Journal of the Korean Data & Information Science Society
2009, **20**(2), 387–399

한국데이터정보과학회지

이동 무선망을 위한 비유사도 기반 비정상 행위 탐지 방법의 설계 및 평가

이화주[1] · 배인한[2]

[12]대구가톨릭대학교 컴퓨터정보통신공학부

접수 2009년 1월 21일, 수정 2009년 3월 17일, 게재확정 2009년 3월 22일

요 약

이동 무선망은 인증의 절도와 침입에 의해 계속 고통을 받고 있다. 그러한 두 문제 모두 2가지 다른 방법: 오용 탐지 또는 비정상 행위 기반 탐지로 해결될 수 있다. 이 논문에서, 우리는 이동 무선망의 이동 패턴과 같은 정상 행위를 효율적으로 식별할 수 있는 비유사도 기반 방법을 제안한다. 제안하는 알고리즘에서, 정상 프로파일은 이동 무선망에서 이동 사용자들의 정상 이동 패턴으로부터 구축되어진다. 구축된 정상 프로파일로부터, 가중 비유사도 측정으로 비유사도가 계산되어진다. 만일 가중 비유사도 측정치가 시스템 매개변수인 비유사도 임계치보다 크면, 경고 메시지가 발생된다. 제안된 방법의 성능은 모의실험을 통하여 평가되었다. 그 결과, 제안하는 방법의 성능이 비유사도 측정을 사용하는 다른 비정상 행위 탐지 방법의 성능 보다 우수함을 알 수 있었다.

주요용어: 비유사도, 비정상 행위 탐지, 이동 무선망, 이동 패턴, 정상 프로파일.

1. 서 론

모바일 컴퓨팅 환경의 특징은 상대의 악의적인 공격에 취약하다. 무선 링크의 사용으로 네트워크는 수동적인 엿듣기와 능동적인 간섭에 이르기 까지 공격당하기 쉽다. 공격자가 물리적인 액세스를 획득해야하거나 방화벽과 게이트웨이에서 방어 라인을 통과해야하는 유선 네트워크와 달리 무선 네트워크에서 공격은 모든 위치로부터 발생할 수 있고, 임의의 노드에서 표적이 될 수 있다. 손해는 비밀정보의 누설, 메시지 손상, 노드 위장 등을 포함할 수 있다 (Zhang 등, 2003).

셀룰러 기반 모바일 무선망에 데이터 서비스 도입으로 사람들은 매일 생활에서 전자 쇼핑과 전자 뱅킹과 같은 중요하고 민감한 일에 셀룰러 전화를 사용하고 있다. 편리하고 인기 있는 새로운 서비스는 중요한 보안 문제가 자연적으로 발생된다. 비록 셀룰러 모바일 망에 많은 인증 프로토콜이 있지만 개방 무선 전송 환경과 모바일 장치의 물리적 취약성으로 보안은 도전할 만한 중요한 분야이다 (Sun 등, 2004).

일반적으로, 어떤 시스템을 보호하기 위하여 서로 보완하는 2가지 방법으로 예방과 탐지가 있다. 암호와 인증과 같은 침입 예방 기법들은 악의의 행위에 의한 공격자들을 방지할 수 있다. 보안 연구의 역사를 통하여 예방 기법으로만 공격자들을 완전히 제거하기 보다는 단지 침입을 줄일 수 있음을 알 수 있다. 아무리 많은 침입 예방 수단이 개발될지라도, 공격자는 시스템에 침입하기 위하여 다소의 보안 결함을 이용할 수 있다. 그러므로 2차 방어선으로 서비스하는 침입 탐지 시스템(IDS, intrusion detection

[1] (712-702) 경북 경산시 하양읍 금락로 5, 대구가톨릭대학교 대학원 컴퓨터정보통신공학과, 박사과정.
[2] 교신저자: (712-702) 경북 경산시 하양읍 금락로 5, 대구가톨릭대학교 컴퓨터정보통신공학부, 교수.
E-mail: ihbae@cu.ac.kr

system)은 신뢰적인 시스템에 꼭 필요하다 (Cai 등, 2006). 일반적으로 2가지 침입 탐지 기법으로 오용 기반 탐지와 비정상 행위 기반 탐지가 있다. 오용 기반 탐지 기법은 알려진 공격 서명과 시스템 취약성을 코드화한다. 만일 오용 탐지 기법이 현재 활동에 반하는 매치를 찾았다면, 경고가 발생된다. 오용 탐지 기법은 새로운 공격을 탐지하는데 비효율적이다. 비정상 행위 탐지 기법은 시스템 상태와 노드 행위의 정상 프로파일을 생성하고 그것들을 현재 활동과 비교한다. 만일 상당한 이탈이 관찰되면, 경고가 발생된다. 비정상 행위 탐지는 알려지지 않은 공격을 탐지할 수 있다. 분명히, 모바일 환경 하에서, 사전에 모든 공격 패턴을 아는 것은 불가능하기 때문에 비정상 행위 탐지 기법은 오용 탐지 기법 보다 더 필수적이다. 그러나 노드의 이동에 기인하여 정상 프로파일 구축은 매우 힘들다. 그러므로 이동 노드들에 대한 정상 프로파일을 구축하는 것은 효율적인 침입 탐지 알고리즘 설계에서 중요하다 (Sun 등, 2004; Hall 등, 2005).

