通信情報保護學會誌
第1卷, 第3號, 1991. 12

CORRELATION 공격

성 수 학*

1. 서 론

Stream 암호 시스템의 구성은 평문을 이진 수열로 부호화하여 이진수열 발생기(RKG, running key generator)에서 생성된 이진 수열과 비트별 XOR하여 이진 수열로 된 암호문을 발생한다(그림 1). 이진 수열 발생기는 최대 주기를 갖는 N개의 선형 쉬프트 레지스터(LFSR, linear feedback shift register)를 비선형 함수 f와 결합하여 구성한다(그림 2). 좋은 암호 시스템이 되기 위해서는 출력 수열의 주기가 크고 선형 복잡도(LC, linear complexity)가 커야 한다. 대표적인 알고리즘으로는, 승산(multiplication) 시스템, J-K 플립-플롭, Geffe 시스템 등이 있다[16]. 출력 수열의 주기와 선형 복잡도는 LFSR의 크기와 비선형 함수에 달려있다. 그러나, 출력 수열과 어떤 LFSR의 수열이 상관관계(correlation)가 있으면 이 LFSR에 대한 correlation 공격이 가능하다. Blaser와 Heinzmann[1]에 의해 최초로 correlation 공격 가능성이 제시되었으며, Siegenthaler[12]에 의해 이론적으로 완성되었다. LFSRi의 키(key)는 초기 벡터(초기값)와 원시다항식인 선형 궤환함수(linear feedback function)로 구성된다. LFSRi의 크기가 r_i이면 최대 주기의 수열을 만들어 낼 수 있는 초기 벡터의 갯수는 $2^{r_i}-1$이며, 원시 다항식의 갯수도 구할 수 있으나, 초기 벡터의 갯수에 비하면 아주 작다[4]. R_i를 원시 다항식의 갯수라고 하면, LFSRi의 키의 갯수는 $R_i(2^{r_i}-1)$이다. 일반적으로 비선형 함수는 알려지므로, N개의 LFSR로 구성된 이진 수열 발생기의 키의 총 갯수는 $\prod_{i=1}^{N} R_i(2^{r_i}-1)$이다. 따라서 공격자는 최악의 경우 $\prod_{i=1}^{N} R_i(2^{r_i}-1)$회의 시행후에 정확한 키를 찾을 수 있다. 그러나, RKG의 출력 수열과 LFSRi의 출력수열 사이에 상관관계가 있으면 LFSRi의 키를 LFSRj($j\neq i$)의 키와 관계없이 찾을 수 있다. 따라서, 상관관계가 있는 경우 키를 찾는 시행 횟수는 $\sum_{i=1}^{N} R_i(2^{r_i}-1)$로 크게 줄일 수 있다. 예를들면, Geffe, Pless, Bruder 시스템의 경우 상관관계가 있으며, LFSR의 크기가 50 이하이면 correlation 공격이 가능하다.

최근에 많은 학자들이 correlation 연구를 두 방향으로 하고 있다. 상관관계를 가지지 않는 비선형 함수(correlation immune) 연구와 Siegenthaler 알고리즘 보다 빠른 알고리즘 개발을 연구 하고 있다.

* 정회원, 배재대학교 응용수학과 조교수.

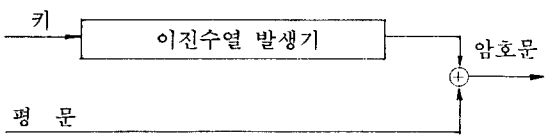

그림 1. Stream 암호 시스템 구성도

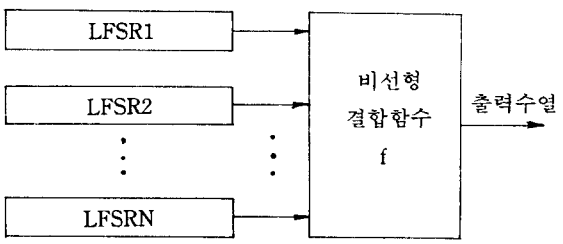

그림 2. 이진 수열 발생기

2. Correlation 공격의 예

Correlation 공격의 간단한 예로 Geffe 시스템을 생각해보자. 3개의 LFSR로 구성된 Geffe 시스템은 그림 3과 같다.

