Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.25, NO.4, Aug. 2015 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2015.25.4.767

수사단서를 이용한 동일 사이버범죄 판단기법

김 주 희†‡
경기지방경찰청 사이버수사대

Technique for Indentifying Cyber Crime Using Clue

Ju Hee Kim†‡
Gyeonggi Provincial Police Agency

요 약

최근 몇 년간 스마트폰의 보급률이 폭발적으로 증가하면서 사이버범죄는 기존의 수사체계의 한계를 넘어서는 새로운 형태의 수사단서들을 쏟아내고 있다. 일선 경찰관서에서는 사건 접수 시 피해자로부터 이러한 형태의 수사단서를 수집하여 방대하게 축적하고 있으나, 이를 체계적으로 관리하고 있지 않아 많은 데이터 속에서 이것이 내포하고 있는 숨은 의미를 지나치는 경우가 많다. 사이버범죄에서 주 범행 도구인 컴퓨터 시스템의 특성상 기계적이고 복잡한 단서가 대량 생성되므로, 수집된 수사단서를 체계적으로 분류, 단순화하여 분석할 필요가 있다. 본 논문에서는 국내에서 발생되는 사이버범죄 유형에 따른 수사단서를 체계적으로 분류, 단순화하여 주요수사단서를 선정하고, 데이터 마이닝 및 시각화를 통해 사건 수사단서 간 상호 연관성을 확인할 수 있었다. 이러한 사이버범죄 데이터 활용을 통해 범죄 조기차단 및 중복수사를 방지하여 수사의 효율성을 증대하고 사이버범죄 예방을 도모하고자 한다.

ABSTRACT

In recent years, as smart phone penetration rate is growing explosively, new forms of cyber crime data is poured out beyond the limits of management system for cyber crime investigation. These new forms of data are collected and stored in police station but, some of data are not systematically managed. As a result, investigators sometimes miss the hidden data which can be critical for a case. Crime data is usually generated by computer which produces complex and huge data and records many logs automatically, so it is necessary to simplify a collected data and cluster by crime pattern. In this paper, we categorize all kinds of cyber crime and simplify crime database and extract critical clues relative to other cases. Through data mining and network-visualization, we found there is correlation between clues of a case. From this result, we conclude cyber crime data mining helps crime prevention, early blocking and increasing the efficiency of the investigation.

Keywords: Cyber Crime, Crime data, Data mining, Visualization, Crime pattern

I. 서 론

범죄자는 범행현장에 흔적을 남기고 수사관은 범죄자가 남긴 수사단서 수집을 통해 범인을 추적한다. 경찰청에서는 이렇게 수집된 수사단서를 형사사법정보시스템에 범죄데이터정보, 범죄수사정보[1] 형태로 저장하며 이를 활용하여 각종 강력범죄, 교통사고, 지능범죄들을 해결하고 있다. 하지만 일반 범죄를 기준으로 구축한 형사사법정보시스템은 수사단서를 체계적으로 관리하지 못하고 있으며, 특히 신종범죄가 빈발하는 사이버범죄에 대해서는 발생유형과 수사단서를 의미 있게 관리하지 못하고 있다.

변화가 빠른 정보통신 환경에서 충분한 보안성 검증 없이 새로운 형태의 서비스가 출시되면, 필연적으

접수일(2014년 12월 12일), 수정일(2015년 7월 8일), 게재확정일(2015년 7월 9일)
† 주저자, favorpol@police.go.kr
‡ 교신저자, favorpol@police.go.kr(Corresponding author)

로 서비스의 허점을 노린 사이버 범죄가 뒤따르기 마련이다. 정보통신 서비스의 허점을 노린 범행 현장에서 수집된 수사단서는 현재의 형사사법정보시스템에서 제대로 축적되지 못하고 있다. 신종 사이버 범죄의 피해자들은 새로운 형태의 수사단서를 수사기관에 제공하나, 이러한 수사단서는 특성에 맞게 저장되지 않고 있어 범죄정보는 넘쳐나지만 정작 피의자 추적에는 제대로 활용되지 못하고 있다.

사이버범죄는 컴퓨터의 빠른 전파성, 시간·공간의 무제약성, 자동성으로 인해 동일 피의자 또는 범죄조직에 의해 동시다발적으로 발생되는 경우가 많다. 일반 범죄는 시간·공간의 한계로 인해 피해자가 소수이지만 사이버 범죄의 피해자는 개별 지역에 국한되어 있지 않고 전국적으로 흩어져 있다. 예를 들어 인터넷 사기 범죄조직이 1시간 동안 100건의 사기사건을 저지를 경우 100명의 피해자는 각자 자신의 거주지 관할 경찰서에 사건을 접수하게 되며, 이로 인해 동일한 사건을 100명의 수사관이 중복수사하게 된다.