본 논문에서는 모바일 무선망의 이동 패턴과 같은 비정상 행위를 효율적으로 식별할 수 있는 비유사도 기반 비정상 행위 탐지 알고리즘을 제안한다. 제안하는 알고리즘에서는 모바일 무선망에서 정상 이동 패턴으로 모바일 노드의 정상 프로파일을 구축하고, 구축된 정상 프로파일로부터 가중치 기반 비유사도 측정을 사용하여 어떤 모바일 노드의 이동 패턴의 정상 행위로부터 비유사도를 계산한다. 만일 계산된 비유사도 값이 비유사도 임계치 보다 크면, 경고 메시지가 생성되어진다. 아니면, 정상 활동으로 인식되어진다.

본 논문의 구성은 다음과 같다. 2장에서는 모바일 무선망에서 침입 탐지 기법에 대한 관련 연구를 살펴보고, 3장에서는 비정상 이동 패턴을 탐지하는데 사용되는 유사도와 비유사도 측정 방법을 설명한다. 4장에서는 본 논문에서 제안하는 비유사도 기반 비정상 행위 탐지 모델을 제안하고, 5장에서는 모의실험을 통하여 제안하는 방법의 성능을 평가한다. 그리고 6장에서 결론 및 향후 연구 내용을 기술한다.

2. 관련 연구

새로운 자동화된 침입 도구가 매일 출현하기 때문에 컴퓨터 시스템에서 침입 횟수는 증가하고 있다. 편리함을 주는 인기 있는 새로운 서비스는 자연적으로 중요한 보안 문제가 발생된다. 비록 셀룰러 모바일 망에 많은 인증 프로토콜이 있지만, 보안은 개방 무선 전송 환경과 모바일 장치의 물리적 취약성으로 인하여 매우 도전적인 연구이다. Sun 등 (2004)에서 사용자 이동 패턴은 그 사용자에 의해 운행된 셀 ID의 고차 Markov 모델로 기술된다. 이동 trie에서 데이터를 파스하고 관련 통계 정보를 저장하기 위하여 Ziv-Lempel 데이터 압축 알고리즘을 이용하고, 최근 정상 프로파일 관리를 위해 이동 trie 갱신에 EWMA(Exponentially Weighted Moving Average)를 적용하였다. Hall 등 (2005)은 응용 계층에서 비정상 행위 기반 침입 탐지를 위한 대중 교통수단을 사용하는 이동 사용자들의 이동 패턴에 기초하는 프로파일 사용의 실현 가능성을 검토하였다. 이화주와 배인한 (2007)은 러프 셋을 사용하여 정상 프로파일을 구축하고, 사용자 프로파일의 나이와 가중 특징 값을 고려한 러프 소속 함수를 사용하여 비정상 행위를 효율적으로 탐지하는 동적 비정상 행위 탐지 알고리즘을 설계하고 성능을 평가하였다.

Zhang 등 (2000, 2003)에서 무선 애드 혹 망을 위한 침입 탐지 에이전트 시스템을 설계하였다. 각 노드의 IDS 에이전트는 독립적으로 실행되고, 국부적 활동을 감시한다. 그것은 국부 추적으로부터 침입을 탐지하고 대응을 시작한다. 만일 비정상 행위가 국부 데이터에서 탐지되거나, 또는 만일 추적이 결정적이 아니고 더 넓은 검색이 요구되면, 이웃 IDS 에이전트들은 광역 침입 탐지 행위에 협동적으로 참여하는 통합 침입 탐지와 대응 메커니즘을 제안하였다. Kachirski와 Guha (2003)에서는 다수의 망 센서 즉, 패킷 단계, 사용자 단계, 시스템 단계 센서들로부터 검사 데이터를 효율적으로 합병하여 침입에 대해 전체 애드 혹 무선망을 분석하고 침입 억제를 시도하는 분산 협동 침입 탐지 시스템을 제안하였다. Cai 등 (2006)에서는 패턴 인식에서 시작된 통계적 방법에 기초하여 모바일 애드 혹 망의 이동성 패턴

과 같은 비정상 행위를 식별할 수 있는 비정상 탐지 알고리즘을 제시하였다.

3. 유사도와 비유사도 측정

유사도는 데이터 마이닝에서 중심 개념이다. 유사도 연구는 주로 두 가지 영역: 객체 유사도와 속성 유사도에 따라 진행되었다. 객체 유사도는 데이터베이스 내의 두 객체들 간의 거리를 정량화하고, 반면에 속성 유사도는 속성들 간의 거리를 나타낸다. 관련된 문제는 데이터의 두 부분집합들의 유사도와 비유사도를 결정하는 것이다. 변화 시점 탐지, 비정상 행위 탐지, 클러스터링을 위하여 데이터의 다른 부분 집합들이 검사될 수 있다. 이것은 적절한 비유사도 측정의 개발을 요구한다. 적절한 비유사도 측정은 많은 요구사항을 가진다. 첫째, 그것은 데이터 집합에 포함된 많은 정보를 가능한 한 고려해야 한다. 둘째, 그것은 영역 지식을 고려하기 위하여 사용자가 조정할 수 있어야 한다. 예를 들어, 몇몇 영역에서 두 데이터 집합의 평균들의 차이는 그것들의 상관관계들의 차이 보다 중요하지 않을 수도 있다. 이 경우에, 평균 차이는 상관관계 차이 보다 더 적은 가중치가 부여된다. 셋째, 많은 영역에서 데이터 모임은 불완전하기 때문에 많은 누락 속성 값을 이끌어 낸다. 따라서 비유사도 측정은 누락 및 손상 데이터에 대해 포용력이 있어야 한다 (Otey와 Parthasarathy, 2006).

