Journal of Convergence for Information Technology
Vol. 10. No. 8, pp. 15-22, 2020
e-ISSN 2586-4440
DOI : https://doi.org/10.22156/CS4SMB.2020.10.08.015

그래프 데이터베이스 환경에서 이상징후 탐지를 위한 연관 관계 분석 기법

정우철[1], 전문석[2], 최도현[1*]
[1]숭실대학교 컴퓨터학과 학생, [2]숭실대학교 컴퓨터학과 교수

Association Analysis for Detecting Abnormal in Graph Database Environment

Woo-Cheol Jeong[1], Moon-Seog Jun[2], Do-Hyeon Choi[1*]
[1]Student, Computer Science, Soongsil University
[2]Professor, Computer Science, Soongsil University

요 약 4차 산업 혁명과 데이터 환경의 급격한 변화는 기존 관계형 데이터베이스(RDB)는 기술적 한계를 드러내고 있다. IDC/금융/보험 등 전 분야에서 비정형 데이터에 대한 새로운 분석방안으로 그래프 데이터베이스(GDB) 기술에 관심이 높아지고 있다. 그래프 데이터베이스는 상호 연동된 데이터를 표현하고 광범위한 네트워크에서 연관 관계 분석에 효율적인 기술이다. 본 연구는 기존 RDB를 GDB 모델로 확장하고, 새로운 이상징후 탐지를 위해 기계학습 알고리즘(패턴인식, 클러스터링, 경로거리, 핵심추출)을 적용하였다. 성능분석 결과 이상 행위 성능(약 180배 이상)이 크게 향상되었고, RDB로 분석 불가능한 5단계 이후 이상징후 패턴을 추출할 수 있음을 확인하였다.

주제어 : 이상징후 탐지, 그래프 데이터베이스, 그래프 분석, 패턴 분석, 관계형 데이터베이스

Abstract The 4th industrial revolution and the rapid change in the data environment revealed technical limitations in the existing relational database(RDB). As a new analysis method for unstructured data in all fields such as IDC/finance/insurance, interest in graph database(GDB) technology is increasing. The graph database is an efficient technique for expressing interlocked data and analyzing associations in a wide range of networks. This study extended the existing RDB to the GDB model and applied machine learning algorithms (pattern recognition, clustering, path distance, core extraction) to detect new abnormal signs. As a result of the performance analysis, it was confirmed that the performance of abnormal behavior(about 180 times or more) was greatly improved, and that it was possible to extract an abnormal symptom pattern after 5 steps that could not be analyzed by RDB.

Key Words : Anomaly Detection, Graph Database(GDB), Graph Analysis, Pattern Analysis, Relattional Database(RDB)

1. 서론

파나마 페이퍼즈(Panama Papers)는 전 세계 전·현직 지도자, 정치인, 유명인사 등 약 1,150만 건의 기밀 문서를 확보하여 공개한 사건으로 관련 PDF 문서, 내부 이메일, 사진, 그리고 모색 폰세카(Mossack Fonseca)의 데이터베이스 기밀 등이 포함됐다[1]. S Srivastava, L Zhuhadar 등은 기존 RDB로 분석이 어려운 파나마 사건(약 총 2.6 테라바이트)의 방대한 자료를 GDB 분석으로 조세회피 근거를 추적할 수 있음을 연구로 증명했다[2][3]. GDB와 같은 그래프 마이닝 기술은 RDB

*Corresponding Author : Do-Hyeon Choi(cdhgod0@ssu.ac.kr)
Received May 7, 2020
Revised August 1, 2020
Accepted August 20, 2020
Published August 28, 2020

보다 높은 쿼리(Query) 성능과 비정형 데이터의 새로운 데이터 분석 방법을 제공한다[4]. 본 연구는 연관 관계 분석에 효율적인 GDB 기반 이상징후 탐지 방법을 제안한다. 본 논문의 구성은 다음과 같다. 2장은 GDB와 기존 RDB의 문제점, 3장은 GDB 기반 이상징후 탐지/분석 방법을 설명한다. 4장 제안 모델을 검증 및 성능분석, 5장 결론으로 마친다.

