라 할 수 있는 통계적 난수성에 대하여 논하고자 한다.

미국의 표준기술원인 NIST는 차세대 암호 알고리즘을 선정하는 AES 프로젝트[1] 수행 과정에서 후보 블록암호 알고리즘의 통계적 난수성 평가를 위한 패키지를 발표하였다[2]. NIST의 난수성 평가 패키지는 1라운드 AES 후보 알고리즘의 안전성을 평가하는 중요한 요소로 작용하였다. 실제로 1라운드 후보 알고리즘들 중 많은 것들이 통계적 난수성 평가를 통과하지 못했다[3]. 이후로 NIST의 통계적 난수성 평가 패키지는 각종 암호 알고리즘의 기본적인 안전성 평가 요소로 자리 잡게 되었다. 그러나 NIST의 패키지는 블록암호에 적합한 형태로 구성되어 있으므로 다른 암호 알고리즘에 적용할 경우 수정이 불가피하다. 키유도함수는 블

※ 본 연구는 지식경제부 및 한국산업기술평가관리원의 IT산업원천기술개발사업의 일환으로 수행하였음(2009-S-039-01, u-City용 Binary CDMA 기반 무선 보안 기술 개발).
† 정 회 원 : 국민대학교 수학과 부교수
†† 준 회 원 : 국민대학교 수학과 이학석사
††† 정 회 원 : 전자부품연구원 센터장
논문접수 : 2009년 11월 10일
수 정 일 : 1차 2010년 1월 8일, 2차 2010년 1월 22일, 3차 2010년 2월 1일
4차 2010년 2월 3일
심사완료 : 2010년 2월 3일

록암호나 해쉬함수를 핵심함수로 사용하고 있지만 구조적으로 이들과 다르기 때문에 통계적 난수성 평가 시에도 좀 더 엄밀한 적용이 요구된다.

키유도함수는 길이가 n 비트인 블록 한 개로 구성된 입력 값이 내부 함수를 통해 여러 개의 블록으로 확장되어 출력되는 다중 블록 출력 함수 구조를 가지고 있다. 3GSM[4], 무선랜[5], Bluetooth[6] 같은 무선 통신환경에서 인증이나 키 분배를 위해 사용되는 키유도함수는 모두 다중 블록 출력 함수 구조로 볼 수 있다.

본 논문에서는 키유도함수로 대표되는 다중 블록 출력 함수에 대한 통계적 난수성 검정 과정을 구체적으로 연구하여 이에 적합한 검정 방법을 제안하고자 한다. 이를 위해 NIST의 패키지를 분석하여 다중 블록 출력 함수에 적합한 입력 데이터 구성 방법을 모색한다. 또한, NIST의 패키지 내에 있는 15 가지 통계적 난수성 평가를 다중 블록 출력 함수 입장에서 분석하여 이에 적합한 검정 파라미터와 검정 데이터 집합 구성 방법을 새롭게 제시한다. 특히, 15 가지 평가 방법 중 블록 내부의 빈도, 겹치지 않는 템플릿 매칭, 겹치는 템플릿 매칭, 점근적 엔트로피, 시리얼, 선형복잡도 검정의 경우 NIST 패키지의 파라미터를 수정하는 것이 타당함을 밝히고, 수정된 파라미터에 의한 검정 방법이 통계학적으로 올바름을 이론적으로 증명한다. 또한, 대표적인 블록암호 기반 키유도함수인 3GSM의 Milenage와 NIST가 제안한 Counter, Feedback, Double Pipeline Iteration 모드에 대하여 수정된 파라미터를 사용한 검정 방법을 적용한 시뮬레이션 결과를 제시한다.

2. NIST의 통계적 난수성 평가 패키지

난수성(randomness)은 각종 암호 알고리즘의 안전성 평가에서 가장 기본적으로 요구되는 성질이다. 블록암호나 스트림암호 알고리즘은 랜덤하지 않은 평문을 랜덤한 암호문으로 변환시킨다. 이 때, 암호문이 지녀야할 기본적인 성질이 난수성이다. 실제로 NIST는 차세대 암호 알고리즘 표준인 AES 공모 과정에서 암호문의 난수성을 평가할 목적으로 평가 패키지를 만들었다[2, 3]. 이와 같이 기본적인 안전성 요구조건 뿐만 아니라 암호 기술을 응용하는 많은 정보보호 시스템에서 난수성을 가정한 파라미터들이 사용된다. 난수(random number)라 불리는 이러한 파라미터들은 구현 시에 의사난수발생기를 통하여 생성되는 것이 일반적이므로 이 경우에도 통계적 난수성 검정은 필수적인 평가 항목이 된다.

난수와 의사난수는 주로 암호시스템 상에서 키로 이용되거나, 암호 프로토콜의 다양한 상황에서 입력 값으로 사용되기도 한다. NIST에서는 암호학적으로 다양한 목적을 위해 사용되는 난수와 의사난수 생성기들의 결과값이 난수성을 만족하고 있는지를 검정하기 위하여 통계적 난수성 평가 패키지를 제안했다. 본 장에서는 NIST에서 제안한 통계적 난수성 평가 패키지의 수행 절차를 분석하고, 패키지 내에 있는 15 가지 검정 항목들에 대해서 간단히 살펴보고자 한다. 또한 각 평가 항목에 사용된 파라미터들의 의미를 분석한다.

2.1 NIST의 통계적 난수성 평가 절차

NIST에서 제안한 알고리즘의 통계적 난수성 평가 과정은 통계학의 가설검정 또는 유의성 검증 절차와 동일하다. 검정 대상이 되는 표본 데이터 집합 내의 수열이 확률적으로 0과 1이 균일한 분포를 따르는 독립 확률과정일 때 만족하는 확률론의 극한 정리(limiting theorem)들을 이용하는 것이 기본적인 방법이다.

알고리즘의 통계적 난수성을 평가하기 위해 우선 적당한 입력 데이터 집합을 구성하고 다음 소절에서 소개할 15 가지 평가 항목들에 대해 다음과 같은 가설 검정 절차를 거친다[2].

① 귀무가설(H_0)과 대립가설(H_1)을 설정한다.
② 유의수준(α)을 지정한다.
③ 수열에 대한 검정 통계량을 계산한다.
④ 결과에 대한 유의확률($p-value$)을 계산한다.
⑤ 지정된 유의수준과 유의확률을 비교한다.

검정 절차를 통해 각 검정 대상 수열에 대한 귀무가설의 기각 여부를 결정하고, 기각된 표본의 개수를 분석하는 방법으로 통계적 난수성을 평가한다.

2.1.1 귀무가설과 대립가설

귀무가설 H_0은 검정의 대상이 되는 가설을 말하고, 대립가설 H_1은 귀무가설이 받아들여지지 않을 때 채택하는 가설을 말한다. NIST의 패키지에서는 귀무가설을 “검정 대상 수열이 랜덤하다”로 설정하고 대립가설을 “검정 대상 수열이 랜덤하지 않다”로 설정한다.

가설검정을 귀무가설과 대립가설 중 하나를 택하는 결정의 방법론으로 생각하면, 검정 결과에 따라 다음 두 가지 오류를 생각할 수 있다.

〈표 1〉 가설 검정의 오류

검정결과 \ 실제상황	H_0가 사실	H_1이 사실
H_0를 선택	옳은 결정	제 2종 오류
H_1을 선택	제 1종 오류	옳은 결정

- 제1종 오류 확률 : Pr(H_1선택 | H_0사실)
- 제2종 오류 확률 : Pr(H_0선택 | H_1사실)

2.1.2 유의수준과 유의확률

가설검정에서 발생하는 두 가지 오류 중에서 NIST의 패키지는 제 1종 오류에 중점을 둔다. 제 1종 오류를 범할 확률의 최대 허용한계를 유의수준이라고 하며, 실제로 유의수

준이 0.01이라면 제 1종 오류를 범할 확률을 최대 0.01까지 허용한다. 보통 유의수준은 $[0.001, 0.01]$의 범위에서 적당한 값으로 설정한다.

유의수준을 설정했다면 그 다음은 검정 대상 수열에 대한 검정 통계량을 계산하고, 이에 대한 유의확률을 계산한다. 이 때, 유의확률은 귀무가설을 기각하려할 때 요구되는 유의수준의 최소값이다. 예를 들어, 유의확률이 0.56이면 유의수준이 최소한 0.56은 넘어야 귀무가설을 기각한다.

유의확률을 $p-value$, 유의수준을 α라고 할 때 계산한 유의확률과 지정된 유의수준과 비교한다. 만약 $p-value \geq \alpha$이면 검정 대상 수열의 귀무가설인 "검정 대상 수열이 랜덤하다"를 채택하고, 만약 $p-value < \alpha$이면 검정 대상 수열의 귀무가설을 기각한다.

2.2 NIST의 통계적 난수성 평가 항목

NIST의 통계적 난수성 평가 패키지는 검정 대상 수열이 극한 정리를 따르는지의 유의성을 검증하는 것으로서 극한 정리의 종류에 따라 15 가지 항목들을 평가하는 것으로 구성되어 있다. 이들을 간략히 소개하면 다음과 같다.

