

주성분분석(PCA)을 이용한 출입인원관리에 대한 보안성 확보 방안

김민수* · 이동휘**

요 약

본 연구는 주성분 분석을 통하여 출입인원에 대한 보안성을 확보방안을 제시하기 위함이다. 데이터를 수집하기 위해 K센터(IPS) 보안등급A~E 출입구역 출입데이터를 바탕으로 BoxPlot와 주성분분석으로 통해 연구결과를 도출하였다. 주성분 분석을 수행하기 전에 공통성의 추출값에 대하여 다중공선성을 측정 한 값인 분산팽창인수(VIF)가 2.902 이하이므로 주성분분석을 해석하는데 문제가 없음을 확인하였다. 이를 바탕으로 주성분 분석을 실시하여 제 1 주성분의 고유값 1.453, 제 2주성분의 고유값 1.283, 제 3 주성분의 고유값 1.142을 바탕으로 보안등급별 인원을 나누어 본 결과를 바탕으로 보안등급별 인원을 Green-list, Blue-list, Red-list, Black-list로 구분하였다.

A Way of Securing the Access By Using PCA

Min Su Kim* · DongHwi Lee**

ABSTRACT

This study aimed at making a way of securing the access by using PCA. We got our result through using Box-Plot and PCA with the access data of the area of security level A~E at K(IPS)center. In order to perform PCA, We confirmed the extracted value of commonality has no problem in performing PCA because VIF is below 2.902. Based on this result, We classified people into Green-list, Blue-list, Red-list, and Black-list in a standard of security level with 1.453, as the eigen value of 1 main element, 1.283, as eigen value of 2 main elementm, 1.142, as the eigen value of 3 main element.

Key words : PCA, Facility Security, Controlling the Access, Machine Security, Security

접수일(2012년 6월 7일), 수정일(1차: 2012년 6월 10일),
게재확정일(2012년 6월 11일)

* 경기대학교 산업보안학과

** University of Colorado Denver, Dept. of Computer
Science and Engineering (교신저자)

1. 서 론

출입관리업무는 경비대상시설에 출입하는 사람, 물건, 차량 등을 관리함으로써 시설내에서의 범죄, 사고 등의 발생을 방지하는 업무이다[1].

즉, 출입을 하는 모든 인원에 대한 관리를 하여 시설내의 각종 정보의 반출이나 위험물 등의 반입으로 인한 범죄나 사고 등의 발생을 방지하는 것이다.

출입통제를 통하여 보안 사고를 방지하기 위해 기업은 관리적, 기술적 측면에서 보안업무를 진행하고 있다.

이를 위한 방법으로 기업은 관리적으로 구역별 보안등급을 나누고 스마트카드, 생체인식 등의 출입통제 기술을 이용하여 출입 승인여부를 적용하는 보안업무를 진행하여 출입을 허가받은 인원에 대한 관리로써 출입자의 시설내에서의 출입 정보를 저장하여 출입관리를 하게 된다[2].

본 논문은 구역별 보안등급이 A~E 등급으로 나누어진 경기도의 K센터에 스마트카드를 이용하여 출입하는 25명의 인원에 대하여 65일간의 출입데이터를 수집하였다.

출입자의 형태를 보면 정상적으로 출입하는 인원, 권한 이상으로 출입을 시도하는 인원, 평소의 출입경로와는 다른 출입형태를 띠는 인원, 실수에 의해 출입을 시도한 인원 등으로 구분할 수 있다.

수집한 출입자의 이동경로 데이터를 주성분 분석을 이용하여 나타난 결과를 바탕으로 이상징후를 보이는 인원에 대한 보안위협 등급을 설정하여 효율적·합리적으로 관리할 수 있는 보안성 확보에 대한 방안을 제시하고자 한다.

2. 관련연구

본 장에서는 출입관리를 통해 보안성 확보를 위한 경비시스템과 출입통제에 사용되는 스마트카드 그리고 보안성 확보 방안을 제시하기 위한 분석방법에 대한 연구들을 살펴보고자 한다.

2.1 출입통제 시스템

2.1.1 기계경비 시스템

기계경비시스템은 외부에서 침입할 수 있는 곳에 각종 감지기를 설치하고 경계하는 시스템으로 로컬 기계경비시스템은 소수의 경비원으로 적합한 감지기 와 감지된 장소의 영상을 볼 수 있는 폐쇄회로(CCTV)를 이용하여 감시 관리하는 시스템이다[3].

