논문 19-44-08-14 The Journal of Korean Institute of Communications and Information Sciences '19-08 Vol.44 No.08
https://doi.org/10.7840/kics.2019.44.8.1547

LoRaWAN 환경에서의 머신러닝을 통한 네트워크 공격 탐지 구현

정 택 현◆, 이 승 호*, 김 기 천°

Implement Detecting Network Attack through Machine Learning in LoRaWAN Environment

Tack-hyun Jung◆, Seung-ho Lee*, Kee-cheon Kim°

요 약

본 논문은 사물인터넷(Internet of Things, IoT)의 시대가 도래함에 따라 주목받고 있는 저전력·장거리 통신 지원 기술인 LoRaWAN(Long Range Wide Area Network) 환경에서 발생하는 네트워크 보안 위협의 종류와 메커니즘을 분석하고, 이러한 위협을 머신러닝을 접목하여 사전에 탐지하는 방법을 제안한다. LoRaWAN은 독자적인 구조로 인해 일반적으로 IPS/IDS에서 활용되던 보안탐지 알고리즘의 적용이 어렵다. 따라서 제안하는 알고리즘은 LoRaWAN 환경에서 발생하는 메시지를 대상으로 머신러닝의 기법 중 군집화를 통한 새로운 접근방법을 제시하며, 메시지 재사용 공격과 Spoofing과 같은 공격의 탐지에서 매우 뛰어난 성능을 보인다. 실제로 여러 번의 실험에서 검증된 100%의 탐지율과 적은 소요시간은 실제 LoRaWAN 환경에 이를 적용하는 데 있어서 큰 가능성을 제시한다.

Key Words : LoRaWAN, Cyber Security, Machine Learning, Anomaly Detection

ABSTRACT

This paper analyzes the type and mechanism of network security threats arising in LoRaWAN (Long Range Wide Area Network), a low-power and long-range communication support technology that is drawing attention as the era of the Internet of Things (IoT) arrives, and proposes a method to detect these threats in advance by applying machine learning. LoRaWAN's unique architecture makes it difficult to apply security detection algorithms that were generally used in IPS/IDS. Therefore, the proposed algorithm presents a new approach through clustering of machine learning techniques for messages occurring in LoRaWAN environments, and performs very excellent performance in detection of attacks such as message reuse attack and spoofing. In fact, the 100% accuracy of detection and the low time required, which have been verified in several experiments suggest great potential in applying it to the actual LoRaWAN environment.

※ 본 연구는 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.2017-0-00275,안전한 IoT 전용망 구축을 위한 LPWAN 침해 방지 기술 개발)을 받아 수행되었습니다.
◆ First Author : Konkuk University Department of IT Convergence Information Security, tackhyun12@konkuk.ac.kr, 학생회원
° Corresponding Author : Konkuk University Department of Computer Engineering, kcim@konkuk.ac.kr, 종신회원
* Konkuk University Department of Computer Engineering, phg0726@konkuk.ac.kr, 학생회원
논문번호 : 201903-027-D-RN, Received March 25, 2019; Revised May 20, 2019; Accepted June 12, 2019

Ⅰ. Introduction

최근 정보통신 기술의 발전에 따라 모든 사물이 네트워크로 연결되는 사물인터넷(Internet of Things, IoT)이 주목받고 있다. IoT는 각종 사물에 센서를 내장하여 수많은 정보를 수집하며, 통신 기능을 내장하여 유무선 통신을 통해 네트워크에 연결한다.

IDC와 가트너와 같은 세계적인 시장 분석기관은 IoT를 차세대 핵심기술로 보고 2020년까지 208억 개 이상의 사물들이 연결될 것으로 예측하였다.[1] 이러한 예측에 따라 IoT를 활용하기 위한 기술 또한 다양하게 발전하고 있는데, 대표적인 예가 무선 통신기술의 발전이다.

IoT 환경에서 많은 사물은 내장된 무선 통신기술을 통해 장거리에 있는 서버와 통신하여 데이터를 교환한다. 이러한 과정에서 서비스가 원활하게 수행되기 위해서는 장거리 통신이 가능하게 하는 기술이 필요하며, 무선 통신 기기 중 다수가 배터리를 통해 전력을 공급받는다. 이러한 특징에 따라 사물 간의 통신에서 저전력·장거리 무선 통신 기술이 발전되고 있다.

LoRaWAN(Long Range Wide Area Network)[2]은 대표적인 장거리 무선 통신기술이다. 타 무선 통신 기술보다 훨씬 긴 도달 범위를 가지므로[2] 많은 리피터 및 AP가 필요 없다는 강점은 인프라 구축 비용을 낮출 수 있다는 점에서 크게 주목받고 있다. 이러한 특성 때문에 도시 및 농촌을 비롯한 모든 유형의 환경에서 끊김 없이 사용할 수 있다. 실제로 농촌/준도시 지역에서는 LoRaWAN의 통신 적용 범위가 15km에 달하며, 도시에서도 2-5km가 넘는 우수한 성능을 보인다.[2]