이러한 중복수사의 방지를 위해 현재 경찰에서는 대규모로 피해가 발생된 인터넷 쇼핑몰사기나 인터넷 물품사기 사건의 경우, '인터넷 물품사기 이송에 관한 규칙'에 의해 범행에 사용된 계좌 개설 금융지점 관할 경찰서로 사건을 이송하거나 상부기관의 집중수사 지시에 따라 특정 경찰관서가 해당 사건을 도맡아 수사한다. 그러나 이는 단순히 인터넷을 통한 전자상거래 사기사건에 국한되어 있을 뿐 그 외의 신종금융범죄, 조건만남 사기사건의 경우 정해진 이송규칙이 없거나, 사건들 간 동일 피의자 사건으로 결정하기 위한 수사단서 선정에 어려움이 있어 집중수사를 못하고 있다. 게다가 최근에는 인터넷 사기 역시 대포통장, 대포폰 등을 활용한 조직적 범죄가 발생되는 경향이 있어, 계좌정보를 기준으로 선정된 집중수사 체계에 한계가 나타나고 있다.

이러한 문제를 해결하기 위해서 우리는 축적된 사이버범죄 수사단서 및 사건정보에 데이터마이닝 기법을 적용하여 적극 활용할 필요가 있다. 수집된 정보나 구축된 데이터베이스에서 의미 있는 정보를 가려내는 기법 중 하나가 바로 데이터마이닝이다. 데이터마이닝은 대량의 데이터로부터 지식을 추출하는 과정 또는 기법을 말하는 것으로, 쌓여가는 데이터 속에서 의사결정에 필요한 중요한 정보를 파악하고, 각 데이터간의 패턴을 인식하는 과정이다[2]. 범죄수사에서도 이러한 기법을 이용하여 주요한 수사단서 간의 결합 또는 분류를 통해 의미 있는 정보를 추출하여 적극 활용한다면 효율적인 수사가 가능하다.

본 논문은 이러한 관점에서 사이버범죄의 유형을 체계적으로 분류하고 각 유형에서 주요한 수사단서를 선정한 후, 수사단서를 결합하였을 때 별개의 사건이 동일 사건으로 군집화 됨을 보이고자 한다.

II. 사이버 범죄의 유형 별 조사 체계 고찰

2.1 사이버 범죄 유형과 수사단서

사이버범죄는 한 가지 수법이 아닌 여러 가지 수법이 결합된 형태로 발생되는 경우가 많으며 이는 유형화를 어렵게 하는 요인으로 작용한다[3]. 게다가 컴퓨터가 일반 대중의 삶에 폭넓게 자리하면서 컴퓨터의 특수성을 이용한 범죄 외에도 차츰 일반 범죄의 발생장소가 사이버 공간으로 옮겨 가고 있다. 경찰청 사이버안전국에서는 2014년 6월부터 사이버범죄의 주요 수단인 정보통신망 침입여부 및 그 내용을 기준으로 Table 1과 같이 분류하고 있다.

정보통신망 이용범죄 중 인터넷 사기, 사이버금융범죄는 다른 범죄와 달리 직접적인 금전 피해를 발생시킨다. 경찰청 사이버안전국의 2013년 사이버범죄 통계자료[1]에 따르면 인터넷 사기는 지난 10년 동안 매년 줄어들지 않고 지속적으로 발생되고 있으며, Fig.1.과 같이 2013년 한해 발생된 사이버범죄 총

Table 1. Cyber Crime types

Network intrusion	• Hacking • Cyberterrism(DDos) • Virus spreading • Etc.
Crime on Internet	• Internet fraud • Online financial crime • Identity theft • Copyright Infringement • Spam mail • Etc.
Illegal contents	• Distributing pornography • Illegal online gambling • Online defamation • Cyber stalking • Etc.

1) 경찰청 사이버안전국, http://cyberbureau.police.go.kr/share/sub3.jsp

86,103건 중 인터넷 사기가 39,282건으로 전체 사이버범죄 중 45.6%나 차지하는 것으로 확인되었다[2]. 인터넷 상 개인 간 물품거래 활성화, 휴대폰 소액결제, 모바일 뱅킹, 전자결제, 가상계좌 등 새로운 형태의 결제시스템의 도입으로 인해 사이버 상 전자상거래는 날로 발전하고 있는 반면, 이로 인한 범죄 역시 끊이질 않고 있다.

인터넷 사기는 다른 범죄의 발생건수에 비해 상당히 많이 발생한다. 그래서 언뜻 많은 수의 범죄자가 범죄를 일으키는 것처럼 보이나 실제로 수사를 하다 보면 사실 소수의 피의자 또는 전문 범죄조직에 의해 사건이 대량으로 발생되는 경우가 많다. 여기서 주목해야 할 점은 동일 조직의 피의자에 의해 발생된 인터넷 사기 사건이라면 그들이 남긴 수사단서 역시 비슷한 패턴 또는 중복되는 단서가 존재할 가능성이 있다는 점이다. 따라서 인터넷 사기와 사이버 금융범죄에는 어떠한 유형들이 있고, 각 유형은 어떤 수사단서를 포함하고 있는지 분석할 필요가 있다.

