情 報 保 護 學 會 誌
第 16 卷 第 4 號, 2006. 8

효율적 센서 데이터 수집 전략과 비정상 데이터 검출에 관한 연구

손 태 식*, 최 욱*

요 약

센서 네트워크는 네트워크 특성상 근본적으로 기존의 네트워크와 다른 많은 제약 사항을 가지고 있다. 이러한 제약 사항으로는 대량의 센서를 위한 비용 문제, 센서 자체의 물리적 취약성 문제 그리고 센서가 취합하는 데이터의 중요도에 따른 보안성 문제 등이 제기될 수 있다. 특히, 본 논문에서는 다양한 센서 네트워크의 기술 이슈 중에서 센서 네트워크의 특정 애플리케이션 지향적 정보 습득 특성에 초점을 맞추었다. 이때 센서 네트워크에서 빼놓을 수 없는 전력 소비 문제가 함께 고려된 센서 네트워크의 효율적인 데이터 수집을 위한 클러스터 기반 지연 적응적 전략과 커버리지 적응적 전략을 소개하였다. 또한 이러한 데이터 습득 과정에서 발생할 수 있는 이상 데이터에 대한 검출 문제를 제시하고 그 대응방안으로서 K-means clustering을 사용한 비교사 학습 기반 방식을 제안하였다.

Ⅰ. 서 론

2004년에 발표된 정통부의 IT839 정책에서 3대 인프라였던 센서 네트워크 기술이 2006년 2월에 발표된 u-IT839에서도 역시 3대 첨단 인프라의 하나로서 포함되어 발표 되었다. 이것은 현재 그리고 앞으로의 정보통신환경이 IP망을 기본으로 센서 네트워크의 특성을 반영하는 핵심 센서 네트워크 기술(location awareness, self-organization, multi-hop routing, clustering, data gathering, data aggregation) 및 다양한 응용서비스를 발굴 개발하는 방향으로 진행될 것임을 암시하고 있는 것이다. 센서 네트워크 환경에서는 대량의 센서로부터의 효율적 정보 수집(data gathering/aggregation)과 원활한 통신 보장(reliable data delivery)이 가장 큰 이슈로서 제기 되며 이때 얼마만큼 적은 전력(power consumption)으로 운용될 수 있는가 하는 것 또한 많은 연구자들 사이에서 하나의 초점이 되고 있다. 이러한 센서 네트워크에 관한 연구에서 필수적으로 고려해야하는 부분 중의 하나가 또한 보안 분야이다. 센서 네트워크의 특성상 센서 네트워크의 센서들은 기본적으로 인간과 환경에 밀접한 데이터를 수집하여 군사 환경, 개인생활, 보건, 재난방재 등의 다양한 응용 서비스를 제공하기 때문에 개인정보보호에 관한 사생활 노출 문제나 악의의 사용자들로부터 부적절한 용도로 유용될 수 있는 등의 문제점을 가질 수 있다. 이외에도 센서 네트워크는 기존의 네트워크 환경과 달리 다양한 보안 취약성 (물리적 취약성, 낮은 성능)을 내포하므로 보안의 필요성은 더더욱 중요해진다.

본 논문의 2장에서는 먼저 현재의 센서 네트워크 기술 및 그 동향에서 대해서 알아본다. 또한 센서 네트워크 보안에 관련된 어떠한 이슈들이 제기되고 있는지도 함께 살펴볼 것이다. 3장에서는 센서 네트워크의 중요한 여러 기술 중에서도 센서 데이터에 관련된 수집/병합(data gathering/aggregation)에 관련된 기술을 살펴본 후 이러한 데이터 수집 과정에서의 보안 이슈와 비정상 센서 데이터 처리 방안을 제시한다. 4장은 본 논문의 결론을 서술한다.

Ⅱ. 센서 네트워크 기술 및 보안 동향

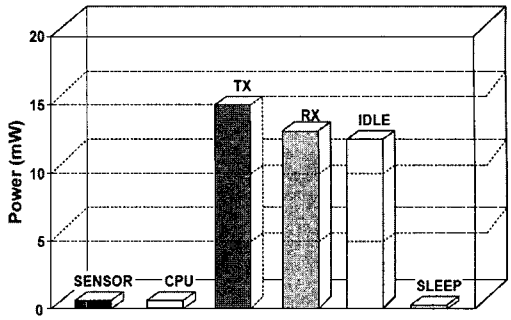
1. 센서 네트워크 기술 동향

Micro Electro Mechanical Systems (MEMS)

* IP Lab, Telecom. R&D Center, SEC(ts.shon, to.choi@samsung.com)

