한국컴퓨터정보학회 동계학술대회 논문집 제27권 제1호 (2019. 1)

무선 센서 네트워크에서 False-Praise 공격 대응을 위한 합의 알고리즘 기반의 신뢰 메커니즘 연구

서태석[O], 조영호(교신저자)[*]
[O]국방대학교 국방관리대학원 국방과학학과 컴퓨터공학/사이버전협동전공
e-mail: {corneli1202, yhcho94}@gmail.com[O*]

A Trust Mechanism with Consensus Algorithm against False-Praise Attacks in WSNs

Taisuk Suh[O], Youngho Cho[*]
[O*]Dept. of Computer Science and Engineering, Korea National Defense University

● 요 약 ●

무선 센서 네트워크(Wireless Sensor Network)는 낮은 배터리, 짧은 통신거리 등의 제한된 센서들의 성능에 기인하여 내부자공격(insider attacks)에 취약한 것으로 알려져 있는데, 내부자 공격에 대응하기 위한 대표적인 방법으로 노드들의 행위 관찰하여 신뢰도를 평가하고 낮은 신뢰도를 갖는 노드들을 제거하는 신뢰메커니즘(Trust Mechanism: TM)이 있다. TM은 평가노드 자신의 직접관찰 정보뿐만 아니라 이웃노드의 간접관찰 정보를 함께 고려하도록 발전되어 왔는데, False-Praise 공격은 의도적으로 거짓 관찰 정보를 평가노드에게 제공하여 TM의 신뢰도 평가 프로세스의 신뢰성을 훼손하는 지능적 공격이다. 본 논문에서는 False-Praise 공격에 대응을 위한 합의 알고리즘을 기반의 개선된 TM 제안하고, 실험을 통해 제안 체계의 성능과 효과를 검증한다.

키워드: trust mechanism, false-praise attack, consensus algorithm, wireless sensor network

I. 서론

WSN(Wireless Sensor Networks)는 수많은 센서들이 무선 애드혹(Ad Hoc) 방식으로 상호 연결된 후,각 센서가 수집한 데이터 정보를 WSN의 센서들이 협력 중계하여 목적지까지 전송하는 무선 네트워크의 일종이다. 중계 협력 전송이 필요한 이유는 센서들이 제한된 배터리 용량과 그에 따른 짧은 무선 통신거리를 갖는 WSN의 고유한 특성에 기인한다. 이에 따라, WSN의 중계 노드들이 내부공격자인 경우 이들은 전송 중인 패킷에 대해 도청, 변조, 드롭 등 다양한 사이버공격을 통해 WSN의 기능과 성능을 심각하게 훼손하고 파괴할 수 있다[1].

WSN에서의 내부자 공격을 방어하는 대표적인 방법으로 각 센서노드의 행위를 관찰하고 그에 따라 신뢰도를 측정하여 낮은 신뢰도를 갖는 노드를 제거하는 **신뢰메커니즘(Trust Mechanism: TM)**이 있다[2, 3].

일반적으로 TM은 내부공격자 탐지를 위해 **① 이웃노드의 행위 관찰/기록 → ② 신뢰도 평가 → ③ 공격자 탐지**의 3단계로 동작한다. 초기의 TM은 평가 주체가 되는 노드(평가노드)가 피평가노드의 행위를 직접 관찰한 정보로만 신뢰도를 평가하였으나, 이후 보다 정확한 신뢰도 평가를 위해 주변의 이웃노드가 관찰한 정보를 추가로 고려하는 방법으로 발전되어 왔다[3, 6]. 이러한 방법은 이웃노드가 정직할 때는 유효하나, 의도적으로 잘못된 정보를 제공할 때에는 오히려 신뢰도 평가 프로세스에 악영향을 미쳐 TM 본래의 기능과 신뢰성을 훼손하는데, 이러한 공격에는 False-praise 공격과 Slandering 공격이 있다[7].

본 논문에서는 이러한 False-Praise 공격에 대응하기 위한 합의 알고리즘을 설계하고, 이를 기반으로 한 기반의 TM을 제안한다. 또한, 제안된 기법이 기존의 TM에 비해 False-Praise 공격 대응력을 현저히 개선한다는 것을 보이기 위해 간단한 실험결과를 제시한다.