유사성 측정의 목표는 동형의 k-차 데이터 집합 X와 Y의 비유사도를 정량화하는 것이다. 일반적으로 두 데이터 집합 X와 Y의 비유사도는 $dissim(X,Y)$로 나타낸다. 가장 폭넓게 사용되는 비유사도 측정은 Jaccard 계수이다 (Henning과 Hausdroft, 2006).

$$dissim_J(X,Y) = \min_{Y \in MP_{i,j}} \left(1 - \frac{X \cap Y}{X \cup Y}\right). \tag{3.1}$$

여기서 X는 사용자 i의 현재 이동 패턴, Y는 사용자 i의 정상 이동 패턴, 그리고 $MP_{i,j}$는 사용자 i의 정상 경로 j에 대한 정상 패턴 프로파일 집합을 각각 나타낸다. 이 거리는 X와 Y 둘 다에는 없고 X 또는 Y에 있는 구성요소의 비율로 직접 계산되어진다. Jaccard 거리가 갖는 중요한 문제가 있다. 어떤 영역이 아주 큰 영역의 부분 집합이면, Jaccard 거리는 아주 커지는 경향이 있다. 예를 들어, 만일 $X \subset Y$, $|X| = 4$, $|Y| = 20$이면, $dissim_J(X,Y) = 1 - 4/20 = 1 - 1/5 = 4/5 = 0.8$이다. 그러나 Jaccard 계수는 데이터 집합 X와 Y의 교집합에 속하지 않는 원소들의 특성을 전혀 고려하지 않으므로 그것들의 비유사도 가중치는 1.0으로 일정하다. 따라서 비유사도 측정의 두 번째 요구사항을 만족하지 않는다.

Hall 등 (2005)은 위치 좌표들로 구성된 이동 순서의 유사도 계산을 위하여 식 (3.2)를 사용한다.

$$sim(X,Y) = \sum_{k=0}^{l-1} w(X,Y,k). \tag{3.2}$$

$$w(X,Y,k) = \begin{cases} 0 & \text{if } i < 0 \text{ or } x_i \neq y_i. \\ 1 + w(X,Y,k-1) & \text{if } x_i = y_i \end{cases}$$

여기서 l과 k는 이동 패턴의 길이와 위치 좌표들의 순서의 인덱스를 각각 나타낸다. 즉 $w(X,Y,k)$는 만일 인덱스 i에서 X와 Y 순서의 좌표들이 같지 않다면 0이다. 아니면 1의 값이 k-1에서 $w(X,Y,k-1)$의 결과에 더해진다. 사용자 프로파일 MP는 정상 이동 패턴의 집합을 포함하는 사용자 관련 정보를 포함한다. 반면에, 그 유사도 측정은 현재 이동 패턴의 위치 좌표들과 정상 이동 패턴의 위치 좌표들을 일대일로 비교하여 결정되어지고, 그 프로파일에 대한 유사도 측정은 관찰된 현재 이동 패턴 X를 그 사용자 프로파일내의 모든 정상 이동 패턴들과 비교하여 계산되어진다. 그것은 식 (3.3)과 같이 정의되어진다.

$$sim_{MP_{i,j}}(X,Y) = \max_{Y \in MP_{i,j}}(sim(X,Y)). \tag{3.3}$$

따라서 비유사도는 식 (3.4)와 같이 정의되어진다.

$$dissim_H(X,Y) = 1 - \frac{sim_{MP_{i,j}}(X,Y)}{the\ maximum\ of\ similarity}. \tag{3.4}$$

여기서 이동 패턴 길이 l의 최대 유사도(the maximum of similarity)는 $\sum_{i=1}^{l} i = l(l+1)/2$이다. Hall의 방법이 갖는 문제는 사용자의 현재 이동 패턴 X에서 정상 이동 패턴 Y와 다른 패턴 원소의 개수만을 고려하고 그 패턴 원소의 속성은 고려치 않는다는 것이다.

본 논문에서는 정상 이동 패턴과 사용자 이동 패턴 간의 비유사도를 측정하기 위하여 식 (3.5)의 가중치 기반 비유사도 측정 방법을 제안한다.

$$dissim_W(X,Y) = \min_{Y \in MP_{i,j}} \left(\frac{\sum_{t=1}^{l} w(x_t, y_t)}{l} \right). \tag{3.5}$$

여기서 $w(x_t, y_t)$는 현재 이동 패턴의 t-번째 원소와 정상 이동 패턴의 t-번째 원소 간의 비유사도의 가중치를 나타내고, $0 \leq w(x_t, y_t) \leq 1$이다.

4. 비유사도 기반 비정상 행위 탐지 방법

본 논문에서 제안하는 비정상 행위 탐지 알고리즘은 다음 가정에 의존한다.

- 첫째, 각 모바일 노드는 이동 패턴과 같은 그것의 정상 활동을 기술한 특정 이동성 데이터베이스를 가지고 있다고 가정한다. 모든 노드 이동성 데이터베이스들은 공격하기 힘든 확실히 안전한 장소에 저장된다. 그리고 각 모바일 노드 내부에 어떤 시점에 그 노드의 정확한 위치를 제공할 수 있는 GPS와 같은 최소한 하나의 장치가 존재한다.
- 둘째, 대부분 모바일 노드들은 아주 규칙적인 경로들을 갖는다고 가정한다. 그러므로 각 각 노드에 대한 정상 프로파일을 구축하는 것이 가능하다.