2. 관련 연구

2.1 그래프 데이터베이스

Google, Amazon, Microsoft, IBM 등 세계적 기업들은 영상처리, 음성인식 등 4차 혁신기술 시장을 독점하기 위해 인공지능 제품 개발과 플랫폼 구축/운영에 집중하고 있다[5]. 비정형 데이터 분석은 NoSQL, NewSQL 같은 실시간 처리의 특징이 있는 데이터베이스가 요구된다[6]. NoSQL 계열의 GDB는 그래프 이론을 기반으로 RDB를 보완하기 위해 개발되었다[7]. Fig 1은 GDB 모델의 예이다. 사용자의 관계를 정점(Node 또는 Vertex), 성질이 비슷한 객체들의 묶음(Group 또는 Label), 객체 간의 관계를 표시하는 간선(Edge)으로 표현한다[8,9].

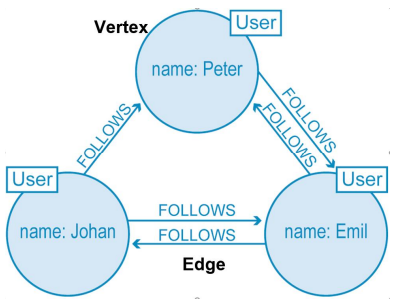

Fig. 1. GDB model expressing user relationship(example)

플랫폼 구성요소는 그래프 저장소, 실시간 그래프 프로세싱 엔진과 내부 전용 쿼리 언어가 필수적이며 RDB와 새로운 모델(키값, 문서, 그래프 등)을 통합하는 멀티모델 데이터베이스 구조가 일반적이다[10]. 내부 기능에는 실시간 그래프 분석과 시각화를 지원하고 다양한 딥러닝 알고리즘을 제공한다. 특히, 키값이나 문서 데이터 모델과 비교하여 연결된 노드나 다중 문서 모델의 관계 분석에 최적화되었다[11]. GDB를 기반 연구 사례에는 패턴 분석을 통한 지식 활용, 그래프 검색, 데이터 시각화 등이 있다[12][13][14]. 그러나 보안 분야에서 실제 실험 데이터를 기반으로 진행된 GDB 연구는 매우 부족하다. 급속도로 변화하는 데이터 분석 시장에서 보안을 위한 그래프 기술의 필요성은 분명 증가할 것이다.

2.2 변화하는 사이버위협과 RDB의 한계

최근 다양한 침해사고 분석 보고서를 통해서 특정 해킹공격의 유사성이 분석되었고 동일 공격 그룹에 의해 수행되었는지 추적하고 있다[15]. 해킹 사례로 Lazarus 그룹은 악성코드를 재사용(변종)하여 필리핀/방글라데시/에콰도르 은행에 약 920억 원을 탈취했다[16]. IBM, Dell, Symantec, Fireeye 등 보안 업계는 신종 악성코드, 패턴 변조 공격, 다중 공격 등 사이버 보안을 위한 위협 인텔리전스(CTI, Cyber Threat Intelligence)의 중요성을 강조하고 있다[17]. CTI 기술은 보안 구조로써 잠재적 혹은 현재 공격과 유사한 APT(Advanced Persistent Threat) 공격 등에 대응할 수 있는 기술이다.

기존 업계에서 사용 비중이 높은 RDB는 정형화된 스키마 구조의 유연성 부족과 비정형 데이터에 대한 성능 저하 등 문제가 존재한다[18]. 테이블(Table) 간에 외부 참조키를 통해 주키 열(Column)에 연결하는 방법은 실제로 많은 계산과 메모리가 필요하다. GDB는 노드(Node)와 관계(Relation)로 연결된 노드의 n depth로 떨어져 있는 다른 노드 검색 성능이 RDB에 비해 훨씬 효율적이다[19]. Fig 2는 기존 RDB와 NoSQL 계열의 GDB의 성능 비교 분석을 나타낸다. 인터넷 기업 라쿠텐에 의하면 MySQL를 NewSQL 계열 데이터베이스로 전환하면 기존 사용 서버 대수를 90%가량 줄일 수 있고, 분산 처리에 샤딩(Sharding)이 필요할 경우 애플리케이션마다 12개월이 걸리는 개발을 한 달 내로 줄일 수 있다고 발표했다[20].

Depth	RDBMS execution time(s)	Neo4j execution time(s)	Records returned
2	0.016	0.01	~2500
3	30.267	0.168	~110,000
4	1543.505	1.359	~600,000
5	Unfinished	2.132	~800,000

Fig. 2. RDB and GDB(Neo4j) Performance Comparison

3. GDB 기반 이상징후 탐지

제안하는 기법은 CTI 위협 대응 플랫폼을 고려하였고, 기존 RDB 데이터 확장과 함께 새로운 GDB 구조를 모델링(Modeling)한다.