LoRaWAN의 또 다른 장점으로는 LoRaWAN에서 동작하는 기기는 ISM 주파수 사용 규정에 따라 기기 출력이 10-25mW 이상을 넘지 않는다는 점이다.[3] 이는 LTE Cat-M 및 협대역 사물인터넷(NB-IoT) 환경에서 동작하는 기기의 배터리 효율성을 훨씬 상회한다. 이러한 배터리 높은 효율성으로 인하여 충전을 하지 않아도 약 10년 정도 연속 사용이 가능하다.[3]

이러한 장점들로 인해 LoRaWAN 기술은 개인 및 공용 네트워크를 포함하여 약 50개국에서 구현되며 큰 인기를 누리고 있지만, 다른 네트워크 기술과 마찬가지로 LoRaWAN에도 기기의 서비스 거부를 유발하거나 배터리 수명을 소진 시키는 등 수많은 보안 위협이 존재한다. 대표적인 보안 위협으로는 메시지 재사용 공격(Replay Attack)이 있는데, 이는 LoRaWAN의 통신 메커니즘을 이용하여 기기의 서비스 장애를 유발하는 공격이다.

이러한 위협에 대응하기 위한 연구도 각지에서 다양하게 진행되고 있지만, 현재로서는 표준 등의 명확한 대응방안이 존재하지 않는다. 시중의 정보보호 솔루션인 IPS(침입방지시스템) 및 IDS(침입탐지시스템) 등은 대부분 TCP, UDP 프로토콜 등 인터넷에서 자주 사용되는 프로토콜을 대상으로 구현되어 있어, 이를 LoRaWAN에 접목하는 것은 사실상 불가능하다. 따라서 본 논문은 LoRaWAN에서 발생할 수 있는 여러 네트워크 침해시도들의 메커니즘을 고찰하고, 이를 머신러닝을 접목하여 탐지하는 기술을 구현하고자 한다.

Ⅱ. Related Works

2.1 LoRaWAN's Security Vulnerabilities

2.1.1 Up-Link Message

LoRaWAN에서 Device는 통신을 위해 Gateway에게 Up-Link 메시지를 보낸다.[4] 이때 Device는 메모리에 자체적으로 counter 변수를 선언하여 메시지의 순번을 관리하며, 안전한 통신을 위한 신뢰성 확보에 사용된다.

[Fig 1]은 Up-Link 메시지의 전달방식을 의미한다. Device가 Gateway에게 Up-Link 메시지를 전달하게 되면, Gateway는 통신을 허용하는 ACK 메시지를 통해 Device에게 응답한다.[4]

이러한 과정이 모두 수행되게 되면 LoRaWAN의

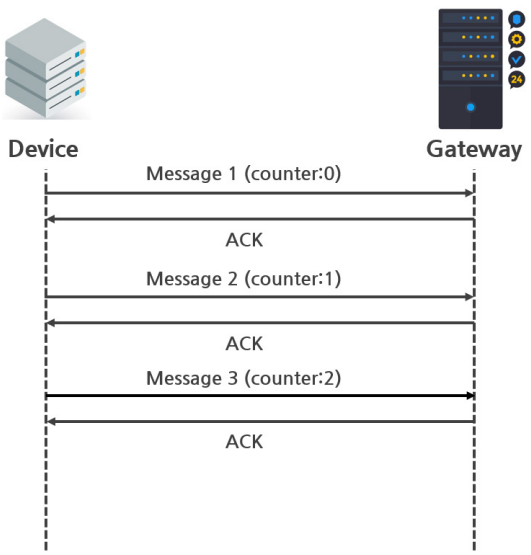

그림 1. Up-Link 메시지 전달 방식
Fig. 1. Up-Link Message Transfer Method

A Class[4] 기반의 통신은 Device로부터 2번의 Down-Link 메시지를 수신할 수 있는 대기 상태가 되고, Device는 해당 시점에서 전송하고자 하는 데이터를 보내게 된다. 이후, 2번의 데이터 통신이 완료되면 통신을 위한 UP-Link 인증과정을 다시 갖는다.

2.1.2 Message Replay Attack

[Fig 2]는 메시지 재사용 공격의 시나리오를 도식화한 그림이다. 공격자(Attacker)는 Device와 Gateway의 사이에서 공격을 수행하는 개념이다.

[Fig 3]은 메시지 재사용 공격이 수행되는 메커니즘을 의미한다. 공격자는 Device의 통신에 장애를 유발하는 것을 목적으로 한다.

먼저, 공격자는 Device가 Gateway에게 보내는 Up-Link 메시지를 가로채어 저장한다. 이후, 공격자는 Device의 메모리가 재부팅 등으로 인해 초기화되기를 기다리거나, 직접 overflow와 같은 공격을 통해 Device의 메모리의 초기화를 유도한다.

Device의 메모리에 초기화가 발생하면, Device가 관리하는 counter 변수는 다시 0으로 재설정된다. 해당 시점에서 공격자는 미리 저장해둔 Up-Link 메시지를 다시 전송한다. Gateway는 공격자가 보낸 UP-Link 메시지를 구분할 수 없으므로, 정상적으로 응답하게 되고, 이에 따라 변수의 매칭 오류가 유발된다. 즉, Device가 보내는 Counter와 Gateway가 최종적으로 수신한 Counter와 일치하지 않게 됨에 따라 서비스의 장애가 발생한다. 이는 일종의 DoS와 유사한 개념이다.