이동성은 모바일 무선망에서 가장 일반적이고 중요한 특징이다. 우리는 모바일 노드의 이동 패턴을 겨냥한 공격의 탐지에 초점이 있다. 가장 중요한 활동적인 공격 중 하나인 어떤 출발 지점부터 어떤 지리적 지역 내의 다음 지점까지 정상 이동 패턴을 변경시키는 것을 탐지하는 것이다. 그림 4.1은 지역 A내의 지점 a_1으로부터 지역 B내의 지점 b_1까지의 몇몇 경로를 보여준다.

그림 4.1에서, 2개의 지리적 지역, A와 B; 4개의 라우트, R1: $a_1 \rightarrow o \rightarrow b_1$, R2: $a_1 \rightarrow o' \rightarrow b_1$, R3: $a_1 \rightarrow m \rightarrow b_1$, R4: $a_1 \rightarrow m' \rightarrow b_1^{'}$. 모든 라우트들은 출발 지점 a_1에서 출발하여 목표 지점 b_1또는 $b_1^{'}$까지 2 홉을 갖는다. 우리는 b_1과 $b_1^{'}$이 매우 가깝고 각 $\alpha = \angle ma_1m'$은 매우 작다고 가정한다.

- **경우 1** : R1이 정상 라우트라고 가정하면, 다른 3개의 라우트들은 비정상 행위로 취급되어야 한다. b_1과 $b_1^{'}$이 가깝고 R2, R3 그리고 R4가 같은 주기(2 홉)내에 도달될 수 있다는 것에 유의한다. 비록 R2가 R1과 같은 패턴 분포를 가질지라도, R2를 비정상 라우트로 분류할 필요가 있다.
- **경우 2** : R3이 정상 라우트라고 가정하면, 분명한 편차에 기초하여 R1과 R2는 비정상 행위로 쉽게 분류할 수 있다. 그러나 R4는 정상 라우트 또는 비정상 라우트로 분류 가능하다. 따라서 이 문제를 해결하기 위한 어떤 정량화된 방법이 있어야 한다.

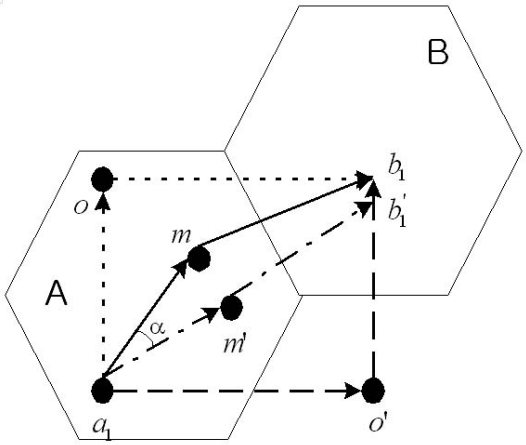

그림 4.1 위협 모델

위의 문제 설명에 기초하여, 경우 2가 어떤 관찰된 클러스터 집합에서 많은 근접 지점들과 이동 패턴들을 갖는 문제를 해결하는 가장 중요한 핵심이라는 것을 쉽게 알 수 있다. 우리는 이 논문에서 이동성 패턴의 근접성을 처리한다.

침입 탐지에서 첫 번째 단계는 효율적인 특징을 추출하는 것이다. 특징은 적합한 탐지 알고리즘을 구축하기 위하여 사용될 수 있는 보안 관련 척도이다. 효율적인 척도는 대상 활동을 반영하기 위하여 선택되어야 한다. 이것에 기초하여, 우리는 모바일 무선망에서 정상 이동 패턴들로 모바일 노드들의 정상 프로파일을 구축한다. 각 노드는 자신의 정상 라우트를 가질 것이라는 가정 하에, 각 노드에 의해 운행된 지리적 지역 내의 장소 번호는 하나의 이상적인 후보 특징이다.

우리는 n 겹치지 않는 속도와 방향 범위, 예를 들어 $V_1, V_2, \ldots, V_n$그리고 $D_1, D_2, \ldots, D_n$으로 속도와 방향을 분류한다. 이동 패턴은 $P_{ij} = V_i D_j$로 정의될 수 있다. 따라서 우리의 분류는 n^2이동 패턴들이 있다. 그것은 2차원 표로 표현할 수 있다.

우리는 특정 노드의 정상 이동 행위에 대한 하나의 정규 프로파일을 구축한다. 우리는 속도와 방향에 따라 표 4.1과 같은 이동 패턴들을 가진다. 여기서 $D_1 = [0, \pi/3]$, $D_2 = [\pi/3, 2\pi/3]$, $D_3 = [2\pi/3, \pi]$, $D_4 = [\pi, 4\pi/3]$, $D_5 = [4\pi/3, 5\pi/3]$, $D_6 = [5\pi/3, 2\pi]$이다. 그러므로 전체 16 이동 패턴이 있다.

표 4.1 이동 패턴 정의

		방향					
		D1	D2	D3	D4	D5	D6
속도	$V_1(1 \sim 3)$	$a(= P_{11})$	$d(= P_{12})$	$g(= P_{13})$	$j(= P_{14})$	$m(= P_{15})$	$p(= P_{16})$
	$V_2(4 \sim 6)$	$b(= P_{21})$	$e(= P_{22})$	$h(= P_{23})$	$k(= P_{24})$	$n(= P_{25})$	$q(= P_{26})$
	$V_3(7 \sim 9)$	$c(= P_{31})$	$f(= P_{32})$	$i(= P_{33})$	$l(= P_{34})$	$o(= P_{35})$	$r(= P_{36})$

본 논문에서 제안하는 이동 패턴을 이용한 비유사도 기반 비정상 행위 탐지 모델의 구조는 그림 4.2와 같다.