기계경비시스템의 분류를 보면 <표 1>과 같다.

<표 1> 기계경비시스템의 분류[3]

구 분	내 용
침입감지 시스템	무인경비의 개념으로 방법 및 방재분야까지 적용범위가 넓어지고 있다.
화상감시 시스템	카메라를 이용하는 방법으로 외부침입자를 감시하거나 건물 및 중요자료의 외부유출을 감시하는 시스템
출입통제 시스템	주요 시설물에 사람과 차량·물품의 반입 반출을 확인하고 인가되지 않은 사람을 통제하는 시스템
외곽침입 시스템	시설물의 외곽부터 침입을 인지하기 위해 설치·감시하는 시스템
생체인식 시스템	출입자의 신체상의 특징으로 신분을 확인하는 시스템
전자보안 검색시스템	주로 국가중요시설고 공항·항만 이용객의 안전을 확보하기 위해 활용되는 시스템

2.1.2 시설경비

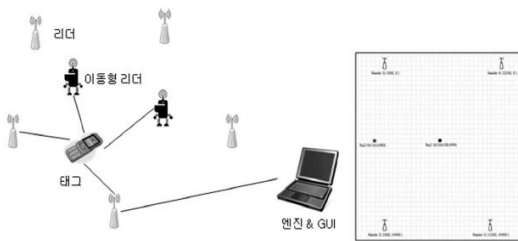
경비대상시설에서의 도난·화재 그밖에 혼잡 등으로 인한 위험발생을 방지하는 업무로 국가중요시설, 빌딩 및 사무소, 금융기관, 백화점 및 상업시설, 호텔 및 숙박시설 등에 대하여 상주경비, 순찰경비, 보안경비의 유형으로 시설보호를 하게 된다.

상주경비는 경비대상시설 내에 필요한 경비원을 24시간 상주시켜 출입관리업무, 순찰업무, 감시업무를 통해 도난, 화재 등의 사고발생을 경계하고 방지하는 경비방법이다. 순찰경비는 경비대상시설에 상주하면서 순찰활동을 하는 경비형태로 상주경비를 구성하는 일부 활동이다. 보안경비는 백화점, 대형 할인마트 등에서 절도를 방지하거나, 경계를 하는 경비방법이다 [4].

2.2 RTLS(Real Time Locating Systems)

2.2.1 RTLS

RTLS는 (그림 1)과 같이 대상의 위치를 실시간으로 파악하는 기술이다. 유사한 기술인 GPS(Global Positioning System)가 있으나 음영 지역에서는 사용할 수 없는 문제가 있어 Wi-Fi, Zigbee, Bluetooth, RFID와 같은 근거리 통신 기술을 이용한 RTLS가 적합하다[5].



(그림 1) RTLS 개념도[5]

2.2.2 RFID 시스템

RFID(Radio Frequency Identification)는 무선을 통해 식품, 동물, 사물 등 다양한 개체의 정보를 관리할 수 있는 차세대 인식 기술이다. 이러한 RFID 기술은 크게 사물의 정보를 포함하고 있는 태그와 그 태그의 정보를 읽어낼 수 있는 판독기로 구성되어 있다. 여기서 태그는 전원으로서는 배터리의 사용여부에 따라 배터리가 요구되는 능동형과 배터리가 요구되지 않는 수동형으로 분류된다[6].

이러한 RFID 태그를 이용하여 사람 및 차량의 출입을 구역별로 통제할 수 있어 출입통제효과를 높일 수 있고 각 구역은 그 직책이나 업무에 해당되는 인원만이 출입할 수 있는 등 차별화된 관리를 할 수 있다[7].

2.3 주성분분석(Principal Component Analysis)

주성분분석은 해석하고자 하는 다차원의 데이터를 거기에 포함된 정보의 손실을 가능한 한 적게 해서 2 혹은 3차원의 데이터로 축약하는 수법이다. 주성분분석을 활용하면 관측대상이 어떠한 위치에 있는지 시각적으로 파악할 수 있게 된다[8].

즉, 복잡한 데이터를 단순화하기 위한 다변량 통계

기법이다[9].

