Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.25, NO.5, Oct. 2015 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2015.25.5.1001

스마트 기기 환경에서 전력 신호 분석을 통한 프라이버시 침해 위협*

조 재 연,† 윤 지 원‡
고려대학교 정보보호대학원

Threatening privacy by identifying appliances and the pattern of the usage from electric signal data*

Jae yeon Cho,† Ji Won Yoon‡
School of Information Security, Korea University

요 약

스마트 그리드 안에서 고안된 스마트 미터는 우리가 사용하는 전력 신호를 실시간으로 데이터화해서 전력 공급단의 메인 서버로 전송한다. 이를 통해 전력 관리의 효율성은 증가한 반면, 사용자의 정보를 담은 데이터의 보안 문제가 새로운 위협으로 부상하였다. 본 논문은 스마트 미터에서 추출한 전력 데이터를 통해 가정 내 기기의 식별 및 기기별 사용패턴에 대한 추론을 보안 관점에서 해석함으로써 스마트 기기 환경에서 데이터 노출의 위협을 지적한다. 주성분분석(Principal Component Analysis)으로 데이터의 특징을 추출하였고 k-근접 이웃(k- Nearest Neighbor)분류기로 기기를 식별하고 기기상태를 추론하였으며, 검증방법으로는 10차 교차검증(10-fold Cross Validation)을 활용하였다.

ABSTRACT

In Smart Grid, smart meter sends our electric signal data to the main server of power supply in real-time. However, the more efficient the management of power loads become, the more likely the user's pattern of usage leaks. This paper points out the threat of privacy and the need of security measures in smart device environment by showing that it's possible to identify the appliances and the specific usage patterns of users from the smart meter's data. Learning algorithm PCA is used to reduce the dimension of the feature space and k-NN Classifier to infer appliances and states of them. Accuracy is validated with 10-fold Cross Validation.

Keywords: Smart meter, Privacy, PCA, k-NN Classifier, 10-fold Cross Validation

I. 서 론

스마트 홈, 스마트 시티와 같은 용어들의 등장과 함께 전 산업계가 '스마트'에 주목하고 있다. 전력, 교통, 수도, 에너지, 커뮤니케이션 등 다양한 산업분야에 스마트 기술을 융합함으로써 우리 삶을 좀 더 윤택하고 효율적으로 만들기 위한 노력들이 활발히 진행 중이다. 대표적인 스마트 기술인 스마트 미터링

접수일(2015년 4월 29일), 수정일(1차: 2015년 8월 17일, 2차: 2015년 9월 23일), 게재확정일(2015년 10월 2일)
* This research was supported by the MSIP(Ministry of Science, ICT & Future Planning), Korea, under the "The Types of employment contract to support master's degree in Information Security" supervised by the KISA(Korea Internet Security Agency).
† 주저자, sjjy811@korea.ac.kr
‡ 교신저자, jiwon_yoon@korea.ac.kr(Corresponding author)

시스템이 가구마다 설치되고, 스마트 플러그 및 미터기를 탑재한 스마트 제품들이 본격적으로 출시되면서 단순히 먼 미래로 보였던 스마트 홈, 스마트 시티로의 전환이 눈앞의 현실로 다가온 것이다.

ICT(Information Communication Technology)를 전체 전력망 시스템에 접목시킨 스마트 그리드 시스템에서는 집 외부에 위치한 스마트 미터를 통해 가정의 전력 사용량이 실시간으로 전력 공급 단의 메인 서버로 전송된다. 즉, 스마트미터는 사용자와 내·외부 통신망의 접점으로서, 전력 공급자와 수요자 간의 상호 피드백을 통해 꼭 필요한 만큼의 전력량을 공급하기 위한 스마트 그리드 시스템의 핵심이라 할 수 있다. 그러나 스마트 미터는 바이러스 및 웜, 악성코드 등의 목표가 되기 쉽다. 관리가 어려운 외부에 설치되고 각종 위협에 대처할 만큼의 보안 장치가 마련되어 있지 않기 때문이다. 실제로 보안 장치 없이 평문으로 전송되는 전력데이터를 분석하여 가정 내의 가전기기의 수를 파악한 프라이버시 해킹 연구가 발표된 바 있다[8].

본 논문에서는 공격자의 입장에서 스마트 미터에서 AMI 관리 서버로 전송되는 시간-전력 데이터를 추출하여, 머신러닝 기법을 통해 가정 내부의 기기 상태 및 사용자의 가전기기 사용패턴을 식별할 수 있음을 보인다. 이는 외부의 데이터를 통해 내부의 정보를 알아내는 공격으로, 공격자로 하여금 사용자의 행동패턴을 추론하는 근거를 제공하기 때문에 스마트 기기 환경의 취약한 보안 허점을 드러낸다. 전력 사용 습관 및 행위 패턴, 빈집 여부 등이 그대로 노출되어 외부 공격자가 내부를 침입하기 위한 근거를 제공할 뿐 아니라, 데이터 위변조 공격도 가능하기 때문에 스마트 미터를 통한 데이터 노출은 광대역의 전력수급에 큰 혼란을 야기할 수 있다.