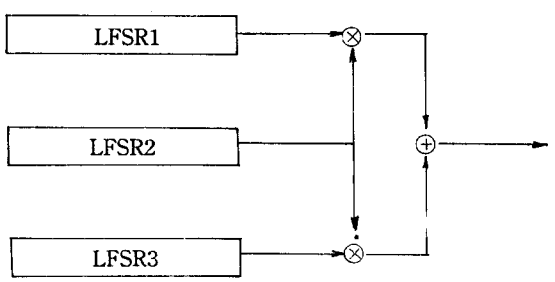

그림 3. Geffe 시스템, ⊗는 비트별 곱셈, •는 역수

비선형 함수 f를 대수적 정규 형태(algebraic normal form)로 나타내면 다음과 같다.

$$f(x_1, x_2, x_3) = x_1x_2 \oplus x_3(x_2 \oplus 1)$$
$$= x_1x_2 \oplus x_2x_3 \oplus x_3$$

각 x_i(i=1, 2, 3)의 입력값에 대한 출력값은 표 1과 같다.

표 1. $f(x_1, x_2, x_3) = x_1x_2 \oplus x_2x_3 \oplus x_3$의 함수값.

x_1	x_2	x_3	$f(x_1, x_2, x_3)$
0	0	0	0
0	0	1	1
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

표 2. f, x_1f, x_2f, x_3f의 함수값

x_1	x_2	x_3	f	x_1f	x_2f	x_3f
0	0	0	0	0	0	0
0	0	1	1	0	0	1
0	1	0	0	0	0	0
0	1	1	0	0	0	0
1	0	0	0	0	0	0
1	0	1	1	1	0	1
1	1	0	1	1	1	0
1	1	1	1	1	1	1

f의 입력값이 같은 확률을 가지면 (각 x_i가 0과 1을 가질 확률이 같다) 출력 수열도 0과 1이 같은 확률로 나타난다. 그러나, 어떤 사람이 x_1의 값을 누출시키면 x_1과 $f(x_1, x_2, x_3)$가 같은 확률은 3/4이므로 출력 수열에 관한 불확실성은 줄어든다. 따라서, 쉽게 Geffe 시스템을 공격할 수 있다. 공격자는 f의 출력값과 각 LFSR의 원시다항식을 안다고 가정하면, 키는 각 LFSR의 초기 벡터(초기값)가 된다. LFSR1의 옳은 초기값(키)에서 나온 수열과 f의 출력 수열을 곱하면 0이 될 확률은 3/8이 된다(표 2). 그러나, 두개의 랜덤 수열을 곱하면

0이 될 확률은 1/4이다(0×0=0, 0×1=0, 1×0=0, 1×1=1). 따라서, LFSR1의 모든 초기값을 변화시켜 나온 수열과 주어진 f의 출력수열을 곱하여 0이 될 확률이 3/8일때의 LFSR1의 초기값이 LFSR1의 키가 된다. x_3와 $f(x_1, x_2, x_3)$가 같은 확률도 위의 경우와 같이 3/4이다. 따라서, 위와같은 방법으로 LFSR3의 키도 찾을 수 있다. x_2와 $f(x_1, x_2, x_3)$가 같은 확률은 1/2이므로 위의 방법을 적용할 수 없다. 그러나, LFSR1과 LFSR3의 키를 찾은후 LFSR2의 모든 가능한 초기값을 변화시켜 나온 출력수열과 주어진 f의 출력수열을 비교하여, LFSR2의 키를 찾을 수 있다. 이와같이, correlation 공격을 이용하여 Geffe 시스템을 공격하는데, 필요한 시행 횟수는 $(2^{r_1}-1)+(2^{r_2}-1)+(2^{r_3}-1)$ (r_i는 LFSRi의 크기)이다. 그러나, 모든 가능한 시행(exhaustive trial, exhaustive search)을 하는데, 필요한 횟수는 $(2^{r_1}-1)(2^{r_2}-1)(2^{r_3}-1)$이다.

3. Correlation Immunity

제 2 절에서 언급한 correlation 공격을 피하기 위해 비선형 함수 f가 상관관계를 가지지 않는 경우를 생각하게 되었다. 이러한 함수를 correlation immune이라 한다. 함수 f의 correlation immunity를 정의하기 위해 그림 4와 같은 이론적인 정보 모델(information-theoretic model)을 생각해 보자.

