논문 19-45-02-15 The Journal of Korean Institute of Communications and Information Sciences '20-02 Vol.45 No.02
https://doi.org/10.7840/kics.2020.45.2.337

편집거리 알고리즘을 활용한 장애 알람 상관분석 방안

김 상 일◆, 김 화 성°

Fault Alarm Correlation Analysis Using Edit Distance Algorithm

Sang-il Kim◆, Hwa-sung Kim°

요 약

네트워크 환경에서 발생한 장애는 링크나 디바이스 간 영향으로 인해 대량의 알람을 발생시킬 수 있다. 알람 상관 처리는 대량의 알람 사이에서 원인 알람을 찾아내 관리 시스템의 부하를 줄이고 정확한 대처를 가능케 하며 대표적인 방식으로 기존의 케이스를 모방하는 CBR(Case-Based Reasoning) 방식이 있다. 본 논문에서는 네트워크 환경에서 CBR 방식을 활용한 알람 상관분석 시 케이스 간 유사도 계산을 위해 편집 거리 알고리즘을 제안하였으며 성능 평가를 통해 상관분석 정확도가 향상됨을 확인하였다.

Key Words : Correlation, CBR, Fault management, Edit Distance, Levenshtein

ABSTRACT

Any failure in the network environment may cause a great deal of alarms due to the effect between links or devices. The correlated processing of alarms reduces the load of the control system and makes it possible to respond accurately by finding out causes for the alarm among a great deal of alarms and there is the CBR (Case-Based Reasoning) method emulating the existing case as a representative method. This paper suggested an algorithm of editing distance for calculating similarity between cases upon the correlated analysis of alarms by using the CBR method in the network environment and confirmed that the accuracy of the correlated analysis was improved through performance evaluation.

Ⅰ. 서 론

네트워크 환경에서 특정 디바이스나 링크의 장애 발생 시, 해당 장애의 영향으로 인해 발생 원인뿐만 아니라 여러 장소에서 알람이 발생할 수 있다. 따라서 관리자는 장애 관리를 위해 여러 알람들의 상관관계를 파악하고 정확한 진단과 액션을 수행하도록 하는 알람 상관 처리 기술이 필요하다. 특히, 실시간 서비스의 제공 중 장애가 발생 시, 원인 파악과 회복을 신속히 제공하기 위해선 자동화된 알람 상관분석 시스템이 요구된다[1]. 알람 상관분석 시 기존 CBR 방식의 경우, 유사도 계산을 위한 두 케이스의 동일하거나 서로 다른 증상의 횟수에 따라 유사도를 산출한다. 하지만 이러한 방식의 경우, 발생 원인에 따라 달라지는 알람의 순서를 고려하지 않기 때문에 진단의 위치가 잘못될 경우가 존재한다. 본 논문에서는 기존 CBR 방식의 유사도 산출방식을 편집 거리 알고리즘으로 수행한다. 편집 거리 알고리즘은 집합 요소들의 순서를 고려하여 집합 간 유사도를 산출하며, 이때 집합의 요소들은 각 증상 알람들이 활용된다. 제안하는 방식의 성능 평가를 위해 컨테이너 가상화 환경에서 네트워크를 구성하였으며 임의적으로 발생시킨 알람들에 대한 분석 정확도를 측정하였다.

Ⅱ. Case-Based Reasoning

CBR은 유사한 과거 케이스의 해결 사례를 기반으로 새로운 문제를 해결하는 알고리즘을 뜻한다[1]. 관련 연구로서, 2005년 발표된 Fathi의 논문[2]에서는 통신 사업자 환경에서 CBR을 활용한 알람 상관분석

※ 이 논문은 2018년도 광운대학교 교내 학술연구비 지원에 의해 연구되었음
※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 ICT혁신선도연구인프라구축 사업의 일환으로 수행하였음 [2019-0-00260, 초연결 공통 네트워킹 서비스 연구인프라 구축]
◆ First Author : (OCID:0000-0001-7784-0663)Kwangwoon University Department of Electronics and Communications Engineering, rlatkd234@kw.ac.kr, 학생회원
° Corresponding Author : (OCID:0000-0001-5893-5691)Kwangwoon University Department of Electronics and Communications Engineering, hwkim@kw.ac.kr, 종신회원
논문번호 : 201912-335-B-LU, Received December 13, 2019; Revised January 3, 2020; Accepted January 20, 2020

시 유사도 계산 과정을 식(1)과 같이 정의하였다.

$$S_{(Case\,1,\ Case\,2)} = \frac{a^{*}card(E)}{a^{*}card(E)+b^{*}card(D)+c^{*}card(U1)+d^{*}card(U2)} \quad (1)$$

- E : Case1, Case2에 대한 동일한 증상 집합
- D : Case1, Case2에 대한 다른 증상 집합
- U1 : Case1에는 나타나지만 Case2에는 나타나지 않는 증상 집합
- U2 : Case2에는 나타나지만 Case1에는 나타나지 않는 증상 집합
- a=1, b=2, c=1/2, d=1/2

해당 식은 Case 1과 2의 유사도 계산 과정을 보이며, Card는 해당 집합의 개수인 Cardinality를 뜻한다. 장애 발생 시의 각 증상들은 수집된 알람을 의미하며 따라서 해당 유사도 계산 과정은 각 알람들의 차집합, 교집합 개수들을 활용하고 있다.