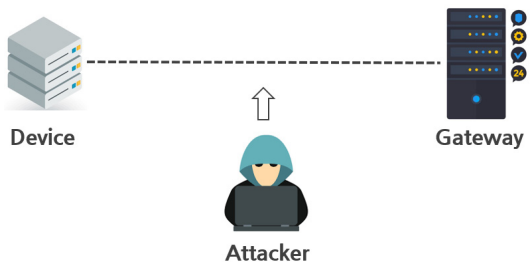

그림 2. 메시지 재사용 공격의 시나리오
Fig. 2. Message Replay Attack's Scenario

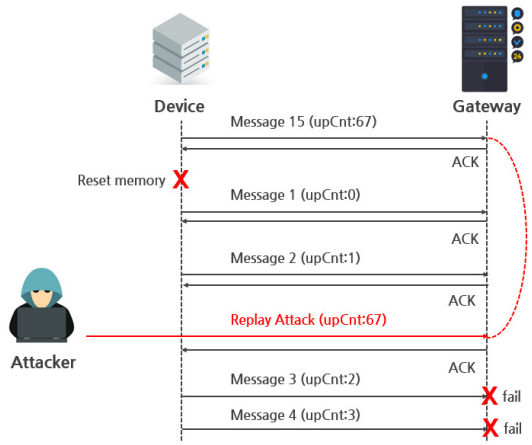

그림 3. 메시지 재사용 공격의 메커니즘
Fig. 3. Message Replay Attack's Mechanism

2.1.3 Why Used Machine Learning?

본 연구는 이러한 공격에 대응하는 방안으로 머신러닝을 활용한 접근방법을 선정하였다. 메시지 재사용을 탐지하는 대표적인 방법으로는 (1)과 같이 메시지에 포함된 타임스탬프 필드와 고유의 Key를 해시 암호화하여, 무결성을 검증하는 방안이 있다.

$$M = hash(time + key_i) \tag{1}$$

[Fig 4]는 (1)의 해시 기반 메시지 재사용 공격 탐지 과정을 도식화한 그림이다. Device는 고유의 키(key_i)와 타임스탬프를 통해 해시값을 생성한다. 이후 Gateway는 미리 가지고 있던, 키(key_i)를 통해 무결성을 검증하는 방식이다. 하지만, 본 연구에서는 해당 방식이 아닌 머신러닝을 통한 접근방법을 선택하였으며, 그 이유는 LoRaWAN의 독자적인 구조[4]를 근거로 한다.

[Table 1]은 LoRaWAN의 통신에서 발생한 로그이다. Device가 한 번의 Up-Link 메시지 이후 두 번의 메시지를 보내는 것을 볼 수 있는데, LoRaWAN의 표준에서 타임스탬프(Timestamp) 필드가 밀리세컨드 단위를 지원하지 않는다는 것을 볼 수 있다.[4]

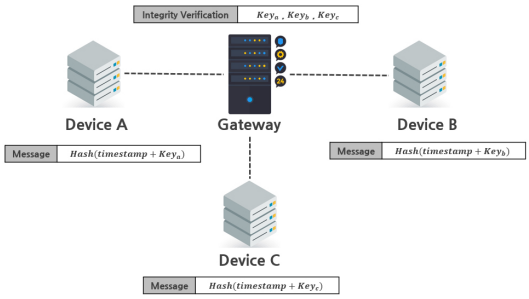

그림 4. 해시 기반 메시지 재사용 공격 탐지
Fig. 4. Hash Based Replay Attack Detection

표 1. LoRaWAN 환경에서의 메시지 로그
Table 1. LoRaWAN's Messages Log

Timestamp	UpCnt	Port	Dev_id	Payload
18:04:03	-	20	14	343437
18:04:03	-	20	14	357429
18:04:01	9	20	14	342174
18:01:58	-	20	14	376456
18:01:58	-	20	14	324785

[Fig 5]는 LoRaWAN의 센서에서 발생한 메시지이다. Time-stamp에 표시된 아라비아 숫자 "3290480372"는 총 10자리로 GMT 기준 "2074년 April 9일 Monday AM 06:19:32"에 해당한다. 만일 밀리세컨드 단위를 표시하기 위해서라면 총 길이가 13자리여야 한다. 따라서 타임스탬프 필드는 중복이 발생하는 문제점이 있다.

중복이 발생하게 되면 해시 기반으로는 메시지 재사용 공격을 효율적으로 탐지할 수 없다. 이에, 문제를 해결하기 위해 메시지별로 Key를 각기 다르게 주는 방안도 고려해 보았지만, 이는 유효한 해결책이 되지 못한다. LoRaWAN이 실제 적용되는 IoT 환경은 매우 많은 수의 Device를 운용하므로 Device들이 모든 메시지 단위로 Key를 별도로 부여하여 통신을 수행한다면, 매우 큰 부하가 발생하며, 키 관리의 측면에서도 문제가 발생할 수 있다. 또 다른 문제점으로는 Device에서 해시 암호화 연산을 수행할 시 배터리와 하드웨어 자원을 더 사용하게 되는 문제점이다. LoRaWAN 환경에 적용되는 Device에는 작은 단위의 센서 등도 포함되어 있어서, 이러한 방법의 적용은 또 따른 문제가 발생할 수 있다.


```
{
  "hits": [
    {
      "messageId": "0A6317F6",
      "deviceEUI": "05D25D35",
      "receivedBy": [
        {
          "messageId": "0A6317F6",
          "deviceEUI": "05D25D35",
          "baseStation": "976C64DFD4A9220F",
          "timestamp": "3290480372",
          "position": {
            "latitude": "52.38338",
            "longitude": "5.21295",
            "altitude": "-29.0"
          },
          "data": {
            "size": "37.0",
```


