무선 센서 네트워크에서 안전한 Top-k 질의 처리 기법*

이명수, 심규선, 박상현, 이상근
고려대학교 정보통신대학 컴퓨터통신공학부
e-mail:{lms9711, bluesks, condols, yalphy}@korea.ac.kr

Secure Top-k Query Processing in Wireless Sensor Networks

Myong-Soo Lee, Kyu-Sun Shim, Sang-Hyun Park, SangKeun Lee
Division of Computer and Communication Engineering, Korea University

요 약

무선 센서 네트워크에서 데이터 전송은 에너지를 소비하는 주 요인이다. 에너지를 줄이기 위한 주 기법 중 하나가 센서 데이터를 애그리게이션하여 전송할 데이터를 줄이는 것이다. 무선 센서 네트워크는 공개된 공간에서 진행되어 외부 공격에 취약점을 가지고 있으며, 기존 몇몇의 연구에서 애그리게이션 기법의 보안 기술을 제안하고 있다. 하지만, 기존의 기술들은 특정 연산자에만 제한되어 있고, 많은 유용성을 가진 top-k 질의에 대해서는 취약점을 가지고 있다. 본 논문에서는 무선 센서 네트워크에서 에너지 효율을 위한 애그리게이션 기법 적용시 top-k 질의를 효율적으로 처리할 보안 기법에 대해 제안한다.

1. 서론

신호 처리, 무선 통신, 무선 기기의 발달에 따라 무선 센서 네트워크는 많은 응용 어플리케이션과 함께 주목을 받고 있다. 무선 센서 네트워크는 일반적으로 하나의 싱크(Sink)과 여러 개의 센서 노드들로 구성되어지는데, 싱크은 여러 센서 노드들로부터 데이터를 전송받는다. 각각의 센서 노드들은 제한된 자원과 배터리로 인해서 에너지 효율을 극대화할 필요가 있다.

센서 네트워크에서 에너지 효율을 극대화하는 수단으로 대표적으로 인 네트워크 애그리케이션 기법 (In-network Aggregation)이 있다 [2]. 이 기법은 각각의 센서에서 생성된 데이터를 연산자(예: 평균, 합계, 최대, 최소, Top-k)를 통해 데이터를 압축하여 베이스 스테이션에 제공하여 전송되는 전체 데이터의 총량을 줄여 에너지 사용량을 최소화한다.

또한, 센서 네트워크에서는 외부 공간에 공개되어 외부자에 의한 공격에 취약점을 가지고 있다. 이에 많은 보안 기법들이 제안되어 왔으며, 특히 애그리게이션 기법을 적용한 보안 기법들이 제안되어 왔다 [1][2]. 하지만, 기존 기법들은 특정 연산자만 적용가능하게 디자인되어 왔으며, 특히 Top-k [4]와 같은 연산자는 거의 처리하지 못하는 단점이 있다.

본 논문에서는 무선 센서 네트워크에서 인네트워크 애그리게이션을 이용한 안전한 top-k 질의 처리 기법에 대해 제안한다. 본 기법은 수동적인 공격에 방어하기 위해 센서 식별자(ID)를 암호화하여 센서의 정보를 은닉화하여 데이터의 기밀성을 유지한다.

* 이 연구에 참여한 연구자의 일부는 '2단계 BK21사업'의 지원비를 받았음.

2장에서는 제안하는 알고리즘의 주 내용을 설명하고, 3장에서는 제안하는 알고리즘의 보안 분석을 수행하며, 4장에서는 결론을 내린다.

2. 안전한 Top-k 질의 처리 기법

무선 센서 네트워크에 접근하는 수동적인 공격자는 센서 중 일부에서 전송되는 데이터를 도청하여 정보를 얻어낸다. 이는 데이터 기밀성을 해치는 행위이며 이를 방지하기 위해 보통 데이터를 암호화한다. 하지만, 무선 센서 네트워크에서는 데이터보다 데이터를 보내는 주체를 숨겨야 할 경우가 많다. 이를 테면, 각 가정의 수도량을 조사하는데 수도량 데이터보다는 각 수도량의 정보의 주체가 누구인지를 숨기는게 개인의 사생활 보호에 더 중요하다.