3.1 이상징후 탐지/분석 시스템 구조

Fig 3은 이상징후 탐지 플랫폼의 내부 구조를 나타낸다. 정보 수집기(Indicator)는 실제 외부 경로로부터 API, 동기화, Crawing, DNS 쿼리 등 수집할 수 있는 모든 정보 유형들을 의미한다. 기존 알려진 악성코드, 봇넷, 취약점 DB, 악성코드 유사도 정보 등이 있다.

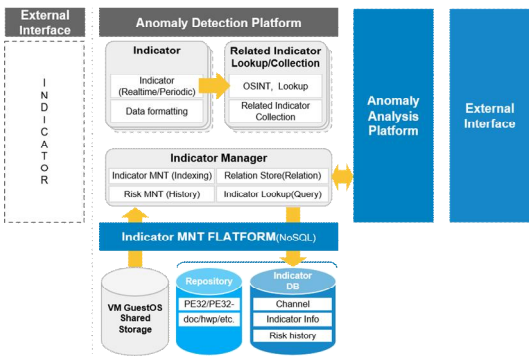

Fig. 3. Anomaly Detection Platform Structure

위협 감시 플랫폼은 실시간으로 수집된 데이터 1차 가공(정형화)을 위한 추가 정보를 검색한다. 정보 수집기 관리자는 하위에 존재하는 기존 RDB 기반의 SQL 엔진과 함께 NoSQL 계열의 그래프 저장소를 독립적으로 저장/실행한다. 내부 저장 플랫폼은 실제 악성코드나 스크립트를 저장하는 공간으로 독립적으로 가상화된 공간을 생성한다. Fig 4는 정보 수집기와 연계되는 이상징후 분석 플랫폼의 내부 구조를 나타낸다. 수집된 정보 유형마다 위협 패턴(Threat)을 검사를 수행한다. 다양한 조합의 위협 시나리오(Intrusion Set)를 설정하여 결과를 도출하고, 위험 수준이 높은 위험들을 예측한다. 하위 구조에는 이상 행위에 대한 위협 수준(Risk)과 우회 기법(Evasion)들을 분석하는 엔진들로 구성된다. 기타 모니터링과 시각화를 위한 외부 인터페이스들로 구성된다.

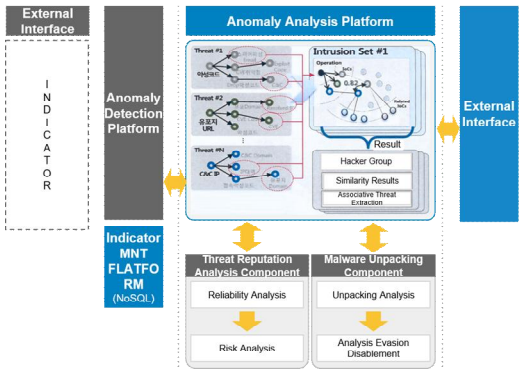

Fig. 4. Anomaly analysis platform Structure

3.2 내부 스키마와 GDB 구조

기존 RDB 기반 정보수집으로 분석할 수 있는 항목은 다음과 같다. Table 1은 이상징후 탐지/분석 테이블의 예를 나타낸다.

Table 1. Anomaly Detection/Analysis Table(Example)

Table Name	Name	Type	Description
observed	sid	Integer	Sequence Number
	rid	Varchar	STIX ID
	ctime	Timestamp	Create Time
	type	Varchar	IP, Domain
	document	JSONB	Body Document
attack-pattern	sid	Integer	Sequence Number
	rid	Varchar	STIX ID
	ctime	Timestamp	Create Time
	indicator	Varchar	Indicator Information
	pattern	JSONB	Pattern Document
intrusion-set	sid	Integer	Sequence Number
	rid	Varchar	STIX ID
	ctime	Timestamp	Create Time
	tid	Integer	threat-actor sid
	threat-actor	Varchar	threat-actor name
	tool	List	Malicious tool
campaign	sid	Integer	Sequence Number
	rid	Varchar	STIX ID
	ctime	Timestamp	Create Time
	indicator	Varchar	Indicator Information
	lifecycle	JSONB	Intrusion Lifecycle
	description	JSONB	Campagn Document

관측데이터, 공격패턴, 악성 도구, 캠페인 테이블은 XML/JSON 기반 STIX(Structured Threat Information eXpression) 스키마를 지원한다. 내부 GDB 형태로 생성하기 위한 식별 정보(ID)와 타임스탬프(Timestamp) 등 기본 데이터를 포함한다. 실제 정보를 연동하는 공격패턴, 생명주기 등은 JSONB 형태로 RDB 관계를 형

성한다. Fig 5는 RDB를 GDB로 확장/생성하기 위한 이상징후 탐지 연관분석 모델링 과정을 나타낸다.