그림 5. LoRaWAN의 센서에서 발생한 메시지 로그
Fig. 5. Message Log from LoRaWAN Sensor

따라서 본 연구에서는 이러한 문제를 머신러닝 기반의 탐지 알고리즘을 통해 해결하고자 한다. 제안하는 알고리즘은 메시지의 여러 Feature를 활용하여 군집화를 수행하고 이상을 탐지한다. 이러한 원리는 Up-Link 메시지뿐 아니라 Down-Link를 포함한 모든 종류의 메시지 재사용을 탐지할 수 있다는 강점이 있으며, Gateway에 구현하기 때문에 Device에는 전혀 영향을 주지 않는다. 따라서, 이는 LoRaWAN 환경에 적용할 수 없었던 기존 보안 솔루션들의 문제점을 보완하는 접근방법이라고 판단된다.

Ⅲ. Proposed Security Solution

3.1 Detection from Replay Attack

[Fig 6]은 제안하는 메시지 재사용 공격의 탐지 절차를 표현한다. Gateway에 내장하여 구현되는 해당 알고리즘은 매 통신이 수행될 때마다 선정된 Feature를 학습하는 과정을 갖는다. 이후, Replay Attack이 발생하면 학습된 데이터를 기반으로 탐지하게 되고, 차단을 수행할 수 있다. 이러한 동작 원리는 Device의 Memory에 문제가 생기거나 공격으로 인해 초기화가

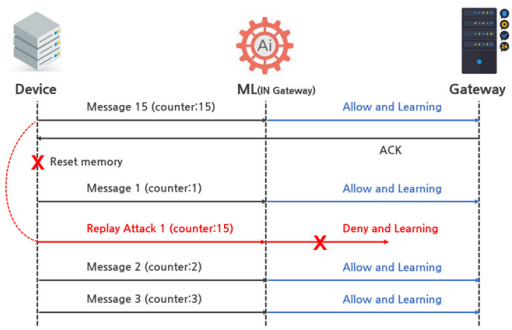

그림 6. 메시지 재사용 공격 탐지 절차
Fig. 6. Message Replay Attack Detection

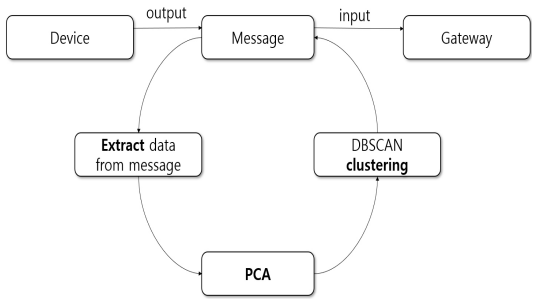

그림 7. 메시지 분류 절차
Fig. 7. Procedure of Message Classification

되어도 영향을 받지 않는다.

[Fig 7]은 제안하는 알고리즘의 전체적인 동작 과정을 의미한다. 메시지가 발생하면, 먼저 메시지 구조에서 데이터를 추출하는 선행 작업을 수행한다. 이후 군집화에 사용될 Feature들을 선정하여 차원축소를 위한 주성분 분석(PCA)[6]을 수행한다. 최종적으로는 DBSCAN[7] 알고리즘을 통해 메시지를 분류하여 재사용 여부를 판단하게 된다.

3.2 Machine Learning

3.2.1 Feature Selection

머신러닝의 접목에 가장 중요한 것은 독립적인 변수로 사용되는 정확한 Feature의 선정이다. Feature는 명확하면서 개별적이고 측정 가능한 경험적(Heuristic) 사실에 근거한 선정일수록 좋은 성능을 기대할 수 있다.

따라서 본 연구에서는 LoRaWAN 환경에서 발생한 로그에서 나타난 여러 필드를 고찰하고 분석한 결과를 바탕으로 Feature 선정을 수행하였다.

LoRaWAN 환경에서 메시지 재사용 공격을 탐지하기 위한 Feature 선정 시에는, time_stamp와 같이 메시지마다 가변하는 필드는 지양하는 것이 좋다. 즉, 재사용이 수행되었을 때 이전에 수행된 원본 메시지와 완전히 같은 값이 활용되는 필드를 기준으로 선정해야 한다.

[Table 2]는 Up-Link 메시지를 기준으로 재사용 공격이 발생한 상황을 기록한 로그이다. time_stamp 필드는 메시지 재사용 공격 발생 시 현재 메시지를 보낸 시간으로 재설정되는 것을 확인할 수 있다. 하지만, UpCnt, Port, Payload 등의 필드는 원본 메시지와 완전히 같은 데이터가 재사용되었다. 따라서 재사용 공격을 탐지하기 위한 목적으로는 이러한 고정적인 필드를 Feature로 선정하는 것이 좋다.