2.2.1 단일 비트 빈도(frequency) 검정

랜덤한 수열이라면 전체 검정 대상 수열 안에서 0과 1의 비율이 유사할 것이라는 사실에 기초하여 난수성을 평가하는 방법이다. 검정의 목적은 이진수열에서 0과 1의 개수가 이상적인 난수열에서의 0과 1의 개수와 근사적으로 같은지 확인하는 것이다. 이상적인 난수열에서의 0과 1의 비율은 각각 1/2로 가정한다.

2.2.2 블록 내부의 빈도 검정

전체 검정 대상 수열을 임의의 M-비트 블록으로 분할했을 때, 하나의 M-비트 블록에서 0과 1의 비율을 분석한다. 검정의 목적은 하나의 M-비트 블록에서 0과 1의 빈도가 이상적인 난수열일 때 기대되는 $M/2$에 근사적으로 같은지 확인하는 것이다.

2.2.3 런(runs) 검정

전체 검정 대상 수열 안에서 총 런(run)의 개수를 분석한다. 런(run)은 같은 비트(0 또는 1)가 연속적으로 나열된 부분수열을 의미한다. 이때, 0이 연속되는 부분수열을 갭(gap), 1이 연속되는 부분수열은 블록(block)이라 한다. 검정 대상 수열 내에서 0에서 1로 혹은 1에서 0으로의 런의 변화가 존재하는 부분을 체인지(change)라고 하는데, 이상적인 난수열은 체인지가 일어날 확률을 1/2로 가정한다. 검정의 목적은 다양한 길이의 블록과 갭의 개수가 이상적인 난수열에 근사한지 확인하는 것이다.

2.2.4 블록 내부의 롱런(long run) 검정

전체 검정 대상 수열을 임의의 M-비트 블록으로 분할했을 때, M-비트 블록에서 1로 구성된 런 중 가장 긴 런의 길이가 이상적인 난수열의 기댓값에 근사한지 검정한다. 1로 구성된 가장 긴 런과 0으로 구성된 가장 긴 런의 검정 결과는 같게 되므로 0으로 구성된 가장 긴 런에 대해서는 따로 검정 하지 않는다.

2.2.5 이진 행렬 위수(rank) 검정

전체 검정 대상 수열의 부분행렬에 대한 위수(rank)의 분포가 이상적인 난수열의 분포에 근사하는지 검정한다. 이 검정의 목적은 고정된 길이의 검정 대상 수열의 부분행렬들 사이에서 선형 종속의 정도를 확인하는 것이다.

2.2.6 이산 푸리에 변환 검정

검정의 주안점은 전체 검정 대상 수열의 이산 푸리에 변환(Discrete Fourier Transform)에서 최고점의 높이이다. 전체 검정 대상 수열 내에서 각각 서로 근접한 반복 패턴과 같은 주기적 특징을 찾는 것이 이 검정의 목적이다.

2.2.7 겹치지 않는 템플릿 매칭(matching) 검정

검정 대상 수열에서 어떤 주어진 특정한 비주기적인 수열(template)의 빈도수를 측정해 이상적인 난수열의 빈도수에 근사하는지를 검정한다. 겹치는(overlapping) 템플릿 매칭 검정과 함께, 특정한 m-비트 패턴을 찾기 위해 m-비트 윈도우(window)를 사용한다. 만약 패턴을 찾지 못한다면, 윈도우를 한 비트 옮겨서 다시 검색한다. 만약 패턴을 찾았다면, 윈도우는 찾은 패턴의 다음 값으로 재설정 되고 다시 검색한다.

2.2.8 겹치는 템플릿 매칭 검정

겹치지 않는 템플릿 매칭 검정과는 다르게 특정한 주기적인 수열의 빈도 수를 측정해 이상적인 난수열의 빈도 수에 근사하는지 검정한다. 특정한 m-비트 패턴을 찾기 위해 m-비트 윈도우를 사용한다. 특정 m-비트 패턴은 1로 구성된 런을 사용하며, 만약 패턴을 찾지 못한다면 윈도우를 한 비트 옮겨서 검색한다. 만약 패턴을 찾았다면 마찬가지로 윈도우를 한 비트 옮긴 후 다시 검색한다.

2.2.9 마우러(Maurer)의 유니버설 검정

검정 대상 수열의 압축성을 통해 난수성을 검정 한다. 검정 대상 수열이 정보 손실 없이 압축이 잘 될 수 있는지 알아본다. 압축성이 지나치게 높거나 낮으면 비난수적인 성질을 보인다고 판단한다.

2.2.10 선형복잡도(linear complexity) 검정

검정 대상 수열의 선형 복잡도가 이상적인 난수열만큼 충분한지 확인하는 검정이다. 스트림 암호 알고리즘에 사용된 논리인 LFSR(linear feedback shift register)을 평가하는 데 사용되었지만 변형시켜 블록암호 알고리즘에 적용 가능 하도록 설계된 검정 방법이다.

2.2.11 시리얼(serial) 검정

전체 검정 대상 수열에서 모든 중복 가능하게 생성한 m-비트 패턴의 빈도에 초점을 맞춘 검정이다. 이 검정의 목적은 검정 대상 수열 안에서 2^m개의 m-비트 중복 패턴이 발생한 개수가 이상적인 난수열의 결과와 근사한지 확인하는 것이다. 이상적인 난수열은 중복 패턴 개수들이 균일하다. 즉, 모든 m-비트 패턴이 발생할 기회는 다른 모든 m-비트 패턴이 발생할 기회와 같다. 만약 $m=1$이라면 이 검정은 단일 비트 빈도 검정과 동일하다.

2.2.12 점근적 엔트로피 검정

시리얼 검정과 유사하게 전체 검정 대상 수열에서 모든 가능한 중복되게 생성한 m-비트 패턴의 빈도에 초점을 맞춘 검정이다. 두 개의 연속적이고 인접한 길이(m과 $m+1$)의 중복 패턴의 빈도가 이상적인 난수열의 기댓값에 근사한지 확인한다.

2.2.13 누적합(cumulative sum) 검정

전체 검정 대상 수열에서 0을 -1로 1을 +1로 변환시켰을 때, 누적합으로 정의한 확률보행(random walk)의 최대값을 이상적인 난수열의 확률보행과 비교해 너무 큰지 혹은 작은지 확인한다. 이상적인 난수열에서 확률보행의 이탈성은 0에 가까우며, 비난수열에서 확률보행의 이탈성은 클 것이다.

2.2.14 랜덤 익스커전(excursions) 검정

누적합 검정처럼 전체 이진수열을 (-1, +1)로 변환시켰을 때, 누적합이 0으로 돌아오는 사이클을 익스커전이라 하고, 이 사이클 수를 조사하여 각 사이클에서 주어진 8개의 상태(state)가 나타난 횟수가 이상적인 난수열의 횟수와 근사한지 확인하는 검정이다.

2.2.15 변형 랜덤 익스커전 검정

랜덤 익스커전 검정과 달리 전체 검정 대상 수열에서 누적합이 주어진 상태에 나타난 총 횟수를 구해서 이상적인 난수열과 비교해 근사한지 확인하는 검정이다.

2.3 NIST 패키지의 파라미터

NIST의 통계적 난수성 평가 패키지 내의 각 평가 항목들에서 사용되는 파라미터는 검정 수행 과정에 큰 영향을 미친다. 파라미터 값을 어떻게 설정 하느냐에 따라 검정 결과는 크게 달라질 수 도 있게 된다. 그렇기 때문에 신뢰할 수 있는 평가 결과를 얻기 위해서는 각 항목들에서 사용되는 파라미터 값의 선택이 매우 중요한 문제가 된다. 키유도 함수의 통계적 난수성 평가시 신뢰 할 수 있는 결과를 얻기 위해서도 파라미터의 선택은 중요하다. 따라서 NIST 패키지의 각 평가 항목들에서 사용되는 파라미터를 분석하고자 한다[2].

2.3.1 단일 비트 빈도 검정

파라미터 ρ는 전체 검정 대상 수열의 길이를 의미하며, 전체 검정 대상 수열이 너무 짧으면 난수성을 충분히 평가할 수 없게 된다. 때문에 ρ는 10^2이상을 선택하도록 권장된다.

2.3.2 블록 내부의 빈도 검정

파라미터 ρ는 전체 검정 대상 수열의 길이를, M은 검정 대상 수열을 분할 할 블록 길이를 그리고 N은 분할된 블록 개수를 의미한다. 이 검정에서 전체 검정 대상 수열은 M으로 분할되기 때문에, 분할된 블록 개수 N은 총 $\left\lfloor \frac{\rho}{M} \right\rfloor$ 개가 된다. ρ는 10^2이상으로, $M \geq 20$, $M > 0.01\rho$, $N < 100$을 만족하는 M, N을 선택하도록 권장된다.

2.3.3 런 검정

파라미터 ρ는 전체 검정 대상 수열의 길이를 의미하며, 10^2이상을 선택하도록 권장된다.

2.3.4 블록 내부의 롱런 검정

파라미터 ρ는 전체 검정 대상 수열의 길이를, M은 검정 대상 수열을 분할할 블록 길이를 의미한다. 그리고 N은 분할된 블록 개수를 의미하며, 전체 검정 대상 수열을 각각 M씩 분할하기 때문에 $N = \left\lfloor \frac{\rho}{M} \right\rfloor$ 가 된다. 이 검정의 검정 통계량은 χ^2-분포를 따르고 M값의 변화에 따라 자유도를 의미하는 K는 변하게 된다. <표 2>는 ρ의 크기에 따라 권장되는 M, N, K값이다.