논문의 구성은 다음과 같다. Ⅱ장에서는 전력 신호를 이용한 가전기기 부하 식별 연구의 큰 맥락을 살펴보고, Ⅲ장에서는 본 논문에서 수행하는 비접촉식 가전기기 부하 식별(Non Intrusive Appliance Load Monitoring)의 관련연구와 본 논문의 의의를 정리한다. Ⅳ장에서는 특성 추출에서 분류 및 검증까지의 분류과정을 설명하고, Ⅴ장에서는 실제 실험 과정에 적용하여 Ⅵ장에서 결과를 정리한다. 마지막으로 Ⅶ장에서는 결론과 보안대책에 대해 논의한다.

Ⅱ. 가전기기 부하 식별

기존의 가전기기 부하 식별에 대한 연구는 데이터의 추출방법에 따라 크게 비접촉식(Non Intrusive Appliance Load Monitoring)과 접촉식(Intrusive Appliance Load Monitoring)으로 나뉜다. 접촉식은 가전기기에 직접적으로 연결하여 전력 사용량 데이터를 추출하고, 비접촉식은 집 밖에 위치한 스마트 미터를 통해서 전체 전력량 데이터를 추출한다. 즉, 비접촉식은 AMI(Advanced Metering Infrastructure)와 같은 하나의 측정 포인트를 가지고, 접촉식은 서브미터기나 스마트 플러그, 기기 안에 임베디드된 미터기와 같이, 데이터가 다양한 측정 포인트로부터 추출된다[1].(Fig.1. 참고) 추출한 데이터를 분석하여 각 기기의 상태를 추론하고 이를 효율적으로 관리하는 것이 바로 개별 가전기기 부하 식별의 목표이다.

외부에 위치한 스마트 미터로부터 데이터를 탈취할 수 있는 공격자의 입장에서는 외부에서 추출한 전체 전력량 데이터를 통해 기기의 식별 및 기기의 상태를 파악해야 하므로, 하나의 측정 포인트를 갖는 비접촉식 부하 식별 방법을 활용한다.

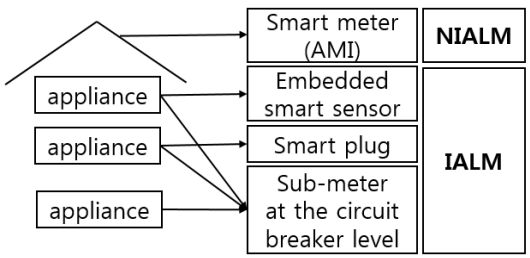

Fig. 1. Distinction between IALM and NIALM

Ⅲ. 관련연구

비접촉식 가전기기 부하 식별은 개별기기의 전력 사용을 전체 전력량 데이터의 변화로 감지한다. 각각의 기기들의 on/off 상태에 따른 전체 전력량 데이터의 특성을 찾아 기기의 상태를 식별하는 것이다[2][3]. 비접촉식 가전기기 부하 식별에 관한 연구는 1990년대 Steady-state based, 즉 기기가 켜져 있거나 꺼져 있는 동안의 특성을 이용하는 방법을 시작으로 전력 신호의 하모닉 성분을 분석하는 Harmonic analysis 기법과 순간적인 변화 시점에서 특성을 추출하는 Transient based 기법 등으

로 전개되어 왔다. Steady-state 기반에서는 각 특성들이 물리적 덧셈 형태로 이루어지기 때문에 그 변화량에 근거를 두고 기기를 식별한다[2]. Harmonic analysis와 Transient based의 경우는 Steady-state based와 대비되는 이벤트 기반의 방법이며, 기기의 on/off 이벤트가 일어나는 순간의 변화량에서 특성을 추출한다[6][7].

기존의 연구에서는 대부분 유효전력, 무효전력 또는 특정 하모닉 성분을 특성으로 선택하는 특성 선택(feature selection) 기법을 주로 사용해왔으나, 본 논문에서는 169개의 특성 정보를 모두 담으면서 연산 속도를 높일 수 있는 주성분 분석 기법을 활용한다. 특정한 특성을 선택하는 것이 아니기 때문에 추출한 데이터의 정보를 그만큼 잘 반영할 수 있으며, 이는 k- 근접이웃 분류기의 성능을 향상시킨다. 또한 기기 식별 분류에 주로 활용되어 왔던 k-근접 이웃 분류 기법을 기기 상태, 즉 기기의 이벤트 식별에 활용함으로써 다양한 데이터에도 적용하기 용이한 전력신호 분류기를 제안한다. 본 논문에서는 기기 식별은 Steady-state based, 기기의 이벤트 식별에는 Transient based로 접근하였다.