표 2. LoRaWAN 환경에서 재사용 공격이 발생한 로그
Table 2. Log of Replay attack in LoRaWAN

Time	UpCnt	Port	Payload	Note
18:20:20	10	20	343437	Failure
18:18:25	9	66	357429	Failure
18:18:12	**8**	**41**	**376456**	**Replay Attack**
18:04:01	0	66	342174	Memory Reset
18:02:32	**8**	**41**	**376456**	**Save Message**
18:01:59	7	20	324785	

3.2.2 PCA

Feature가 많을수록 데이터의 의미를 제대로 표현하는 특징을 추려내는 작업이 필요하다. 이러한 작업을 차원 축소(Dimensionality Reduction)라고 하며, 가장 대표적으로 데이터에서 주성분을 추려내는 작업을 주성분 분석(Principal Component Analysis, PCA)이라고 한다.

주성분 분석은 변수들의 Scale이 서로 많이 다른 경우 값이 큰 특정 변수가 전반적인 결과를 좌우한다.[6] 따라서 Feature로 사용된 변수들의 Scale을 반드시 고려해야 한다. LoRaWAN 환경의 대부분의 필드는 Scale 차이가 큰 특징이 있으므로 주성분 분석의 수행 시 공분산 행렬이 아닌 상관계수 행렬(2)을 활용한다.[8]

$$p[X,Y] = \frac{Cov[X,Y]}{\sqrt{Var[X] \cdot Var[Y]}} \quad (2)$$

[Fig 8]은 실제 LoRaWAN 환경에서 선정한 Feature를 대상으로 주성분 분석을 수행한 결과이다. 총 5개의 주성분 예시가 생성되었는데, 이러한 예시 중 가장 적합한 성분을 선정해야 한다. 선택 시 고려해야 하는 사항으로는 크게 2가지가 있는데, 첫 번째로는 누적 기여율(Cumulative Proportion)이 0.85 이상인 것을 선정하는 것이 좋다. 그리고 상관계수 행렬의 표준편차(Standard deviation)가 1이므로 표준편차의 수치가 1 이거나 0.7 보다 작은 것은 선정하지 않는 것이 좋다.

```
> pca <- prcomp(LoRa_dataset, scale = T) # 상관계수행렬 PCA
> summary(pca)
Importance of components:
                          PC1    PC2    PC3      PC4      PC5
Standard deviation     1.7344 1.0084 0.9875 0.887010 2.965e-19
Proportion of Variance 0.6016 0.2034 0.1950 0.000000 0.000e+00
Cumulative Proportion  0.6016 0.8050 1.0000 1.000000 1.000e+00
```

그림 8. 주성분 분석 결과
Fig. 8. Results of PCA

3.2.3 Clustering

주성분 분석의 결과는 밀도 기반 군집화 모델인 DBSCAN[7] 알고리즘을 통해 활용한다. DBSCAN은 군집의 개수를 지정하지 않아도 효율적으로 군집화를 수행하는 특징이 있으며, 같은 데이터 집합에 대해 여러 번 반복 수행하더라도 항상 같은 결과를 내놓는 일관성이 보장된다는 특징이 있다.[9]

따라서 본 연구에서는 이러한 특징에 따라 군집화를 수행하였을 때, 최소 점의 개수를 매우 낮게 지정하여 중복되는 값에 의해 군집의 생성을 탐지하는 것

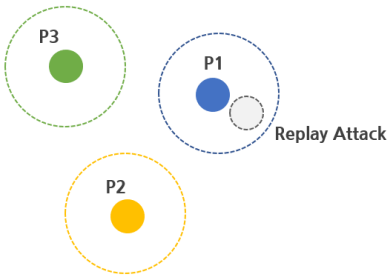

그림 9. DBSCAN 군집화를 통한 탐지 예시
Fig. 9. Detection Example of DBSCAN Clustering

을 목표로 한다.

[Fig 9]는 DBSCAN 군집화를 통한 재사용 공격 탐지의 예시이다. 만일 군집화를 수행함에서, minPts를 2로 설정하게 되면, 각각의 점(p)의 epsilon(eps) 반경 내에 또 다른 점이 1개 이상 추가로 있어야 군집이 형성된다. 본 연구는 이러한 메커니즘을 활용하여, 군집이 기본적으로는 형성될 수 없도록 구현한다. 이는 eps와 minPts를 매우 낮게 설정하는 것을 통해 군집이 형성되지 않도록 할 수 있다. 따라서, 이러한 원리를 이용하여 구현하게 되면 군집의 생성은 완전히 같은 메시지를 재사용한 경우에만 나타날 것이고, 해당하는 메시지를 탐지 혹은 차단을 수행할 수 있다.

[Algorithm 1]은 이러한 동작 원리를 구현하기 위한 명세이다. 먼저, DBSCAN 알고리즘의 epsilon 반경을 매우 낮은 수치로 설정하는 것과 MinPts를 2로 설정하는 것으로 군집이 기본적으로는 생성될 수 없도록 한다. 이후, 군집화를 수행하고 그 결과에서 군집이 생성된다면, 이를 메시지 재사용 공격으로 간주할 수 있다.

한번 군집을 생성한 이후부터는 패킷 발생 단위로 기존 군집을 기준으로 새로운 패킷의 점(p)를 추가하고 epsilon 반경 내에 다른 점(p)의 존재 여부를 확인하는 것으로 군집화가 수행된다. 이러한 과정을 반복적으로 수행하면, 궁극적으로 패킷 단위의 네트워크 공격 탐지가 수행된다.