〈표 2〉 수열 길이에 따른 파라미터 값

ρ	M	N	K
$128 \leq \rho < 6{,}272$	8	16	3
$6{,}272 \leq \rho < 750{,}000$	128	49	5
$750{,}000 \leq \rho$	10^4	$\lfloor \rho/M \rfloor$	6

2.3.5 이진 행렬 위수 검정

파라미터 ρ는 전체 검정 대상 수열의 길이를 의미 하며 전체 검정 대상 수열을 $R \times Q$ 부분 행렬로 분할시킨 후 검정 한다. 이때 R은 부분 행렬의 행의 수를, Q는 열의 수를 의미하고, 분할된 행렬 개수를 의미하는 N은 $\left\lfloor \frac{\rho}{RQ} \right\rfloor$ 가 된다. 보통 R과 Q는 32를 선택하도록 권장되며, 그렇지 못할 경우에는 적어도 $\rho \geq 38RQ$를 만족하도록 권고된다.

2.3.6 이산 푸리에 변환 검정

파라미터 ρ는 전체 검정 대상 수열의 길이를 의미하며, 10^3이상을 선택하도록 권장된다.

2.3.7 겹치지 않는 템플릿 매칭 검정

파라미터 ρ는 전체 수열 길이를, m은 템플릿의 길이를 의미하고 B는 해당되는 m-비트 템플릿을 뜻한다. 검정은 전체 검정 대상 수열을 검정할 블록 길이인 M으로 분할하고, 분할된 블록을 각각 독립적으로 처리한다. 분할된 블록 개수를 의미하는 N은 $\left\lfloor \frac{\rho}{M} \right\rfloor$ 이고, m 값에 따른 템플릿의 개수는 <표 3>과 같다. m은 9나 10을, $N \leq 100$을 선택하도록 권장된다.

〈표 3〉 템플릿 길이에 따른 파라미터 값

m	템플릿의 개수	m	템플릿의 개수
1	0	9	148
2	2	10	284
3	4	11	568
4	6	12	1,116
5	12	13	2,232
6	20	14	4,424
7	40	15	8,848
8	74	16	17,622

2.3.8 겹치는 템플릿 매칭 검정

파라미터 ρ는 전체 검정 대상 수열의 길이를, m은 템플릿의 길이를 의미하며 실제 검정에서 m은 모두 1인 런의 길이가 된다. 파라미터 B, M, N은 겹치지 않는 템플릿 매칭 검정과 동일한 의미를 가지며, 검정 통계량은 자유도가 K인 χ^2-분포를 따른다. 일반적으로 ρ는 10^6 이상을, m은 9 혹은 10 을 선택하도록 권장되며, 그렇지 않을 경우 파라미터들은 다음 조건을 만족하도록 권장된다.

· $\rho \geq MN,\ N \bullet (\min \pi_i) > 5$

· $\lambda = (M - m + 1)/2^m \approx 2,\ m \approx \log_2 M$

· $K \approx 2\lambda$

〈표 4〉 수열 길이에 따른 파라미터 값

ρ	M	$N_I = 10 \bullet 2^M$
≥ 387,840	6	640
≥ 904,960	7	1,280
≥ 2,068,480	8	2,560
≥ 4,654,080	9	5,120
≥ 10,342,400	10	10,240
≥ 22,753,280	11	20,480
≥ 49,643,520	12	40,960
≥ 107,560,960	13	81,920
≥ 231,669,760	14	163,840
≥ 496,435,200	15	327,680
≥ 1,059,061,760	16	655,360

2.3.9 마우러의 유니버설 검정

파라미터 ρ는 전체 검정 대상 수열의 길이를, M은 검정할 블록 길이를 의미한다. 이 때, N_I는 초기 수열에서의 블록 개수를 의미하고, N_t는 초기 수열 다음 검정 수열의 블록 개수를 의미한다. N_t는 $N_t = \lceil \rho/M \rceil - N_I \approx 1000 \bullet 2^M$을 만족하도록 권장되며, 표 4는 ρ에 따라 권장하는 파라미터 M, N_I의 값을 나타낸다.

2.3.10 선형복잡도 검정

파라미터 ρ는 전체 검정 대상 수열의 길이를 의미하고 M은 검정 대상 수열을 분할할 블록 길이를 의미한다. 검정은 전체 검정 대상 수열을 M으로 분할하여 총 $N = \left\lfloor \frac{\rho}{M} \right\rfloor$ 개의 블록을 독립적으로 처리한다. 검정 통계량은 자유도가 K인 χ^2-분포를 따르고 ρ, M, N은 $\rho \geq 10^6,\ 500 \leq M \leq 5000,\ N \geq 200$의 조건을 만족하도록 권장된다.

2.3.11 시리얼 검정

파라미터 ρ는 전체 검정 대상 수열의 길이를, m'는 검정 대상 수열을 분할할 블록 길이를 의미한다. 검정은 전체 검정 대상 수열을 m'-비트 블록으로 분할하여, $2^{m'}$개의 m'-비트 패턴과 일치하는 블록을 검색한다. m'과 ρ는 $m' < \lfloor \log_2 \rho \rfloor - 2$ 를 만족하도록 권장된다.

2.3.12 점근적 엔트로피 검정

파라미터 ρ는 전체 검정 대상 수열의 길이를 의미하고 m'은 첫 번째 블록의 길이를, $m'+1$은 두 번째 블록의 길이를 의미한다. m'과 ρ는 $m' < \lfloor \log_2 \rho \rfloor - 2$의 조건을 만족하도록 권장된다.

2.3.13 누적합 검정

파라미터 ρ는 전체 검정 대상 수열의 길이를 의미 하며 10^2 이상을 선택하도록 권장된다.

2.3.14 랜덤 익스커전 검정

파라미터 ρ는 전체 검정 대상 수열의 길이를 의미 하며 10^6 이상의 값을 선택하도록 권장된다.

2.3.15 변형 랜덤 익스커전 검정

파라미터 ρ는 전체 검정 대상 수열의 길이를 뜻하고, 랜덤 익스커전 검정과 같이 10^6 이상의 값을 선택하도록 권장된다.

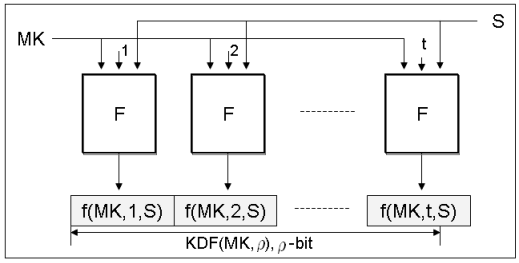
3. 키유도함수를 위한 통계적 난수성 평가

NIST의 통계적 난수성 평가 패키지는 입출력 길이가 동

일한 블록암호에 적합한 형태로 구성되어 있다. 하지만 키 유도함수는 보통 다중 블록 출력 함수의 구조를 갖기 때문에 그대로 NIST의 패키지에 적용하기에는 문제가 있다. 따라서 본 장에서는 NIST의 평가 패키지를 분석하여 키유도함수의 환경에 맞는 통계적 난수성 평가 방법을 제시하고자 한다. 우선 키유도함수의 구조인 다중 블록 출력 함수에 대해 간단히 살펴보고, 다중 블록 출력 함수의 환경에 적합한 데이터 집합 구성방법을 제시한다. 그리고 NIST의 패키지를 다중 블록 출력 함수의 입장에서 분석하여 검정 항목 내 파라미터 설정 방법을 제안하고, 수정된 파라미터에 의한 검정 방법이 이론적으로 올바름을 밝힌다. 또한 통계적 난수성 평가 기준이 되는 최대 기각수를 결정하는 방법을 세밀하게 분석하고자 한다.

3.1 다중 블록 출력 함수

다중 블록 출력 함수는 블록 하나로 이루어진 입력값 x가 t개($t \geq 2$)의 블록 $(z_1, z_2, \cdots, z_t)$로 확장되어 출력되는 함수를 뜻한다. 즉, 길이가 n-비트인 입력이 다중 블록 출력 함수를 통해 tn-비트의 길이로 확장된다[8]. 키유도함수는 보통 다중 블록 출력 함수의 구조를 가지고 있으며 내부 핵심함수로 블록암호와 해쉬함수를 사용한다. 블록암호 기반 키유도함수의 대표적 표준으로는 NIST에서 제안한 Counter, Feedback, Double-Pipeline Iteration 모드 [7]와 3GSM의 Milenage가 있다[4].



(그림 1) Counter Mode 블록도

3.1.1 키유도함수의 Counter 모드

블록암호 운영모드 중 Counter를 사용한 키유도함수로서 블록암호의 출력은 카운터 값과 함께 계산된다. 즉, 1개의 블록인 입력 S가 카운터 값과 키 MK를 통해 t개의 블록으로 확장된다. 내부함수를 $f(\cdot)$로 표현할 경우 키유도함수는

$$KDF(MK,\rho) = [f(MK,1,S) \| \cdots \| f(MK,t,S)]_\rho$$

이다. 즉, 의사난수함수 F의 각각의 출력을 연접시켜 원하는 ρ-비트 길이의 키를 생성한다.