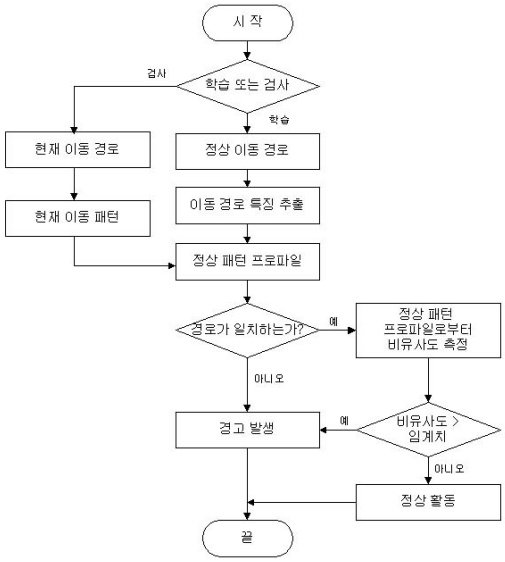

그림 4.2 비유사도 기반 비정상 행위 탐지 방법의 구조

먼저 어떤 모바일 노드의 정상 이동 경로로부터 이동 특징을 추출하여, 그 노드의 정상 패턴 프로파일을 구축한다. 그리고 그 노드의 어떤 이동 활동이 발생하면, 그 경로가 정상 프로파일의 경로와 일치하는지를 검사한다. 만일 일치하지 않으면 경고가 발생되고, 아니면 가중치 기반 비유사도 측정, 식 3.5를 사용하여 정상 프로파일로부터 그 이동 활동에 대한 비유사도를 계산한다. 만일 계산된 비유사도 값이 유사도 임계치보다 크면 경고를 발생시키고, 아니면 정상 이동 활동으로 식별된다.

어떤 모바일 노드가 s 지점에서 d 지점까지 그림 4.3과 같은 2가지 정상 경로를 갖고, 각 정상 경로 당 2개의 이동 패턴을 갖는다면, 그 노드의 정상 이동 활동으로부터 표 4.2와 같은 정상 프로파일을 구축할 수 있다.

표 4.2 정상 패턴 프로파일의 예

				구간(k)			
				1	2	3	4
경로(Y)	I	패턴	1	a	a	p	p
			2	a	b	q	p
	II	패턴	1	p	p	a	a
			2	p	q	g	a

표 4.2의 정상 프로파일로부터 그 모바일 노드의 4 구간 동안의 이동 패턴을 알 수 있다. 그 모바일 노드의 첫 번째 경로의 첫 번째 이동 패턴은 a, a, p, p이다. 즉, 그 모바일 노드는 1, 2 번째 구간에는 이동속도 V1과 이동방향 D1으로 이동하였고, 3, 4 번째 구간에는 이동속도 V1과 이동방향 D6로 이동한 것을 알 수 있다.

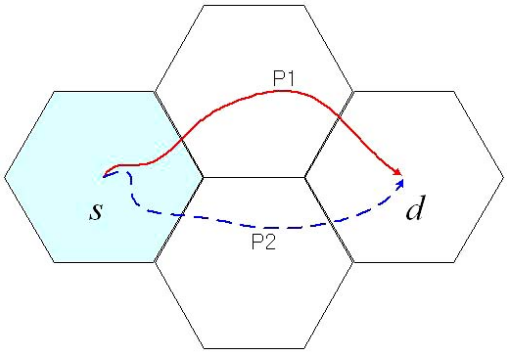

그림 4.3 모바일 노드의 정상 경로의 예

• **예 1**

첫 번째 예는 s에서 출발하여 정상 경로를 거쳐 d에 도착한 어떤 모바일 노드가 이동 패턴 X={ a, b, q, r }을 가지는 경우, 그 이동 패턴은 정상 이동 패턴에 속하지 않는다. 비유사도 측정으로 Jaccard 계수를 사용한 비유사도 값은 식 (3.1)을 사용하여 다음과 같이 계산되어진다.

$$dissim_J(X,Y) = min\left(1-\frac{1}{5}, 1-\frac{3}{5}\right) = \frac{2}{5} = 0.4.$$

이전의 Jaccard 계수를 사용한 비유사도 계산에서 사용된 동일한 예를 Hall의 방법과 본 논문의 가중치 기반 방법으로 비유사도를 계산하면 다음과 같다. 먼저 현재 이동 패턴의 Hall의 유사도 값은 식 (3.3)으로부터 다음과 같이 구해진다.

$$sim_{MP_{i,j}}(X,Y) = max(4,9) = 9.$$

따라서 Hall의 최대 유사도는 $\sum_{i=1}^{4} i = (4\times 5)/2 = 10$이므로 비유사도 값은 식 (3.4)으로부터 다음과 같이 구해진다.

$$dissim_H(X,Y) = 1 - \frac{9}{10} = 0.1.$$

표 4.3은 이동 패턴 원소 간 비유사도 가중치의 예를 보여준다. 여기서 x_i는 어떤 모바일 노드의 현재 이동 패턴에서 현재 구간에서의 이동 패턴, y_i는 그 모바일 노드의 정상 이동 패턴에서 현재 구간의 이동 패턴을 각각 나타낸다. 만일 x_i가 y_i가 같지 않으면, 비유사도 가중치가 할당된다. x_i와 y_i가 이동 방향은 같으나 이동 속도가 다른 경우는 이동 속도는 같으나 이동 방향이 다른 경우 보다 더 작은 비유사도 가중치를 할당한다. 예를 들어, 이동 방향은 같지만 이동 속도가 한 단계 다르면 비유사도 가중치는 0.2이다. 그러나 이동 속도는 같지만 이동 방향이 한 구간 다르면 비유사도 가중치는 0.4이다. 이러한 이동 패턴 원소 간 비유사도 가중치는 시스템 매개변수로 비유사도 기반 비정상 행위 탐지 방법이 사용되는 시스템 환경에 따라 사용자가 적절히 조정할 수 있다.