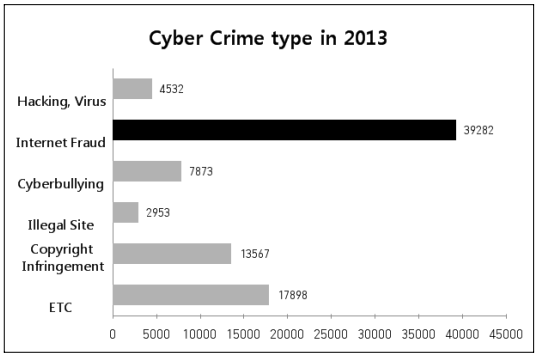

Fig. 1. Cyber Crime record in 2013

2.1.1 인터넷 사기

인터넷 사기란 정보통신망(컴퓨터 시스템)을 통하여, 이용자들에게 물품이나 용역을 제공할 것처럼 기망하여 피해자로부터 금품을 편취한 행위를 말한다. 인터넷 사기의 유형으로는 직거래 사기, 쇼핑몰 사기, 게임 사기, 인터넷 조건만남사기, 기타인터넷 사기 등이 있다. 피의자는 사기 범행을 위해 피해자와 연락하는 통신수단과, 편취한 재물을 보관할 매개체가 필요하다. 통신수단은 주로 전화번호, 인터넷 계정, 스마트폰 메신저, SNS 등이 있으며, 주로 사용되는 재물보관 매체로는 금융계좌, 가상계좌, 사이버머니, 게임 아이템 등이 있다.

2.1.2 사이버 금융범죄

사이버 금융범죄란 정보통신망(컴퓨터 시스템)을 통하여, 타인을 기망·공갈함으로써 재산상의 이익을 취하거나 제3자에게 재산상의 이익을 취하게 하기 위하여, 자금을 송금·이체하도록 하는 행위 또는 개인정보를 알아내어 자금을 송금·이체하는 행위를 말한다. 사이버 금융범죄의 유형으로는 피싱, 파밍, 스미싱, 메모리해킹, 기타 사이버금융범죄가 있다. 사이버금융범죄의 경우 단순 인터넷 사기와는 달리 범행 전 정보통신망을 통해 기술적 범행이 선행되어야 하는데 이는 주로 타인계정 침입, 악성코드 유포를 통해 이루어진다. 사이버 금융범죄의 수사단서로는 위와 같은 기술적 범행에 사용되는 악성코드, 범행을 통해 편취한 금원을 보관할 계좌 또는 사이버머니가 충전 가능한 게임계정 등의 매체정보가 있으며, 그 밖에도 스미싱에 사용된 문자메시지, 파밍에 사용된 도메인 등이 있다.

2.1.3 그 외 정보통신망 이용사기

최근에는 단순 스미싱을 넘어서 다른 유형의 범죄와 결합한 형태의 사건이 많이 발생되고 있다. 예를 들어 피해자의 스마트폰에 악성앱을 유포, 설치하여 신체 주요 부위의 노출을 유도하여 촬영한 노출사진을 빌미로 금품을 요구하여 협박하는 몸캠피싱[3]이 바로 그것이다. 몸캠피싱은 사이버 금융범죄로 분류하기에는 무리가 있으나 정보통신망을 이용하여 금융피해를 발생시키는 행위이므로 이 역시 데이터 마이닝 대상 범주에 속한다고 볼 수 있다. 그 밖에도 피의자는 위에서 언급한 단서 외에 다양한 물적, 인적 단서를 남기게 되는데 그 내용은 Table 2와 같다.

2) 2013년 발생된 범죄 발생유형의 경우 〔Table 1〕에 따른 분류 이전으로, 인터넷 사기는 사이버금융범죄와 인터넷 물품사기를 포함하고 있으며 스미싱 사기는 기타 범죄에 포함되었다.

3) 몸캠피싱(phishing cam) : 불상의 여성이 불특정 남성을 상대로 스마트폰 음란 영상 채팅을 유도하여 피해자 몰래 스마트폰에 악성코드를 설치하여 피해자의 주소록을 탈취한 후 영상을 유포하겠다고 협박해 돈을 요구하는 범죄수법

Table 2. Clues of Internet Fraud and New Financial Fraud

	Internet Fraud	New Financial Fraud, Phishing Cam
Physical clue	account, site, ID, SNS profile, goods, location, phone number, telephone conversation, IP, money, ATM code, CCTV, sms, Date/Time, crime site	apk sample, apk name, apk hash, phone number, sms, C&C IP, domain, account, IP, telephone conversation, site, ID, phone number, money, ATM code, CCTV, Date/Time, crime site
Human clue	suspect victim testifier (usually owner of physical clue)	

2.2 주요 수사 단서 선정

피의자가 남긴 수사단서는 다양하나 실제 수사기관의 형사사법정보시스템에 저장되어 하나의 데이터베이스 스키마로서 관리되는 단서는 그 중 일부에 불과하다. 이러한 수사단서를 형사사법정보시스템에 모두 입력하여 관리하기 어려운 이유 중 하나는 데이터 수집의 비정확성 때문이다. 수사기관이 수집하는 데이터는 범죄자나 사건 관련자들로부터 수집된 정보이거나 과거 발생된 자료이다. 그런데 범죄자들은 유리한 입장에 있기 위해 거짓말을 하고, 허위 정보를 유출하기도 한다. 이는 피해자의 경우도 마찬가지이다. 자신의 피해 품목을 과대 계상한다든지, 순간적으로 목격한 피의자를 다른 형태로 기술한다든지 하는 경우가 많다. 이렇게 부정확한 상황에서 과거 발생된 사안을 현재의 입장에서 수집하기 때문에 범죄나 용의자 데이터에 대한 정보가 부정확한 경우가 많다 [2].