IV. 분류모델

분류에 앞서 본 논문에 쓰인 모든 파라미터를 정리하면 다음과 같다.

Table 1. Notations

Notations	Contents
Σ	covariance matrix
λ	eigenvalues of covariance matrix Σ
U	eigenvectors for each λ
Λ	diagonal matrix whose diagonal elements are made up with λ
D	distance function for pair of vectors
F_x	feature vectors for data point
x, y	elements that make up F
d	number of principal components
k	number of neighbors for k-NN

4.1 특성 추출: 주성분분석(Principal Component Analysis)

주성분분석은 차원이 큰 특성들의 정보를 그대로 유지한 채로 새로운 주축을 재설정하여 차원을 감소시키는 방법이다. 이 때, 각 변량들을 효율적으로 나타내기 위해서 특징벡터들을 상관(correlation)이 없는 주축을 기준으로 재배치하게 된다. 그 방법으로는 변동량, 즉 특성 간의 공분산이 가장 높은 축을 주축 u_1으로 설정하고, 그 다음으로 높은 축을 주축 u_2로 한다. 이때 주축은 서로 수직이고, 서로 상관이 존재하지 않는다. 이와 같은 과정을 수행하기 위해 각 특성 간의 대칭적인 공분산 행렬 Σ를 고유분석(eigen-analysis)이나 SVD(Singular Value Decomposition) 등으로 분해하여 주성분을 추출한다. 다음은 SVD를 활용하여 공분산의 λ값이 가장 큰 순서대로 d개의 주성분을 추출하고, 각 변량들을 새로운 d차원 공간에 사영시키는 과정을 나타낸다.

1. $\Sigma = U\Lambda U^T$ 를 만족하는 고윳값 λ_i와 그에 따른 고유벡터 u_i를 구하여 행렬 Λ와 U를 구성한다.
2. λ_i가 큰 순서대로 U에서 d개의 고유벡터 u_i를 선택하여 $U' = [u_1, u_2, ..., u_d]$ 를 구성한다.
3. 전체 특징 벡터에 U'^T을 곱하여 축소된 d차원의 새로운 특징벡터 공간을 구성한다.

4.2 분류기: k-근접 이웃 분류기(k- Nearest Neighbor Classifier)

k-근접 이웃 알고리즘은 분류기법의 하나로, 트레이닝 집합과의 유사성(similarity)을 바탕으로 변량의 target value를 결정하는 방법이다. 우선, 트레이닝 집합으로 학습된 분류기는 새로이 들어오는 변량과 기존 트레이닝 집합 간의 유사성을 계산한다. 그 후, 가장 인접한 k개의 이웃 변량의 라벨 중 다수의 값으로 target value를 결정하게 된다. 본 논문에서는 기본적인 유클리디안 거리로 개체들 간의 유사성을 계산하였다. 주성분분석을 통해 얻어진 변량의 특징 벡터를 $F_x = [x_1, x_2, ..., x_d]$ 로 나타냈을 때, 두 특징벡터 F_x와 F_y간의 유클리디안 거리 D는 다음과 같이 나타낼 수 있다.

$$D(F_x, F_y) = \sqrt{(x_1 - y_1)^2 + ... + (x_d - y_d)^2} \quad (1)$$

테스트 벡터로부터 (1)의 값이 가장 작은 10개의 이웃 변량을 선택하고, 그 중 가장 많은 라벨로 target value를 결정한다.

4.3 검증: Hold-out 10차 교차검증

검증을 위해 교차검증을 활용하는 이유는 실험에 쓰인 트레이닝 집합에만 적용되는 결과가 아닌 다양한 트레이닝 집합에 대한 정확도를 검증하기 위함이다. Hold-out 교차검증에서는 데이터의 일부를 트레이닝 집합으로 사용하고 나머지는 Hold-out 집합으로 사용한다. 트레이닝 집합을 사용하여 주성분 추출 및 분류기를 만들고 이를 Hold-out 집합에 적용하여 정확도를 측정한다. Hold-out 집합은 주성분 분석 및 분류기 생성에 관여하지 않기 때문에 Hold-out 집합으로 측정된 분류기의 수행능력은 새로운 데이터에서 PCA-KNN 알고리즘의 수행능력을 평가하는 것보다 정확한 방법이 된다. 10차 교차검증에서는 전체 집합을 임의로 10개로 나누고, 각각의 10개의 부분집합을 Hold-out 집합으로 돌아가면서 활용하여, 나머지 9개의 부분집합을 트레이닝 집합으로 사용하게 되는 것이다. 그 결과, 10회의 실험에서 도출된 정확도의 평균으로 전체 정확도를 산정한다.