표 4.3의 이동 패턴 원소 간 비유사도 가중치를 사용한 가중치 기반 비유사도 값은 식 (3.5)로부터 다

음과 같이 계산되어진다.

$$dissim_W(X,Y) = min\left(\frac{0.2+0.2+0.4}{4}, \frac{0.4}{4}\right) = \min(0.2, 0.1) = 0.1.$$

만일 비유사도 임계치(ε)를 0.2로 설정하면, $dissim_J(X,Y) > \varepsilon$이므로 그 이동 패턴은 비정상 행위로 식별되고, 경고 메시지가 발생된다. 반면에 $dissim_H(X,Y), dissim_W(X,Y) \leq \varepsilon$이므로 그 이동 패턴은 정상 행위로 식별되어진다.

• **예 2**

두 번째 예는 어떤 모바일 노드가 다른 이동 패턴 X={a, b, k, p}를 가진 경우, 그 이동 패턴 역시 정상 이동 패턴에 속하지 않는다. 현재 이동 패턴의 Jaccard의 비유사도 값은 다음과 같다.

$$dissim_J(X,Y) = \min\left(1-\frac{2}{4}, 1-\frac{3}{5}\right) = \frac{2}{5} = 0.4.$$

그리고 Hall의 유사도 값과 비유사도 값은 각각 다음과 같다.

$$sim_{MP_{i,j}}(X,Y) = \max(4,8) = 8.$$

$$dissim_H(X,Y) = 1 - \frac{8}{10} = 0.2.$$

본 논문에서 제안하는 가중치 기반 비유사도 값은 다음과 같다.

$$dissim_W(X,Y) = \min\left(\frac{0.2+1.0}{4}, \frac{1.0}{4}\right) = \min(0.3, 0.25) = 0.25.$$

첫 번째 예와 같은 비유사도 임계치를 갖는다고 가정하면, $dissim_W(X,Y) > \varepsilon$이므로 그 이동 패턴은 비정상 행위로 식별되고, 경고 메시지가 발생된다. 반면에 $dissim_H(X,Y) \leq \varepsilon$이므로 그 이동 패턴은 정상 행위로 식별되어진다. 따라서 비유사도들 간에는 $dissim_H(X,Y) < dissim_W(X,Y) < dissim_J(X,Y)$이다. 따라서 Jaccard의 방법은 낮은 탐지율을, Hall의 방법은 높은 거짓 경고율을 가질 가능성이 있다.

5. 성능 평가

본 논문에서 제안하는 비유사도 기반 비정상 행위 탐지 방법의 성능을 평가하기 위하여 다음 2가지 척도를 사용한다.

- 탐지율(detection rate): 비정상 행위로 측정된다. m'개의 비정상 행위에 대하여 n개가 비정상인 것으로 탐지되면, 탐지율은 n/m'으로 정의된다.
- 거짓 경고율(false alarm rate): 정상 행위로 측정된다. m개의 정상 행위에 대하여 n개가 비정상인 것으로 식별되면, 거짓 경고율은 n/m으로 정의된다.

우리는 모의실험을 통하여 비유사도 임계치에 따른 제안하는 가중치 기반 비유사도를 사용한 비정상 행위 탐지 방법의 성능을 분석하고 평가한다. 모의실험에서 사용자 이동 패턴이 다음과 같은 경우 그 모바일 노드의 활동은 정상으로 식별되어진다.

표 4.3 이동 패턴 원소 간 비유사도 가중치의 예

		x_i																	
		a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r
	a	0.0	0.2	0.4	0.4	0.6	0.8	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.4	0.6	0.8
	b	0.2	0.0	0.2	0.6	0.4	0.6	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.6	0.4	0.6
	c	0.4	0.2	0.0	0.8	0.6	0.4	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.8	0.6	0.4
	d	0.4	0.6	0.8	0.0	0.2	0.4	0.4	0.6	0.8	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	e	0.6	0.4	0.6	0.2	0.0	0.2	0.6	0.4	0.6	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	f	0.8	0.6	0.4	0.4	0.2	0.0	0.8	0.6	0.4	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	g	1.0	1.0	1.0	0.4	0.6	0.8	0.0	0.2	0.4	0.4	0.6	0.8	1.0	1.0	1.0	1.0	1.0	1.0
	h	1.0	1.0	1.0	0.6	0.4	0.6	0.2	0.0	0.2	0.6	0.4	0.6	1.0	1.0	1.0	1.0	1.0	1.0
y_i	i	1.0	1.0	1.0	0.8	0.6	0.4	0.4	0.2	0.0	0.8	0.6	0.4	1.0	1.0	1.0	1.0	1.0	1.0
	j	1.0	1.0	1.0	1.0	1.0	1.0	0.4	0.6	0.8	0.0	0.2	0.4	0.4	0.6	0.8	1.0	1.0	1.0
	k	1.0	1.0	1.0	1.0	1.0	1.0	0.6	0.4	0.6	0.2	0.0	0.2	0.6	0.4	0.6	1.0	1.0	1.0
	l	1.0	1.0	1.0	1.0	1.0	1.0	0.8	0.6	0.4	0.4	0.2	0.0	0.8	0.6	0.4	1.0	1.0	1.0
	m	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.4	0.6	0.8	0.0	0.2	0.4	0.4	0.6	0.8
	n	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.6	0.4	0.6	0.2	0.0	0.2	0.6	0.4	0.6
	o	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.8	0.6	0.4	0.4	0.2	0.0	0.8	0.6	0.4
	p	0.4	0.6	0.8	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.4	0.6	0.8	0.0	0.2	0.4
	q	0.6	0.4	0.6	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.6	0.4	0.6	0.2	0.0	0.2
	r	0.8	0.6	0.4	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.8	0.6	0.4	0.4	0.2	0.0