2.3.1 주성분분석의 계산

변수의 수가 p 개(x_1, x_2, \dots, x_p), 관측대상의 수가 n 개인 다변량 데이터를 기초로 p 개보다 적은 m 개의 새로운 변수 z_1, z_2, \dots, z_m 을 생성한다.

$$z_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p$$

$$z_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p$$

.....

$$z_m = a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mp}x_p$$

- ① z_1 은 x_1 에서 x_p 까지의 정보가 최대한 집약되도록 한다.
- ② z_2 는 x_1, x_2, \dots, x_p 의 정보가 z_1 의 다음에 가능한 한 많이 집약되도록 한다. 또한 z_1 과는 독립이 되도록 한다.
- ③ z_3 는 x_1, x_2, \dots, x_p 의 정보가 z_1 과 z_2 의 다음에 가능한 한 많이 집약되도록 한다. 또한 z_1 및 z_2 와는 독립이 되도록 한다.
- ④ 이하 z_4 에서 z_m 까지 마찬가지이다.

이와 같은 성질을 충족시키도록 $a_{11}, a_{12}, \dots, a_{mp}$ 를 산출하고자 하는 것이 주성분 분석이 계산이다[8].

3. 제안하는 방법

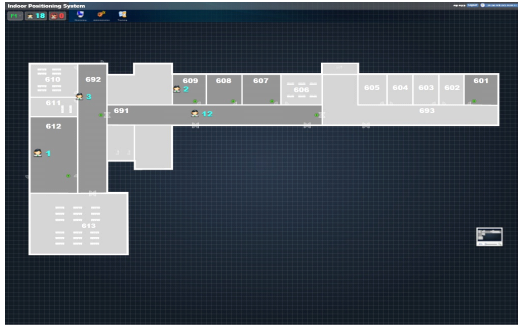
보안 구역에 대해 보안등급별로 출입을 허가받은 인원을 구분하여 연구기간 동안의 데이터를 이용하여 Box-Plot 및 주성분분석을 실시하였다.

3.1 연구대상 및 분석 모형

이 연구는 2012년 1월 경기도 소재의 K센터를 출입하는 내·외부 인원 중 보안 등급별 인원을 무작위로 5명씩 선정하여 총 25명을 대상으로 65일간 출입 경로를 살펴보았다.

3.1.1 IPS(Indoor Position System)

K센터의 종합관제실에서 IPS의 구성화면을 살펴보면 (그림 2)와 같다.



(그림 2) K센터의 IPS

각 호실의 보안 등급을 살펴보면 <표 2>와 같다.

<표 2> 호실별 보안 등급

보안구역 등급	호 실
Level A 구역	612 - 613호
Level B 구역	610 - 611호
Level C 구역	606 - 609호
Level D 구역	602 - 605호
Level E 구역	601호

3.1.2 보안구역 등급

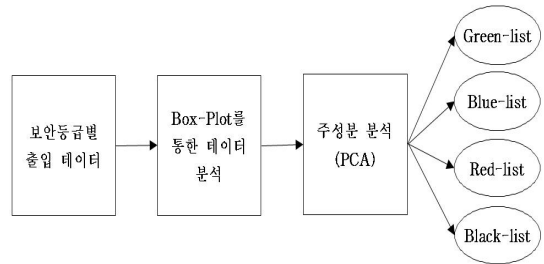
1일의 출입경로 데이터는 본인 등급을 기준으로 최상위, 혹은 최하위 등급으로 이동한 보안등급 데이터를 활용하였다. 보안 등급별 출입 가능 지역을 <표 3>와 같이 정리하였다.

<표 3> 보안등급

등급	경로 제한
Level A	모든 보안 구역(Label A ~ E) 출입 가능
Level B	Level A 구역을 제외한 보안 구역 출입 가능
Level C	Level A, B 구역을 제외한 보안 구역 출입 가능
Level D	Level A, B, C 구역을 제외한 보안 구역 출입 가능
Level E	Level E 구역만 출입 가능

3.1.3 연구 분석 모형

K센터 IPS의 출입데이터를 바탕으로 본 연구를 수행할 연구 분석 모형을 보면 (그림 3)과 같다.



(그림 3) 연구 분석 모형

위의 보안 등급별 데이터를 바탕으로 입력된 자료 처리는 SPSSWIN 12.0을 활용하여, BoxPlot와 주성분 분석으로 통해 연구결과를 도출하였다.

3.2 보안등급별 이동 상황

3.2.1 보안등급별 이동 데이터 탐색

65일간 보안등급별 인원의 이동경로를 종합하여 <표 4>과 같이 정리하였다.