Algorithm 1 : DBSCAN-Based Replay Attack Detection

```
1. Input : PCA_Result p[], Eps e, MinPts m=2,
           Array arr
   data_matrix <- as.matrix(p[])
2. dbscan.result <- dbscan(data_matrix, e, m)
   cluster_data <- data.frame(dbscan.result$cluster)
   abnormal <- cluster_data[dbscan.result$cluster != 0 , 1]
3. Output : abnormal
```

Ⅳ. Evaluation

4.1 Evaluation Scenario

본 장에서는 실제 LoRaWAN 환경에서 제안하는 알고리즘을 적용하여 메시지 재사용 공격을 탐지하는 성능평가를 수행하고 그 결과를 분석하고자 한다.

[Fig 10]은 성능평가의 시나리오를 정의한다. Device가 정상적인 Up-Link 메시지를 전송하고, ACK를 수신하는 과정에서 임의로 메모리를 초기화 시킨다. 이후, 메시지 재사용 공격을 수행하고 이를 탐지함에서, 정확도와 소요시간을 평가한다. 실험은 메시지의 수를 점차 증가시키는 방안으로 총 4회를 수행하였다.

[Table 3]은 성능평가가 수행된 환경이다. LoRaWAN 환경에 실제 사용되는 Gateway는 하드웨어 성능이 좋은 편이 아니므로, GPU 연산이 아닌 CPU 연산을 수행한다. 동일한 이유로 낮은 사양의 단말기를 성능평가 환경으로 사용하였다.

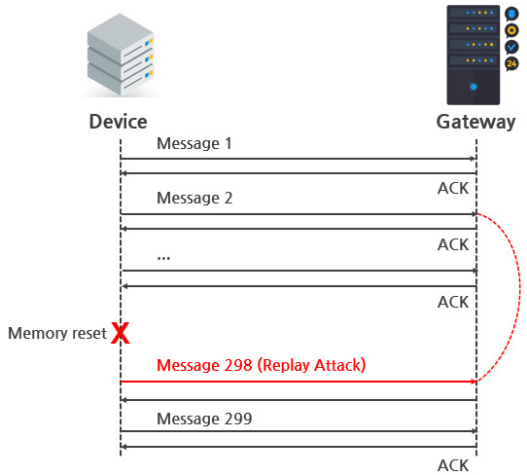

그림 10. 성능평가 시나리오의 예시
Fig. 10. Example of Performance Evaluation Scenario

표 3. 성능평가 환경
Table 3. Performance Evaluation Environment

	Product
CPU	Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz
Memory	8GB
GPU	Intel(R) UHD Graphics 620 (Not Used)

4.2 Results and Discussion

표 4. 평가결과에 대한 요약
Table 4. Results Summary of Evaluation

Case	Number of Messages	Replay Attack	Time Required for ML	Accuracy
1	300	1	0.002007008 sec	100%
2	10,000	5	0.008011818 sec	100%
3	50,000	5	0.03701806 sec	100%
4	450,000	5	0.5061278 sec	100%

4.2.1 Results of Case 1

[Fig 11]은 약 300건의 메시지를 전송하는 과정에서, 1회의 재사용 공격을 수행하였을 때의 성능평가 결과이다. 그림에 생성된 1번 군집이 메시지 재사용 공격을 의미한다. 따라서 정확도는 100%가 되며, 알고리즘을 수행하는 데 0.002007008초가 소요되었다.

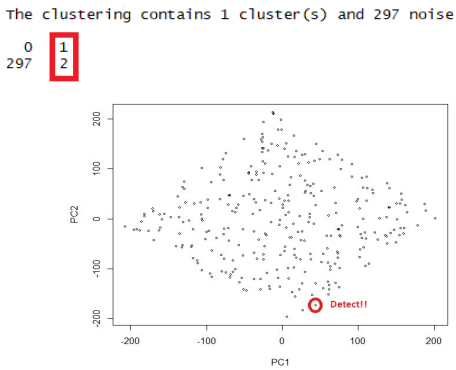

그림 11. Case 1의 군집화 결과
Fig. 11. Cluster Results of Case 1

4.2.2 Results of Case 2

[Fig 12]는 약 10,000건의 메시지를 전송하는 과정에서, 5회의 재사용 공격을 수행하였을 때의 성능평가 결과이다. 그림에 생성된 1번에서 5번 군집은 메시지 재사용 공격을 의미한다. 따라서 정확도는 100%가 되며, 알고리즘을 수행하는 데 0.008011818초가 소요되었다.

4.2.3 Results of Case 3

[Fig 13]은 약 50,000건의 메시지를 전송하는 과정에서, 5회의 재사용 공격을 수행하였을 때의 성능평가 결과이다. 그림에 생성된 1번에서 5번 군집은 메시지 재사용 공격을 의미한다. 따라서 정확도는 100%가 되며, 알고리즘을 수행하는 데 0.03701806초가 소요되었다.