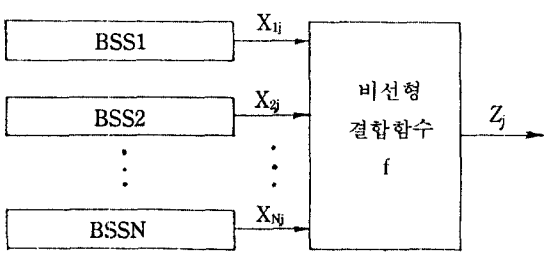

그림 4. 이론적 정보 모델, BSS(binary symmetric source)

그림 4에서 X_{ij}는 BSSi에서 생선된 j번째 출력값이다. 즉, 비선형 결합 함수 f의 입력 수열은 독립이고 0과 1을 가질 확률이 같다. f의 출력 수열 Z_j가 각 BSS의 j번째 출력값의 결합으로 이루어져 있을때 f는 메모리를 가지지 않는 (memoryless) 함수라 정의하고, Z_j가 j이전의 BSS의 출력값과 관련이 있을때, f는 메모리를 가진다고 정의한다. 우선 f가 메모리를 가지지 않는 경우에 대해 생각해 보자.

정의 3. 1[11] f의 출력 수열이 어떤 $m(1\leq m<N)$개의 입력 수열과 독립일때 f는 order m의 correlation immune을 가진다고 정의한다. 식으로 쓰면,

$$P(Z=1 \mid X_{i1}=a_1, \cdots, X_{im}=a_m)=P(Z=1),$$
$$1\leq i_1< \cdots < i_m\leq N,\ a_i=0 \text{ 또는 } 1.$$

즉,

$$I(Z; X_{i1}, \cdots X_{im})=0$$

정의 3. 1에서 f가 메모리를 가지지 않기 때문에 타임 파라메터(time parameter)를 언급하지 않았다. 왜냐하면, f의 출력수열은 항상 현재의 입력 수열에만 의존하기 때문이다. 만일 f가 상수이면, f는 order N의 correlation immune을 가지게 된다. 그러나, 이것은 암호학 관점에서 아무 소용이 없기 때문에 m=N인 경우는 생각하지 않는다.

예1. $f(x_1, x_2)=x_1\oplus x_2$이면, f는 order 1의 correlation immune 함수이다. 왜냐하면, X_1과 $X_1\oplus X_2$는 독립이고, X_2와 $X_1\oplus X_2$도 독립이기 때문이다.

예 2. $f(x_1, x_2, x_3)=x_3(x_1\oplus x_2)$일때, f는 correlation immune 함수가 아니다. 왜냐하면, X_3와 $X_3(X_1\oplus X_2)$는 독립이 아니다.

Correlation immune 함수를 만드는 방법에 관한 논문으로는 Siegenthaler[11], Xian[14], Xiao-Massey[15] 등이 있다. Correlation immune order가 m이면, 적당한 m+1개의 입력과 f의 출력수열은 독립이 아니므로 correlation 공격이 가능하다. 따라서, immune order가 클수록 correlation 공격을 하기가 어렵다. 그러나, immune order가 클수록 f의 비선형성(nonlinear order)은 떨어진다.

정리 3. 2[11,15] k와 m이 각각 f의 비선형성 or-

der와 correlation immune order일때,

(1) $k+m \leq N$

(2) 출력수열 Z가 0과 1을 취할 확률이 같으면, $k+m \leq N-1$

정리 3. 2로 부터 함수 f의 correlation immune order가 크다고 암호학적으로 좋은 함수가 될 수 없음을 알 수 있다.

예 3. $f(x_1, \cdots, x_N)=x_i \cdots x_N$이면, 비선형성 order는 N이나 correlation immune 함수가 아니다.

예 4. $f(x_1, \cdots, x_N)=x_1 \oplus \cdots \oplus x_N$이면, correlation immune order는 N−1이나, f는 선형함수이다.

예 5. $f(x_1, \cdots, x_N)=x_1x_2 \oplus x_3 \oplus \cdots \oplus x_N$이면, 비선형성 order는 2이고, correlation immune order는 N−3이다.