기술의 진보와 단거리 무선통신 기술의 발전이 다기능 소형 센서의 출현을 가능케 하였다. 이러한 센서들은 데이타 처리부, 무선 통신부, 및 (멀티) 센싱 유닛으로 구성되어 서로 간의 자치적인 협동을 통한 최적의 네트워크를 구성하기도 하며 또한 센서 데이타 공조 처리를 통해 특정 상황에 대처하기 위한 적절한 action을 trigger시키는 actuator서의 기능도 수행 할 수도 있다. 무선 센서 네트워크는 대역폭, 에너지, 저장 공간, 및 계산/통신 역량 등에서 높은 제한성과 네트워크를 구성하는 노드(센서)들의 단위면적당 밀도와 그 규모(네트워크의 크기)가 높다는 특징으로 인해 기존 네트워크와는 구별 된다 [1]. 일반적으로 센서 네트워크는 인간의 감각기관을 대신하여 접근 및 활동이 용이하지 않은 실내외의 장소에 환경/장비 monitoring, smart home, 침입탐지, 감시, 우주 탐사 등을 위한 광범위한 분야에서 사용되어진다.

배터리 파워에 의해 작동되는 센서들은 장기간 동안 주어진 task를 수행하여야 하는 것이 센서 네트워크의 기본적인 특성이며 더욱이 센서가 특정 지역에 대규모로 전개 될 경우 배터리의 재충전은 환경적 상황/비용/실용성 등의 문제로 가능하지 않은 것이 일반적이다. 이를 고려해 볼 때 극도로 제한된 에너지 자원의 효율적 활용은 알고리즘이나 프로토콜 설계시 최대의 network lifetime을 얻기 위해 고려되어져야 하는 매우 중요한 요소 중의 하나이다. 센서의 행위 중에서 무선 데이터 전송이 에너지 소비가 가장 높다 (그림 1 참조). 이러한 문제는 높은 밀도로 전개된 센서들의 중복된 센싱 데이터의 전송, 여러 센서의 동시 전송에 의한 신호간섭 등의 요소로 인해 상대적으로 높은 재전송율 등에 의해 불필요한 에너지의 소비를 더욱 높여 네트워크 lifetime에 크게 영향을 미친다. 그러므로 중복된 데이터의 제거/전송전 데이터 통합 (aggregation)을 통해 데이터 전송 빈도를 줄이는 것이 불필요한 에너지 자원 사용의 줄여 많은 양의 에너지를 절약 할 수 있게 된다. 중복된 데이터의 제거/통합시 데이터의 정확성을 잃지 않는 것이 무엇보다도 중요하며 이를 고려한 많은 데이터 gathering (processing) 프로토콜이 라우팅 및 topology control의 견지에서 많이 디자인 되어 오고 있다 [2-5]. 이러한 데이터 gathering은 단순 데이터 collection에서 점차 상황 인식 (context-aware)을 통한 추론적 데이터 collection을 통해 좀더 효율적인 e데이터 gathering 기술로 발전하고 있다.



[그림 1] 각 센서 state에서의 에너지 소비량 비교 (전송 범위 = 20m, data rate = 2.4Kbps)[2]

2. 센서 네트워크 보안 이슈

앞서 언급했듯이 센서 네트워크는 일반적인 네트워크 기능을 포함하는 작은 센서들과 액추에이터로 조합된 이종 시스템들로 구성된 네트워크라고 할 수 있다. 이것은 센서 네트워크가 전형적인 네트워크 시스템들과 비교하여 제한된 메모리와 저장 공간, 전력 제약, 네트워크 지연, 물리적 공격에 노출 등의 다양한 제약 조건을 가진 특별한 형태의 네트워크임을 의미한다. 이러한 제약 조건 때문에, 기존의 네트워크 시스템에 적용되던 보안 기술들을 그대로 센서 네트워크에 적용하는 것은 쉽지 않다. 그러므로 센서 네트워크가 가지는 다양한 제약 조건을 파악하고 그 제약 조건들이 어떠한 보안 이슈를 가지고 있는지 알아보는 것이 필요하다.

먼저, 센서 네트워크에서의 제한된 자원 사용 문제는 효율적인 보안 서비스를 제공하는데 어려움을 준다. 일반적으로 보안 관련된 기법들은 기본적인 네트워크 연산보다 보다 많은 컴퓨팅 파워를 요구하기 때문에 센서 네트워크의 제한된 메모리 크기, 저전력, 코드 사용 공간 등은 센서 네트워크의 보안을 매우 취약하게 만드는 결정적인 요소 중의 하나이다.

또한 대규모의 센서 네트워크 구축하는 경우 이질적인 환경에 대량의 센서가 전개되기 때문에 센서간 신뢰성 있는 통신을 보장, 채널 충돌 최소화, 지연 시간 최소화 등의 문제를 해결하는 것도 센서 네트워크의 보안을 위해서 중요한 요소이다. 만약 센서 네트워크에서 노드간 통신의 신뢰성을 보장할 수 없다면 네트워크 보안 서비스 중 가장 기본적인 암호화 통신을 위한 사전 키분배, 보안 속성 교환, 암호화 씨드(seed) 값 전송 등에 정확성을 제공하기 어렵다.