보안등급별 구역에 대하여 총 출입시도는 25×65 = 1625번으로 'Label A = 217', 'Label B = 312', 'Label C = 326', 'Label D = 373', 'Label E = 381'의 출입 시도가 있었다.

<표 4> 보안등급별 이동 데이터

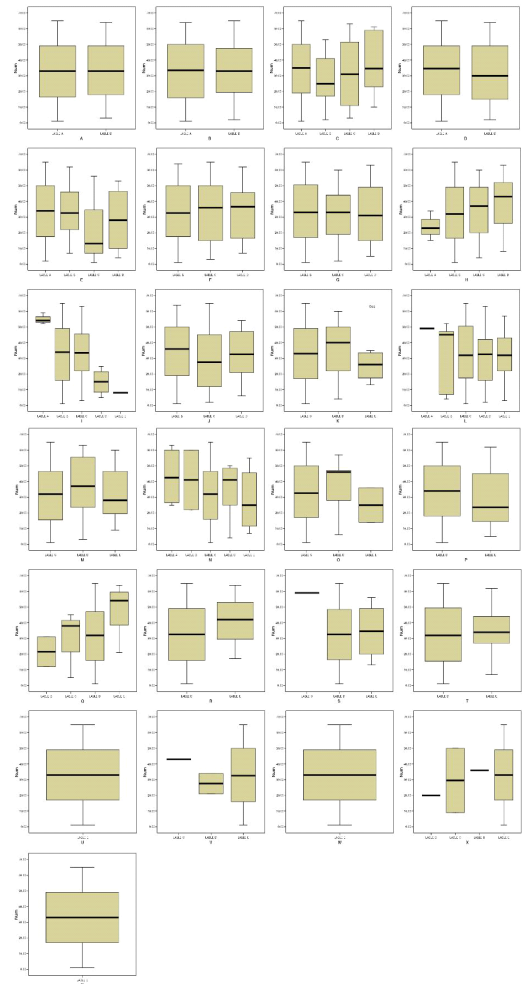
	Level A		N	Level A	
	44	-		-	-
A	Level B	21	O	Level B	4
	Level C	-		Level C	44
	Level D	-		Level D	9
	Level E	-		Level E	7
	Level A	42		Level A	-
B	Level B	23	P	Level B	-
	Level C	-		Level C	58
	Level D	-		Level D	5
	Level E	-		Level E	2
	Level A	39		Level A	-
C	Level B	13	P	Level B	-

	Level C	7		Level C	-
	Level D	6		Level D	53
	Level E	-		Level E	12
D	Level A	44	Q	Level A	-
	Level B	21		Level B	-
	Level C	-		Level C	5
	Level D	-		Level D	53
	Level E	-		Level E	7
E	Level A	45	R	Level A	-
	Level B	20		Level B	-
	Level C	-		Level C	-
	Level D	-		Level D	62
	Level E	-		Level E	3
F	Level A	-	S	Level A	-
	Level B	40		Level B	-
	Level C	17		Level C	4
	Level D	8		Level D	57
	Level E	-		Level E	4
G	Level A	-	T	Level A	-
	Level B	39		Level B	-
	Level C	16		Level C	-
	Level D	10		Level D	56
	Level E	-		Level E	9
H	Level A	3	U	Level A	-
	Level B	47		Level B	-
	Level C	10		Level C	-
	Level D	5		Level D	-
	Level E	-		Level E	65
I	Level A	-	V	Level A	-
	Level B	42		Level B	-
	Level C	18		Level C	1
	Level D	4		Level D	2
	Level E	1		Level E	62
J	Level A	-	W	Level A	-
	Level B	41		Level B	-
	Level C	14		Level C	-
	Level D	10		Level D	-
	Level E	-		Level E	65
K	Level A	-	X	Level A	-
	Level B	-		Level B	1
	Level C	45		Level C	5
	Level D	13		Level D	1
	Level E	7		Level E	58
L	Level A	-	Y	Level A	-
	Level B	-		Level B	-
	Level C	43		Level C	-
	Level D	12		Level D	-
	Level E	10		Level E	65

M	Level A	-	평균	Level A	217
	Level B	-		Level B	312
	Level C	39		Level C	326
	Level D	15		Level D	373
	Level E	11		Level E	381

3.2.2 보안등급별 이동 데이터 Box-Plot

<표 4>의 보안등급별 이동 데이터를 Box-Plot로 나타내면 (그림 4)와 같다.