그러나, 비선형 함수 f가 메모리를 가지면 정리 3. 2는 성립하지 않는다(정리 3. 5 참조). 메모리를 가지지 않는 비선형 함수 f는 $GF(2)^N \rightarrow GF(2)$인 함수로 생각할 수 있다. $GF(2)^N$상의 N차원 벡터 x를 $x=(x_0, x_1, \cdots, x_{N-1})$로 쓰자.

함수 f의 Walsh 변환 S_f

$$S_f(w)=\Sigma f(x)(-1)^{x \cdot w}$$

로 정의된다. 여기서, $x \cdot w=x_0w_0 \oplus \cdots \oplus x_{N-1}w_{N-1}$이다.

역 Walsh 변환에 의해 f를 구할 수 있다.

$$f(x)=2^{-N} \sum_{w} S_f(w)(-1)^{x \cdot w}$$

Xiao와 Massey[15]는 order m인 correlation immune 함수를 Walsh 변환으로 해석할 수 있음을 보였다.

정리 3. 3[15] f가 order m인 correlation immune이 될 필요충분 조건은 w의 Hamming weight $1 \leq W(w) \leq m$에 대해, $S_f(w)=0$이다.

예 6. $f(x_1, x_2, x_3)=x_1 \oplus x_2 \oplus x_3 \oplus x_1x_2 \oplus x_1x_3 \oplus x_2x_3$이면, $S_f(1, 0, 0)=S_f(0, 1, 0)=S_f(0, 0, 1)=0$이나, $S_f=(1, 1, 0)=-2$이다. 정리 3. 3에 의해 f의 correlation immune order는 1이다.

이젠, 비선형 결합 함수 f가 메모리를 가지는 경우에 대하여 생각해 보자. s_0을 초기 메모리 값, $X_i^j=(X_{i1}, X_{i2}, \cdots, X_{ij})$, $Z^j=(Z_1, Z_2, \cdots, Z_j)$로 두자. 즉, X_i^j는 BSSi의 처음부터 j번째까지의 출력 수열이고, Z^j는 처음부터 j번째 까지의 함수 f의 출력 수열이다. 그러면 Z_j는 다음과 같이 쓸수 있다.

$$Z_j=f(X_1^j, \cdots, X_N^j, s_0)$$

메모리를 가지는 경우도 메모리를 가지지 않는 경우와 같이 correlation immune 정의를 자연스럽게 확장할 수 있다.

정의 3. 4[9, 10] f의 출력 수열이 어떤 $m(1 \leq m<N)$개의 입력 수열과 독립일때 f는 order m의 correlation immune을 가진다고 정의한다. 즉,

$$I(Z^j ; X_{i1}^j, \cdots, X_{im}^j)=0, \quad 1 \leq i_1 < \cdots < i_m \leq N$$

비선형 결합 함수 f가 메모리를 가질때 중요한 결과는 다음과 같다.

정리 3. 5[9, 10] $Z_j=\sum_{i=1}^{N} X_{ij}+f'(X_1^{j-1}, \cdots, X_N^{j-1}, s_0)$

이면,

(1) Z_j는 독립이고

(2) $P(Z_j=0)=P(Z_j=1)$

(3) f는 최대의 correlation immune order N−1을 가진다.

위의 정리에서 f'에 대한 제약 조건이 없기 때문에 쉽게 적용 가능하며, 비 선형성 order를 최대로 되게 할 수 있다.

예 7. 두 정수 a, b는 이진수로 $a=a_{n-1}2^{n-1}+\cdots+a_1 2+a_0$, $b=b_{n-1}2^{n-1}+\cdots+b_1 2+b_0$로 표현되며, 두 수의 합(실수합) $z=a+b$는 다음과 같이 구할 수 있다.

$$z_0=a_0 \oplus b_0$$
$$z_1=a_1 \oplus b_1 \oplus a_0b_0$$
$$z_2=a_2 \oplus b_2 \oplus a_1b_1 \oplus a_1a_0b_0 \oplus b_1a_0b_0$$
$$\vdots$$

즉, $z_j=a_j \oplus b_j \oplus c_{j-1}$

$$c_j=a_jb_j \oplus (a_j \oplus b_j)c_{j-1}$$

따라서, 두 정수의 실수합은 정리 3. 5의 조건을

만족한다. 따라서 두 랜덤한 정수를 더하는 알고리즘은 correlation immune order가 최대인(N−1)이 되며, 또한 선형 복잡도도 주기와 거의 같음을 알 수 있다[9, 10]. 그래서 두 정수의 실수합은 아주 좋은 RKG이다.