II. 배경지식 및 관련연구

2.1 WSN에서의 내부자공격과 신뢰 기반 방어

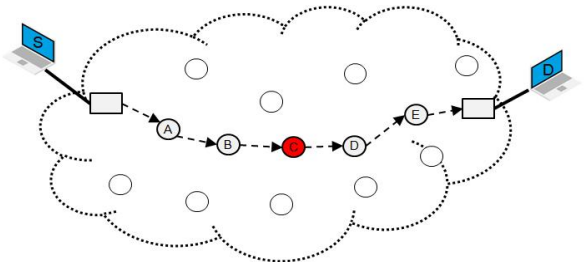

Fig. 1. WSN에서의 정보 전달(소스노드 S → 목적지 D)

WSN은 다수의 무선 센서장치(노드)들이 상호 중계를 통해 소스노드(**S**)가 생성한 데이터패킷을 목적지(**D**)에 전송함으로써 전장감시와

같은 고유의 목적을 달성한다. 센서노드는 일반적으로 값싸고 소형으로 제작되기 때문에 여러 성능 제한이 있어 동료 노드들의 중계 전달이 반드시 필요한 고유 특성이 있으며, 이러한 중간 노드(예를 들면, Fig.1의 노드 **C**)가 내부공격자일 때 패킷 드롭, 패킷 변조, 도청 등 여러 가지 사이버공격을 수행할 수 있다. 또한, 이들은 WSN의 인증된 멤버로서 암호화, 인증체계, 접근제어와 같은 일반적인 보호 체계로는 대응하기 어렵다[1]. 이에 따라, WSN의 내부 노드들의 행위를 관찰하여 신뢰도(trustworthiness)를 측정하는 신뢰모델과 이를 기반으로 내부공격자에 대응하는 신뢰메커니즘(Trust mechanism)이 활발히 연구되고 있다[2, 3].

- **신뢰 메커니즘(Trust Mechanism: TM) 소개**

신뢰 메커니즘은 일반적으로 다음과 같은 3 단계로 동작하여 내부 공격자를 방어한다.

1) 관찰/기록단계 : WSN의 각 노드는 주변 노드의 행위(패킷 전달 행위 등)를 관찰하여 설계된 기능대로 수행하는지의 여부를 기록한다. 이때, 사용되는 대표적인 관찰 메커니즘으로 Watchdog[4]가 있다. Watchdog는 피평가노드의 행위가 정상일 경우에는 성공(**s**)으로, 비정상일경우에는 실패(**f**)로 기록한다.

2) 신뢰도평가단계 : 1단계에서 기록된 행위 이력 정보를 활용하여 피평가노드의 신뢰도를 평가하는데, 신뢰도를 수치로 평가하기 위한 여러 신뢰모델이 제안되었다. 대표적으로 베타신뢰모델[5]이 있는데, 수식 (1)은 **A**가 **B**에 대한 관찰정보인 누적성공회수 as, 누적실패회수 af를 바탕으로 계산한 신뢰값 $T_{A\to B}$이며 [0, 1]사이의 수치로 평가된다. 이때, 1에 가까울수록 신뢰도가 높다는 것을 의미한다.

$$T_{A\to B}(as, af) = \frac{as+1}{as+af+2} \qquad (1)$$

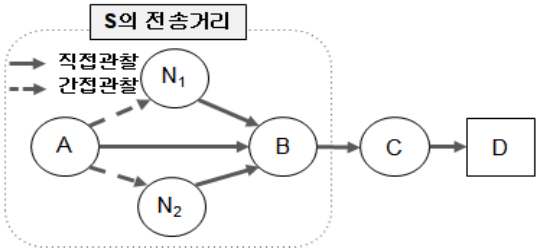

Fig. 2. 직접관찰과 간접관찰을 활용한 신뢰도 평가

한편, Fig.2에서와 같이 평가노드 **A**의 직접관찰 정보 외에 이웃노드들, 즉 N_1과 N_2의 **B**의 행위에 대한 간접관찰정보를 추가로 고려하여 신뢰평가의 신뢰성을 높이려는 연구가 있으며, 이를 Reputation System이라 하며[6], 아래 (2)과 같은 방법으로 최종 신뢰값을 계산할 수 있다; w는 **A**의 측정값에 대한 가중치이며, 우항의 함수 $f(\cdot)$는 활용 가능한 이웃노드들의 간접관찰정보를 활용하여 계산되는 신뢰값이며 이를 계산하는 다양한 방법이 있다.