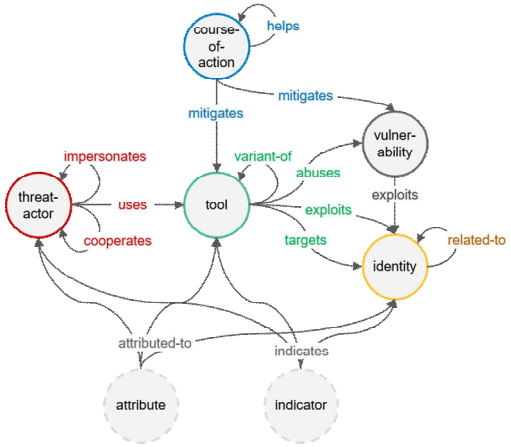

Fig. 5. Association analysis Modeling Process

정보 수집기(indicator)로부터 다양한 정보들의 속성들이 연계되고, 관계 분석을 통해 위협대상으로부터 악성코드를 식별하고 악성코드의 종류와 위협 수준을 판단한다. Table 2는 GDB 기반으로 변환되어 노드가 연결된 테이블 예를 나타낸다. 초기 테이블 생성 이후 내부 정보가 실시간으로 갱신된다. 분석 결과들은 정보 수집기 관리자를 통해 내부 데이터베이스에 저장된다. 위 Fig 5의 모델링 과정에 테이블 변환 연계와 노드 세부 정보를 확인할 수 있다. 위협대상, 악성코드, 위협, 식별자, 시나리오, 속성, 수집기 테이블 등이 존재한다.

Table 2. GDB based Table(Example)

Table Name	Name	Type	Description
threat-actor	uses 〉	tool	Hack S/W, Malware
	impersonates 〉	threat-actor	Bypass / Indirect
	cooperates 〉	threat-actor	Attacker partnership
Tool	abuses 〉	vulnerability	Use of Vulnerabilities
	targets 〉	identity	Target attack
	variant-of 〉	tool	Hack S/W, Malware
	exploits 〉	identity	Information used for attack (or possibility)
Vulnerability	exploits 〉	identity	Information used for attack (or possibility)
identity	related-to 〉	identity	Relationship information between identifiers
course-of-action	mitigates 〉	tool	Specific malicious tool measures
	mitigates 〉	vulnerability	Vulnerability Removal / mitigation
	helps 〉	course-of-action	Complementary measures
attribute	attributed-to 〉	identity	Additional Attribute Information
	attributed-to 〉	tool	Additional Attribute Information
	attributed-to 〉	threat-actor	Additional Attribute Information
indicator	indicates 〉	threat-actor	Specific Attack Types
	indicates 〉	tool	Specific Attack Types
	indicates 〉	identity	Specific Attack Types

3.3 GDB 기반 이상징후 탐지 방법

순위분석(Ranking), 패턴인식(Pattern Detection), 군집분석(Clustering), 경로추출(Path Analysis), 핵심추출(Extract Framework) 알고리즘을 적용한다. Fig 6은 순위분석 알고리즘의 예를 나타낸다.

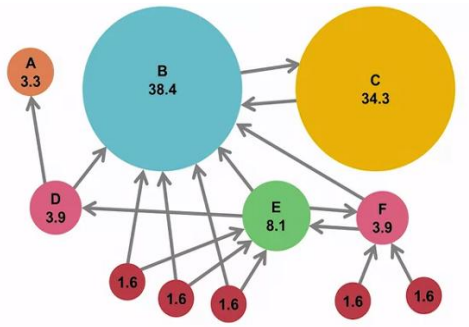

Fig. 6. Ranking analysis Algorithm

그래프 노드 사이의 중심성 척도를 분석한다. 각 노드에서 연관 관계에 대한 임계치를 설정하고, 연결된 우선순위를 추가로 측정한다. 위협대상(A ~ F 대상)에서 비교적 집중되는 위협 항목에 대해 파악할 수 있다. Fig 7은 패턴인식 알고리즘의 예를 나타낸다. 지도 학습(정형 패턴)에서 특징을 추출하는 패턴인식 기법을 선행 수행한다. 예로 GDB 기반 알고리즘 분석에 필요한 바이러스 패턴 정보를 사전 학습에 활용될 수 있다.