이러한 데이터의 특징 때문에, 본 논문에서는 전송되는 데이터의 주체인 센서 식별자를 암호화한다. 식별자를 암호화하기 위해 RSA 암호화 기법[3]을 사용한다. 센서 노드 모두는 같은 공개키를 가지고 싱크는 개인키를 가진다. 암호화 함수는 $Enc(id) = id^e = c(\mathrm{mod}\, n)$이고 대응하는 복호화 함수는 $Dec(c) = c^d = id(\mathrm{mod}\, n)$이다. 암호화는 각각의 센서에서 이루어지고 복호화는 싱크에서만 이루어진다. 그림 1은 본 기법의 알고리즘을 설명하고 있다. 각각의 센서에서 전송하는 데이터는 암호화된 식별자, 각각의 센싱 데이터, 그리고 데이터 생성 시간으로 이루어진다. Top-k에서의 애그리게이션 함수는 데이터의 수를 치 최대 k개만 유지하도록 합친다.

그림 1은 알고리즘의 실행 예제이다. 본 예제에서는 7개의 센서 노드와 하나의 싱크가 있다. 본 예제의 질의는 "가장 큰 온도를 감지한 두 개의 노드를 찾아라"와 같은 Top-k 질의[4]이다. 여기서 k는 두 개의 노드를 뜻하며,

파라미터:
결과 값에 포함될 데이터의 갯수 **k**,
센서 식별자 ***id***,
센서의 메시지 ***m***, 센싱 시각 ***t***,
개인키 정수 ***d***, 공개키 정수 ***e***,
$n = p*q,\ e*d = 1(\mathrm{mod}(p-1)(q-1)$
암호화:
$Enc(id) = id^{e} = c\ (\mathrm{mod}\ n)$
복호화:
$Dec(c) = c^{d} = id\ (\mathrm{mod}\ n)$
전송할 데이터의 메시지:
$Mes(\{c_1 \ldots c_k\}, \{m_1 \ldots, m_k\}, t)$
애그리게이션 함수:
$Mes1(\{c_1 \ldots c_k\}, \{m_1 \ldots, m_k\}, t) + Mes1(\{c_1 \ldots c_k\}, \{m_1 \ldots, m_k\}, t) + \ldots\ldots = Mes_a(\{c_1 \ldots c_k\}, \{m_1 \ldots, m_k\}, t)$

(그림 1) 식별자 암호화 알고리즘

본 그림에서 보는 바와 같이 Top-k 질의의 결과값은 가장 높은 온도인 25℃와 26℃를 가진 2번, 8번 노드이다.

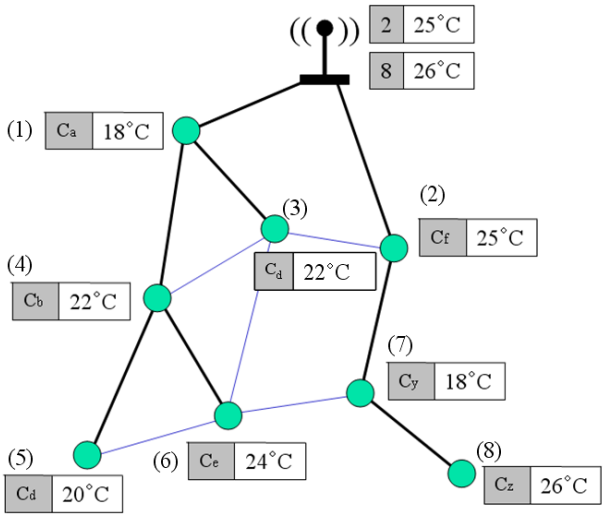
이와 같은 Top-k 질의를 처리하는데 중간노드, 1번이 외부자에 의해 공격당해 포획되었다고 생각해보면, 외부자는 1번 노드 뿐만 아니라 1번 노드의 하위노드의 정보까지 알 수 있다. 정확히는 4번 노드가 애그리게이션한 정보인 4, 5번 노드의 데이터를 보내고, 3번 노드가 자신의 정보를 보내 총 3, 4, 5번 노드의 정보를 모두 알 수 있다.