세 번째로 센서 네트워크의 하드웨어적 특성에 기인

하는 문제가 보안에 영향을 주는 요소로서 대량 센서 관리의 어려움, 물리적 공격에 취약한 환경에 노출, 접근성 부족 등을 들 수 있다. 센서 네트워크는 좁은 실내의 편리성을 제공하는 것은 물론 광역의 환경에 대량으로 전개되어 재난 방지, 군사용등으로도 응용 될 수 있다. 그러므로 후자와 같은 대량의 센서들을 관리의 어려움은 바로 낮은 보안성이라는 문제로 직결 될 수 있다. 또한 광역의 센서 전개의 경우 접근이 어렵다는 문제 역시 보안에 치명적이다. 또한 외부 환경에 쉽게 노출되어 물리적 공격에 취약한 점도 기존의 PC기반 네트워크 환경에 비할 때 보안 위협요소를 증가시켰다.

그러므로 위와 같은 센서 네트워크의 고유 특성과 그 보안 이슈들은 기존의 네트워크 환경에 비해 센서 네트워크는 보다 신뢰성 있고 이상적인 보안 서비스를 필요로 함을 알 수 있다. 그러한 보안 서비스를 충족시키기 위한 보안 요구사항으로는 기본적으로 데이터 기밀성 보장, 데이터 인증, 데이터 무결성 보장, 데이터 가용성 보장, 데이터 신선도(freshness) 보장, 생존성/강건성 제공 등이 있다. 하지만, 전형적인 네트워크 환경에 적용되던 보안 요구사항 형태에서 벗어나 센서 고유 특성에 적합한 요구사항을 수립하는 것이 필요하다. 먼저 데이터 기밀성과 무결성 그리고 인증 요구사항의 경우 센서 네트워크의 다양한 이질성과 제한된 자원 문제를 고려한 효율적이고 경량화된 암호학적 기법들이 제공 되야 한다. 물론 이러한 경량화 된 암호학적 기법은 단순히 빠르고 효율적인 암/복호화 외에도 적은 키 크기를 가지고 안전하게 분배할 수 있는 키 교환을 비롯해 암호화 전반에 필수적으로 요구된다. 또한 센서 네트워크 환경에서 센서들의 생존성(survivability)과 네트워크 자체의 강건성(robustness) 요구사항을 만족시키기 위해서는 물리적으로 취약한 센서 노드의 문제를 개선하는 것이 필요하며 하나의 센서 노드의 취약성이 전체 센서 네트워크에 보안에 영향을 미치지 않게 구성하는 것 또한 중요하다[7-9].

Ⅲ. 센서 데이터 수집 및 보안 연구

1. 데이터 습득(Data Gathering)

센서 네트워크를 설치하는 주된 목적은 특정상황에서 사물이나 현상을 관찰/감시하고 그것에 관한 정보를 수집하기 위함 이다. 그래서 센서들은 끊임없이 주어진 task특성에 기반을 둔 특정 스케줄에 따라 주위를 센싱하고 얻어진 데이터들을 데이타 sink (i.e., gathering point)까지 전달한다. 데이터 습득 문제는 센서를 얼마만큼 선택하서 얼마나 효율적으로 그 선택된 센서로부터 데이터를 수집하는 것이다. 데이터 sink까지 직접통신이 불가능할 경우 멀티 홉 라우팅 경로를 설정하여 데이터를 sink까지 전달하게 된다. 이렇게 데이터를 sink까지 전달하는 동안 데이터 aggregation을 통해 자원의 사용을 최소화하기 위해 센서간 협동이 필요하며 이를 위해 data-centric 라우팅을 이용하기도 한다. 데이터 reporting의 빈도는 위에서 논의한 바와 같이 센싱 task 특성에 의해 결정되어지며 연속형 (continuous), 주문형 (on-demand), event-driven 및 이들의 hybrid형으로 구분되어진다[6]. 연속형 데이터 reporting은 센서들이 어느 특정 주기로 계속해서 데이터를 전송하는 형태를 의미하며 event-driven은 센서가 어떠한 특정한 event가 감지되었을 때에 한해서만 데이터를 전송하는 형태이다. 주문형은 사용자의 요구가 있을 경우에만 센서가 데이터를 전송하는 것이다. 위와 같은 데이터 delivery 모델을 요약하면 아래의 표 1과 같다.

[표 1] 데이터 delivery 모델

Type	Reporting Type	Type of Query	Time Criticality
Continuous	Periodic	Strong Persistent	Low
Event-Driven	Upon Detecting	Weak Persistent	High
On-Demand	User Request	One Time	High

네트워크 lifetime의 연장과 주어진 task의 성공적인 수행을 위해 효율적이고 효과적인 data gathering을 위해서는 아래와 같은 4가지의 기술적 issue들이 먼저 고려되어져야 한다.