(그림 4) 데이터 Box-Plot

4. 주성분 결과 분석

주성분분석을 수행하여 얻은 추출값은 각 변수들이 나머지 다른 변수들에 얼마나 많이 영향을 받는가를 측정할 수 있는 공통성을 살펴보면 <표 5>와 같다.

<표 5> 공통성

인 원	초 기	추 출
A	1.000	.750
B	1.000	.745
C	1.000	.699
D	1.000	.802
E	1.000	.797
F	1.000	.604
G	1.000	.739
H	1.000	.738
I	1.000	.669
J	1.000	.586
K	1.000	.681
L	1.000	.754
M	1.000	.749
N	1.000	.699
O	1.000	.549
P	1.000	.644
Q	1.000	.813
R	1.000	.660
S	1.000	.699
T	1.000	.760
V	1.000	.694
X	1.000	.632

공통성의 추출값이 1에 가까울수록 다중공선성이 존재한다고 볼 수 있기 때문에 다중공선성을 측정하여야 한다.

다중공선성을 고려하지 않고 그 결과를 해석하게 되면 잘못된 결론을 내리게 되는 문제가 발생하게 된다. 여기서 다중공선성이란 입력변수들 간의 상관정도가 높은 상태를 말한다.

다중공선성을 살펴보면 <표 6>과 같다.

<표 6> 다중공선성 통계량

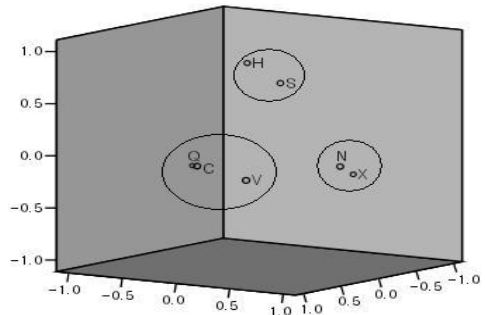
인 원	공선성 통계량	
	공차한계	VIF

A	.345	2.902
B	.441	2.269
C	.618	1.619
D	.345	2.898
E	.358	2.791
F	.590	1.694
G	.765	1.308
H	.544	1.838
I	.665	1.505
J	.766	1.305
K	.488	2.051
L	.455	2.200
M	.505	1.982
N	.530	.1887
O	.666	1.502
P	.605	1.653
Q	.612	1.633
R	.736	1.359
S	.733	1.363
T	.763	1.311
V	.675	1.481
X	.721	1.387

추출방법 : 주성분 분석

VIF(Variance Inflation Factor) 즉, 분산팽창인수가 10이하이면 다중공정성이 약하다고 판단할 수 있는데 <표 6>에서 보는바와 같이 VIF가 2.902 이하이므로 주성분분석을 해석하는데 문제가 없음을 볼 수 있다.

보안등급별 Box-Plot 데이터를 바탕으로 주성분 분석 결과를 성분도표로 나타내면 (그림 5)와 같다.



(그림 5) 주성분 분석 성분도표

주성분의 산출을 하기 위해 성분행렬을 살펴보면,

<표 7>과 같다.

<표 7> 성분행렬

	성분		
	1	2	3
C	-.306	-.485	.411
H	-.389	.665	.319
N	.669	.009	.236
Q	-.499	-.139	-.544
S	-.061	.545	.408
V	.111	-.507	.513
X	.704	.177	-.300

제 1주성분은 $(-.306)^2 + (-.389)^2 + (.669)^2 + (-.499)^2 + (-.061)^2 + (.111)^2 + (.704)^2 = 1.453$ 의 값을, 제 2주성분은 $(-.485)^2 + (.665)^2 + (.009)^2 + (-.139)^2 + (.545)^2 + (-.507)^2 + (.177)^2 = 1.283$ 의 값을, 제 3주성분은 $(.411)^2 + (.319)^2 + (.236)^2 + (-.544)^2 + (.408)^2 + (.513)^2 + (-.300)^2 = 1.142$ 의 고유값을 가지고 있고 <표 8>과 같다.