1. 현재 사용자 이동 패턴이 그 사용자의 정상 프로파일내의 어떤 이동 패턴과 일치하는 경우
2. 현재 사용자 이동 패턴이 그 사용자의 정상 프로파일내의 어떤 이동 패턴과 1개 또는 2개의 구성요소에서 속도와 방향이 다른 경우
3. 현재 사용자 이동 패턴이 그 사용자의 정상 프로파일내의 어떤 이동 패턴과 1개의 구성요소에서 속도와 방향이 1 단계 다른 경우
4. 현재 사용자 이동 패턴이 그 사용자의 정상 프로파일내의 어떤 이동 패턴과 1개의 구성요소에서 속도가 2 단계 다른 경우

상기의 이동 패턴을 제외한 모든 사용자 이동 패턴은 비정상 활동이라 가정한다. 표 5.1은 모의실험에서 사용된 매개변수와 값을 보여준다.

표 5.1 모의실험 매개변수와 값

매개변수	값
사용자 이동 횟수	150
정상 경로 개수	2
경로 당 정상 패턴의 개수	2
정상 패턴의 구간의 개수	6

그림 5.1은 기준(Baseline)으로 82번의 정상 활동과 18번의 비정상 활동을 갖는 100번의 사용자 이동 활동을 확률적으로 발생시킨 후, 비유사도 임계치 0.1을 적용하여 사용자의 실제 활동 대 비유사도 기반 비정상 활동 탐지 기법들의 예측 활동의 4가지 경우: N/N (Normal/Normal), N/A (Normal/Abnormal), A/N (Abnormal/Normal), A/A (Abnormal/Abnormal)의 발생 횟수를 보여준다. 본 논문에서 제안하는 비유사도 가중치를 사용한 비정상 행위 탐지 기법이 사용자의 정상 활동을 정상으로 확인하는 N/N이 다른 방법들에 비해 매우 높고, 사용자의 정상 활동을 비정상 활동으로 탐지하는 N/A가 다른 방법에 비해 매우 낮음을 알 수 있다.

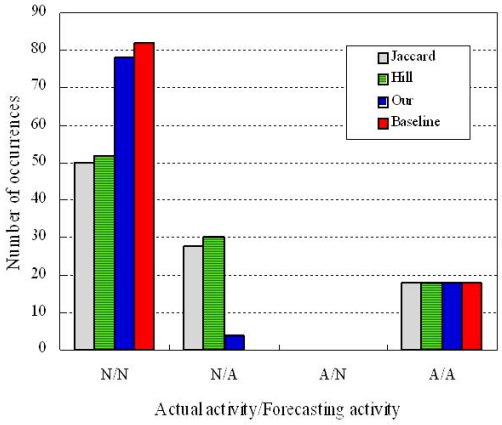

그림 5.1 사용자의 실제 활동 대 예측 활동의 발생 횟수

그림 5.2는 비유사도 임계치에 따른 비유사도 기반 비정상 행위 탐지 방법들의 탐지율과 거짓 경고율의 모의실험 결과를 보여준다. 제안하는 비유사도 가중치를 사용하는 비유사도 기반 비정상 행위 탐지 방법이 비유사도 임계치에 관계없이 거짓 경고율이 다른 방법들에 비해 성능이 아주 우수하고, 제안하는 방법의 탐지율은 비유사도 임계치가 0.1, 0.15인 경우 다른 방법과 같은 성능을 보인다. 그리고 비유사도 임계치가 큰 경우(0.2), 거짓 경고율은 좋아지고 탐지율은 떨어지는 상반관계에 의하여 A/N의 경우가 2번 발생하여 탐지율은 0.923으로 다른 방법들에 비해 조금 낮으나, N/N의 경우가 다른 방법들 보다 16~21번 많다는 것을 알 수 있다(표 5.2 참조). 아울러, 비유사도 임계치를 0.15로 설정하였을 때, 제안하는 방법이 최적의 성능(탐지율 1.0(100%), 거짓 탐지율 0.0(0%))을 갖는다는 것을 확인하였다.

표 5.2 사용자 실제 활동과 예측 활동의 비교 (비유사도 임계치: 0.2)

비유사도 기반 비정상 행위 탐지 방법	Jaccard	Hill	Our
정상 활동 횟수		124	
비정상 활동 횟수		26	
N/N	108	103	124
N/A	16	21	0
A/N	0	0	2
A/A	26	26	24

6. 결 론

본 논문에서는 이동 무선망의 비정상 행위를 효율적으로 식별할 수 있는 사용자 이동 패턴을 이용한 비유사도 기반 비정상 행위 탐지 알고리즘을 제안한다. 제안하는 알고리즘에서는 이동 무선망에서 정상 이동 패턴으로 모바일 노드의 정상 프로파일을 구축하고, 구축된 정상 프로파일로부터 가중치 기반 비유사도 측정을 사용하여 어떤 모바일 노드의 이동 패턴의 정상 행위로부터 비유사도 값을 계산한다. 만일