1) 데이터 중복 제어: 센서 네트워크의 특징 중의 하나는 네트워크상 센서 노드 밀도가 높다는 것이며 이로 인해 두 센서 노드간의 거리가 상대적으로 짧아 두 노드에서 생성된 센싱 결과가 동일한 정보를 포함하는 데이터의 중복성이 높아진다. 이러한 중복성 문제 해결을 하지 않게 되면 불필요한 데이터 전송으로 센서노드 에너지 사용이 높아져 데이터

reporter로서의 수명은 짧아지게 된다. 두 센서 노드에서 수집된 데이터간의 시공간적 (spatio-temporal) 상관관계의 연구가 이러한 센서 데이터 중복성을 최소화 하는데 이용 되어 불필요한 전송으로 인해 사용 되는 에너지의 양을 줄여 센서 노드의 데이터 reporter/forwarder로서의 역할 수행을 최대화 할 수 있다.

2) 신뢰성 있는 event 검출: 센서는 주어진 task에 따라 랜덤하게 또는 주기적으로 발생되는 event들을 monitoring한다. 이러한 event의 검출은 time-critical 또는 non-time-critical한 특성을 지닌다. 그래서 센서들은 검출된 event가 즉각적인 reporting을 필요로 한다면 센싱된 결과를 에너지의 사용이 다소 높다하더라도 주어진 시간내에 가장 높은 우선순위를 가지고 사용자에게 전달할 수 있어야 한다. 이러한 신뢰성 있는 event 검출 수행 능력은 센서가 event의 검출한 결과를 검출한 순간에 report하여 데이터 sink노드까지 성공적으로 데이터 전송을 할 수 있느냐 없느냐에 달려 있다.

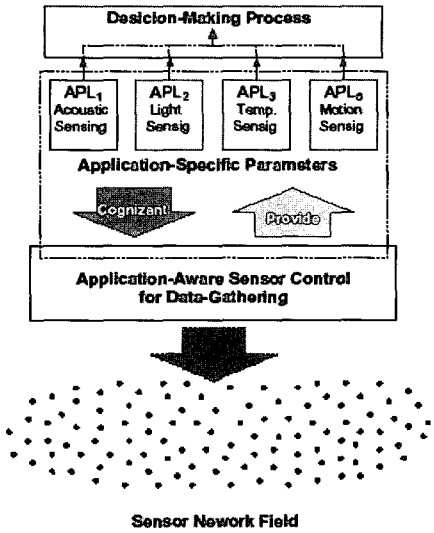
3) 시간에 따라 변화는 사용자 요구에 대한 적응력: 사용자의 관심의 정도는 사건의 형태나 시간에 따라 변화하는 그 사건의 상황(상태)에 따라 변화 할 수 있다. 센싱 커버리지, 데이터 reporting latency, 또는 event detection 실패율 등이 통상 사용자와 센서들 간의 energy-saving negotiation 매개변수로 사용되며 이러한 시스템 매개 변수에 효과적으로 적응하여 energy conservation을 최대화 할 수 있어야 한다.

4) Information security: 무선 센서 네트워크는 우리의 일상생활, military, 그리고 disaster등의 환경에서 mission critical한 정보 수집을 위한 data acquisition network으로 활용되어진다. 수집된 센싱 정보가 잘못 사용되어지거나 또는 왜곡 되어져 전송되어지면 치명적인 결과를 가져올 수 있다. 이로 인해 센서 데이터 보안은 신뢰성 있는 센싱을 제공하기 위한 가장 중요한 요소 중의 하나이다. 고도화된 센서 security를 제공하기 위해서는 비정상적인 센서 정보 검출 방법 및 현재의 다양한 security 해결책의 활용방법에 관한 지속적인 연구가 필요하다.

1.1 Application-aware 데이터 습득

센서 네트워크에서는 수행 되어지는 task의 형태에 따라 얻고자 하는 최적의 결과치를 얻기 위하여 특정 시스템 파라미터들을 중요도에 따라 우선 순위화 함으로써 네트워크 구성을 선택할 수 있다. 이미 언급한 바와 같이 일반적으로 센서 네트워크는 데이터 acquisition network으로서 오랫동안 그 역할을 수행 하여야 하는 특성을 갖는다. 특히 배터리 파워를 이용하여 동작하는 센서들이 지리적 위치적으로 특수한 곳에 전개 되어 질 경우 이러한 네트워크 수명의 연장은 네트워크 유지비용 등을 고려할 때 더욱 중요한 문제로 대두 된다. 앞에서 언급한 바와 같이 센서들이 행하는 특정 task의 형태에 따라 센싱 커버리지, 데이터 reporting latency와 같은 네크워크 시스템 파라미터들의 중요성은 주어진 task를 완수 하는데 있어서 다를 수 있다. 이러한 관점에서 볼때 센서의 극히 제한된 자원을 최적으로 사용하기 위해선 사용 되어지는 application을 인식하고 그의 특성에 따라 센서들의 제어가 이루어져야 한다.