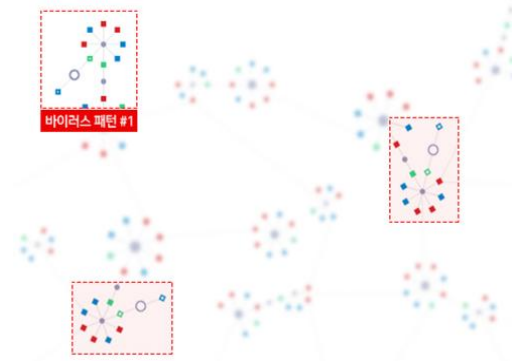

Fig. 7. Pattern Detection Algorithm

Fig 8은 군집 분석 알고리즘의 예를 나타낸다. 이상징후에 대한 유사성 있는 데이터를 묶고, 작은 그룹들로 세분화하여 연관 노드에 대한 특정 그룹을 유추한다. 일반적으로 새로운 데이터가 인접 데이터들과 비교하여 어느 분류에 속할지 과반수로 결정하는 알고리즘을 사용할 수 있다.

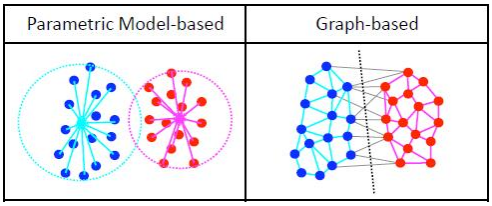

Fig. 8. Clustering Algorithm

Fig 9는 경로분석 및 핵심추출 알고리즘의 예를 나타낸다. 군집화 알고리즘에서 나타난 노드의 시작과 끝의 거리 척도와 최적화 알고리즘을 통해 모든 이상징후 경로들을 예측하고, 이상징후 위협 수준이 높은 요소(강조된 노드)를 추출한다. 이상징후 탐지/분석의 결과로는 위협대상, 악성코드, 위협의 종류, 시나리오(유사성)를 중요 위협의 세부 요인으로 추출한다.

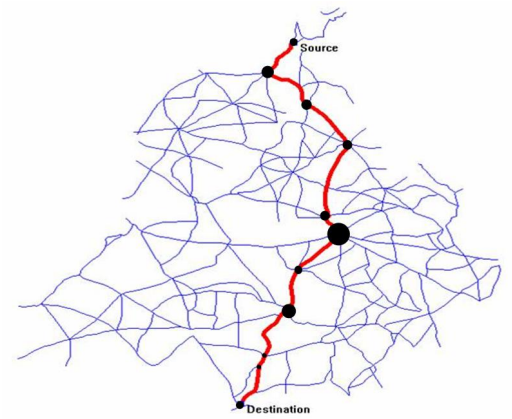

Fig. 9. Path analysis and Extract Framework Algorithm

4. GDB 기반 모델 검증

Fig 10은 이상징후 탐지/분석을 검증하기 위한 시나리오 예를 나타낸다. 알려진 공격 패턴(Attack Pattern)과 침입 시나리오(Intrusion Set)의 유사도와 상호관계를 예측 분석한다. 특정 대상(Threat Actor)으로 유입되는 트래픽에서 정보들을 위협을 식별(Identity)하고, 위협(Treatment)으로 분류된 새로운 패턴들을 추출한다.

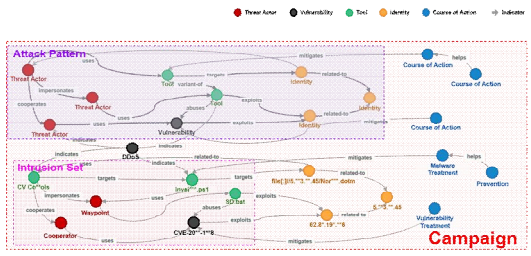

Fig. 10. Anomaly Detection and Analysis(Example)

4.1 순위 분석

Fig 11은 정보 수집기(ID)에서 연결된 전체 노드를 나타낸다.

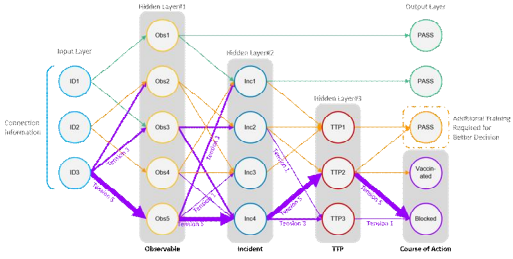

Fig. 11. Ranking analysis Algorithm(Proposed)

입/출력 노드 사이에 존재하는 속성을 하나의 계층으로 구분한다. 접근성(Tension)이 높은 속성들은 접근 경로나 정규식 필터링 검사 등 결과에 따라 임계치 한계의 강도를 결정한다. ID 3의 경우 키워드와 핵심 문장을 추출한 결과 노드 연관성이 순위(Ranking)가 가장 높은 노드를 강조한 예이다.