V. 실 험

5.1 데이터 추출

전력 신호로부터 데이터를 추출할 때는 보통 샘플링 주파수가 높을수록 각 기기의 전력 사용량의 변화를 세밀하게 표현하기 때문에 식별에 유리해진다. 논문에 쓰인 데이터는 Yokogawa WT1803 Power Analyzer를 통해 20Hz로 샘플링 했으며, 이는 실제 국내 스마트미터 기기들이 지원하는 50Hz~60Hz의 샘플링 주파수보다 조금 낮은 수준이다. 본 실험은 도청을 통해 얻을 수 있는 실제 데이터보다 낮은 주파수의 데이터로도 유의미한 기기 식별 결과를 도출할 수 있음을 보인다. 공격을 위해 추출된 특성은 총 169개로 기본적인 유효전력(P)과 무효전력(Q), 전압(Urms), 전류(Irms) 등을 바탕으로 연산되었다 (Table2. 참고).

1)기기의 식별 실험의 ⅰ)에서는 TV, Audio, 셋톱박스가 on 상태일 때의 데이터를 추출해서 비슷한 양으로 샘플링 했다. 각 기기의 데이터를 비슷한 양으로 샘플링하기 위해서 셋톱박스는 2회, 나머지 TV와 오디오는 1회에 걸쳐 추출하였다. 또한, ⅱ)에서는 기기들이 동시에 동작하는 경우에 대한 분류

Table 2. Method of extracting features

Feature	Equation	Feature	Equation
Urms	$\sqrt{\frac{1}{T}\int_0^T u(t)^2 dt}$	Irms	$\sqrt{\frac{1}{T}\int_0^T i(t)^2 dt}$
P	$\frac{1}{T}\int_0^T u(t)i(t)\,dt$	S	Urms × Irms
Q	$\sqrt{S^2-P^2}$	PF	P/S
Deg	Φ	U+peak	max(Urms)
U-peak	min(Urms)	I+peak	max(Irms)
I-peak	min(Irms)	P+peak	max(P)
P-peak	min(P)	CfU	max(Urms)/ Urms
CfI	max(Irms)/Irms		
U	harmonic of Urms	I	harmonic of Irms
P'	harmonic of P	S'	harmonic of S
Q'	harmonic of Q	PF'	harmonic of PF
Deg'	harmonic of Deg		

- $u(t)$: Voltage at t
- $i(t)$: Current at t
- $P = S * \cos\Phi$
- Harmonic factor : factors with other than fundamental frequency

를 수행하기 위하여 TV, 셋톱박스, 오디오가 동시에 동작하는 데이터를 추출했다.

2)기기의 상태 추론 실험에서는 기기 중 오디오의 경우에 대해서, 상태가 변하는 순간의 데이터를 추출하여(20Hz) 각 상태별로 라벨링 하였다.

5.2 특성 추출

Fig. 2.에서 볼 수 있듯이, 데이터가 가지고 있는 특성 중 유효전력과 무효전력 간에는 큰 상관이 존재한다. 각 특성이 변량들의 정보를 반영한다고 했을 때, 큰 상관을 가지는 두 특성은 변량을 나타내는 데 있어 효율성이 떨어진다고 할 수 있다. 주성분 분석은 이와 같이 큰 상관을 가지는 특성들을 상관을 갖지 않는 새로운 주축으로 재설정하여 특성 공간의 효

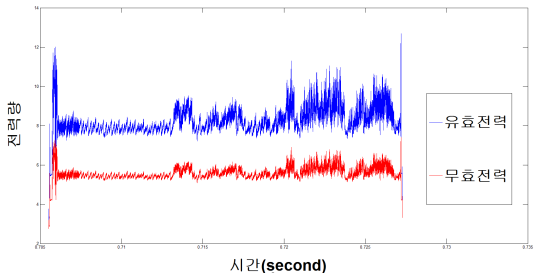

Fig. 2. Correlation between real power and reactive power (Audio case)

율성을 높인다.

특성 추출에 주성분 분석을 활용했던 기존의 연구들은 주성분의 개수 d를 임의로 설정하였거나 뚜렷한 기준을 제시하고 있지 않다[4][5]. 본 연구에서는 주성분의 개수 d를 변화시키면서 정확도가 가장 높을 때의 값을 제시하기로 한다.

5.3 분류

분류는 크게 1)기기의 식별과, 2)기기의 상태 추론으로 나뉜다. Table 3. 은 본 논문에 쓰인 데이터가 담고 있는 기기 종류와 각 기기별 상태의 세부사항을 나타낸다. 본 논문에서는 이해를 돕기 위해 1) 기기의 식별 실험을 다시 ⅰ)기기가 동시에 동작하지 않는 경우와 ⅱ)동시에 동작하는 경우로 나누어 실험한다. ⅰ)은 Ⅱ장에서 살펴보았던 IALM에 해당하고 ⅱ)는 스마트 미터로부터 탈취한 데이터로 집안 내부의 기기정보를 식별하기 할 수 있는 NIALM에 해당한다. 또한 2)기기의 상태 추론에서는 오디오 데이터가 작동하는 동안의 데이터를 이용하여, 오디오의 세부적인 on/volume 조절/off 상태를 식별한다. 이와 같은 분류를 통해, 스마트 미터로 측정된 데이터로 어떤 기기가 켜져 있는 지를 구분할 수 있는지, 각 기기별 데이터로부터 기기의 세부적인 상태를 추론할 수 있는지를 실험하는 것이다. Fig.3. 는