$$T_{R:A\to B} = wT_{A\to B} + (1-w)f(T_{N1\to B}, T_{N2\to B}, \ldots) \qquad (2)$$

3) 공격자탐지단계 : 마지막 단계에서는 T와 공격자탐지를 위해 설정된 임계치 θ_T를 비교하여, $T < \theta_T$ 일 경우 피평가노드를 내부공격자로 간주하여 네트워크에서 가상적으로 또는 실제로 제거한다.

2.2 TM에 대한 지능적공격: False-Praise 공격

직접과 간접관찰을 활용하는 TM의 동작 방식을 공략하여 본래의 정확한 신뢰도 평가를 하지 못하도록 하는 지능적 공격방법이 있는데, 대표적인 공격이 False-Praise 공격[7]이다. False-Praise 공격은 이름에서 알 수 있듯이 낮은 평가를 받아야 할 내부 노드의 신뢰도를 의도적으로 높이기 위해 거짓으로 관찰정보를 제공하는 것을 말한다. 예를 들어, Fig. 2에서 **B**가 패킷 드롭 공격자이고, N_2가 False-Praise 공격자라 하자. **A**가 패킷을 **B**에게 전송하고, **B**가 **C** 또는 **D**에게 전송하지 않고 드롭한 경우에도 N_2는 **B**가 **A**의 **TM**에 의해 공격자로 탐지되는 것을 방해하기 위해 **B**가 성공적으로 패킷을 전달했다고 정보를 거짓으로 제공한다면, **A**는 **B**의 신뢰도를 정확히 평가하기 어려우며, 조기에 공격자로 탐지하여 제거하지 못할 수 있다. 유사하게, Slandering 공격은 네트워크 내에서 나쁜 의도를 가진 노드들이 특정 노드들의 신뢰도를 낮추기 위해서 잘못된 정보를 보내는 공격을 말한다. 본 연구에서는 False-praise 공격에 대한 대응을 위해 합의 알고리즘 기반의 개선된 TM을 제안한다. 이는 Slandering 공격에도 동일한 대응효과를 갖는다.

III. 합의 알고리즘 기반의 TM 제안

3.1 False-Praise 공격 대응 아이디어: 합의 알고리즘

2장에서 기술한 것과 같이, 기존 TM의 구조에서는 평가노드의 이웃노드 중에서 False-Praise 공격자가 있을 경우 이러한 공격자가 제시한 관찰정보를 의심없이 있는 그대로 수용하여 피평가노드의 신뢰값을 계산하는 과정에서 문제가 발생한다. 따라서, 본 연구에서는 평가노드의 직접관찰 정보와 이웃노드들의 간접관찰 정보를 모두 고려하여 피평가노드의 각 패킷 전달 행위에 대한 **합의 프로세스(consensus process)**를 도입하여 이를 TM과 결합하는 새로운 합의 알고리즘 기반의 TM을 제안한다. 즉, Fig. 2에서 **B**는 패킷 드롭 공격자이고 N_2는 **B**를 돕는 False-Praise 공격자라고 했을 때, 평가노드 **A**는 자신의 직접관찰 정보와 믿을 수 있는 이웃노드 N_1의 간접관찰 정보를 함께 고려하여 **B**의 패킷 전송 행위에 대한 합의를 성공 또는 실패로 결정하여 **B**의 신뢰도를 계산한다면, N_2의 False-Praise 공격 행위를 무력화하고 **B**의 패킷 드롭 공격을 보다 빨리 탐지할 수 있다. 위의 Fig. 3은 이러한 과정을 기존의 TM과 비교하여 예를 들어 설명한 것이다. 일반적인 TM과 합의 알고리즘 기반의 TM-C 모두 세 단계로 구성되나, 후자는 합의 알고리즘을 기반으로 신뢰도 값을 계산한다. 이에 대한 세부 동작설명은 아래에서 기술한다.