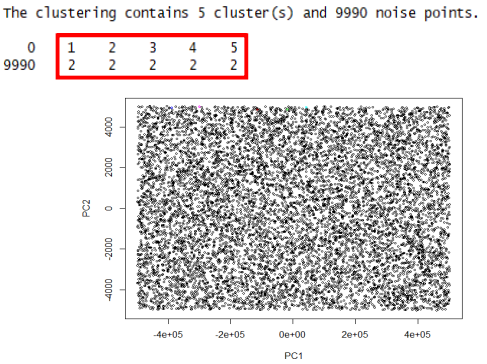

그림 12. Case 2의 군집화 결과
Fig. 12. Cluster Results of Case 2

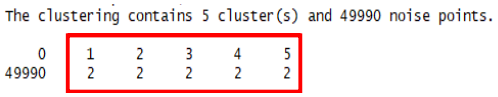

그림 13. Case 3의 군집화 결과
Fig. 13. Cluster Results of Case 3

4.2.4 Results of Case 4

[Fig 14]는 약 450,000건의 메시지를 전송하는 과정에서, 5회의 재사용 공격을 수행하였을 때의 성능평가 결과이다. 그림에 생성된 1번에서 5번 군집은 메시지 재사용 공격을 의미한다. 따라서 정확도는 100%가 되며, 알고리즘을 수행하는 데 0.5061278초가 소요되었다.

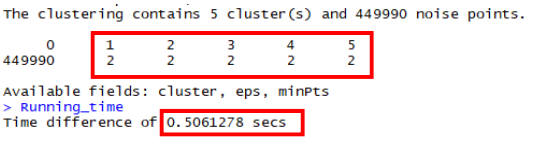

그림 14. Case 4의 군집화 결과
Fig. 14. Cluster Results of Case 4

4.2.5 Additional Comments

제안된 알고리즘은 Feature 선정에 따라 Flooding 형태의 공격에 대한 탐지 가능성도 갖는다. 배터리 소진을 목표로 하는 단순 Flooding 공격의 경우, 유사한 메시지가 연속적으로 발생하는 특징이 있다.

이러한 메커니즘은 DBSCAN의 Eps와 MinPts의 조절을 통해 유사한 메시지를 대상으로 군집화를 수행할 수 있으며, 이를 탐지할 수 있다.

[Fig 15]는 약 300건의 메시지를 전송하는 과정에서, 5회의 유사한 Flooding 공격을 수행하였을 때의 성능평가 결과이다. 그림에 생성된 1번 군집은 이러한 공격의 정확한 탐지를 의미한다. 알고리즘을 수행하는

```
The clustering contains 1 cluster(s) and 295 noise points.

  0   1
295   5

Available fields: cluster, eps, minPts
> Running_time
Time difference of 0.002122164 secs
```

그림 15. 유사한 Flooding 메시지 대상 군집화 결과
Fig. 15. Clustering result of similar flooding messages

데 0.002122164초가 소요되었다.

한번 군집을 생성해두면, 이후부터는 생성된 군집에 1개의 메시지 단위로 점(p)을 추가하기 때문에, 실제 알고리즘에 소요시간은 더욱 짧을 것으로 예상한다.

데이터를 무한히 누적하는 방식보다는 일정한 임계치(Threshold)를 설정하는 방식으로 지연(Latency)을 방지하는 것이 좋다. [Fig 16]은 이러한 방법의 적용을 의미한다. 그림에서는 데이터의 수가 임계치(500,000)에 도달하면 가장 오래된 데이터를 순서대로 제거한다. 이를 통해 연산으로 인한 지연을 통제할 수 있다. 하지만 이러한 방법은 탐지율과 높은 상관관계를 가지기 때문에 반드시 시스템의 성능과 통신의 양 등 많은 변수요인을 고려하여 설정해야 한다.

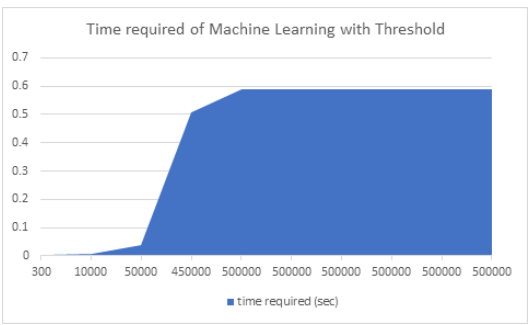

그림 16. 임계값을 적용한 상황에서 머신러닝에 필요한 시간
Fig. 16. Time Required of ML with Threshold

V. Conclusion

본 논문에서는 LoRaWAN 환경에서 발생하는 보안 위협을 DBSCAN 알고리즘을 접목하는 방법으로 위협 탐지에 접근하였다. LoRaWAN은 독자적인 구조의 특성에 의해 일반적으로 IPS/IDS에서 활용하던 보안탐지 알고리즘의 적용이 어렵다. 따라서 제안하는 알고리즘은 머신러닝을 결합한 새로운 접근법을 제시하며, 메시지 재사용 공격과 메시지 Spoofing 공격과 같은 보안 위협을 탐지하는데 높은 성능을 보인다. 실제로 여러 번의 실험에서 100%의 높은 공격 탐지율과 적은 소요시간이 검증되었다. 이러한 높은 성능과 새로운 접근법은 실제 LoRaWAN 환경에 이를 적용하는데 큰 가능성을 제시한다.

References

[1] F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B, Martinez, J. Melià-Seguí, and T. Watteyne, "Understanding the limits of LoRaWAN," *IEEE Commun. Mag.*, Jan. 2017.