이와 같은 알람의 개수를 활용한 유사도 계산 과정은 복잡한 연산과정이 없어 빠른 추론이 가능하지만 발생한 장애에 대한 정확한 알람이 도착하기 전까지는 해당 장애를 유추할 수 없다. 예를 들어, 하나의 네트워크 서비스를 수행 중인 4개의 네트워크 기능 중 하나의 장애 발생 시, 해당 장비나 소프트웨어의 결함을 감지하기 전까지는 발생한 증상 알람들만으로 4개의 기능 중 발생 원인을 유추할 수 없다.

Ⅲ. 편집 거리 알고리즘을 통한 케이스 매칭

본 논문은 기존 CBR 방식의 유사도 계산 시 정확도 개선을 위해 편집 거리 알고리즘을 제안한다. 편집 거리 알고리즘은 주로 배열이나 문자열의 유사도를 계산하기 위해 사용되며 Levenshtein 거리 알고리즘으로도 불린다. 편집 거리 알고리즘은 두 문자열의 비교 시, 교체하거나 삭제하여야 하는 문자열의 수에 따라 유사도를 측정한다[3]. 상관분석 과정 시 비교할 문자열은 각 알람들의 ID를 활용할 수 있다. 예를 들어, 2, 31, 24 34의 ID를 갖는 알람들을 갖는 케이스와 31, 24, 5, 34의 알람을 갖는 케이스는 2의 편집 거리를 갖는다. 편집 거리 알고리즘은 식 (2)와 같이 정의된다.

$$D_{a,b}(i,j) = \begin{cases} If\ \min(i,j) = 0 : \max(i,j) \\ Else : \min\begin{cases} D_{a,b}(i-1,j)+1 \\ D_{a,b}(i,j-1)+1 \\ D_{a,b}(i-1,j-1)+1_{a_i \neq b_j} \end{cases} \end{cases} \quad (2)$$

편집 거리 알고리즘은 이전 문자까지의 비교 결과가 다음 문자까지의 거리 계산에 활용된다. 식(2)에서 a와 b 문자열 비교 시, i와 j는 두 문자열의 인덱스가 되며, a의 i 번째 문자와 b의 j 번째 문자가 같을 시 D(i, j)은 D(i-1, j-1) 와 같다.

제안하는 상관분석 시스템은 과거의 케이스 별 편집 거리를 산출하고, 가장 낮은 거릿 값을 갖는 케이스의 회복 액션을 선택하도록 한다. 편집 거리 알고리즘은 기존 CBR 방식과 비교하여 각 케이스의 동일 및 차별된 증상의 횟수뿐만 아니라 순서도 고려할 수 있다. 따라서 비슷한 알람들을 수신하지만 발생 원인이 다른 케이스 또한 구별하여 액션을 선택하도록 할 수 있어 기존 CBR 방식보다 높은 정확도를 가질 수 있다.

Ⅳ. 시뮬레이션 결과

그림 1과 같이 실험을 진행한 네트워크는 8개의 네트워크 기능(NF) 들과 링크로 이루어져 있으며 각 컨테이너에 설치된 네트워크 기능들은 초당 약 50회의 요청을 주고받으며 장애가 발생 시 알람을 전달한다. 테스트를 위해 각 네트워크 기능들과 링크들에 대한 컨테이너 및 소프트웨어 정지, 스트레스 유발로 인한 과부하 장애를 랜덤으로 주입하였다. 전달된 알람들이 하나의 케이스로 규정되기 위해 일정 간격의 대기시간이 요구된다. 케이스 규정을 위한 대기시간이 짧으면 충분한 알람이 수집되지 않아 기존 케이스들과의 유사도가 현저히 낮아진다. 반대로 대기시간이 길 시, 상관분석 시간이 길어져 장애에 대한 대처가 늦어진다. 따라서 정확하고 빠른 장애대응을 위해선 적절한 대기시간의 선정이 중요한 평가 요소가 된다. 그림 2는 대기시간에 따른 기존 CBR 방식(Edit distance