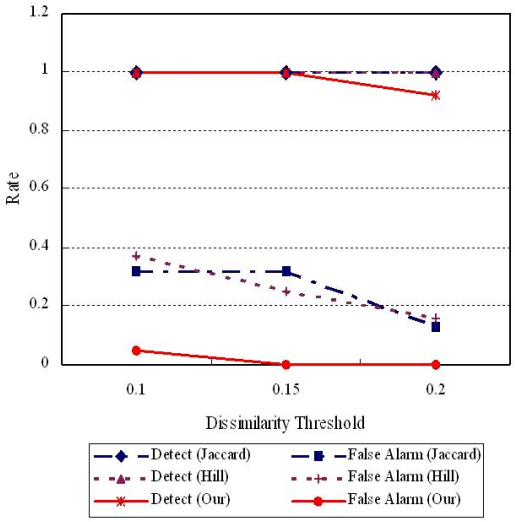

그림 5.2 비유사도 임계치에 따른 탐지율과 거짓 경고율

계산된 비유사도 값이 비유사도 임계치보다 크면, 경고 메시지를 생성한다. 아니면, 정상 활동으로 식별되어진다. 모의실험 결과, 본 논문에서 제안하는 비유사도 가중치를 사용한 비정상 행위 탐지 기법이 사용자의 정상 활동을 정상으로 확인하는 N/N이 다른 방법들에 비해 매우 높고, 사용자의 정상 활동을 비정상 활동으로 탐지하는 N/A가 다른 방법에 비해 매우 낮음을 알 수 있다. 그리고 비유사도 임계치에 관계없이 거짓 경고율이 다른 방법들에 비해 성능이 아주 우수하고, 탐지율은 다른 방법과 거의 같은 성능을 보였다. 아울러, 비유사도 임계치를 0.15로 설정하였을 때, 최적의 성능(탐지율 1.0(100%), 거짓 탐지율 0.0(0%))을 갖는다는 것을 확인하였다.

향후 연구 내용으로는 퍼지 논리를 사용하여 비유사도 가중치를 자동적으로 계산하는 퍼지 논리를 사용한 비유사도 기반 비정상 행위 탐지 방법에 관한 것이다.

참고문헌

이화주, 배인한 (2007). 사용자 프로파일 나이를 고려한 동적 비정상 행위 탐지 방법의 설계 및 평가. <한국데이터정보과학회지>, **2**, 315-326.

Cai, C., Guizani S., Ci, S. and Al-Fuquaha, A. (2006). Constructing an efficient mobility profile of ad-hoc node for mobility-pattern-based anomaly detection in MANET. *GLOBECOM '06*, 1-5.

Henning, C. and Hausdort, B. (2006). Design and dissimilarity measures: a new dissimilarity between species distribution areas. *Data Science and Classification*, Springer, 29-37.

Hall, J., Barbeau, M. and Kranakis, E. (2005). Anomaly-based intrusion detection using mobility profiles of public transportation users, wireless and mobile computing. *WiMob '2005*, **2,** 17-24.

Kachirski, O. and Guha, R. (2003). Effective intrusion detection using multiple sensors in wireless ad hoc networks. *HICSS'03*, 57.

Otey, M. E. and Parthasarathy, S. A. (2005). Dissimilarity measure for comparing subsets of data: application to multivariate time series. *The Fifth IEEE International Conference on Data Mining*, 101-112.

Sun, B., Yu, F., Wu K. and Leung, V. C. M. (2004). Mobility-based anomaly detection in cellular mobile networks, *WiSe '04*, 61-69.

Zhang, Y. and Lee, W. (2000). Intrusion detection in wireless ad-hoc networks. *MobiCom '2000*, 275-283.
Zhang, Y., Lee, W. and Huang, Y-A. (2003). Intrusion detection techniques for mobile wireless networks. *ACM/Kuluwer Mobile Networks and Applications*, **9,** 545-556.

Journal of the Korean Data & Information Science Society
2009, **20**(2), 387–399

한국데이터정보과학회지

Design and evaluation of a dissimilarity-based anomaly detection method for mobile wireless networks

Hwa-Ju Lee[1] · Ihn-Han Bae[2]

[12]Department of Computer and Information Communication, Catholic University of Daegu

Received 21 January 2009, revised 17 March 2009, accepted 22 March 2009

Abstract

Mobile wireless networks continue to be plagued by theft of identify and intrusion. Both problems can be addressed in two different ways, either by misuse detection or anomaly-based detection. In this paper, we propose a dissimilarity-based anomaly detection method which can effectively identify abnormal behavior such as mobility patterns of mobile wireless networks. In the proposed algorithm, a normal profile is constructed from normal mobility patterns of mobile nodes in mobile wireless networks. From the constructed normal profile, a dissimilarity is computed by a weighted dissimilarity measure. If the value of the weighted dissimilarity measure is greater than the dissimilarity threshold that is a system parameter, an alert message is occurred. The performance of the proposed method is evaluated through a simulation. From the result of the simulation, we know that the proposed method is superior to the performance of other anomaly detection methods using dissimilarity measures.

Keywords: Anomaly detection, dissimilarity, mobile wireless networks, mobility patterns, normal profile.

[1] Graduate student, Department of Computer and Information Communication, Catholic University of Daegu, Gyeongbuk 712-702, Korea.
[2] Corresponding author: School of Computer and Information Communication, Catholic University of Daegu, Gyeongbuk 712-702, Korea. E-mail: ihbae@cu.ac.kr