본 논문에서 제안한 알고리즘을 통해 위의 예제에서 데이터 기밀성을 보장할 수 있다. 즉, 1번 노드가 외부자에 의해 포획되었다고 하여도 각 데이터의 정보는 공개되지만, 각 데이터의 주체가 모두 암호화 되어 공격자는 온도의 정보는 알 수 있으나 암호화된 식별자값 C_i를 가지게 되어 어디서 어떤 온도를 가질 수 있는지 알 수 없다. 물론 4번의 노드가 포획되었을 경우에는 하위 노드에 대한 정보가 4번 노드가 가지고 있고 보내는 정보는 자신의 정보이므로 이 경우에는 공격자가 4, 5, 6번 노드의 정보를 모두 알 수 있다. 하지만, 그 외의 경우 많은 정보를 은닉화함으로서 공격의 피해를 최소화할 수 있다.

3. 보안 분석

본 장에서는 제안한 알고리즘이 데이터의 기밀성을 유지하는데 얼마나 효율적인지 분석해 본다. RSA 암호화 기법을 사용한 알고리즘에서 복호화 할 수 있는 복호화 개인키가 없이 암호를 풀 수 있는 방법은 거의 없음을 많은 논문에서 제시하고 있다. 하여 본 알고리즘에서 정보의 기밀성의 가장 큰 문제는 잠식당한 노드가 하위 트리의 구조와 정보를 알고 있음으로서 전달받은 정보를 통해 암호화한 정보를 유추할 수 있는 경우이다. 앞 장에서도 설명한 바와 같이, 그림 2에서 4번 노드가 잠식당할 경우 공격자는 4, 5, 6번 노드의 정보와 주체를 모두 알 수 있다.

앞에서 언급한 바와 같이 정보를 유출할 가능성에 대해 알아보기 위해 다음과 같이 센서 네트워크의 분석을 수행한다. 트리의 종류가 m원 트리라고 할 때, 센서 노드의 총 개수는 $\sum_{i=0}^{h-1} m^i = \frac{m^h - 1}{m-1}$이다. 이때 h는 트리의 높이



(그림 2) Top-k 질의 실행 예제

이며, 또한, 단말 노드의 개수는 m^{h-1}이며, 단말 노드의 부모 노드의 개수는 m^{h-2}이다. 앞에서 언급한 정보 누출은 단말의 부모노드가 포획되었을 경우만 가능하며, 정보가 누출될 확률은 다음과 같다.

$$P(x) = \frac{(m-1)(m^{h-2}-1)}{m^h - 1} \qquad (1)$$

(1)의 식에서 m이 5이고 h가 3인 총 노드의 개수가 124개인 센서 노드에서는 정보가 누출될 확률은 12.9%에 불가하다. 또한 각 노드에서 공격자가 얻을 수 있는 정보는 총 데이터 중 자식 노드의 정보만을 포획할 수 있으므로 실제 원하는 데이터를 얻을 수 있는 확률은 더 감소한다.

4. 결론

본 논문에서는 무선 센서 네트워크에서 Top-k 질의 처리시 데이터를 기밀성있게 유지할 수 있는 기법에 대해 제안하였다. 본 기법은 데이터 자체보다 데이터의 주체에 대한 암호화를 수행함으로서 센서 네트워크 데이터 특정에 특화된 기법을 제공하였다. 트리 구조의 특성상 네트워크 위상에 따른 데이터의 누수가 일정 부분 발생하지만 그 정보 누출량이 많지 않음을 보안 분석을 통해 알아보았다.

참고문헌

[1] A. Perrig, J. Stankovic, and D. Wagner, “Security in wireless sensor networks,” Communications of the ACM, vol.47, pp.53~57, 2004
[2] B. Przydatek, D. Song, and A. Perrig, “SIA: Secure Information Aggregation in Sensor Networks,” ACM SENSYS, pp.255~265, 2003.
[3] R. L. Rivest, A. Shamir, and L. M. Adleman, “A Method for Obtaining Digital Signatures and Public-Key Cryptosystems,” Communications of the ACM, vol.21, pp.120~126, 1978.
[4] M. Wu, J. Xu, X. Tang, and W.-C. Lee, “Monitoring Top-k Query in Wireless Sensor Networks,” The IEEE International Conference on Data Engineering, pp.143~145, 2006.