Table 3. Category and states of appliances

Appliances	States
TV	on, vol1, vol2, vol3, chan, off
Audio	**on, vol1, vol2, vol3, off**
Set Top Box	on, chan1, vol1, chan2, vol2, standby, off

- vol = volume change
- chan = channel change
- standby = steady-state

Fig. 3. Step of experiments

실험1)과 실험2)에 쓰인 머신러닝 과정을 도식화한 결과이다. 실험에 대한 결과는 전체적인 머신러닝 과정을 PCA-KNN 분류기로 명명하고 이에 대한 정확도를 검증한다. 단, d값에 대한 정확도 비교를 위하여 이웃 수 k는 10으로 고정하였다.

5.3.1 Appliance Identification

5.3.1.1 기기가 동시에 동작하지 않는 경우

본 실험에서 추출된 169개의 특성을 모두 활용하게 되면, 연산 복잡도가 매우 높아지는 결과를 초래한다. 이에 따라 주성분분석을 적용하여 d차원으로 특징 공간을 축소하였다. 10차 교차검증(10-fold Cross Validation) 결과, 3차부터 99% 이상의 정확도를 가졌고, 가장 정확도가 높았던 주성분 수는 7이었다.

10차 교차검증은 Hold-out 방법에 따라 트레이닝 집합을 랜덤하게 10등분하여 트레이닝 집합과 테스트 집합을 따로 구성한다. 트레이닝 집합을 주성분분석하여 U'(4.1 참고)를 도출하였고, 트레이닝 집합으로 구한 변환벡터를 테스트 집합에 적용하여 결과를 예측했다. Fig.4.에 d를 1부터 차례대로 증가시키면서 10차 교차검증의 결과를 나타내었다. k를 변화시킴에 따라 97%가 넘는 높은 정확도를 보였고 Fig.4.는 변화 추이를 자세히 관찰하기 위하여, 정

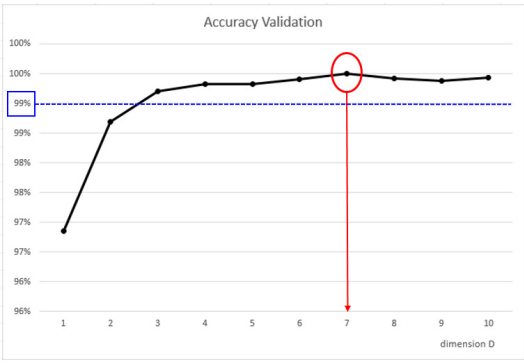

Fig. 4. Accuracy with dimension d (5.3.1.1)

확도를 나타내는 y축의 시작점을 0%가 아닌 97%로 고정했다. d가 3일 때부터 99% 이상의 정확도를 보였고 d가 7일 때 가장 높은 정확도 99.50%를 기록했다.

99% 이상의 정확도를 보이기 시작하는 d가 3인 경우를 특성공간에 나타내면 Fig.5.과 같이 나타난다. TV는 빨간색, 셋톱박스는 파란색, 오디오는 노란색으로 표시되었다. 셋톱박스의 경우 두 번에 걸쳐 측정되었으며, 한 번에 측정된 벡터들이 아님에도 불구하고 매우 견고한 군집을 형성하는 것을 확인할 수 있다. 그 후, Fig.6.과 같이, 유클리디안 거리를 기준으로 테스트 벡터에 대한 10개의 근접한 이웃 변량들의 라벨을 확인한다. Fig.6.의 테스트 벡터의 경우 8개의 이웃 변량이 TV class에 있으므로 테스트 벡터의 target value는 TV로 분류된다.

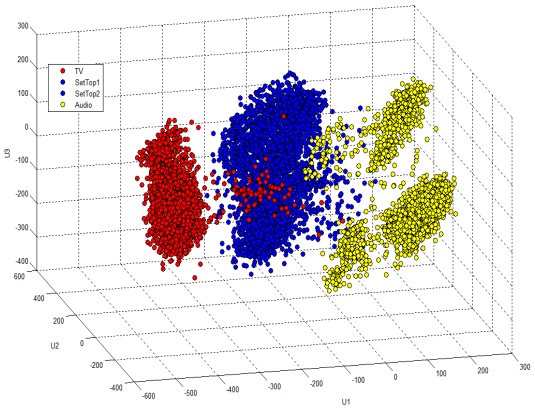

Fig. 5. Result of PCA with 3 principal components (5.3.1.1)