4.2 패턴 탐지

Fig 12는 새로운 위협 패턴을 추출하는 과정을 나타낸다. 정상 패턴 노드는 제외하고 가장 위험성이 높은 위협 패턴을 추출할 수 있다. 예로 ID 3 노드 연결을 의심 패턴 1로 정의한 후 학습된 데이터베이스로부터 위협 패턴으로 의심되는 노드 연결들을 검사할 수 있다. 각 계층에 존재하는 속성들은 GDB 테이블에 새로 시그니처(Signature)로 저장한다.

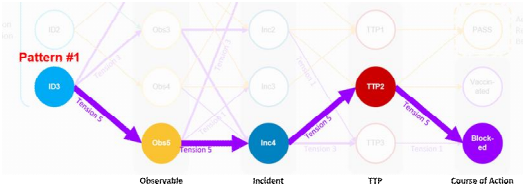

Fig. 12. Pattern Detection Algorithm(Proposed)

4.3 클러스터링 및 군집화 분석

Fig 13은 위협 패턴을 클러스터링(Clustering)하는 예를 나타낸다. 추출된 수만 개의 노드를 모두 분석하는 것은 비효율적이다. 순위 분석과 패턴 탐지 결과 추출된 특정 시나리오(Campaign 1~n)를 재분류한다. 이는 특정 패턴의 대표적인 특징을 추출하고, 명확한 분류 기준이 존재하지 않는 위협 패턴에 대해 유사성(근접한 분포)을 분석하는 작업이다.

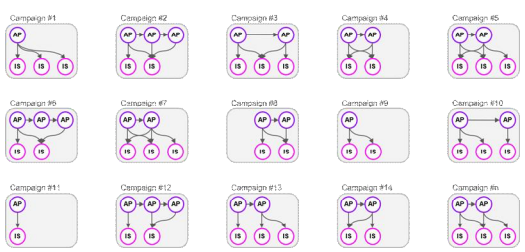

Fig. 13. Clustering Algorithm(Proposed)

4.4 경로 및 핵심추출

Fig 14는 경로 및 핵심추출 과정의 예이다. GDB 연결 노드 그룹과 유사도를 검사하고, 가능성 있는 전체 경로를 위협에 대한 핵심 그룹으로 묶는다. 위험도가 비교적 높은 패턴 1(경로 4)은 경로 1, 3, 2를 향후 위협 패턴으로 예측 분석을 수행할 수 있다.

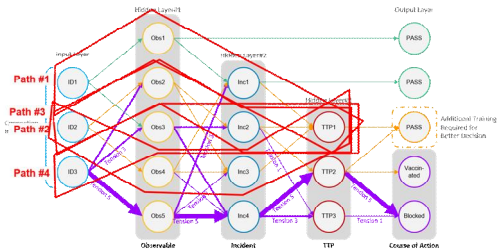

Fig. 14. Path and Extract Framework Algorithm(Proposed)

4.5 GDB 테이블 성능 및 비교 분석

성능 분석 환경에 RDB는 MariaDB, GDB는 Neo4j를 구축하여 비교 분석했다. 사용자 정보 1백만 건과 친구 정보 5백만 건 데이터를 입력하고, SQL과 Cypher 언어로 생성된 스키마의 5개의 쿼리 진행 및 처리 속도를 측정했다. Table 3은 RDB와 GDB 스키마의 성능분석 결과를 나타낸다.

Table 3. RDB and GDB Performance Test

Depth	Execution time(s)		Records returned	Performance difference
	RelationalDB	Graph DB		
2	0.105	0.21	~1900	5.00x
3	55.715	0.323	~125,000	172.49x
4	3080.505	1.932	~700,000	1594.46x
5	Unfinished	3.617	~913,000	unlimited

테스트 결과, SQL 쿼리의 실행/접근/응답 등 전체적으로 성능이 효율적으로 나타났다. 이상 행위 3단계 수준부터 성능 차이가 172배 이상 차이가 존재하고, 이후 단계부터는 기하급수적으로 늘어남을 확인했다. 기본적으로 GDB의 행위 분석 방법이 이상 행위 패턴을 추출하는데 RDB 보다 훨씬 효율적이기 때문이다. Table 4는 RDB와 제안하는 GDB의 비교 결과를 나타낸다.