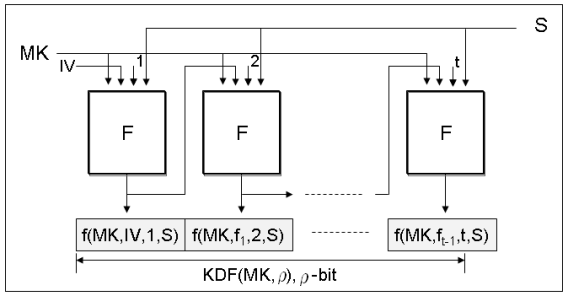
대표적으로 무선 LAN상의 단말과 AP사이의 통신에서 일어나는 "4-way handshake"단계에서의 키유도함수의 구조가 이와 유사하다.

3.1.2 키유도함수의 Feedback 모드

이 함수는 블록암호 운영모드 중 CBC 모드와 Counter 모드를 합성한 것으로 입력값 S가 키 MK, 카운터값과 함께 계산되어 그 다음 블록에 계속 영향을 미쳐 t개의 블록으로 확장된다.

$$f_i = f(MK, f_{i-1}, i, S), f_0 = IV$$
$$KDF(MK,\rho) = [f(MK,f_0,1,S) \| \cdots \| f(MK,f_{t-1},t,S)]_\rho$$

이 함수의 블록도는 (그림 2)와 같다.

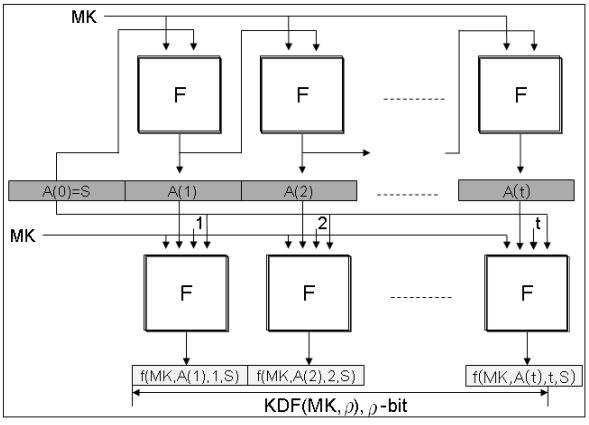


(그림 2) Feedback Mode의 블록도

3.1.3 Double-Pipeline Iteration(DPI) 모드

이 키유도함수의 구조는 2단계로 구성되고, CBC 운영모드와 Counter 운영모드를 단계적으로 합성한 것이다. 입력값 S는 키 MK를 통해 2번 암호화되어서 t개의 블록을 출력한다. DPI의 블록도는 (그림 3)과 같으며 다음과 같이 정의된다.

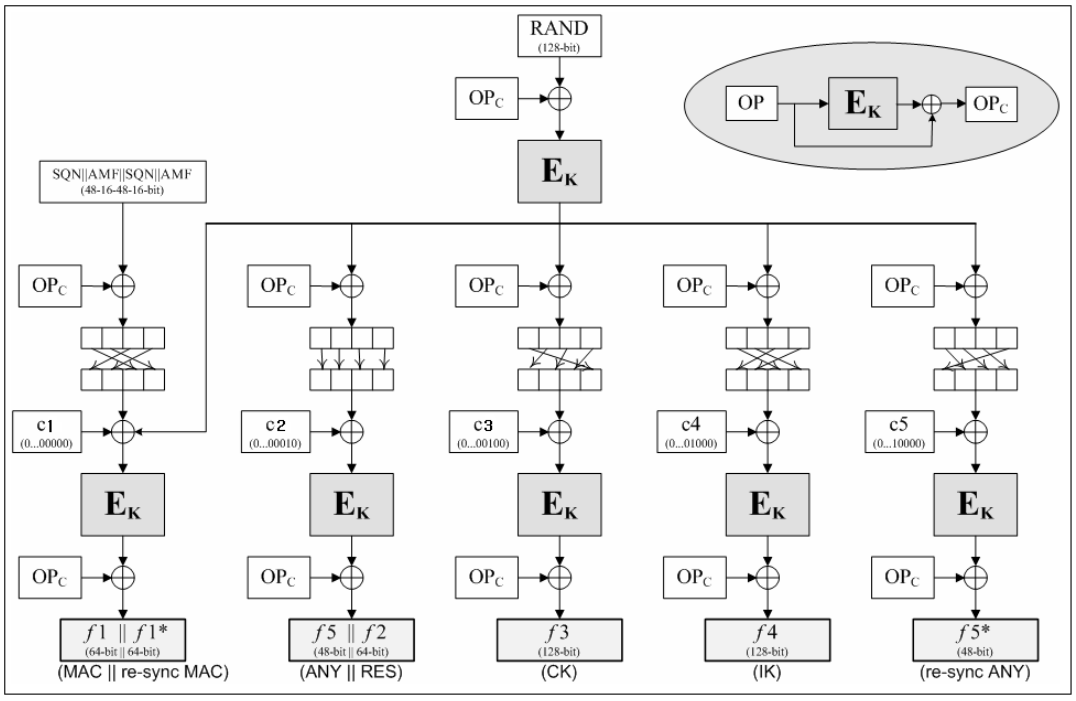
$$A(i) = f(MK, A(i-1)), A(0) = S$$
$$KDF(MK,\rho) = [f(MK,A(1),1,S) \| \cdots \| f(MK,A(t),t,S)]_\rho$$



(그림 3) Double-Pipeline Iteration Mode의 블록도

3.1.4 3GSM의 Milenage 모드

Milenage는 3GSM의 키유도함수로서 네트워크 인증을 위한 MAC값을 생성하는 f1, 사용자 인증을 위한 RES (response)



(그림 4) Milenage의 블록도

값을 생성하는 f2, 인증 후 데이터 암호화를 위한 키 CK (cipher key)를 생성하는 f3, 데이터 무결성 확인을 위한 키 IK(integrity key)를 생성하는 f4, 익명성을 위한 키 AK (anonymity key)를 생성하는 f5로 구성되어있다. (그림 4)는 Milenage의 블록도를 나타낸다.

3.2 입력 데이터 구성 방법

통계적 난수성을 평가하기 위해, 암호 알고리즘의 결과값을 그대로 NIST의 패키지에 적용시키는 것은 해당 암호 알고리즘의 안전성을 분석하는 데 충분하지 못하다. 차세대 암호 알고리즘을 선정하는 AES 프로젝트 수행 과정에서, 후보 알고리즘의 통계적 난수성 평가를 위해 NIST는 동일한 입출력 길이를 가진 블록암호에 적합한 입력 데이터 구성 방법을 제안하였다[3]. 데이터 구성 방법은 총 9 가지이며, 128비트 입력에 128비트를 출력하는 알고리즘을 기준으로 구성되었기 때문에 다중 블록 출력 함수의 구조를 지닌 키유도함수에 적합하지 않다. 따라서 본 절에서는 키유도함수에 적합한 입력 데이터 구성 방법을 새롭게 제시하고자 한다.

키유도함수의 출력값들은 통계 검정의 입력 데이터가 되어 난수성이 충분히 평가되어야만하고, 각각의 출력값 사이의 독립성이 평가되어야한다. 따라서 출력값 사이의 독립성을 평가하기 위해 n-비트 입력에 대해, n-비트 t개의 출력값 $(z_1,\cdots,z_t)$ 각각을 연접해서 데이터를 구성한다. 길이가 n-비트인 블록 t개를 길이가 tn인 하나의 블록 $(z_1\|z_2\|\cdots\|z_t)$으로 간주하며, 또한 표본 데이터 집합의 개수도 n-비트 입출력 함수를 검정할 때와 동일하거나 그 이상의 값으로 설정해 난수성을 충분히 평가한다. 이러한 기준이 만족되도록 다음과 같이 입력 데이터를 구성한다.

3.2.1 n-비트 키 쇄도(avalanche) 입력 데이터 집합

알고리즘의 민감함을 검정하기 위하여 입력 파라미터들(키 혹은 입력문)의 변화가 출력문에 어떤 영향을 주는지 s개의 검정 대상 수열들을 통해 분석된다.

① n-비트 키 K_i를 랜덤하게 l개 생성한다. $(1 \le i \le l)$
② 각각의 랜덤키 K_i의 1비트만을 치환(1은 0으로, 0은 1로 변형) n개의 키 K_{ij}(키 K_i의 j번째 비트를 치환)를 생성한다. $(1 \le i \le l, 1 \le j \le n)$
③ 각각의 키 K_{ij}로 모두 0으로 구성된 n-비트 입력문 I에 대한 키유도함수의 tn-비트 출력문 $Z_{ij} = KDF_{K_{ij}}(I)$을 계산한다. $(1 \le i \le l, 1 \le j \le n)$
④ 각각의 키 K_i로 모두 0으로 구성된 n-비트 입력문 I에 대한 키유도함수의 tn-비트 출력문 $Z_i = KDF_{K_i}(I)$을 계산한다. $(1 \le i \le l)$
⑤ Z_{ij}와 Z_i를 XOR시켜 검정 대상 수열 Z^*를 생성한다.

$$Z_i^* = Z_i \oplus Z_{i1} \| Z_i \oplus Z_{i2} \| \cdots \| Z_i \oplus Z_{in}, (1 \le i \le l),$$
$$Z^* = Z_1^* \| Z_2^* \| \cdots \| Z_l^*$$

⑥ ①~⑤ 과정을 s번 반복해 길이가 $l \cdot n \cdot t \cdot n$-비트인 검정 대상 수열을 s개 생성하여 통계적 난수성 검정의 입력 값으로 한다. 이 때, s는 표본 데이터 집합의 개수이고

ltn^2 은 검정 대상 수열의 길이다.