4. 빠른 Correlation 공격

Siegenthaler가 제안한 correlation 공격은 RKG의 출력 수열과 LFSR의 출력 수열 사이에 상관관계가 있으면, 각 LFSR의 키를 독립적으로 찾을 수 있으며, 각 LFSR의 키를 찾는 방법은 모든 가능한 시행(exhaustive trial)을 하는 것이다. 따라서, Siegenthaler 방법으로는 LFSR의 크기가 50 이상이면 공격 불가능하다. Meier와 Staffelbach는 LFSR 수열의 linear recurrence 관계를 이용하여, RKG 수열에 대한 일차 방정식을 얻어 LFSR의 초기값을 빨리 찾아내는 두 개의 알고리즘을 제안했다. 두 알고리즘은 각각 correlation 확률이 0.75 근방에 있을때, 0.5 근방에 있을때 유용한 알고리즘이며, LFSR의 탭(tap)의 수가 작을때(10이하일때) 가능하며, 이때 알고리즘의 복잡도는 O(r)이다(r는 LFSR의 크기). LFSR의 크기가 1,000인 경우도 공격가능하다. 그러나, 탭의 수가 크면 알고리즘 복잡도는 r에 대한 지수(exponential) 함수가 되어, Meier-Staffelbach 알고리즘은 탭의 수가 작을때만 유용하다. Chepyzhov와 Smeets[2]는 lowweight checks를 찾는 빠른 알고리즘을 이용하여, LFSR의 초기값을 찾는 알고리즘을 개발했다. 탭의 수가 LFSR의 크기에 따라 선형으로 증가하면, 알고리즘의 복잡도는 모든 가능한 시행을 하여 찾는 방법보다 훨씬 좋음을 보였다.

이제까지 언급한 RKG와 다른 RKG에 대한 공격 알고리즘을 살펴보자. 최대 주기를 생성하는 하나의 LFSR과 비선형 feedforward state filter 함수로 이루어진 RKG(그림 5)에 대한 공격 알고리즘이다. 주어진 RKG의 LFSR(초기값은 다를 수 있음)과 같은 여러개의 LFSR을 degenerate 함수로 결합하여, 나온 수열이 주어진 RKG의 출력 수열과 같게