3.2 합의 알고리즘 기반의 TM 동작 설명(3단계)

앞에서 설명한대로, 기존의 일반적인 TM은 ① 행위관찰/기록→ ② 신뢰도평가 → ③ 공격탐지의 3단계로 동작하는데, 제안한 합의 알고리즘은 ① 관찰/기록 단계의 변경이 필요하며, 그에 따라 ②

<table>
<tr><th>동작단계</th><th>기존 신뢰메커니즘(TM)</th><th>합의 알고리즘 기반의 신뢰메커니즘(TM-C)</th></tr>
<tr><td>1 관찰/기록</td><td>
<table>
<tr><th colspan="2">t</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th><th>…</th></tr>
<tr><td rowspan="2">직접 관찰</td><td>A</td><td>f</td><td>s</td><td>f</td><td>f</td><td>s</td><td>s</td><td>s</td><td>s</td><td>s</td><td>s</td><td>…</td></tr>
<tr><td rowspan="2">간접 관찰</td><td>N_1</td><td>s</td><td>s</td><td>f</td><td>f</td><td>f</td><td>f</td><td>f</td><td>s</td><td>s</td><td>f</td><td>…</td></tr>
<tr><td>N_2</td><td>s</td><td>s</td><td>s</td><td>f</td><td>f</td><td>s</td><td>f</td><td>f</td><td>f</td><td>s</td><td>…</td></tr>
</table>
</td><td>
<table>
<tr><th colspan="2">t</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th><th>…</th></tr>
<tr><td rowspan="2">직접 관찰</td><td>A</td><td>f</td><td>s</td><td>f</td><td>f</td><td>s</td><td>s</td><td>s</td><td>s</td><td>s</td><td>s</td><td>…</td></tr>
<tr><td rowspan="2">간접 관찰</td><td>N_1</td><td>s</td><td>s</td><td>f</td><td>f</td><td>f</td><td>f</td><td>f</td><td>s</td><td>s</td><td>f</td><td>…</td></tr>
<tr><td>N_2</td><td>s</td><td>s</td><td>s</td><td>f</td><td>f</td><td>s</td><td>f</td><td>f</td><td>f</td><td>s</td><td>…</td></tr>
<tr><td colspan="2">합의결과</td><td>s</td><td>s</td><td>f</td><td>f</td><td>f</td><td>s</td><td>f</td><td>s</td><td>s</td><td>s</td><td>…</td></tr>
</table>
</td></tr>
<tr><td>2 신뢰도평가</td><td>$T_{최종}(A \rightarrow B) = wT_{A \rightarrow B} + (1-w)\left\{\frac{T_{N1 \rightarrow B} + T_{N2 \rightarrow B}}{m}\right\}$
* T는 수식 (1) 로 계산</td><td>$T_{최종}(A \rightarrow B) = \frac{as+1}{as+af+2}$</td></tr>
<tr><td>3 공격탐지</td><td colspan="2">동일 (2장 3) 공격자탐지단계 참조)</td></tr>
</table>

Fig. 3. 기존 TM과 합의 알고리즘 기반의 TM(TM-C)의 동작 비교

신뢰도평가 단계에 변화를 준다. 지금부터 합의 알고리즘 기반 TM의 3단계 동작을 설명하며, 일반적인 TM의 동작단계에 대비하여 변경된 부분을 위주로 설명한다.

• [1단계] 행위관찰/기록(합의 알고리즘) : Fig. 3의 관찰/기록단계에서 비교한대로, 기존의 TM은 평가노드 A의 직접관찰결과는 기록하여 저장하고, 이웃노드 N_1과 N_2로부터 간접관찰 정보를 제공 받는다. 반면, **TM-C**에서는 직접관찰결과와 간접관찰결과를 합의과정을 통해 성공(s) 또는 실패(f)로 확정하고 그 결과에 따라 2단계에서 신뢰값을 계산한다. 다음은 majority voting 방식을 고려한 합의 알고리즘이다. 이웃노드(간접관찰노드)의 수가 n 이고, 직접관찰에 대한 가중치가 w일 때, 합의를 위해 판별식 **(3)**의결과를 구한다.