이런 한계점을 극복하기 위해서는 수집된 수사단서에 대한 무결성 및 신뢰성 입증이 우선이다. 예를 들어 은행에서 발급된 계좌이체확인증, 인터넷 거래정보내역, 통신회사에서 발급된 SMS 수신내역, 발신통화내역, 전문수사관에 의해 작성된 악성코드 분석보고서 등을 통해 피해자 또는 피의자로부터 수집된 단서 중 입증할 증거가 존재하여 신뢰성이 보장된 단서에 대해서는 이를 관리 가능한 형태의 정보로 변환하여 저장하여야 한다.

수사관이 수사 활동 중에 얻게 되는 수사단서도 있다. 공문을 통해 타 기관으로부터 수집한 수사단서, 탐문수사 및 소재수사 등으로 알게 된 수사단서, 피해금원 추적 중 알게 된 범인의 자금 세탁 경로 및 그 이면에 감춰진 지하조직 등의 정보이다. 이런 종류의 정보는 관리가 어려울 수 있다. 예를 들어 피의자의 가정환경, 피의자의 출신학교, 공범 피의자들의 주요 배회지, 그 외 피의자의 프라이버시 등은 전자적인 기호로 표현하기 어렵다. 그럼에도 불구하고 수사단서로서의 의미가 있는 정보에 대해서는 이 역시 전자 정보로 변환하는 노력이 필요하다.

의미 있는 범죄정보 추출을 위해서는 이렇게 수집된 수사단서 중 주요한 데이터를 선택, 데이터 전처리, 데이터 변환을 거친 후 데이터 마이닝을 하게 되는데, 본 연구에서는 Table 2에서 언급한 수사단서 중 유일한 값 또는 고유한 값으로 인식될 수 있는 정보를 선택하여 본격적으로 데이터 마이닝을 하고자 한다. 우선 고유한 값을 가지는 수사단서로는 계좌번호, 출금지점 금융기관코드, 전화번호, IP, 주민등록번호(피의자 또는 참고인), apk hash 값 등이 있다. 또한 고유한 값은 아니나 고유한 값의 수사단서와 결합하여 사용할 수 있는 수사단서인 apk명, 사이트, 아이디를 주요 수사단서로 선택하고자 한다. Table 3은 각 사이버범죄 유형별 주요 수사단서를 선정한 결과이다.

Table 3. Critical clues for investigation

	Internet Fraud	New Financial Fraud	Phishing Cam
account	√	√	√
ATM code	√	√	√
phone	√	√	√
IP	√	√	√
suspect testifier	√	√	√
apk name	-	√	√
apk hash	-	√	√
site	√	√	√
ID	√	√	√

2.3 동일 범죄 여부 판단 방안

현재 경찰청에서 시행하고 있는 '인터넷물품사기의 이송규칙'에 의하면 인터넷물품사기는 범행에 사용된 계좌 개설 금융지점 관할경찰서로 사건을 이송하여 집중 수사한다. 하지만 최근 발생되고 있는 인터넷물품사기는 1명의 피의자에 의해 발생되기 보다는 분업화된 범죄 조직에 의해 발생되며, 하나의 계좌가 아닌 여러 개의 계좌를 이용하는 경향이 있다. 하나의 거대 물품사기 조직이 여러 개의 대포통장을 이용하여 범행을 하면, 동일 피의자 사건임에도 범행수단인 계좌의 개설금융지점에 따라 서로 다른 경찰서에서 동일 사건에 대해 수사하게 된다.

이는 사이버금융범죄에서도 역시 동일한 경향을 보인다. 사이버금융범죄의 경우에는 이송규칙이 따로 없어 해당 사건을 접수한 경찰서에서 수사를 하고 있다. 하지만 사건을 실제로 수사하다 보면 서로 다른 피해자의 계좌에서 동일한 피의자의 접속 IP, 동일한 포털 계정이 사용되는 경우가 많다.

또한 파밍 사기를 당한 피해자의 피해금원은 금융기관에서 개설된 계좌가 아닌 게임 사이트에서 생성된 가상계좌로 이체되는가 하면, 동일금융계좌가 조건만남사기와 파밍 사기에 중복되어 사용되는 경우도 종종 발생된다. 때로는 이러한 계좌가 중국 환전상의 계좌로 밝혀지기도 한다.

이와 같이 현재 사건 병합의 기준이 되고 있는 계좌는 더 이상 유일한 단서가 아니며, 동일 계좌에 따른 이송규칙에 의해 오히려 동일 인터넷사기 범죄가 전국에 흩어져 있는 여러 명의 수사관에 의해 중복수사되는 결과를 초래하고 있으며, 신종금융사기 범죄 역시 병합기준이 없는 탓에 중복수사가 이루어지고 있다. 본 연구에서는 이러한 중복수사를 방지하기 위해 1개의 수사단서에 한정하지 않고 Table 3에서 선정한 주요수사단서간의 결합을 통해 사건 간 하나 이상의 중복되는 수사단서를 가지는 경우 동일 범죄로 판단할 것을 제안한다. 또한 동일 범죄로 판단된 사건의 실제 연관성 역시 확인해 보고자 한다.