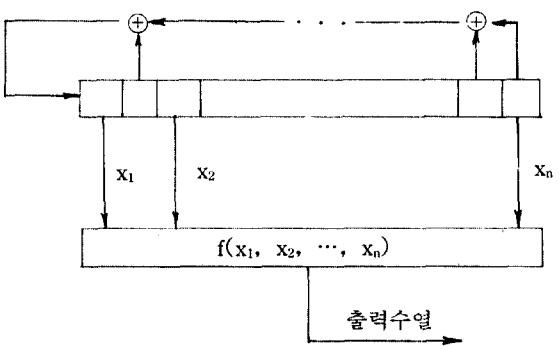

그림 5. 하나의 LFSR로 구성된 RKG

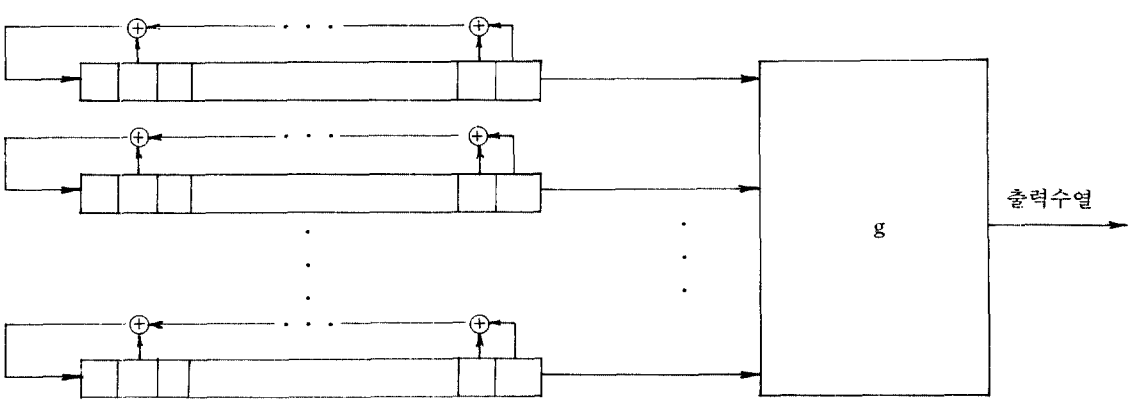

그림 6. 동치 시스템

동치시스템을 구성한다(그림 6). Siegenthaler[13]는 LFSR의 출력 수열과 RKG의 출력 수열 사이의 CCF (cross correlatin function)의 피이크(peak) 값을 이용하여 동치 시스템의 초기값을 구했다. 그러나, 이 공격은 LFSR의 크기가 크면 CCF의 피이크 값을 구하는데, 계산량이 지수 함수로 증가하기 때문에 작은 크기의 LFSR에만 적용가능하다. Forre[3]는 Meier와 Staffelbach가 correlation 공격에 적용하기 위해 개발한 알고리즘을 변형하여 크기가 큰 LFSR에 적용 가능한 알고리즘을 개발했다.

5. 결 론

여러개의 LFSR을 비선형 함수로 구성한 RKG (running key generator)를 설계할때, 앞에서 살펴본 여러 correlation 공격을 피할 수 있게 해야 한다. LFSR의 크기를 크게, 탭의 수를 많이, 비선형 함수가 상관관계를 가지지 않게 구성하면 좋다. 또한, 메모리를 가지게 비선형 함수를 구성하면 메모리를 가지지 않을 때 보다 correlation 공격을 피할 수 있다.

그러나, RKG의 설계자는 여러 공격 알고리즘에 피할 수도 있지만 경제적인 관점에서 설계해야 하기 때문에 두 상반된 개념을 절충하여 설계하여야 한다.

참 고 문 헌

1. W. Blaser, P. Heinzmann, "New cryptographic device with high security using public key distribution," Proceedings of IEEE student paper contests, pp. 145-153, 1982.

2. V. Chepyzhov, B. Smeets, "On a fast correlation attack on certain stream ciphers," Abstracts of EUROCRYPT'91.

3. R. Forre, "A fast correlation attack on nonlinearly feedforward filtered shift register sequences," Abstracts of EUROCRYPT'89.

4. S. W. Golomb, "Shift Register Sequences," San Francisco, LA, Holden Day, 1967.

5. S. Lloyd, "Properties of binary functions," Abstracts of EUROCRYPT'90.

6. W. Meier, O. Staffelbach, "Fast correlation attacks on certain steam ciphers," Journal of Cryptology, Vol 1, No. 3, pp. 159-176, 1989.

7. W. Meier, O. Staffelbach, "Nonlinearity criteria for cryptographic functions," Proceedings of EUROCRYPT'89.

8. W. Meier, O. Staffelbach, Correlation properties of combiner with memory in stream ciphers," Journal of Cryptology, to appear.

9. R. A. Rueppel, "Correlation immunity and the summation generator," Proceeding of CRYPTO'85.

10. R. A. Rueppel, "Aralysis and Design of Stream Ciphers," Springer-Verlag, 1986.

11. T. Siegenthaler, "Correlation immunity of nonlinear combining functions for cryptographic applications," IEEE Trans. Inform. Theory, Vol 30, pp. 776-780, 1984.

12. T. Siegenthaler, "Decrypting a class of stream ciphers using ciphertext only" Vol. 34, No. 1, pp. 81-85, 1985.

13. T. Siegenthaler, "Cryptanalysts representation of nonlinearly filtered ML-sequences," Proceedings of EUROCRYPT'85

14. Y. Y. Xian, "On the correlation-immunity of boolean functions," Proceedings of IEEE Singapore ICCS'88.

15. G. Xiao, J. L. Massey, "A spectral characterization of correlation immune combining functions," IEEE Trans. Information Theory, Vol. 34, No. 3, pp. 569-571, 1988.

16. 한국전자통신연구소, "현대암호학", 1991.

□ 著者紹介

成 洙 學(正會員)

1982年 慶北大學校 數學科(學士)
1985年 KAIST 應用數學科(碩士)
1988年 KAIST 應用數學科(博士)
1988年～1991年 韓國電子通信研究所 先任研究員
1991年～現在 培材大學校 應用數學科 助教授