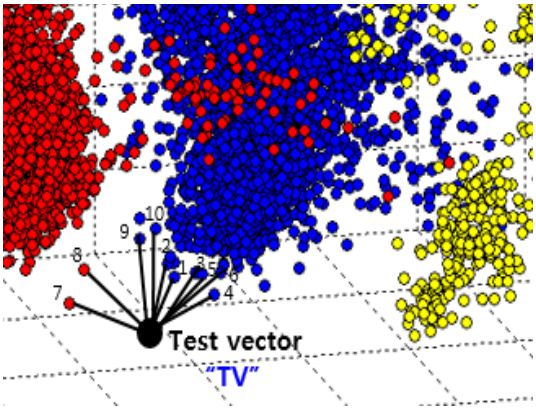

Fig. 6. k-NN with 10 neighbors (5.3.1.1)

5.3.1.2 기기가 동시에 동작하는 경우

기기가 동시에 동작하는 경우에는 각 기기 별 특성 뿐 아니라 동시 동작하는 경우의 특성이 추가된다. ⅰ)과 달리 k-근접 이웃 분류기로 NIALM을 수행하기 위해서는 나올 수 있는 모든 경우의 수를 고려하여 모든 동시 동작 상황에 대해 트레이닝 시켜야 한다. 현재 TV, 셋톱박스, 오디오 총 3개의 기기에 대한 데이터를 가지고 있고, 각 기기들이 함께 동작할 때의 전력 데이터를 측정한다. 이에 따라 기기 수 3개에 대해서, 하나의 기기만 동작하는 경우, 두 개의 기기가 동시 동작하는 경우, 세 개의 기기 모두 동작하는 경우로 나누어 총 7가지의 경우에 대한 전체 데이터 집합을 구성하였다. 기기들의 동시 동작 상태를 반영한 라벨은 Table 4.와 같다. 주성분 분석을 통해 축소된 특성공간에 나타난 결과는 Fig.7.과 같고, 총 7개의 라벨 값을 바탕으로 k-근접이웃 분류로 교차검증을 수행하여 구한 정확도는 차원 수 d에 따라 Fig.8.과 같이 나타난다. k를 변화시킴에 따라 94%가 넘는 높은 정확도를 보였고 Fig.8.은 변화 추이를 자세히 관찰하기 위하여, 정확도를 나타내는 y축의 시작점을 0%가 아닌 93.60%로 고정했음에 유의한다. d를 2에서 10까지 변화시키면서 관

Table 4. Labelling for multi-appliances

Label	Appliances
1	TV
2	Set-top box
3	Audio
4	TV, Set-top box
5	Set-top box, Audio
6	TV, Audio
7	TV, Set-top box, Audio

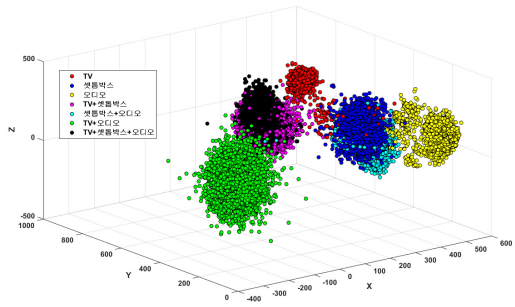

Fig. 7. Result of PCA with 3 principal components (5.3.1.2)

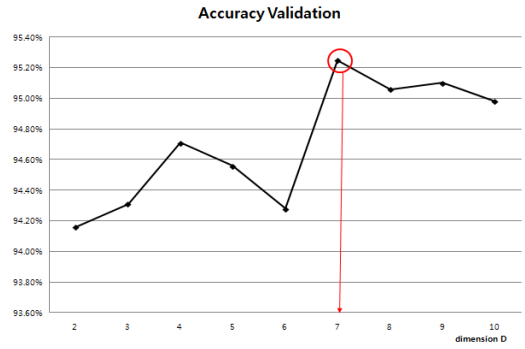

Fig. 8. Accuracy over dimension d (5.3.1.2)

찰한 결과, d가 7일 때 최댓값 95.25%의 정확도를 기록했다. 10차 교차검증을 바탕으로 수행했기 때문에 더 정확한 검증 값을 추출할 수 있었다.

5.3.2 Event detection

실험2)는 오디오의 전력 신호를 분석하여 오디오의 이벤트를 추론한다. 주성분분석, k-근접 이웃 분류기로 분류하고 10차 교차검증으로 정확도를 확인하는 점에서 실험1)과 유사하지만, 특성을 추출하는 방법에서 차이가 있다. 실험1)의 데이터는 기기가 켜질 때부터 꺼질 때까지의 연속적인 데이터를 활용했다면, 실험2)는 오디오의 'on', 'off', 'volume change' 와 같은 이벤트를 식별하기 위해서, 오디오가 켜질 때(on), 꺼질 때(off), 볼륨을 높이는(vol) 순간에 해당하는 변량 값을 각 상태별 데이터로 추출했다. Fig.9.은 오디오 데이터의 유효전력을 시간에 따라 나타낸 것으로, 각 이벤트에 따라 변량들이 급작스럽게 변화하는 모습을 확인할 수 있다. 본 실험에서 쓰인 데이터는 같은 오디오 모델로 5회에 걸쳐서 측정했으며, 총 169개의 특성에 대하여 이벤트가 일어난 시각의 초당 20개(20Hz)의 변량 값을 가져