[2] D.-Y. Kim and S.-H. Kim. "LoRaWAN technology for internet of things," *J. Platform Technol.*, 2015.

[3] I. You, S. Kwon, G. Choudhary, V. Sharma, and J. T. Seo, "*An enhanced LoRaWAN security protocol for privacy preservation in IoT with a case study on a smart factory-enabled parking system*(2018)," Retrived May 21, 2019, from https://www.ncbi.n lm.nih.gov

[4] *LoRaWAN 1.1 Specification*, LoRa Alliance, Inc.

[5] X. Yang, E. Karampatzaki, C. Doerr, and F. Kuipers, "Security vulnerabilities in LoRaWAN," *Conf. IEEE*, 2018.

[6] L. Ni and S. Jinhang, "The analysis and research of clustering algorithm based on PCA," *13th IEEE Int. Conf. Electronic Measurement & Instruments*, Oct. 2017.

[7] A. Lulli, M. Dell'Amico, P. Michiardi, and L. Ricci, "NG-DBSCAN: Scalable density-based clustering for arbitrary data," *VLDB*, pp. 158-167, New Delhi, India, 2016.

[8] IT Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Philos. Trans. A Math. Phys. Eng. Sci.*, Apr. 2016.

[9] I. Cordova and T. S. Moh, "Dbscan on resilient distributed datasets," *Int. Conf. HPCS*, Amsterdam, Netherlands, 2015.

[10] A. Rahmadhani and F. Kuipers, "*Understanding collisions in a LoRaWAN*(2017)," Retrived May 8, 2019, from https://wiki.surfnet.nl.

[11] B. Reynders, Q Wang, and S Pollin, "A LoRaWAN module for ns-3: Implementation

and evaluation," *WNS3*, pp. 61-68, Surathkal, India, 2018.

[12] I. Butun, N. Pereira, and M. Gidlund, "Security risk analysis of LoRaWAN and future directions," *Future Internet*, 2018.

[13] J.-H. Kim, J.-Y. Choi, and C.-S. Park, "Anomaly detection in networks using big data," *Korea Intell. Inf. Syst. Soc. Assoc. Spring Conf.*, 2017.

[14] G.-S. Kim and I.-R. Jeong, "Practical privacy-preserving DBSCAN clustering over horizontally partitioned data," *Korea Inst. Inf. Secur. and Cryptology*, pp. 105-111, 2010.

[15] ST. Mai, I. Assent, and M. Storgaard, "AnyDBC: An efficient anytime density-based clustering algorithm for very large complex datasets," *KDD*, San Francisco, USA, 2016.

[16] K. M. Kumar and A. R. M. Reddy, "A fast DBSCAN clustering algorithm by accelerating neighbor searching using Groups method," *Pattern Recognition,* vol. 58, pp. 39-48, Oct. 2016.

[17] D.-W. Kim and M.-M. Han, "A data fusion algorithm based on inference for the next-generation intrusion detection," in *Proc. KIIS Spring Conf.*, vol. 26, no. 1, 2016.

[18] J.-W. Lee and K.-C. Kim, "A study on lightweight intrusion prevention system algorithm in long-range wide area network environment," *Korea Telecommun. Assoc. Winter General Presentation*, 2018.

[19] A. Girma, M. Garuba, and R. Goel, "Advanced machine language approach to detect DDoS attack using DBSCAN clustering technology with entropy," *Inf. Technol.-New Generations*, pp. 125-131, Jul. 2017.

[20] K.-H. Kim, J.-H. Park, D.-Y. Hwang, S.-J. Na, and J.-W. Lee, "Counter measure techniques through threat analysis of replay and bit-flipping attacks on LoRaWAN," *J. Inf. Sci.*, vol. 35, no. 1, 2017.

[21] Z. Ghahramani, *Probabilistic machine learning and artificial intelligence*, Nature, May 2015.

[22] C. Szepesvári, *Algorithms for Reinforcement Learning*, Morgan & Claypool Publishers, 2009.

[23] D. M. Raihi, S. Machani, M. Pei, and J. Rydell, "TOTP: Time-based one-time password algorithm," *Internet Eng. Task Force*, May 2011.

[24] A. Lavric and V. Popa, "Internet of things and LoRaTM low-power widearea networks challenges," *9th Int. Conf. ECAI*, Jul. 2017.

[25] G. Thomas, *Mathematics for Machine Learning*, Jan. 2018.

정 택 현 (Tack-hyun Jung)

2018년~현재 : 건국대학교 IT융합정보보호학과 석사과정

<관심분야> 통신공학, 사이버 보안, 기계학습, 소프트웨어 공학

[ORCID:0000-0002-9172-0817]

이 승 호 (Seung-ho Lee)

2019년~현재 : 건국대학교 IT융합정보보호학과 석사과정

<관심분야> 통신공학, 사이버 보안, 기계학습, 소프트웨어 공학

[ORCID:0000-0002-8142-5136]

김 기 천 (Kee-cheon Kim)

1992년 : Northwestern Univ. 공학박사

1998년~현재 : 건국대학교 컴퓨터공학과 교수

<관심분야> 통신공학, 사이버 보안, 미래인터넷, IoT

[ORCID:0000-0003-3445-3334]