3.2.2 평문 쇄도 입력 데이터 집합

평문 쇄도 입력 데이터 집합은 키 쇄도 입력 데이터 집합과 동일한 방법으로 파라미터만 변형(키는 입력문으로, 입력문은 키로)시켜서 생성한다.

3.2.3 입출력 상관관계 입력 데이터 집합

입력과 출력 쌍들의 상관관계를 알아보기 위해 s개의 검정 대상 수열을 생성하여 분석한다.

① n-비트 키 K를 랜덤하게 생성한다.
② n-비트 입력문 블록 I_i를 랜덤하게 l개 생성한다. $(1 \le i \le l)$
③ 키 K로 입력문 블록 I_i에 대한 키유도함수의 tn-비트 출력문 $Z_i = KDF_K(I_i)$를 각각 계산한다. $(1 \le i \le l)$
④ $I_i^* \oplus Z_i$를 계산하여 검정 대상 수열 Z^*를 생성한다. $Z^* = I_1 \oplus Z_1 \| I_2 \oplus Z_2 \| \cdots \| I_l \oplus Z_l$. 이 때, I_i^*는 I_i를 t개 연접한 tn-비트 수열이다. $I_i^* = I_i \| \cdots \| I_i$
⑤ ①~④ 과정을 s번 반복해 길이가 $l \cdot t \cdot n$-비트인 출력문 수열을 s개 생성하여 통계적 난수성 검정의 입력 값으로 한다. 이 때, s는 표본 데이터 집합의 개수이고 ltn 은 검정 대상 수열의 길이다.

3.2.4 출력문 블록과 체이닝 모드 입력 데이터 집합

알고리즘의 구조적 특이성이 존재하는지 알아보기 위해 블록암호의 CBC모드와 유사한 과정을 통해 s개의 검정 대상 수열을 생성해 분석한다.

① n-비트 키 K를 랜덤하게 생성한다.
② 모두 0인 n-비트 입력문 I와 초기벡터 IV를 생성한다.
③ tn-비트 길이의 출력문 $Z_1 = KDF_K(IV \oplus I)$, $Z_{i+1} = KDF_K(Z_i \oplus I^*)$, $(1 \le i \le l-1)$을 계산하여 검정 대상 수열 Z^*를 생성한다. $Z^* = Z_1 \| Z_2 \| \cdots \| Z_l$. 이 때, I^*는 I를 t개 연접한 tn-비트 수열이다. 즉, $I^* = I \| \cdots \| I$ 이다.
④ ①~③ 과정을 s번 반복해 길이가 $l \cdot t \cdot n$-비트인 검정 대상 수열을 s개 생성하여 통계적 난수성 검정의 입력 값으로 한다. s는 표본 데이터 집합의 개수이고 ltn 은 검정 대상 수열의 길이다.

3.2.5 랜덤한 입력과 랜덤한 키 입력 데이터 집합

랜덤한 입력과 랜덤한 키에 의해 생성된 출력의 통계적 난수성을 검정하기 위해서, s개의 검정 대상 수열을 구성해 분석한다.

① n-비트 키 K를 랜덤하게 생성한다.
② n-비트 입력 블록 I_i를 랜덤하게 l개 생성한다. $(1 \le i \le l)$
③ 키 K로 입력문 블록 각각에 대한 길이가 tn-비트인 키유도함수의 출력문 $Z_i = KDF_K(I_i)$, $(1 \le i \le l)$를 계산하고, 검정 대상 수열 Z^*를 생성한다. $Z^* = Z_1 \| Z_2 \| \cdots \| Z_m$ 이다.
④ ①~③ 과정을 s번 반복해 길이가 $l \cdot t \cdot n$-비트인 검정 대상 수열을 s개 생성하여 통계적 난수성 검정의 입력 값으로 한다. s는 표본 데이터 집합의 개수이고 ltn은 검정 대상 수열의 길이다.

3.2.6 저밀도(low density) 입력 데이터 집합

알고리즘의 구조적 특이성 및 저밀도 입력과 출력과의 상관관계를 알아보기 위해 s개의 검정 대상 수열을 생성하여 분석한다.

① n-비트 키 K를 랜덤하게 생성한다.
② n-비트인 저밀도 입력 블록 I_i를 $\frac{n^2+n+2}{2}$개 생성한다$\left(1 \le i \le \frac{n^2+n+2}{2}\right)$. 저밀도 입력 블록은 모두 "0"으로 구성된 입력문 블록$\left(\binom{n}{0}=1\right)$과 한 개만 "1"이고 나머지 $n-1$개는 "0"으로 구성된 입력 블록$\left(\binom{n}{1}=n\right)$과 두 개만 "1"이고 나머지 $n-2$개는 "0"으로 구성된 입력 블록$\left(\binom{n}{2}=\frac{n^2-n}{2}\right)$을 더해, 총 $1+n+\frac{n^2-n}{2}$개로 구성된다.
③ 키 K로 입력 블록 각각에 대한 길이가 tn인 키유도함수의 출력 $Z_i = KDF_K(I_i)$를 계산하고 검정 대상 수열 Z^*을 생성한다.

$$Z^* = Z_1 \| Z_2 \| \cdots \| Z_{\frac{n^2+n+2}{2}}, \left(1 \le i \le \frac{n^2+n+2}{2}\right)$$

④ ①~③ 과정을 s번 반복해 s개의 길이가 $\frac{n^2+n+2}{2} \cdot t \cdot n$-비트인 출력문 수열을 생성해 통계적 난수성 검정의 입력값으로 한다. s는 표본 데이터 집합의 개수이고 $\frac{tn(n^2+n+2)}{2}$은 검정 대상 수열의 길이다.

3.2.7 저밀도 키 입력 데이터 집합

저밀도 키 입력 데이터 집합은 저밀도 입력 데이터 집합과 동일한 방법으로 파라미터만 변형(키는 입력으로, 입력은 키로)시켜서 생성한다.

3.2.8 고밀도(high density) 입력 데이터 집합

알고리즘의 구조적 특이성 및 고밀도 입력과 출력과의 상관관계를 알아보기 위해 s개의 검정 대상 수열을 생성하여

분석한다.

① n-비트 키 K를 랜덤하게 생성한다.

② n-비트인 고밀도 입력 블록 I_i를 $\frac{n^2+n+2}{2}$개 생성한다$\left(1 \le i \le \frac{n^2+n+2}{2}\right)$. 고밀도 입력 블록은 모두 "1"로 구성된 입력 블록$\left(\binom{n}{0}=1\right)$과 한 개만 "0"이고 나머지 $n-1$개는 "1"로 구성된 입력 블록$\left(\binom{n}{1}=n\right)$, 그리고 두 개만 "0"이고 나머지 $n-2$개는 "1"로 구성된 입력 블록$\left(\binom{n}{2}=\frac{n^2-n}{2}\right)$을 더해서 총 $1+n+\frac{n^2-n}{2}$개로 구성된다.

③ 키 K로 입력 블록 각각에 대한 길이가 tn인 키유도함수의 출력 $Z_i = KDF_K(I_i)$를 계산하고 검정 대상 수열 Z^*을 생성한다.

$$Z^* = Z_1 \| Z_2 \| \cdots \| Z_{\frac{n^2+n+2}{2}}, \left(1 \le i \le \frac{n^2+n+2}{2}\right)$$

④ ①~③ 과정을 s번 반복해 s개의 길이가 $\left(\frac{n^2+n+2}{2} \cdot t \cdot n\right)$-비트인 출력 수열을 생성해 통계적 난수성 검정의 입력값으로 한다. s는 표본 데이터 집합의 개수이고 $\frac{tn(n^2+n+2)}{2}$은 검정 대상 수열의 길이가 된다.

3.2.9 고밀도 키 입력 데이터 집합

고밀도 키 입력 데이터 집합은 고밀도 평문 입력 데이터 집합과 동일한 방법으로 파라미터만 변형(키는 평문으로, 평문은 키로)시켜서 생성한다.

3.3 키유도함수를 위한 파라미터 수정

NIST의 통계적 난수성 평가 패키지의 각 평가 항목은 검정의 입력 파라미터에 매우 민감하다. 신뢰할 수 있는 평가 결과를 얻기 위해서는 평가하고자 하는 알고리즘의 환경에 적합하게 검정 대상 수열의 길이, 표본 데이터 집합의 개수, 블록의 길이, 템플릿의 길이 등과 같은 파라미터를 설정해야 한다. NIST의 통계적 난수성 평가 항목의 파라미터는 입출력 블록 길이가 동일한 블록암호에 적합하게 구성되어 있으므로 출력 길이가 확장되는 키유도함수의 통계적 난수성 평가에 적용하기 위해서는 파라미터의 수정이 불가피하다. 본 소절에서 우리는 블록암호의 통계적 난수성 평가를 위한 NIST 패키지와 일관성 있는 평가 방법을 키유도함수에 적용하기 위한 파라미터 구성 방법에 대하여 논한다.