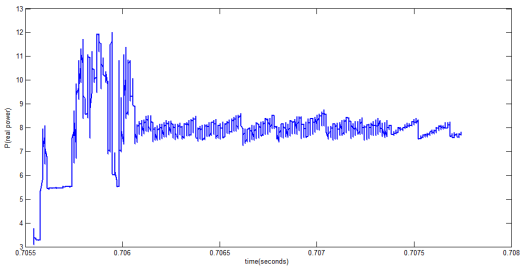

Fig. 9. Real power over time (sec) (Audio case)

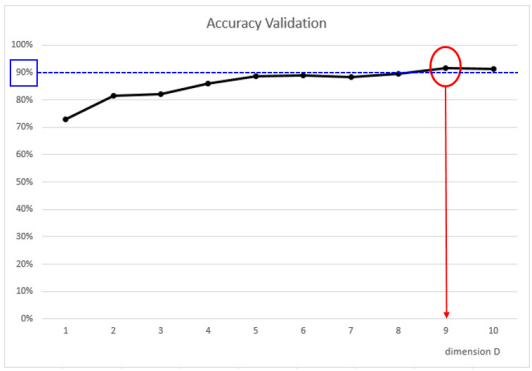

Fig. 10. Accuracy over dimension d (5.3.2)

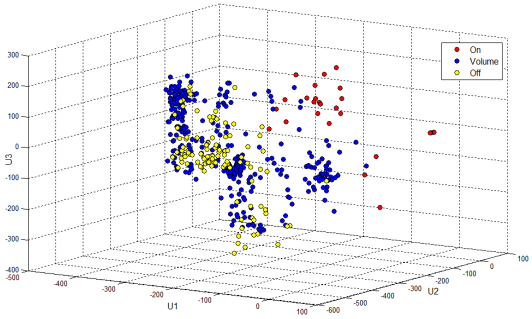

Fig. 11. Result of PCA with 3 principal components (5.3.2)

왔다.

Fig.10.에서 볼 수 있듯이, d를 변화시키며 교차검증을 수행한 결과 d가 9일 때부터 90% 이상의 정확도를 가졌고 d가 9일 때 최고의 정확도인 91.56%를 기록하였다.

참고를 위해 Fig.11. 에 주성분의 개수를 3으로 가정하여 변환한 결과를 나타냈다. on은 빨간색, volume change는 파란색, off는 노란색으로 표시되었다. 3차원의 경우는, 82.22%의 정확도를 보이는 만큼 뚜렷한 군집이 나타나지 않는다. 90% 이상의 정확도로 기기 상태를 추론하기 위해서는 9차 이상의 특성 공간이 필요하다.

실험1)과 달리 실험 2)의 이벤트 기반 실험에서는 몇 가지 어려움이 있었다. 첫째, 이벤트가 발생하는 시각의 초당 데이터 수가 20개씩이므로(20Hz) steady-state의 값을 제외한 나머지 데이터의 양이 상대적으로 매우 적어 충분한 데이터를 확보하기 힘들었다. 둘째, 볼륨 조절 이벤트는 시작점과 볼륨 변화 정도, 수행 환경에 따라 결과가 상이할 수 있어서

분류가 까다로웠다. 결과적으로, 9차원으로 주성분분석 후 k-근접 이웃 분류의 결과를 검증하여 91.56%의 정확도를 확인했다.