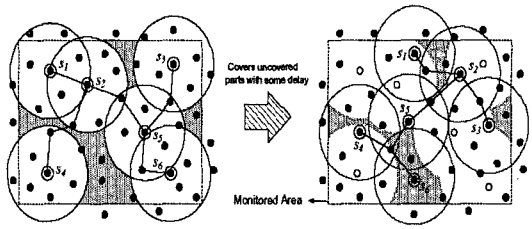
Application-aware 데이터 습득에서 센싱 커버리지를 네트워크 lifetime의 연장을 위한 negotiation 매개변수로 하는 예를 들어 보기로 하자. 먼저 장기간 동안 특정 지역에 습도에 대한 통계적 수치를 연구하는 목적으로 센서 네트워크를 전개 하였다고 가정해 보자. 이 경우 센서들의 데이터 reporting의 결과가 센서 네트워크가 커버하는 지역의 약 80%만을 커버해도 과학적 계산치의 통계적 수치를 얻는데 충분하다면 매 센서 데이터 reporting 회마다 약 80%만을 커버 할 수 있도록 센서를 제어하는 것이 센서 네트워크 수명을 더욱 연장할 수 있는 효율적인 데이터 acquisition 방법이 될 수 있을 것이다. 반면에 time-critical한 event 검출에서는 센싱 데이터가 보고 될 때 마다 센서 네크워크가 커버하는 지역의 100%를 커버 하여 발생된 event의 검출을 지연 없이 보고 받는 것이 에너지 사용량 보다 더욱 중요 하다고 할 수 있다. 이처럼 에너지 사용의 최적화를 위해 사용된 application의 특성을 기반한 파라미터의 제어는 커버리지만이 아닌 데이타 reporting latency, 요청된 센싱 정확도를 얻기 위해 센싱 결과를 전송해야 하는 센서의 수 등 여러가지에 활용 되어질 수 있다. 예를 들면, 서로 다른 지역에 전개된 센서 네트워크에서 동일 레벨의 센싱 정확도를 얻기 위해 필요한 센서의 수는 application의 특성이나 주변 상황에 따라서 다를 수 있다. 이러한 센서 제어는 각 application마다



〔그림 2〕. Application-aware 데이터 수집을 위한 센서 제어 개념도

가 아닌 한 application 내에서 시간에 따라 변화는 상황 및 센싱된 결과에 따라 유동적으로 센서를 제어하는 방법으로도 설계 되어져야 할 필요가 있다.

그림 2는 무선 센서 네트워크에서 에너지 사용을 최적으로 하며 사용자의 요구를 만족시키는 application-aware 데이터 gathering을 위한 센서 제어 개념도를 보인다. 그림 3은 센싱 커버리지와 데이터 reporting latency간의 trade-off를 기반으로 한 Application-aware 데이터 gathering 예를 보인다. 여기서 작은 원안의 검은 점, s1, s2, s3, …, s6은 요구 되는 센싱 커버리지를 만족시키기 이해 현재 선택된 센서들을 나타내며 반면에 그림 3 (b)에서 보여지는 속이 빈 점들은 전 센싱 결과 reporting 라운드에서 선택된 센서들을 나타낸다. 큰 원은 현재 선택된 센서들의 센싱 범위를 나타낸다. 그림 3 (a)에서 보여지는 첫번째로 선택된 6개의 센서들이 요구된 센싱 커버리지를 만족 시키나 모니터 되어 져야 하는 전체 영역을 커버하지는 못하고 그 커버 되지 못한 영역은 그림 1-3 (b)에서 보이는 것처럼 두번째로 선택된 6개의 센서들에 의해 커버 되어 진다. 이처럼 전체의 영역은 두개의 연속적인 데이터 reporting 라운드에 의해 커버 되어져 전체적인 센싱 결과를 얻는 데는 어느 정도의 (고정적) 지연이 발생하나 결론적으로 한 reporting 라운드에 단지 6개의 active 센서만을 유지함으로써 센서의 에너지 자원 사용을 줄여 궁극적으로는 네트워크의 lifetime을 연장 시킨다.



〔그림 3〕 센싱 커버리지와 데이터 reporting latency간의 trade-off를 기반으로 한 Application-aware 데이터 gathering 예

에너지 절약형 데이터 습득 방법의 논의에 이어 다음은 데이터 습득시의 보안 문제에 대해서 논의하기로 한다.