Table 4. RDB and GDB Comparison Analysis

	RDB	GDB
Model	Single	Multi
Data Type	Document(Only)	Unstructured data
Anomaly Detection	1~4 level(limit)	5 level ~
optimization	Simple query	Complex query
Machine learning	Classification	Clustering
Etc	Universal	Grape Visualization Support

GDB는 RDB를 확장하여 데이터 타입 분석에 한계가 없고 복잡한 쿼리 분석에 적합하다. 기존 RDB에서 적합한 지도 학습의 분류 알고리즘보다 제안하는 클러스터링 알고리즘 등을 기반으로 그래프 네트워크를 구성하는데 최적화되었다. 가장 큰 차이점은 이상징후 5단계 이상 레벨의 쿼리 분석에서, 기존 RDB는 무한 행(Hang)에 빠져 결과를 출력할 수 없다는 사실이다.

5. 결론

본 연구는 새로운 이상징후 탐지/분석을 위한 GDB 기반 장애 분석 방법을 설명했다. 기존 RDB의 한계를 해결하기 위한 기술로 GDB는 비정형 데이터를 통계 분석뿐만 아니라 위협 패턴을 추출/예측할 수 있다. 현실 모형에 가까운 직관적인 형태의 데이터 모형은 개발에 큰 업무 증가율과 시스템의 이해가 빠르다는 큰 장점이 있다. 또한 데이터(Data)를 활용하는 주체인 실무진들이 폭넓게 참여할 수 있다는 가능성을 제시한다. RDB는 범용성과 호환성이 높다는 장점이 있지만, GDB는 플랫폼 구축과 제품 개발 이후 시장 활성화 초기라는 단점이 있다. 그러나 이는 시간이 해결해 줄 것이며 향후 다양한 분야에서 RDB로 분석이 어려운 분야에서 새로운 분석 방법과 전용 그래프 시각화 등 다양한 장점을 제공해 줄 것이다. 본 연구는 지도 학습 수준에서 발전된 비지도 학습을 수행하는 딥러닝 머신러닝(Machine Lerarning) 알고리즘들을 조합/활용하는 수준까지 개선할 계획이다.

REFERENCES

[1] J. Y. Kim & K. H. No. (2019). Construction of Knowledge Base Based on Graph Database for College Student Career Advice Using Public Data, *Journal of the Institute of Electronics Engineers of Korea, 56(10)*, 41-48. DOI : 10.5573/ieie.2019.56.10.41

[2] S. Srivastava & A. K. Singh. (2018). Graph Based Analysis of Panama Papers, *In 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC) IEEE*, 822-827. DOI : 10.1109/PDGC.2018.8745785

[3] L. Zhuhadar & M. Ciampa. (2019). Leveraging learning innovations in cognitive computing with massive data sets: Using the offshore Panama papers leak to discover patterns. *Computers in Human Behavior, 92,* 507-518. DOI : 10.1016/j.chb.2017.12.013

[4] S. M. Bae, J. H. Kim, J. M. Yoo, S. R. Yang & J. J. Jung. (2019). Structural Analysis and Performance Test of Graph Databases using Relational Data. *Journal of Korea Multimedia Society, 22(9)*, 1036-1045. DOI : 10.9717/kmms.2019.22.9.1036

[5] K. Y. Lee, H. R. Kim & J. S. Kim. (2017). AI Platform Solution Service and Trends. *Journal of Korea Bigdata Society, 2(2)*, 9-16. DOI : 10.36498/kbigdt.2017.2.2.9

[6] K. T. Song & S. H. Park (2017). A Recent Trend of Database for Big Data Handling using Key-value database, *Journal of Knowledge Information Technology and Systems, 12(1)*, 47-57. DOI : 10.34163/jkits.2017.12.1.005

[7] N. Roy-Hubara, L. Rokach, B. Shapira & P. Shoval. (2017). Modeling graph database schema. *IT Professional(Magazin) IEEE, 19(6)*, 34-43. DOI : 10.1109/MITP.2017.4241458

[8] R. Angles, M. Arenas, P. Barceló, A. Hogan, J. Reutter & D. Vrgoč. (2017). Foundations of modern query languages for graph databases, *ACM Computing Surveys (CSUR), 50(5)*, 1-40. DOI : 10.1145/3104031