VI. 실험 결과

본 논문은 스마트 미터 데이터를 주성분분석과 k-근접 이웃 알고리즘으로 분석하여 내부 기기들의 동작 상태를 예측하는 데 목적이 있었다. 결과적으로, 실험1)에서 기기가 동시에 동작하지 않는다는 가정 하에서는 가장 높은 정확도를 가지는 주성분 수 7을 활용하여 99.50%의 정확도로 가정 내 기기를 분류할 수 있었고, 기기가 동시 동작하는 환경에서는 주성분 개수 7에서 95.25%의 정확도를 기록했다. 이에 이어 실험2)에서는 Transient based로 접근하여 오디오의 세부 상태를 주성분 9개에서 91.56%로 추론할 수 있었다. 이와 같이, 다수의 특성을 지닌 전력 데이터는 주성분 분석과 k- 근접 이웃 분류기를 통해 빠르고 꽤나 정확하게 분류가 가능함을 확인했다. 이에 따라, 기계식 전력량계와 달리 스마트 미터는 유효전력, 무효전력, 역률, Peak 등 분류에 쓰일 수 있는 다수의 특성을 지닌 데이터를 메인 서버와 쌍방으로 주고받기 때문에, 집안 내부 기기의 사용 상태를 예측하기 위한 취약 포인트가 될 수 있다는 결론에 이른다. 따라서 스마트 미터와 메인 서버간의 도청공격을 방지할 수 있는 보안 조치가 필수적이다. 추가적으로, 분석에 쓰이는 데이터의 양도 그 취약성에 영향을 미쳤다. 실험1)의 ii)에서 데이터의 양에 따른 분석 소요시간과 정확도를 살펴보면,(Table6. 참고) 데이터가 많을수록 소요시간도 증가하지만, 정확도도 높아짐을 알 수 있다. 따라서 공격자가 분석할 수 있는 데이터의 양을 제한하는 조치가 필요하다. 본 논문의 실험 결과는 대용량의 전력량 데이터를 그대로 반영했다고 볼 수는 없다. 실제 공격자는 더 많은 양의 데이터를 추출할 수 있고, 훨씬 더 빠른 연산속도를 지닌 기기를 사용할 수 있기 때문에 더 정확한 분류 결과를 도출할 가능성이 있다.

Table 5. Computing time and accuracy by amount of data

Amount of data	Computing time (sec)	Accuracy (%)
1/2	32.597	90.30
1	35.964	92.56

VII. 결 론

결론적으로, 스마트 미터기로부터 추출한 데이터는 가정 내 기기의 상태를 드러낼 수 있고, 이는 가전기기의 사용이 일상화된 현대 사용자들의 행동 패턴을 그대로 반영한다. 따라서 스마트 기기 환경의 안전한 정착을 위해서는, 각 스마트 기기들이 다루게 될 정보의 중요성을 인지하고 이에 대한 보안대책이 마련되어야 한다. 공격자로 하여금 평문으로 된 스마트 미터 데이터를 탈취할 수 없도록 각 가정과 서버간의 E2E 암호화를 철저히 하고, 중요 데이터가 저장되어 있는 데이터베이스가 탈취되는 경우에 대비하여 데이터의 저장 기간을 최소화 하고 데이터의 주기적인 삭제를 통해 추출할 수 있는 데이터의 양을 제한해야 한다.

References

[1] Ridi, Antonio, Christophe Gisler, and Jean Hennebert, "A survey on intrusive load monitoring for appliance recognition," Pattern Recognition (ICPR), 2014 22nd International Conference on. IEEE, pp. 3702-3707, Aug. 2014.

[2] Zoha, Ahmed, et al, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," Sensors 2012, vol. 12, no. 12, pp. 16838-16866, Dec. 2012.

[3] Hart, George W, "Nonintrusive appliance load monitoring," Proceedings of the IEEE, vol. 80, no. 12, pp. 1870-1891, Dec. 1992.

[4] Zufferey, Damien, et al, "Machine learning approaches for electric appliance classification," Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on. IEEE, pp. 740-745, Jul. 2012.

[5] Kato, Takekazu, et al, "Appliance recognition from electric current signals for information-energy integrated network in home environments," Ambient Assistive Health and Wellness Management in the

Heart of the City. Springer Berlin Heidelberg, vol. 5597, pp. 150-157, Jul. 2009.

〔6〕 Leeb, Steven B., Steven R. Shaw, and JJames L. Kirtley Jr, "Transient event detection in spectral envelope estimates for nonintrusive load monitoring," Power Delivery, IEEE Transactions, vol. 10, no. 3, pp. 1200-1210, Jul. 1995

〔7〕 Najmeddine, Hala, et al, "State of art on load monitoring methods," Power and Energy Conference, 2008. PECon 2008. IEEE 2nd International. IEEE, 2008, pp. 1256 - 1258, Dec. 2008.

〔8〕 Brinkhaus, S., et al, "Smart hacking for privacy," 28th Chaos Communication Congress. Retrieved from: http://events. ccc. de/congress/2011/Fahrplan/attachments/1968_28c3-abstract-smart_hacking_for_privacy. pdf, Dec. 2011.

〈저 자 소 개〉

조 재 연 (Jae yeon Cho) 학생회원
2011년 2월: 고려대학교 수학과 졸업
2014년 3월~현재: 고려대학교 정보보호대학원 석사과정
〈관심분야〉 IoT, 머신러닝, 정보보호, 금융보안

윤 지 원 (Ji Won Yoon) 종신회원
2008년 11월: University of Cambridge 전자공학과 박사 졸업
2008년 2월~2009년 5월: University of Oxford, 로봇연구소 박사후과정
2009년 5월~2011년 5월: University of Dublin 통계학과 연구원 및 강사
2011년 7월~2012년 8월: IBM 연구소 정규 연구원
2012년 9월~현재: 고려대학교 정보보호대학원 조교수
〈관심분야〉 신호정보처리, 응용통계, 도감청 탐지기술