Ⅲ. 사이버 범죄 조사 체계 개선 방안

3.1 자료 수집 및 정제

본 논문에서는 수사단서 결합을 통한 동일범죄 판단실험을 위해 2013. 11월경부터 2014. 1월경사이 실제 일선 경찰서 사이버수사팀에 접수된 인터넷 물품사기사건 중 피의자 AXX와 연관된 사건정보와 국내 최대 인터넷 사기 검색 사이트인 더치트[4], 국내 포털 사이트인 네이버 카페 중고나라(cafe.naver.com/ joonggonara)에서 위 사건과 관련한 사건 정보를 검색, 수집하였으며, 이를 바탕으로 각 경찰서의 사건번호(C)를 기준으로 인적사항과 전화번호, 계좌번호 등은 모두 개인정보보호를 위해 가명으로 변경하고 날짜는 재구성하여 Table 4와 같이 총 13개 사건의 사건정보를 정리하였다.

실제 위 사건들은 접수될 당시 인터넷물품사기 사건으로 취급되어 최초 접수관서의 담당 수사관들에 의해 범행에 사용된 계좌의 개설금융지점 관할경찰서(취급관서)로 이송되었으며, 이를 이송 받은 취급관서의 수사관들은 이송 받은 사건간의 연관성을 자체 고려하여 여러 개의 접수사건을 하나의 사건번호(C)로 병합, 입건하였다. 그 결과 Table 4에서는 하나의 사건에 여러 개의 금융계좌, 전화번호, 아이디 정보를 포함하고 있다. 또한 Table 4의 사건정보는 사건 접수 시 수집된 수사단서와 사건종결 이후 확인된 수사단서가 모두 결합된 정보이다. 접수 시 수집된 단서는 계좌번호, 아이디, 전화번호이고, 수사종결 후 수집된 단서는 피의자, 참고인(범행에 사용된 매체의 명의자)등이다.

Table 4에서 수집된 정보들 중 신종금융범죄, 몸캠피싱에서도 공통적으로 수집되는 수사단서와 Table 3에서 선정된 주요수사단서 중 그 값이 유일하고, 디지털정보로 변환가능하며, 사법기관 또는 민간사이트에서 관리가 되고 있는 항목 중 각각의 사건을 구별할 수 있는 주요수사단서 항목을 추출하여 다시 Table 5와 같이 병합 정리하였다. Table 5에서 선정된 피의자, 참고인, 계좌, 계좌주, 전화번호, 인터넷 계정은 인터넷물품사기 뿐만 아니라, 그 외 신종금융사기, 몸캠피싱 사건에서도 반드시 수집되는 정보이며, Table 3에서 선정한 주요 수사단서 중 IP, 출금지점코드(ATM CODE)는 수사과정에서 수사관이 이와 같은 정보를 수집하게 되더라도 사법정보시스템에 정형화된 기록으로 남기거나 이를 관리하지 않고 있으므로 Table 5에서는 포함하지 않았으며, 계좌 명의자로 확인된 참고인에 대해서는 중복

4) 2006년 1월 4일 비영리로 개설된 국내 최초의 사기피해정보공유 사이트이며, 사기피해사례 공유를 통한 사기피해 재발방지 및 피해자간 공동대응을 목적으로 운영되고 있다.

Table 4. Crime Data(raw data)

Case	Date	Dept.	Suspect	Testifier	Account	Account Owner	Phone	Site	ID
C1	2014-01-23	Police Station A	AXX BXX CXX DXX EXX	FXX GOO	WOORI 2009196992***	GOO	010-5814-XXXX 010-2982-XXXX 010-5833-XXXX 010-2652-XXXX	Naver	wlsxxx wlgxxx dltxxx kejxxx
C2	2013-11-04	Police Station B	AXX	BXX	NH 1390911329***	BXX	010-4092-XXXX	Naver	
C3	2013-12-28	Police Station C	AXX JXX	BXX HXX	IBK 322920932*** KB 292602029***	HXX	010-5099-XXXX 010-9445-XXXX	Naver	solxxx kimxxx
…	…	…	…	…	…	…	…	…	…

Table 5. Crime data(only critical clue)

Case	Suspect	Testifier	Account	Account Owner	Phone	ID
C1	AXX BXX CXX DXX EXX	FXX	WOORI 2009196992***	GOO	010-5814-XXXX 010-2982-XXXX 010-5833-XXXX 010-2652-XXXX	wlsxxx wlgxxx dltxxx kejxxx
C2	AXX		NH 1390911329***	BXX	010-4092-XXXX	
C3	AXX JXX	BXX	IBK 322920932***	HXX	010-5099-XXXX 010-9445-XXXX	solxxx kimxxx
…	…	…	…	…	…	…

단서 제거를 위해 참고인 항목에서 제외하였다.

3.2 데이터 마이닝 기법 선택

데이터 마이닝은 대량의 데이터 분석을 통해 우리가 알지 못했던 의미나 유용한 패턴을 얻는 과정으로 개념/클래스 서술, 빈발패턴/연관성/상관성마이닝, 분류와 예측, 군집분석, 이상치분석, 전개분석 등의 기능이 있다.