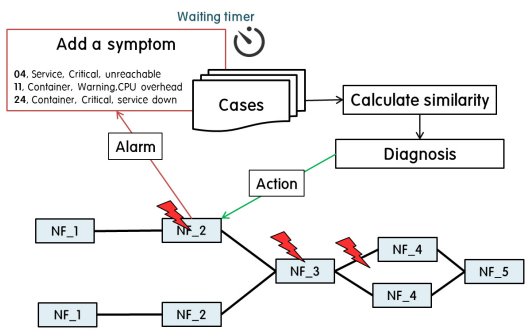

그림 1. 모의실험 시 네트워크 환경
Fig. 1. Network environment in simulation

algorithm) 과 제안하는 방식(Symptom-based algorithm)의 상관분석 정확도를 보이고 있다. 해당 정확도는 각 링크와 네트워크 기능들에 250회의 알람을 주입 후 정확히 진단 한 비율을 보인다. 특히 제안하는 알고리즘은 대기시간이 낮아 충분한 수의 알람이 모이지 않는 경우 15% 이상의 높은 정확도를 갖는다. 편집 거리를 통한 유사도 비교 시, 충분한 수의 알람이 모이지 않아도 알람의 순서에 따라 기존 케이스와 매칭 될수 있는 기회가 존재하기 때문이다.

예를 들어, NF_3에서 컨테이너의 동작 정지로 인한 장애는 해당 컨테이너가 설치된 하드웨어의 장애 감지 시스템에 따라서 알람의 전달이 늦을 수 있다. 편집 거리 알고리즘은 해당 알람이 도착하지 않아도 NF_1,2,4,5에서 수신된 알람의 순서에 따라 NF_3의 장애를 추론해 낼 수 있다.

그림 3은 상관분석 시 또 다른 평가 요소로서 수행 시간의 측정 결과를 보이고 있다. CBR 방식의 유사도 계산 시간은 실시간으로 도착하는 알람의 횟수보다 저장된 케이스의 수에 따라 결정된다. 케이스의 수가 많을 시 참조할 데이터가 많아 상관분석 정확도가 늘어나지만 각 유사도 계산을 위해 케이스를 검색 및 계산하는 CPU 활동이 요구되기 때문에 두 케이스는 Big O(n^2)의 시간 복잡도를 갖는다. 이때, 제안하는 편집 거리 알고리즘은 케이스의 증상 알람의 개수에 따라 추가적인 연산이 요구되기 때문에 기존 CBR 방식보다 높은 연산 시간을 갖지만, 문자열과 달리 장애 발생 시의 알람 수는 비교적 짧은 길이를 갖기 때문에 소요 시간에 큰 차이는 없다.

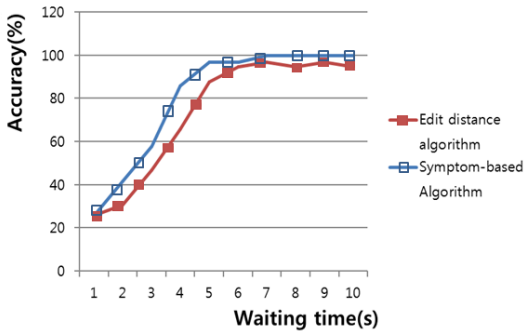

그림 2. 대기시간에 따른 상관분석 정확도 비교
Fig. 2. Comparison of Correlation Analysis Accuracy According to Latency

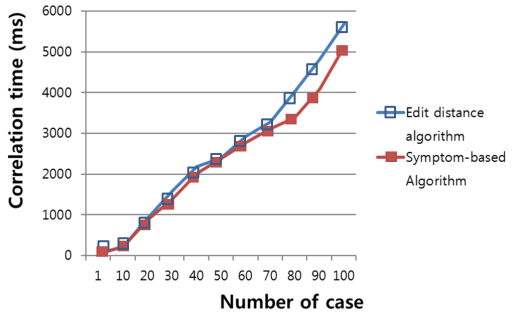

그림 3. 케이스 수에 따른 상관분석 소요시간
Fig. 3. Correlation analysis time by number of cases

V. 결 론

본 논문에서는 네트워크 환경에서 CBR 방식을 활용한 알람 상관분석 시 케이스 간 유사도 계산을 위해 편집 거리 알고리즘을 제안하였으며 해당 알고리즘을 통해 기존 CBR 방식을 이용할 경우 순서를 고려하지 않아 발생하는 문제를 해결하였다. 또한 실험을 통해 비교 알고리즘의 추가로 인한 CPU의 추가 연산 시간도 기존 알고리즘과 큰 차이가 없다는 것을 확인할 수 있었다.

References

[1] G. Jakobson and M. Weissman, “Alarm correlation,” *IEEE Network*, vol. 7, no. 6, pp. 52-59, Dec. 1993.

[2] N. Amani, M. Fathi, and M. Dehghan, “A case-based reasoning method for alarm filtering and correlation in telecommunication networks,” *IEEE Canadian Conf. Electrical and Comput. Eng., 2005,* Saskatoon, Sask., Canada, May 2005.

[3] E. S. Ristad and P. N. Yianilos, “Learning string-edit distance,” *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 20, no. 5, pp. 522-532, 1998.