2. 데이터 습득 보안 문제

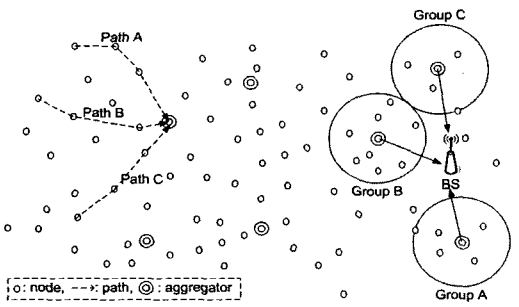
센서 네트워크에서 대두되는 보안 문제 중 가장 중요한 요소 중의 하나가 수집하는 데이터 자체에 대한 보안 문제이다. 센서 네트워크의 주된 응용은 인간과 환경에 밀접한 정보를 수집하여 다양한 서비스를 제공하기 때문에 각 센서들이 수집하는 데이터들이 위변조 될 경우 심각한 보안 문제를 야기 할 수 있다. 이러한 보안 문제에 대한 일반적인 솔루션으로 비도가 우수한 암호학적 기법들을 사용하여 데이터 자체에 대한 인증이나 무결성 보장 기능을 제공할 수 있지만, 센서 네트워크의 저전력/저성능이라는 한계특성으로 높은 비도를 가지는 암호화 기법을 사용하는 것이 항상 좋은 해결 방법이 될 수는 없다. 그러므로 센서 네트워크의 고유 특성을 감안한 새로운 보안 기능이 요구된다. 이러한 보안 기능을 위한 요구사항은 각 센서로부터 수집된 데이터를 수신하는 기지국에서 또는 센서 노드 자체에서 수신된 데이터가 정당한 데이터인지를 판별할 수 있는 기능을 제공하는 것이다.

일반적으로 널리 알려진 감지된 센서 데이터(Sensed Data)에 대한 이상 유무 판단은 각 센서 또는 전달/수집 노드(Relay/Aggregator Node)에서 데이터를 수신하면 각 노드들은 이미 알고 있는 정해진 임계값(Static Threshold Value)과 일정 오차 이상인 경우에 이상 데이터로 판단한다. 하지만 이런 방법은 변화하는 환경에 민감하게 대응하지 못할뿐 더러 많은 오류(false-positive)를 유발 할 수 있다. 즉, 정해진 임계값(Static Threshold Value)을 사용하는 방법의 문제점은 현재 수집되고 있는 센서 데

이터의 이상 유무를 판단하기 위해서 필요한 임계값을 결정하는 것이 쉽지 않은데서 기인한다. 그러므로 본 논문에서는 정적인 임계값에 의존적인 문제를 해결할 수 있는 학습 기반의 센서 이상 데이터 유무 판정 방법에 대해서 살펴본다.

3. 이상 데이터 검출 방안

센서 노드로부터의 이상 데이터 탐지 방안은 이상 데이터를 탐지하는 기법 자체에 대한 기술 개발은 물론이고 탐지되는 소스 데이터를 어떻게 수집하는지도 중요한 문제가 된다. 다음 그림 4에서는 센서 데이터 라우팅 경로를 통한 두 가지 센서 데이터 수집 방법을 보여준다. 먼저 그림 4의 왼쪽 방법은 센서 노드로부터의 데이터를 기지국에 전달하는 전달노드에서 수집하여 그 이상 유무를 판정하는 방법으로서 전달노드는 유사한 센서 데이터를 전달하는 상이한 세 개의 경로 A, B, C를 알고 있다고 가정 한다. 그리고 각 경로들의 고유값을 통해서 특이 값을 전달해오는 경로로부터의 데이터를 이상 데이터로 판정하는 방법을 적용할 수 있다. 또한 그림 4의 오른쪽 방법에서는 센서 환경에서 노드들을 일정 그룹으로 분산하여 분산된 노드들로부터 데이터를 수집한 전달 노드의 값들을 비교함으로써 역시 상이한 값을 발생하는 그룹을 비정상 데이터를 생성한 노드가 포함된 그룹으로 보고 이상 유무를 판정할 수 있다.



〔그림 4〕 이상 데이터 검출을 위한 센서 데이터 습득 방안

다음에서는 이러한 센서 데이터 수집 후에 실제 이상 유무를 판정할 수 있는 그림 5와 같은 K-means clustering 알고리즘(표 2 참고) 응용 방법을 제안한다. 먼저 비교사 학습 방법의 하나인 K-means clustering 알고리즘〔10-11〕은 표 2와 같은 알고리즘으로 동작한다. 이러한 K-means clustering 기법을 센서 네트워크에 적용하는 방법은 먼저 센서 환

〔표 2〕 K-Means 알고리즘 동작 단계

1) 초기화 단계:
생성할 클러스터의 개수 K를 정하고, K개의 각 클러스터에 대하여 클러스터의 중심을 초기화한다.

$$\{z_1(l), z_2(l), \ldots, z_k(l)\}$$

2) 개체 분산 단계:
개체들을 각 클러스터에 분산시킨다.

$$x^{(p)} \in S_j^{(l)} \quad if \quad \left\| x^{(p)} - z_j(l) \right\| < \left\| x^{(p)} - z_i(l) \right\|$$
$$for\ all \quad i = 1,2,\ldots K, \quad i \neq j$$