본 연구에서는 형사사법시스템에서 처리되지 않던 수사단서를 주요수사단서로 선정한다. 그리고 사건 간 중복 값을 측정하여 동일 범죄군으로 판단하고자 한다. 주요수사단서로 선정된 수사단서들은 연속형보다는 범주형 데이터가 대부분을 차지하고 있다. 이러한 수사단서의 특성에 따라 데이터 마이닝 기법 중 비지도 학습(unsupervised learning) 및 여러 유형의 데이터 속성을 다룰 수 있는 계층적 군집화 분석 방법(Hierarchical clustering)을 이용하여 실험한다.

본 연구에서 실험할 피의자 AXX과 연관된 사건정보 데이터베이스의 특성은 하나의 사건에 여러 개의 수사단서가 종속되어 있으며, 각 수사단서는 서로 다른 속성(attribute)으로 존재하고 있다. 각 수사단서 값은 연속형 변수(int)가 아닌 고유한 변수(varchar)이며 k개의 사건 개별 데이터의 총 개수는 'k × 각 사건에 종속된 수사단서의 개수'가 된다. 군집 분석 시 이러한 데이터 형태를 데이터 행렬(Data Matrix)[5]이라고 하는데, 데이터 행렬 간

5) 데이터행렬(Data matrix, 객체×변수의 구조): 이것은 사람의 경우 나이, 키, 몸무게, 성별, 인종 등과 같은 p개의 변수들을 나타낸다. 구조는 관계형 테이블로 column과 row의 속성이 다르며, n × p 행렬의 형태이다[4].

연관성을 측정하기 위해 가장 효과적인 군집화 기법은 바이클러스터링(biclustering)이다.

바이클러스터링은 1972년 J.A. Hartigan에 의해 'Direct Clustering'라는 개념으로 처음 소개가 된 군집화 기법[6]으로, 전통적인 군집분석이 Fig.2.(a)와 같이 특징적인 객체군 혹은 속성군을 찾는 분석방법인 것과 달리, 바이클러스터링은 객체와 속성을 동시에 하나의 군집으로 묶어 Fig.2.(b)와 같이 특성화된 그룹을 추출하는 기법이다. 따라서 전통적인 군집분석에서는 군집 내에 속한 객체들이 모든 속성에 대해 비슷해야 하는 반면, 바이클러스터링에서는 일부 속성에 대해 매우 유사한 성향을 보이는 객체군이라면 나머지 속성들에 대해서는 유사성을 보이지 않더라도 해당 객체-속성군을 의미가 있는 군집으로 간주하고 추출해낸다[5].

바이클러스터링은 유전자정보공학(bioinfomatics)에서 사용되기도 하는데 주로 DNA 마이크로어레이(DNA Microarray)[6]의 분석에 사용된다. 각 유전자는 여러 개의 DNA 속성(condition)으로 구성되어 있으며, 각각의 유전자는 일부 DNA에서 서로 동일한 DNA 속성을 가지기도 하는데 이러한 유전자 간 DNA 연관성 분석을 위해 바이클러스터링 기법을 적용하기도 한다.

이러한 바이클러스터링 개념을 유전자 데이터를 표현에 처음 적용한 사람은 Cheng and Church[7]이다. Tanay et al.는 Cheng and Church가 제시한 개념에 대해 더 나아가 그래프-이론적으로 접근하였다. Fig.3.의 오른쪽 그래프에서와 같이 DNA 마이크로어레이에서 Gene(객체)와 Conditions(속성)를 이분그래프(bipartite graph)[7]로 표현하고 각각의 Gene과 Condition을 유전자 특성에 맞게 연결한 후, 많은 Gene이 연결되어 있는 바이클러스터링 군집(가중치 있는 군집)에 대해서는 하위그래프(Sub-graph)로 따로 추출하여 중요한 유전자 군집으로 다루었다[8].

Fig. 3.은 이스라엘의 텔아비브 대학교(Tel-Aviv University) Benny Chor 교수가 학생들에게 유전자의 바이클러스터링 군집화와 이분그래프 개념 설명을 위해 작성한 강의 자료로, 실제 유전자 데이터 행렬의 바이클러스터링 과정을 자세히 설명하고 있다.[8]

Data Matrix M의 데이터행렬로 구성된 Gene(객체)에서 각 행의 DNA는 A-F까지의 Condition(속성)을 가지게 되는데, 바이클러스터링 기법을 적용하게 되면 각 DNA가 서로 완벽하게 일치하지 않더라도 일부 Condition이 동일한 값을 가지는 경우 Sub-Matrix를 이루게 된다. 이를 Graph G와 같이 이분그래프화 하게 되면, Gene집합과 Condition집합 간 연결되는 경우 연결선(edge)을 그려주고, 연결성이 높은 바이클러스터링 군집에 대해서는 Sub-graph 로 추출하여 의미있는 정보로 활용

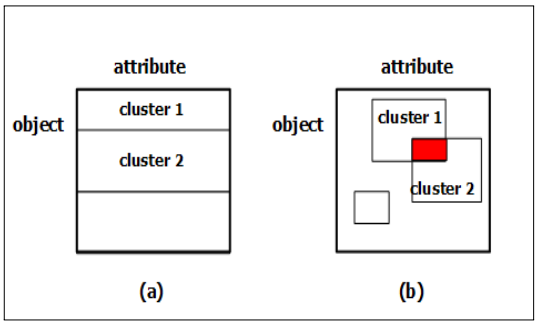

Fig. 2. Comparision of traditional clustering(a) and biclustering(b)