$$D = wO_{직접} + \frac{(1-w)}{n}\sum_{i=1}^{n} O_{간접(N_i)} \quad (3)$$

이때, $O_{직접}$과 $O_{간접(Ni)}$은 각각 직접관찰결과와 이웃노드 i의 간접관찰결과를 뜻하며, 성공(s)은 1, 실패(f)는 -1로 계산한다. D값이 구해지면 합의결과 C는 다음 수식 **(4)**에 따라 최종적으로 결정된다.

$$C = \begin{cases} s & \text{for } D \geq 0 \\ f & \text{for } D < 0 \end{cases} \quad (4)$$

• [2단계] 신뢰도평가단계 : TM-C는 기존 TM과 다르게 1단계에서 합의한 관찰결과를 바탕으로 신뢰도값을 계산하기 때문에, (1)과 같은 신뢰도 모델의 수식으로 평가가 가능하며, Fig. 3의 2단계인 신뢰도평가는 기존 TM에 비해 간소화된다.

• [3단계] 공격탐지단계 : 기존 TM모델과 동일하게 수행된다(2장 **3) 공격자탐지단계** 참조).

Ⅳ. 제안 알고리즘 성능실험 및 분석

4.1 실험목적 및 방법

본 실험의 목적은 제안한 합의 알고리즘 기반의 TM이 False-Praise 공격에 대해 기존의 알고리즘 보다 더 좋은 대응 능력을 보이는 것이다. 시뮬레이션 실험을 위해 노트북 PC(CPU 1.6GHZ, RAM 4GB)에서 Python 3.7로 3장과 아래에 기술된 네트워크 모델, 공격모델, TM, 합의 알고리즘 등을 구현하였다.

• **네트워크 모델 :** Fig.2와 같이 4개 노드(**A, B, N_1, N_2**)로 구성한다. **A**는 평가노드, **B**는 피평가노드, N_1과 N_2는 이웃노드임. **A**는 자신의 패킷을 목적지로 전송하기 위해 **B**에게 전달(중계 요청)하고, 이웃노드인 N_1과 N_2는 **A**의 패킷에 대한 **B**의 패킷 전달행위를 관찰할 수 있다.

• **공격모델 :** 노드 **B**와 N_2는 내부공격자이며, 상호 협력할 수 있다. **B**는 **A**가 자신에게 전송한 패킷

중 70%를 확률적으로 드롭하는 패킷 드롭 공격(Grayhole 공격)을 수행하고, N_2는 **B**가 패킷 드롭을 수행하는 경우에도 **A**에게 **B**가 모두 성공적으로 전달했다고 알리는 False-Praise공격을 수행한다.

• **TM 및 합의 알고리즘 :** TM은 수식 (1)의 베타신뢰모델과 수식(2)의 Reputation 체계를 적용한다. 기존 TM과 합의 알고리즘 기반의 TM을 구현하고, 각각 TM_{Beta}와 TM_{Beta}-C로 명칭한다. **B**의 초기 신뢰값은 0.99로, 공격탐지를 위한 θ_T 는 0.3~0.8 사이의 값을 사용하며 가중치 w는 0.7을 사용하였다.

실험은 평가노드인 A가 패킷드롭공격자 B를 **TM_{Beta}**와 **TM_{Beta}-C**로 얼마나 빨리 탐지하는지를 측정하여 비교하고, 탐지가 느린 쪽을 기준으로 실험을 종료한다. 실험 회수는 총 100회를 실시하여 평균값을 계산하고 그 결과로 탐지성능을 분석하였다.