3.3.1 검정 대상 수열의 길이

NIST 패키지에서는 신뢰할 수 있는 실험 결과를 얻기 위해서 검정 대상 수열의 길이는 최소한 10^6은 넘어야 할 것을 권고하고 있다. 이 권고 사항은 통계적 난수성 검정에서 일반적으로 적용되는 기준으로 볼 수 있으므로 입출력이 n-비트 길이인 함수이거나 출력 길이가 tn-비트로 확장되는 키유도함수인 경우 모두 검정 대상 수열의 길이가 10^6 이상이 되면 통계적 난수성 검정에 무리가 없음을 말해주는 것이다. 하지만 우리는 함수의 출력이 t배 확장되는 키유도함수의 경우 검정 대상 수열의 길이를 t배로 확장시키는 것이 NIST 패키지와 일관성 있는 검정 방법을 유지하기 위해서 필요한 조치임을 알 수 있다.

3.3.2 표본 데이터 집합의 개수

표본 데이터 집합의 개수는 보통 유의수준의 설정 값과 관련이 있다. NIST 패키지에서는 유의수준을 $[0.001, 0.01]$의 범위 내의 값을 선택하기를 권고하고, 표본 데이터 집합의 개수는 최소한 유의 수준의 역수(inverse)인 α^{-1}을 넘도록 권고한다[2]. 만약 유의수준이 너무 작아 기준이 엄격해지면 기준을 통과하지 못하지만 실제로 난수성을 만족하는 표본들이 발생할 수 있고, 그렇지 않다면 실제로 난수성을 만족하지 못하는 표본들도 통과하는 오류가 발생할 수 있다. 때문에 키유도함수의 경우 역시 $[0.001, 0.01]$의 범위 내의 값을 선택하고 표본 데이터 집합의 개수는 α^{-1}을 넘도록 한다.

3.3.3 블록 길이

NIST의 통계적 난수성 평가 항목 중 몇 가지는 검정 대상 수열을 블록 별로 분할하기 때문에 분할하는 길이의 선택이 중요하다. 출력이 t배 확장되는 키유도함수의 경우 블록 길이도 t배 늘려준다.

① 블록 내부의 빈도 검정

만약 입출력이 n-비트인 알고리즘이라면, 보통 알고리즘의 한 블록당 비트 수 M을 블록 길이로 사용한다. 예를 들어, AES는 128을 DES는 64를 사용한다. 하지만 만약 출력이 t배 확장되는 알고리즘이라면, 블록 길이는 $t \cdot M$으로 설정한다. 예를 들어, AES를 핵심함수로 사용하는 Milenage인 경우 블록 길이는 $128 \cdot 5 = 640$이 된다.

출력이 t배 확장되는 키유도함수를 위한 블록 내부의 빈도 검정은 검정 대상 수열을 분석하여 겹치지 않는 블록에서 이상적인 1의 빈도로부터의 편차를 χ^2-분포를 이용해 계산한다. 검정 내에서 전체 검정 대상 수열은 tM-비트 크기의 블록들 $\left\lfloor \frac{t\rho}{tM} \right\rfloor = \left\lfloor \frac{\rho}{M} \right\rfloor = N$ 개로 나뉘어져 검정된다. i번째 블록에서 1의 비율을 π_i라 할 때, 랜덤성 가설하에 다음 합이 자유도 N인 χ^2-분포를 갖는다는 사실을 이용하여 $p-value$ 값이 계산된다.

$$\chi^2(obs) = 4(tM)\sum_{i=1}^{N}\left[\pi_i - \frac{1}{2}\right]^2$$

우리는 키유도함수를 위한 검정 대상 수열의 길이를 NIST 패키지에서 보다 t 배로 증가시켰기 때문에 위에서 보는 바와 같이 χ^2-분포의 자유도 N은 NIST 패키지 내의 자유도와 동일하다. 그러므로 블록 길이와 검정 대상 수열의 길이를 동시에 t 배로 증가시키는 것이 NIST 패키지와 일관성 있는 검정을 수행하는 방법이 됨을 알 수 있다.

② 선형 복잡도 검정

NIST 패키지에서 입출력이 n-비트인 알고리즘의 선형복잡도 검정은 검정 대상 수열의 길이가 ρ인 경우 이를 M-비트 길이의 블록 N개로 분할함으로써 검정을 시작한다. 즉, $\rho = MN$이다. 출력이 t 배 확장되는 키유도함수를 위한 선형복잡도 검정은 ρ와 M을 t 배 증가시켜 각각 $t\rho$와 tM으로 수정한 파라미터를 사용한다.

NIST 패키지의 선형복잡도 검정에서는 $\rho \geq 10^6$, $500 \leq M \leq 5000$, $N \geq 200$ 을 만족하는 파라미터 설정을 권고하고 있다. 키유도함수를 위한 선형복잡도 검정에서 이 권고 사항을

$$t\rho \geq 10^6,\ 500 \leq tM \leq 5000,\ N \geq 200$$

으로 수정하면 NIST 패키지와 일관성 있는 검정 방법이 됨을 알 수 있다.

3.3.4 템플릿의 길이

난수성 평가 항목 중에서 몇 가지는 템플릿을 통해 검정 대상 수열에 특정한 패턴이 얼마나 반복되는지를 검정한다. 이런 평가 항목의 경우 검정 결과는 템플릿의 길이에 매우 큰 영향을 받는다. NIST의 패키지에서 템플릿 길이는 검정 대상 수열이 ρ-비트일 경우, 보통 $\lfloor \log_2 \rho \rfloor$ 와 근사한 값으로 설정된다. 따라서 출력이 t 배 확장되어 검정 대상 수열이 $t\rho$-비트인 키유도함수의 경우 템플릿의 길이는

$$\begin{aligned} \lfloor \log_2 t\rho \rfloor &= \lfloor \log_2 t + \log_2 \rho \rfloor \\ &= \lfloor \log_2 t \rfloor + \lfloor \log_2 \rho \rfloor + \delta\ (\delta = 0 \text{ 또는 } 1) \end{aligned}$$

으로 설정하는 것이 합리적이다. 여기에서 δ는 경우에 따라 0 또는 1의 값을 가지는 양이므로 실제적인 응용 환경에서는 무시할만하여 항상 0으로 설정해도 무방하다.

① 겹치지 않는 템플릿 매칭 검정

NIST의 패키지에서 입출력 길이가 n-비트인 알고리즘의 겹치지 않는 템플릿 매칭 검정은 전체 길이가 ρ인 검정 대상 수열을 M-비트 길이의 블록 N개로 분할하여 각 블록 내에 $\lfloor \log_2 M \rfloor$ 보다 작거나 같은 m-비트 길이의 겹치지 않는 템플릿의 비율을 계산한다. 즉, $m \leq \lfloor \log_2 M \rfloor$ 이다.

출력이 t 배 확장되는 키유도함수를 위한 겹치지 않는 템플릿 매칭 검정은 전체 길이가 $t\rho$인 검정 대상 수열을 길이가 tM-비트인 블록 N개로 나누어 $(m + \lfloor \log_2 t \rfloor)$-비트 길이의 템플릿 비율을 계산한다. 블록 길이와 검정 대상 수열의 길이를 동시에 t 배 증가시켰으므로 χ^2-분포의 자유도 N은 NIST 패키지 내의 자유도와 동일하다. 또한, t 배 증가한 블록 내에서 템플릿의 빈도를 계산하기 때문에 템플릿의 길이는 $m + \lfloor \log_2 t \rfloor$ 로 증가시키고 이 값은

$$\begin{aligned} \lfloor \log_2 tM \rfloor &= \lfloor \log_2 t + \log_2 M \rfloor \\ &= \lfloor \log_2 t \rfloor + \lfloor \log_2 M \rfloor + \delta \\ &\geq \lfloor \log_2 t \rfloor + m \end{aligned}$$

이므로 NIST 패키지와 일관성 있는 검정을 수행하는 방법이 됨을 볼 수 있다.

② 겹치는 템플릿 매칭 검정

입출력이 n-비트인 알고리즘의 겹치는 템플릿 매칭 검정은 겹치지 않는 템플릿 매칭 검정처럼 전체 길이가 ρ인 검정 대상 수열을 M-비트 길이의 블록 N개로 분할하여 검정을 시작한다. 겹치지 않는 템플릿 매칭 검정과는 달리 분할한 블록 내에 겹치는 $\lfloor \log_2 M \rfloor = m$ 을 만족하는 m-비트 길이의 템플릿 비율을 계산한다.

출력이 t 배 확장되는 키유도함수를 위해 ρ와 M을 t 배 증가시킨 $t\rho$와 tM으로 수정하고, 템플릿의 길이를 $\lfloor \log_2 t \rfloor$ 만큼 증가시킨 $m + \lfloor \log_2 t \rfloor$ 로 수정한다. NIST 패키지의 겹치는 템플릿 매칭 검정에서 템플릿 길이는 $\lfloor \log_2 M \rfloor = m$ 을 만족하도록 권고하며, 키유도함수를 위해 수정한 템플릿 길이는

$$\begin{aligned} \lfloor \log_2 tM \rfloor &= \lfloor \log_2 t \rfloor + \lfloor \log_2 M \rfloor + \delta \\ &= \lfloor \log_2 t \rfloor + m + \delta \end{aligned}$$

이므로 NIST 패키지와 일관성 있는 검정이 수행될 수 있음을 알 수 있다. 여기에서 δ는 0 또는 1이므로 무시할 수 있는 양이다.