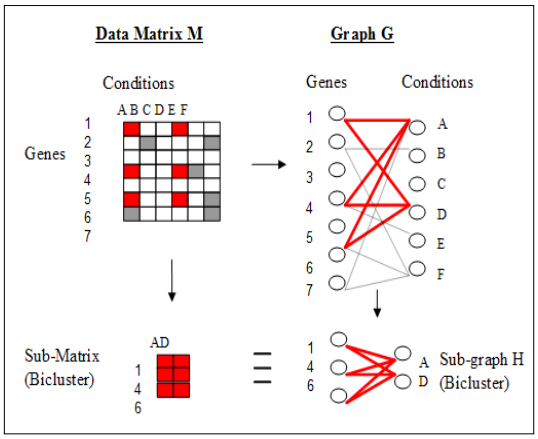

Fig. 3. Biclustering data matrix of genes

6) DNA 마이크로어레이(DNA microarray): DNA Chip이라고도 부르며, 복잡한 DNA구조를 고정된 chip에 미세한 간격의 점으로 배열의 형태로 부착한 조합으로 인종, 암세포 등의 DNA 분석에 널리 이용되고 있다. (http://en.wikipedia.org/wiki/DNA_microarray)

7) 집합 내 교집합이 없는 서로소 집합(U, V)에 대해 서로 접점이 있을 경우 U의 원소와 V의 원소 간에만 연결을 하는 무방향 그래프.

8) Http://www.cs.tau.ac.il/~bchor/CG09/CG10_mic roarrays.ppt (Benny Chor, School of Computer Science, Tel-Aviv University)

할 수 있다.

사이버범죄에서도 유전자와 같이 각 사건(객체)은 다양한 수사단서(속성)를 포함하고 있으며, 수사단서가 100% 일치하는 사건은 극히 일부에 불과하다. 본 연구에서는 사건에 대해 바이클러스터링 군집화 기법을 적용하여 중복된 수사단서가 있는 경우 동일 범죄로 판단하기로 한다. 그러나 Fig.3.의 Graph 바이클러스터링을 보면 객체와 속성(node)이 늘어날수록 연결선(edge)도 기하급수적으로 늘어나게 되므로, 대량의 사건 정보를 투입할 경우 복잡도가 높아져 사건의 군집파악이 어려울 수 있다. 이에 본 연구에서는 Fig. 4.와 같은 과정을 거쳐 사건 간 동일 수사단서를 가지는 경우 수사단서(속성)의 표현은 생략하고 사건 간 연관성에 대해서만 연결선(edge)으로 표현하기로 한다.

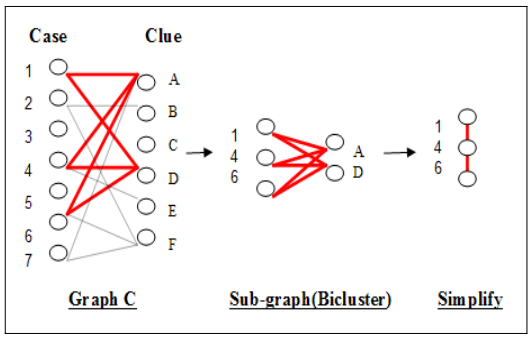

Fig. 4. Simplify after biclustering cases and clues

3.3 수사단서를 활용한 집중 수사 체계 제안

Table 5에서 선정한 사건의 주요수사단서를 활용하여 동일 수사단서를 가진 사건에 대해 수사단서가 중복될수록 서로에 대한 중복 값을 두어 그 값의 높고 낮음에 따라 동일 사건 여부를 판단한다. 사건 간 중복 값을 계산하기 위해서는 Algorithm 1을 이용하였다.

Table 6은 실제 Table 5에서 제시된 사건 13개에 대해 Algorithm 1 을 적용하여 각 사건 간 중복되는 단서 하나당 1씩 더하여 계산한 표이다. 각 사건은 서로에 대해 같은 중복 값을 가지므로 음영이 있는 경우에 한해서만 계산하였으며, 1의 결과 값을 가지는 사건들은 피의자 AXX 외에는 서로 중복단서가 존재하지 않음을 의미한다.

Algorithm 1
Require: char Case[i][j]
(i: total count of case)
Ensure:

```
Let D be 2-dimensional arrays indexed from 1 to i
for a in range(i):
    for b in range(i-1):
    tot = 0
        for c in range(j-1):
      if Case[a][c+1] == Case[b+1][c+1]:
            tot = tot + 1
        D[a][b+1] = tot
return D
```

Table 6에서 중복단서 값을 보면 동일 피의자 AXX 정보 외에도 절반 이상의 사건들은 서로 중복 수사단서를 가지고 있는 것으로 확인된다. 그 중 C4-C8 사이의 중복단서 값은 5로 가장 많은 중복 값을 가지는 것으로 확인되었으며, 5개의 중복 값은 모두 피의자 정보가 중복되는 것으로 확인되었다.