4.2 실험결과 분석

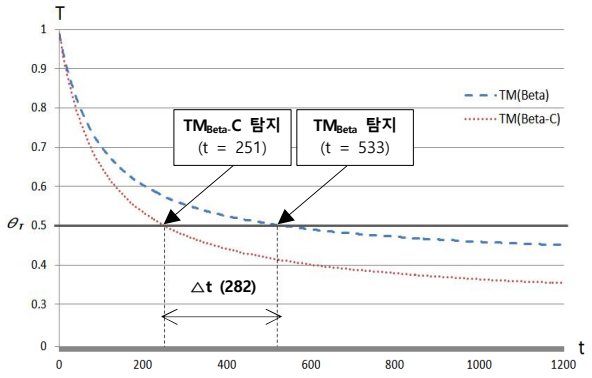

Fig. 4. 공격 탐지 속도 비교 (θ_T = 0.5)

Table 1. 공격탐지능력 비교(TM_{Beta} vs. TM_{Beta}-C)

θ_T	공격탐지소요시간(t)		탐지성능 비교	
	TM_{Beta}	TM_{Beta}-C	단축(△t)	개선(%)
0.8	50	40	10	25
0.7	102	77	25	32
0.6	208	134	74	55
0.5	533	251	282	156
0.4	미탐지	607	비교불가	-
0.3	미탐지	미탐지	-	-

• **결과 1 : False-Praise 공격 하에서 합의 알고리즘 기반의 TM_{Beta}-C이 기존의 TM_{Beta} 보다 패킷 드롭 공격을 빠르게 탐지한다.** Table 1에서 볼 수 있듯이, 다양한 θ_T 에 따른 공격탐지소요시간의 측정결과를 보면 TM_{Beta}-C이 TM_{Beta} 보다 최소 25% ~ 156% 이상의 탐지속도의 개선효과를 보였다. 특히, θ_T = 0.4 인 경우 TM_{Beta}-C만 공격을 탐지할 수 있었다. 이는 합의 알고리즘이 False-Praise 공격에 잘 대응하여 패킷드롭공격자의 신뢰도를 정확히 평가한 결과이다. 참고로, Fig.4은 θ_T = 0.5일 때 TM_{Beta}와 TM_{Beta}-C의 공격탐지시점과 성능차이를 보여준다.

• **결과 2 : 탐지 임계치(θ_T) 값이 작을수록 TM_{Beta}-C 의 False-Praise 공격에 대한 대응 효과가 높았다.**

또한, θ_T = 0.3인 경우에서와 같이, 너무 낮은 θ_T 값을 사용하면 공격자의 패킷 드롭율에 따라 공격을 탐지하지 못할 수 있음을 알 수 있다.

V. 결론 및 향후 연구계획

본 논문에서는 WSN에서 False-Praise 공격에 대한 대응을 위해 합의 알고리즘 기반의 신뢰메커니즘(TM)을 제안하고, 실험을 통해 공격탐지 성능이 개선됨을 보였다. 향후에는 다양한 실험을 통해 알고리즘의 성능을 개선하고, TM 운용 과정에서 생성되는 정보들을 블록체인(Blockchain)에 저장하여 센서들이 공유 활용토록 하여 WSN 보안을 강화하도록 하는 연구를 수행할 것이다.

REFERENCES

[1] E. Shi, and A. Perrig, "Designing secure sensor networks." IEEE Wireless Communications 11.6, pp.38-43, 2004.

[2] Y. Yu., et al. "Trust mechanisms in wireless sensor networks: Attack analysis and countermeasures." Journal of Network and computer Applications 35.3, pp. 867-880, 2012.

[3] R. Román., et al. "Trust and reputation systems for wireless sensor networks." Security and Privacy in Mobile and Wireless Networking, pp. 105-128, 2009.

[4] M. Sergio et al. "Mitigating routing misbehavior in mobile ad hoc networks," Mobicom, ACM, 2000.

[5] A. Josang, and R. Ismail, "The beta reputation system", In Proceedings of the 15th bled electronic commerce conference, Vol. 5, pp. 2502-2511, June. 2002.

[6] Y. Cho., et al. "Insider threats against trust mechanism with watchdog and defending approaches in wireless sensor networks." Security and privacy workshops (SPW), 2012 IEEE Symposium on. IEEE, pp. 134-141, 2012.

[7] C. Esposito et al "Information theoretic-based detection and removal of slander and/or false-praise attacks for robust trust management with Dempster-Shafer combination of linguistic fuzzy terms." Concurrency and Computation: Practice and Experience 30.3 : e4302, 2018.