[9] B. M. Sasaki. (2018). *Graph Databases for Beginners: Why Graph Technology Is the Future.* Neo4j (Online). https://neo4j.com/

[10] J. Pokorný. (2019). Integration of relational and graph databases functionally. *Foundations of Computing and Decision Sciences, 44(4)*, 427-441. DOI : 10.2478/fcds-2019-0021

[11] K. Wongsuphasawat et al. (2017). Visualizing dataflow graphs of deep learning models in tensorflow. *IEEE transactions on visualization*

and computer graphics, 24(1), 1-12.
DOI : 10.1109/TVCG.2017.2744878

[12] J. Y. Kim, K. H. Ro. (2019). Construction of Knowledge Base Based on Graph Database for College Student Career Advice Using Public Data, *Journal of the Institute of Electronics and Information Engineers, 56(10)*, 41-48.
DOI : 10.5573/ieie.2019.56.10.41

[13] U. C. Park. (2020). Is-A Node Type Modeling Methodology to Improve Pattern Query Performance in Graph Database. *Journal of The Korea Society of Computer and Information, 25(4)*, 123-131.
DOI : 10.9708/jksci.2020.25.04.123

[14] U. C. Park. (2017). Visualization of Recommendation Items Based on Graph Database, *Journal of Korean Institute of Information Technology, 15(6)*, 1-9.
DOI : 10.14801/jkiit.2017.15.6.1

[15] S. M. Park & J. I. Lim. (2017). Study On Identifying Cyber Attack Classification Through The Analysis of Cyber Attack Intention. *Journal of The Korea Institute of Information Security and Cryptology, 27(1)*, 103-113.
DOI : 10.13089/JKIISC.2017.27.1.103

[16] M. Abd Majid & K. Z. Ariffi. (2019). Success Factors for Cyber Security Operation Center (SOC) Establishment. *Conference: Proceedings of the 1st International Conference on Informatics, Engineering, Science and Technology*.
DOI : 10.4108/eai.18-7-2019.2287841

[17] W. Tounsi & H. Rais. (2018). A survey on technical threat intelligence in the age of sophisticated cyber attacks, *Journal of ScienceDirect(Computers & security), 72*, 212-233.
DOI : 10.1016/j.cose.2017.09.001

[18] J. S. Lee & S. C. Hong (2014). Study on the Application Methods of Big Data at a Corporation-Cases of A and Y corporation Big Data System Projects. *Journal of Internet Computing and Services, 15(1)*, 103-112.
DOI : 10.7472/jksii.2014.15.1.103

[19] D. Fernandes & J. Bernardino. (2018). Graph Databases Comparison: AllegroGraph, ArangoDB, InfiniteGraph, Neo4J, and OrientDB, *Conference: 7th International Conference on Data Science, Technology and Applications*, 373-380.
DOI : 10.5220/0006910203730380

[20] Ryutaro Yada. (2012). How Rakuten Reduced Database Management Spending by 90% through Clustrix implementation, *Database Platform Group Global Infrastructure Development Dept. Rakuten*, tech showcase(Onlie). https://global.rakuten.com/

정 우 철(Woo-Cheol Jeong) [정회원]

· 2008년 2월 : 동서울대학교 컴퓨터소프트웨어학과 졸업
· 2010년 8월 : 숭실대학교 컴퓨터학과(공학석사)
· 2016년 3월: 숭실대학교 컴퓨터학과 박사 수료

· 관심분야 : Mobile, Network Security, Big data, Deep Learning
· E-Mail : jwc0116@gmail.com

전 문 석(Moon-Seog Jun) [정회원]

· 1981년 : 숭실대학교 전산학과 학사
· 1986년 : University of Maryland 전산학과 석사
· 1989년 : University of Maryland 전산학과 박사

· 1989년 ~ 1991년 : New Mexico State University Physical Science Lab 책임연구원
· 1991년 ~ 현재 : 숭실대학교 컴퓨터학과 정교수
· 관심분야 : 정보보안, PKI, 전자여권, 암호학
· E-Mail : mjun@ssu.ac.kr

최 도 현(Do-Hyeon Choi) [정회원]

· 2008년 2월 : 동서울대학교 컴퓨터소프트웨어학과 졸업
· 2010년 8월 : 숭실대학교 컴퓨터학과(공학석사)
· 2016년 3월 : 숭실대학교 컴퓨터학과(공학박사)

· 관심분야 : Mobile, Network Security, PKI, Virtualization
· E-Mail : cdhgod0@ssu.ac.kr