3) 새로운 클러스터의 중심 계산

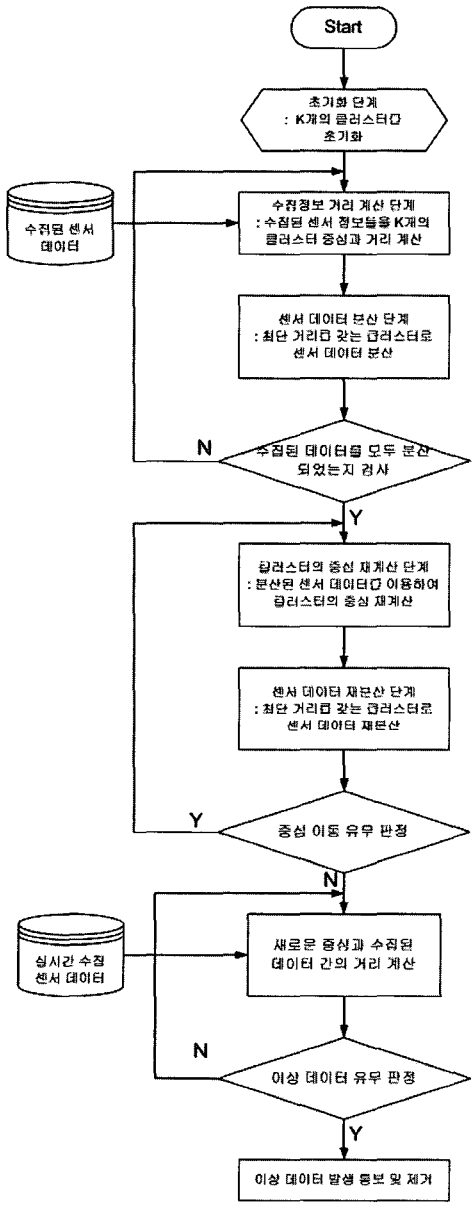
$$J_j = \sum_{x^{(p)} \in S_j^l} \left\| x^{(p)} - z_j(l+1) \right\|^2, \quad j = 1,2,\ldots,K$$

J_j 를 최소화 시키는 $z_j(l+1)$ 는 간단히 클러스터 j에 속하는 모든 개체들의 평균값을 취함으로써 구할 수 있다.

$$z_j(l+1) = \frac{1}{N_j} \sum_{x^{(p)} \in S_j^l} x^{(p)}$$

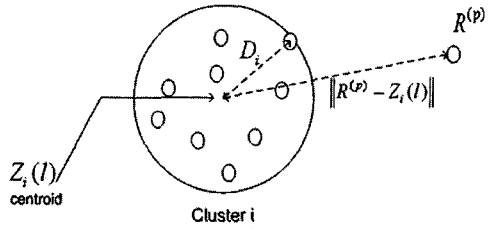
4) 수렴 여부의 확인
K-means 알고리즘은 더이상 각 클러스터의 중심에 변화가 생기지 않을 때 종료된다. 그렇지 않다면 2번의 과정이 반복된다.

경에서 센서들로부터 수집된 데이터를 클러스터링하고 해당 클러스터의 중심으로부터 멀리 떨어진 위상을 가지는 수집 노드의 값을 이상 데이터로 판정하는 방법이다. 이 방법에서는 먼저 Relay Node가 각 센서 노드로부터 데이터를 수집하고, Relay Node는 수집된 데이터를 바탕으로 K-means clustering 알고리즘을 사용하여 클러스터들을 구성한다. 여기서 K-means clustering 알고리즘을 적용한 센서 환경의 이상 데이터 판정 방법의 전체 구성은 그림 5와 같다. 본 논문에서 제안하는 K-means clustering을 이용한 알고리즘은 대상 센서 네트워크 노드의 수집 정보의 특성에 따라 구성할 클러스터의 개수 K를 결정한다. 클러스터의 개수를 결정하는 방법은 다음과 같은 세 가지 방법을 적용할 수 있다. 첫 번째는 대상 환경에 관한 사전 지식을 바탕으로 대상 환경에 설치할 센서들이 수집할 값을 예측하고 이 예측을 통해 클러스터의 개수 및 각 클러스터의 중심값을 설정 할 수 있다. 두 번째로 첫 번째 방법과 유사하나 단순히 적용된 환경의 특성에 비추어 클러스터의 개수를 설정하고 중심값 설정 등의 과정은 K-means clustering 알고리즘에 의해서 결정하는 방법이 가능하다. 마지막으로 K-means clustering 알고리즘이 비교사 학습(unsupervised)의 특성을 이용하여 충분히 많은 클



〔그림 5〕 K-means 알고리즘을 응용한 이상검출 흐름도

러스터 생성을 방임하고 사후 레이블링을 통해 클러스터가 포함하는 데이터의 정상/비정상 유무를 결정하여 실제 비정상 센서데이터 선별에 적용하는 것이다. 이렇게 초기화 된 클러스터를 사용하여 클러스터의 중심을 변화가 없을 때까지 재계산한다. 만약 클러스터 중심의 변화가 없다면 그림 6과 같은 클러스터 중심과 최외각 센서간의 중심거리를 구할 수 있으며 중심거리를 얼마만큼 벗어나느냐에 따라 수집된 데이터가 이상 데이터로 판정 될 수 있다.