③ 점근적 엔트로피 검정

NIST 패키지에서 입출력이 n-비트인 알고리즘의 점근적 엔트로피 검정은 길이가 ρ-비트인 검정 대상 수열에서 총 2^m개인 m-비트 템플릿들의 엔트로피를 이용하여 $p-value$를 계산한다. 출력이 t 배 확장되는 키유도함수를 위해 ρ를 t 배 증가시킨 $t\rho$로 수정하고 템플릿 길이를 $m + \lfloor \log_2 t \rfloor$ 로 수정한다.

NIST 패키지의 점근적 엔트로피 검정에서는 템플릿의 길이 m이 $m < \lfloor \log_2 \rho \rfloor - 2$를 만족하도록 권고하고 있다. 키유도함수를 위한 점근적 엔트로피 검정에서 수정한 템플릿 길이는

$$m + \lfloor \log_2 t \rfloor < \lfloor \log_2 \rho \rfloor - 2 + \lfloor \log_2 t \rfloor \leq \lfloor \log_2 t\rho \rfloor - 2$$

이므로 NIST 패키지와 일관성 있는 검정 방법임을 알 수 있다.

④ 시리얼 검정

NIST 패키지에서 입출력이 n-비트인 알고리즘의 시리얼 검정은 ρ-비트인 검정 대상 수열에서 2^m개의 m-비트 템플릿, 2^{m-1}개의 $(m-1)$-비트 템플릿, 2^{m-2}개의 $(m-2)$-비트 템플릿의 비율을 이용하여 $p-value$를 계산한다. 출력이 t배 확장되는 키유도함수를 위해 ρ를 t배 증가시킨 $t\rho$로 수정하고 템플릿 길이를 $m + \lfloor \log_2 t \rfloor$로 수정한다.

NIST 패키지의 시리얼 검정에서는 템플릿의 길이 m이 점근적 엔트로피 검정과 같은 조건을 만족하도록 권고하고 있으므로 키유도함수를 위한 시리얼 검정 역시 NIST 패키지와 일관성 있는 검정 방법임을 알 수 있다.

3.4 통계적 난수성 평가 기준 분석

NIST 패키지에서 통계적 난수성 평가 기준으로 기각된 표본 개수를 분석하는 방법을 적용했다. 키유도함수를 위한 통계적 난수성 평가 역시 검정 대상 수열의 기각 여부를 결정하고 기각된 표본 데이터 집합의 개수를 분석하는 방법을 평가 기준으로 적용하고자 한다.

먼저 난수성 평가의 기준이 되는 최대 허용 기각수(이하 최대기각수)를 정해진 방법에 따라 계산하고, 해당 알고리즘의 기각된 표본 데이터 집합의 총 개수를 계산한다. 만약 어떤 통계적 난수성 평가 항목에 대해 기각된 표본 데이터 집합의 개수가 최대기각수를 넘는다면 알고리즘은 해당 통계적 난수성 평가 항목에 대해서 통계적 비난수성을 보인다고 한다. 만약 기각된 수열 개수가 최대기각수를 넘지 않는다면 알고리즘은 해당 통계적 난수성 평가 항목에 대해서 통계적 난수성을 보인다고 한다. 평가 기준이 되는 최대 기각수를 설정하는 방법에는 이항분포를 이용하는 방법과 정규분포를 이용하는 방법 두 가지가 있으며, 본 논문의 시뮬레이션 결과 분석에는 오차가 더 적은 이항분포를 이용하는 방법을 사용하고자 한다.

3.4.1 이항분포를 이용하는 방법

어떤 시행의 결과가 “성공”과 “실패” 두 가지 가능한 결과만을 가질 경우 이를 베르누이 시행이라 한다. 이 때 성공 확률을 p, 실패 확률을 $1-p$로 나타내며, 일반적으로 성공 확률이 p인 베르누이 시행을 r번 반복 할 때, 성공 횟수 X의 분포를 파라미터가 r과 p인 이항분포라 하고 $B(r,p)$로 나타낸다. 이항분포 $B(r,p)$의 확률밀도함수는 다음과 같다.

$$P(x=k) = \binom{r}{k} p^k (1-p)^{r-k} .$$

알고리즘의 통계적 난수성 평가에서 검정 대상 수열은 “채택” 혹은 “기각” 두 가지 가능한 결과만을 가지기 때문에 통계적 평가 과정을 이항분포 $B(s,p)$를 따르는 시행으로 볼 수 있다. 이 때, s는 표본 데이터 집합의 개수이고 “기각” 확률 p는 유의수준과 동일하다고 해석한다. 최대기각수를 계산하기 위해 정규근사 신뢰수준 0.001을 기준으로 하여 표본 데이터 집합의 개수 s에 대한 각 기각수의 누적 이항분포를 계산한다. <표 5, 6, 7, 8>은 표본 데이터 집합의 개수에 따른 누적이항분포를 계산한 것이다.

누적이항분포가 0.001이 되는 정확한 k값을 찾기 어려우므로 k값은 근사값으로 구하려 한다. <표 9>는 검정 대상 수열 개수에 따른 최대기각수를 나타낸 것이다.

〈표 5〉 $s=128$일 때 누적이항분포

표본데이터집합 개수	채택 개수	기각 개수	$P(s,k)$	누적 이항분포
128	125	3	0.09719	0.04028
128	124	4	0.03067	0.00960
128	123	5	0.00768	0.00192
128	122	6	0.00159	0.00032
128	121	7	0.00028	0.00004

〈표 6〉 $s=300$일 때 누적이항분포

표본데이터집합 개수	채택 개수	기각 개수	$P(s,k)$	누적 이항분포
300	293	7	0.02127	0.01147
300	292	8	0.00787	0.00360
300	291	9	0.00257	0.00102
300	290	10	0.00075	0.00026
300	289	11	0.00020	0.00006

〈표 7〉 $s=500$일 때 누적이항분포

표본데이터집합 개수	채택 개수	기각 개수	$P(s,k)$	누적 이항분포
384	376	8	0.02489	0.01644
384	375	9	0.01050	0.00594
384	374	10	0.00397	0.00196
384	373	11	0.00136	0.00059
384	372	12	0.00042	0.00016

〈표 8〉 $s=1000$ 일 때 누적이항분포

표본데이터집합 개수	채택 개수	기각 개수	$P(s,k)$	누적 이항분포
1000	982	18	0.006927	0.006905
1000	981	19	0.003616	0.003289
1000	980	20	0.001791	0.001497
1000	979	21	0.000844	0.000652
1000	978	22	0.000379	0.000273

〈표 9〉 이항분포를 따르는 최대기각수

표본데이터 집합 개수	유의 수준	신뢰 수준	최대 기각수
128	0.01	0.001	5
300	0.01	0.001	9
384	0.01	0.001	10
1000	0.01	0.001	20

3.4.2 정규분포를 이용하는 방법

정규분포는 특성값이 연속적인 무한모집단 분포의 일종으로서 가장 대표적인 표본분포이다. 정규분포의 밀도곡선은 한 점을 중심으로 좌우 대칭이고 대칭점에서 그 높이가 가장 높으며 대칭점에서 멀어질수록 점점 낮아지는 성질이 있다. 원래 검정 대상 수열의 채택 및 기각에 대한 확률은 이항분포를 따르지만 이항분포 $B(s,p)$의 s가 크면 확률의 계산이 힘들어지므로 정규분포를 이용해 신뢰구간을 측정한다. 검정 대상 수열의 길이가 ρ이고 표본 데이터 집합의 개수를 s라고 하면, 검정 대상 수열에 대한 귀무가설이 기각된 표본 데이터 집합의 개수는 이항분포 $B(s,p)$를 따른다. 만약 s가 충분이 크다면 이항분포 $B(s,p)$는 근사적으로 정규분포 $N(sp, sp(1-p))$를 따른다. 대체로 $s>30$일 때, 이를 다시 쓰면,

$$\frac{X-sp}{\sqrt{sp(1-p)}}=\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{s}}}\sim N(0,1)$$

이 된다.

검정 대상 수열의 기각과 채택 비율을 추정하는 신뢰구간을 설정하기 위해 신뢰수준 α_c는 0.001로 고정한다. 그러면 99.9% 신뢰수준의 신뢰구간은 양측검정일 경우

$$\left(p-z_{\alpha_c/2}\sqrt{\frac{p(1-p)}{s}}\ ,\ p+z_{\alpha_c/2}\sqrt{\frac{p(1-p)}{s}}\right)$$

이고, 최대값을 고려하는 단측 검정일 경우

$$\left(0\ ,\ p+z_{\alpha_c}\sqrt{\frac{p(1-p)}{s}}\right)$$

이다. 이 때, 기각수가 적으면 적을수록 좋은 결과이기 때문에 최대값만 고려하는 단측 검정으로 신뢰구간을 설정하자. 그러면 최대기각수는

$$s\left(p+z_{\alpha_c}\sqrt{\frac{p(1-p)}{s}}\right)$$

〈표 10〉 정규분포를 따르는 최대기각수

표본데이터 집합 개수	$\alpha=0.01$	신뢰 구간	최대 기각수
128	1.28	4.657	4
300	3.00	8.170	8
384	3.84	9.689	9
1000	10.00	19.439	19

이 된다. $z_{0.001}\fallingdotseq 3$이고 p는 유의수준으로 본 논문에서는 0.01로 설정한다. 정규분포에 대한 최대기각수는 다음 <표 10>과 같다.