<표 8> 고유값

성분	초기 고유값			회전 제곱합 적재값		
	전체	%분산	%누적	전체	%분산	%누적
1	1.453	20.758	20.758	1.380	19.715	19.715
2	1.283	18.325	39.083	1.262	18.032	37.747
3	1.142	16.318	55.401	1.236	17.654	55.401

이러한 주성분 분석을 토대로 보안등급별 정의를 하면 <표 9>와 같이 총 4단계의 보안등급으로 정의하였다.

<표 9> 보안 등급 정의

보안 등급	정 의
Green-list	보안위협률이 없는 인원
Blue-list	보안위협률이 미미한 인원
Red-list	보안위협률이 높은 인원
Black-list	보안위협률이 심각한 인원

<표 9>의 보안 등급 정의를 바탕으로 주성분분석을 수행하여 나누어진 보안 등급 인원을 살펴보면

<표 10>과 같다.

<표 10> 보안 등급 인원

보안 등급	인 원
Green-list	나머지 인원
Blue-list	C, V, Q
Red-list	H, S
Black-list	N, X

5. 결 론

본 연구는 주성분 분석을 통하여 출입인원에 대한 보안성을 확보방안을 제시하기 위함이다.

이를 위해 K센터를 출입하는 내·외부 인원 중 보안 등급별 인원내에 대해 총 25명을 대상으로 65일간의 출입경로데이터를 이용하였다.

데이터를 수집하기 위해 K센터의 IPS의 보안등급 A~E 출입구역 출입데이터를 바탕으로 BoxPlot과 주성분분석으로 통해 연구결과를 도출하였다.

하지만 이를 위해 우선되어야 하는 것이 공통성의 추출값이 1에 가까울수록 다중공선성이 존재한다고 볼 수 있기 때문에 다중공선성을 측정하여야 한다.

분산팽창인수(VIF)가 10이하이면 다중공정성이 약하다고 판단할 수 있는데 VIF가 2.902 이하이므로 주성분분석을 해석하는데 문제가 없음을 볼 수 있다.

주성분 분석결과를 살펴보면 제 1 주성분의 고유값 1.453, 제 2주성분의 고유값 1.283, 제 3 주성분의 고유값 1.142을 바탕으로 보안등급별 인원을 나누었다.

보안등급별 인원은 Blue-list C·V·Q, Red-list H·S, Black-list N·X로 나머지 인원은 Green-list로 나타났다.

참고문헌

- [1] 안황권, “시설경비론”, 진영사, p.103, 2009.
- [2] 정태갑 외, “RFID를 이용한 출입관리시스템 설계 및 구현”, 한국멀티미디어학회, Vol.2005, No.1, pp. 741-742, 2005.

- [3] 김찬선, “기계경비콘텐츠 연구동향”, 한국콘텐츠학회, 제8권, 제3호, pp.17-19, 2010.
- [4] 박영만, “시설경비콘텐츠 연구동향”, 한국콘텐츠학회, 제8권, 제3호, pp.21-23, 2010.
- [5] 김잠제 외, “위치 측정 정확도 향상을 위한 RTLS의 이동형 리더 선택”, 한국정보과학지, Vol.16, No.1, p.45, 2010.
- [6] 송태승 외, “능동형RFID 태그에서 자체 시험 모드를 순차적으로 적용한 적합성 평가방법”, 대한전자공학회, Vol.45, No.6, p.807, 2008.
- [7] 정태황, “RFID의 보안업무 적용환경과 적용방안에 관한 연구”, 한국경호경비학회, 제21호, p.167, 2009.
- [8] 노형진, ‘SPSS에 의한 다변량 데이터의 통계분석’, 도처출판 효산, 2007.
- [9] Raychaudhuri. S, Stuart J. M and Altman R. B, “Principal Component Analysis to Summarize Microarray Experiments : Application to Sporulation Time Series”, Pacific Symposium on Biocomputing 5, pp.452-463, 2000.

[저 자 소 개]



김 민 수 (Min-Su Kim)

2004년 2월 컴퓨터공학사
2012년 2월 경호안전학석사
2012년 현재 경기대학교
산업보안학과 박사과정

email : fortcom@hanmail.net



이 동 휘 (DongHwi Lee)

2000년 경기대학교 컴퓨터과학과
(이학사)
2003년 경기대학교
정보보호기술공학과
(공학석사)
2006년 경기대학교 정보보호학과
(정보보호학박사)

2011년~현재 University of Colorado
Denver, Dept. of Computer Science
and Engineering

email : dhclub@naver.com