〔그림 6〕 클러스터 중심거리 계산을 통한 이상 데이터 검출

V. 결 론

센서 네트워크는 향후 유비쿼터스 네트워크 환경을 선도해나가는 핵심기술로서 자리 잡아 갈 것이다. 하지만 전형적인 기존의 유/무선 네트워크와 달리 센서 네트워크 고유 특성으로 인해 다양한 고려사항이 존재하는 것도 사실이다. 본 논문에서는 현재의 센서 네트워크 동향을 센서 데이터 수집 및 병합 전략을 중심으로 다루었다. 이때 저전력 소비 문제는 물론이고 클러스터링과 지역을 기반으로 하는 전략으로 기존의 알고리즘들보다 효율적인 운용이 가능한 알고리즘을 소개하였다. 또한 이러한 센서 데이터에 관련된 기반 기술과 함께 수집되는 데이터들 사이의 문제점을 탐지할 수 있는 이상 센서 데이터를 판정하기 위한 K-means 기반 학습 알고리즘의 응용 방안을 제안하였다. 향후에는 제안된 알고리즘의 실제 검증을 통해 센서 네트워크 환경에서 발생할 수 있는 개인정보 노출과 같은 데이터 민감성 보장에 관한 대응 방안을 마련할 수 있을 것이다.

참 고 문 헌

〔1〕 I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless Sensor Networks: a Survey. Computer Networks, vol. 38(4):393-422, 2002

〔2〕 A. Gosh and S. K. Das. A Distributed Greedy Algorithm for Connected Sensor Cover in Dense Sensor Networks. In Proceedings of Int'l Conference on Distributed Computing in Sensor Networks (DCOSS), 2005

〔3〕 D. Tihan and N. D. Georganas. A Coverage-Preserving Node Scheduling Scheme for Large Wireless Sensor

Networks. In Proceedings of ACM Workshop on Wireless Sensor Networks and Applications (WSNA), pp. 32-41, 2002

[4] O. Younis and S. Fahmy. HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks. IEEE Transactions on Mobile Computing, vol. 3(4):366-379, 2004

[5] C. Intanagonwiwat, R. Govindan, and D. Estrin. Directed Diffusion:A Scalable and Robust Communication Paradigm for Sensor Networks. In Proceedings of ACM Mobile Computing and Networking (MOBICOM), pp. 56-67, 2000

[6] S. Tilak, N. B. Abu-Ghazal도, and W. Heinzelman. A Taxonomy of Wireless Micro-Sensor Network Models. ACM Mobile Computing and Communications Review, vol. 6(2):28-36, 2002

[7] Perrig, J. Stankovic, and D. Wagner. Security in wireless sensor networks. Commun. ACM 47(6):53-57, 2004

[8] J.ohn Paul Walters, Zhengqiang Liang, Weisong Shi, and Vipin Chaudhary, Wireless Sensor Network Security: A Survey, http://www.cs.wayne.edu/~weisong/papers/walters05-wsn-security-survey.pdf

[9] Haowen Chan: Perrig, A, "Security and privacy in sensor networks", Computer Volume 36, Issue 10, Oct. 2003 Page(s): 103 - 105

[10] Steinbach, M., Karypis, G., Kumar, V., "A Comparison of Document Clustering Tech," U of Minnesota, Technical Report #00-034, 2000.

[11] Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, Angela Y. Wu, "An Efficient k-Means Clustering Algorithm: Analysis and Implementation", IEEE Transactions on Pattern Analysis and Machine Intelligence archive, Volume 24 , Issue 7 (July 2002), pp 881 - 892, 2002

〈著 者 紹 介〉

손 태 식 (SHON TAESHIK)

정회원

2005년 08월: 고려대 공학박사

2004년 ~ 2005년 : Research Scholar, Univ. of Minnesota

2005년 8월 ~ 현재 : 삼성전자 통신연구소

관심분야 : Anomaly Detection, 802.11/16 Security, Sensor Network Security, Mobility, VoIPSec

최 욱 (CHOI WOOK)

2005년 05월 : 텍사스 주립대 공학박사

2005년 08월 ~ 현재: 삼성전자 통신연구소

관심분야 : Wireless Mesh Networks, Smart Envirionment, Wireless sensor and ad hoc networks, Multi-radio access technology