표본 데이터 집합의 개수에 따른 최대기각수를 살펴보면 위의 이항분포를 이용한 것과 약간의 오차가 있는 것을 볼 수 있다. 이것은 정규분포로 계산한 것은 이항분포를 근사시킨 값이기 때문에 발생한 것이다. 만약 표본 데이터 집합의 개수가 더 커진다면 그 오차는 줄어들 것이다.

4. 시뮬레이션 수행 및 결과 분석

본 논문에서 제안한 다중 블록 출력 함수의 구조를 가진 키유도함수에 대한 통계적 난수성 평가 방법을 실제로 시뮬레이션을 통해 적용해 보고자 한다. 시뮬레이션은 블록암호 기반 키유도함수의 대표적 표준인 3GSM의 Milenage와 NIST가 제안한 Counter, Feedback, DPI 모드에 대하여 실시하고, 수행 결과를 분석하고자 한다.

4.1 시뮬레이션 환경

Milenage와 Counter, Feedback, DPI 모드의 내부 핵심함수는 128비트 길이의 입출력인 AES를 사용하고자 한다. 각 알고리즘의 입력은 128비트 평문블록이고 출력은 128비트 키에 대해 128비트 블록을 5개 연접해 5배로 확장된 640비트 수열이다.

통계적 난수성 평가를 위한 입력 데이터는 3.2절에서의 방법으로 구성한다. 검정 대상 수열 길이는 저밀도 및 고밀도 입력 데이터 집합들의 경우 5,284,480 비트이고, 나머지 입력 데이터 집합들은 5,242,880 비트로 한다. 통계적 난수성 평가를 위한 유의수준은 0.01로 설정하고, 각 입력 데이터 집합의 표본 데이터 집합의 개수는 1,000개로 한다. 검정 대상 수열의 통계적 난수성 혹은 비난수성을 판단할 기준인 최대기각수는 이항분포를 적용한 20개를 사용한다. 만약 기각수가 20개를 넘으면 해당 평가 항목에 대해 검정 대상 수열은 통계적 비난수성을 보이고 해당 알고리즘은 비난수성을 보이는 것으로 판단한다.

<표 11>은 시뮬레이션할 키유도함수의 수정된 파라미터 값이다. 내부 함수를 AES로 사용할 것이기 때문에 기존 AES의 입력 파라미터 값을 기준으로 새롭게 설정하였다.

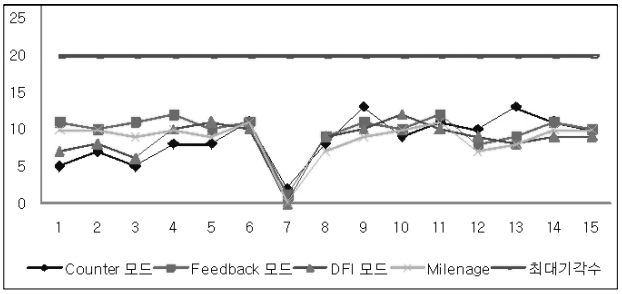
〈표 11〉 각 평가 항목의 수정된 파라미터 값

평가 항목	기존 파라미터 값		수정 파라미터 값	
	블록길이	템플릿길이	블록길이	템플릿길이
블록 내부의 빈도	128	-	640	-
겹치지 않는 템플릿 매칭	2^{17}	9	$5 \cdot 2^{17}$	11
겹치는 템플릿 매칭	1032	9	5160	11
점근적 엔트로피	-	10	-	12
시리얼	-	16	-	18
선형 복잡도	500	-	2,500	-

4.2 시뮬레이션 결과 분석

키유도함수의 시뮬레이션 수행 결과는 (그림 5)와 같다. (그림 5)의 x축은 단일 비트 빈도 검정부터 변형 랜덤 익스커전 검정까지의 통계적 난수성 평가 항목들이고, y축은 해당 평가 항목에 대한 알고리즘의 기각수이다. Counter모드, Feedback모드, DPI모드 그리고 Milenage를 시뮬레이션 하여 통계적 난수성 평가를 수행하였다. 검정을 실시한 키유도함수들은 모든 통계적 난수성 평가 항목에서의 기각수가 최대기각수를 넘지 않으므로 평가 기준에 따라 통계적 난수성을 만족한다고 할 수 있다.

한편, Counter 모드의 경우 구조적 특성 때문에 입력 데이터 구성 시 충돌쌍이 발생하여 통계적 난수성 검정을 통과하지 못할 수가 있다. 하지만 키 유도 함수는 출력을 키로 사용하는 것이 목적이고 입력은 서로 달라야 하는 것이 원칙이므로 충돌쌍을 제외할 수 있는 입력 데이터 집합을 사용하였다. Counter 모드에서 충돌쌍을 피할 수 있는 방법으로는 회전(rotation) 연산과 상수를 적절히 추가해주는 방안이 널리 사용된다.



(그림 5) 키유도함수의 통계적 검정 결과

5. 결 론

본 논문에서는 다중 블록 출력 함수의 형태를 가진 키유도함수의 통계적 난수성 평가 방법을 새롭게 제안하였다. 또한 대표적인 키유도함수를 시뮬레이션 하여 새로운 평가 방법을 적용하고, 그 결과를 분석해 보았다. NIST의 통계적 난수성 평가 패키지를 세밀히 분석하여 키유도함수의 평가에 적용하기 위한 파라미터의 수정 및 평가 기준을 합리적으로 제안하였고, 3GSM의 Milenage와 NIST의 세 가지 모드에 대한 통계적 난수성 평가를 실시하였다.

키유도함수는 정보보호 관점에서 매우 중요한 역할을 하지만, 그 중요성에 비해 활발한 연구가 진행되고 있지 않다. 따라서 본 논문은 키유도함수의 기본적인 안전성 분석 방법인 통계적 난수성 평가 방법을 연구해 좀 더 개선시킨 것에 그 의미가 있다고 하겠다. 향후 기존에 있던 안전성 평가 방법을 개선하는 것에 그치지 않고 키유도함수에 대한 새로운 안전성 평가 방법에 대한 연구가 필요할 것으로 사료된다.

참 고 문 헌

[1] J. Nechvatal, E. Barker, L. Bassham, and W. Burr, “Report on the Development of the Advanced Encryption Standard (AES),” Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology, 2000.

[2] A. Rukhin, J. Soto, J. Nechvatal, M. Smid, E. Barker, S. Leigh, M. Levenson, M. Vangel, D. Banks, A. Heckert, J. Dray, and S. Vo, “A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications,” NIST SP800-22, 2008.

[3] J. Soto, “Randomness Testing of the AES Candidate Algorithms,” NIST, 1999.

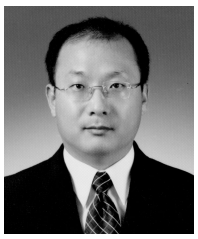
[4] 3GPP TR 35.909 v8.0.0 : “3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; 3G Security; Specification of the MILENAGE Algorithm Set; An example algorithm set for the 3GPP authentication and key generation functions f1, f1*, f2, f3, f4, f5 and f5*; Document 5: Summary and results of design and evaluation,” 2008.

[5] IEEE 802.11i, “IEEE Standard for Information technology－Telecommunications and information exchange between systems－Local and metropolitan area networks－Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 6: Medium Access Control (MAC) Security Enhancements,” 2004.

[6] IEEE 802.15.1™, “IEEE Standard for Information technology－Telecommunications and information exchange between systems－Local and metropolitan area networks－Specific requirements Part 15.1: Wireless medium access control (MAC) and physical layer (PHY) specifications for wireless persHY) sarea networks (WPANs),” 2002.

[7] L. Chen, “Recommendation for Key Derivation Using Pseudorandom Functions,” NIST SP800-108, 2008.

[8] H. Gilbert, “The Security of “One-Block-to-Many” Modes of Operation,” FSE 2003 LNCS 2887, pp.376-395, 2003.



강 주 성

e-mail : jskang@kookmin.ac.kr
1989년 고려대학교 수학과(학사)
1991년 고려대학교 수학과(이학석사)
1996년 고려대학교 수학과(이학박사)
1997년~2004년 한국전자통신연구원 선임연구원
2001년~2002년 벨기에 루벤대학 COSIC 방문연구원
2004년~현 재 국민대학교 수학과 부교수
관심분야 : 암호이론, 정보보호이론, 응용확률론 등

이 옥 연

e-mail : oyyi@kookmin.ac.kr
1988년 고려대학교 수학과(학사)
1990년 고려대학교 수학과(이학석사)
1996년 University of Kentucky 수학과 (이학박사)
1999년~2001년 한국전자통신연구원 선임연구원, 팀장
2001년~현 재 국민대학교 수학과 부교수
관심분야 : 정보보호, 이동통신, 암호론 등

염 지 선

e-mail : bohun86@hanmail.net
2008년 국민대학교 수학과(학사)
2010년 국민대학교 수학과 이학석사
관심분야 : 암호이론, 정보보호이론 등

조 진 웅

e-mail : chojw@keti.re.kr
1986년 광운대학교 전자통신공학과(학사)
1988년 광운대학교 전자통신공학과(석사)
2001년 광운대학교 전자통신공학과(박사)
1999년 (日本) ETL 연구소, Fellow Ship 초빙연구원
1993년~현 재 전자부품연구원(KETI) 센터장
관심분야 : 근거리 무선통신 및 네트워킹, 통신 SoC 설계, 정보보안