하지만 Table 6에서 C4-C8의 중복 값은 수사종결 이후에야 알 수 있는 정보들로 여죄 확인에는 도움이 되나 사건을 최초 접수하는 단계에서는 피의자 정보를 알 수 없으므로 동일 범죄 여부를 판단하기 어렵다. 이에 수사종결 후에 확인되는 피의자, 참고인 정보와 기존의 동일사건 판단 근거인 계좌정보를 제외하고 나머지 수사단서인 계좌명의자, 전화번호, 아이디에 대해 각 사건 간 연관성을 계산하면 Table 7과 같다.

Table 6. Numbers of duplicate clue between cases

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
C1	0	2	2	3	4	2	2	4	1	1	2	2	2
C2	0	0	2	1	3	2	2	1	1	1	1	1	1
C3	0	0	0	2	2	2	2	2	1	1	1	2	2
C4	0	0	0	0	1	2	1		1	1	1	2	4
C5	0	0	0	0	0	3	2	1	1	1	3	1	1
C6	0	0	0	0	0	0	2	1	1	1	2	1	1
C7	0	0	0	0	0	0	0	1	1	1	1	1	1
C8	0	0	0	0	0	0	0	0	1	1	2	3	3
C9	0	0	0	0	0	0	0	0	0	1	1	1	1
C10	0	0	0	0	0	0	0	0	0	0	1	1	1
C11	0	0	0	0	0	0	0	0	0	0	0	2	1
C12	0	0	0	0	0	0	0	0	0	0	0	0	4
C13	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 7. Duplicates between cases except suspect, testifier and account

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
C1	0	0	0	0	1	0	0	0	0	0	0	0	0
C2	0	0	0	0	1	0	1	0	0	0	0	0	0
C3	0	0	0	0	0	0	0	0	0	0	0	0	0
C4	0	0	0	0	0	0	0	1	0	0	0	0	0
C5	0	0	0	0	0	1	0	0	0	0	2	0	0
C6	0	0	0	0	0	0	0	0	0	0	1	0	0
C7	0	0	0	0	0	0	0	0	0	0	0	0	0
C8	0	0	0	0	0	0	0	0	0	0	0	0	0
C9	0	0	0	0	0	0	0	0	0	0	0	0	0
C10	0	0	0	0	0	0	0	0	0	0	0	0	0
C11	0	0	0	0	0	0	0	0	0	0	0	0	0
C12	0	0	0	0	0	0	0	0	0	0	0	0	2
C13	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 7을 보면 사건 수사 중에 밝혀진 피의자, 참고인, 계좌정보를 제외하고도 일부 사건에서 수사단서가 중복되고 있음을 확인할 수 있다. 하지만 이러한 표를 이용한 사건 간 연관성 파악은 사건의 수가 늘어날수록 오히려 비효율적이며, 전체 사건 간의 연결고리를 추적하기에는 직관성이 떨어지는 단점이 있다.

이에 Table 5에서 확인된 13개의 사건에 대해 Python NetworkX[9]를 이용하여 Fig.5.과 같이 그래프로 나타냈다. 각각의 사건은 노드(node)로 표현되며, 사건 간 수사단서가 중복될 경우 선(edge)으로 연결하였다. 그 결과 A(C1-C2-C5-C6-C7-C11)와 B(C4-C8-C13-C12)에서 군집이 확인되었다. 결과적으로 각 군집 내 사건은 서로 다른 형태의 수사단서를 공유하고 있으나 시각화를 통해 동일 범죄의 군집인 것으로 확인되었다. 실제로도 해당 사건의 공범 피의자들은 각자 다른 전화번호와 아이디를 사용하며, 여러 개의 계좌를 공동으로 사용한 것으로 확인되었다.

이러한 조사를 통해 동일 피의자 또는 조직이 범행할 때 이용한 계좌 정보 외에 전화번호, 아이디 등의 수사단서를 통해서도 동일 범죄 파악이 가능함을 알 수 있다. 이와 같은 동일 범죄 파악기법을 사건 접수초기단계에 활용한다면 13개의 경찰관서에서 별

9) Networkx는 그래프 등을 그려주는 시각화 툴로 데이터 구조 및 데이터 간 역학관계 연구 및 복잡한 네트워크를 위한 함수개발을 위한 파이썬(python) 기반의 소프트웨어 패키지이다. http://networkx.lanl.gov

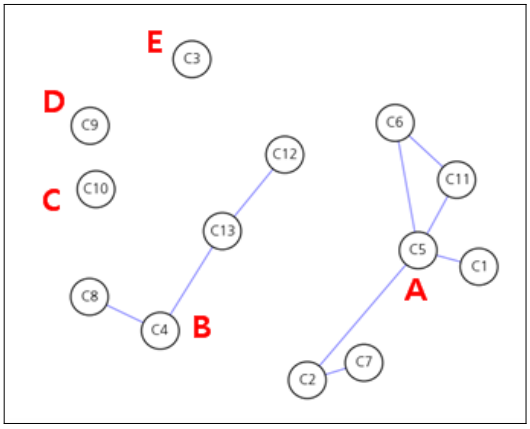

Fig. 5. Clustering cases by clue(except account)

도로 처리되던 사건을 최소 5개의 사건(A, B, C, D, E)으로 줄일 수 있으며 중복수사의 수고를 덜 수 있을 것이다.

3.4 실제 데이터 적용 사례

3.4.1 다수의 대포통장 이용 인터넷 사기사건 연관성 분석

본 논문에서는 수사단서 연관성에 대한 네트워크 추적기법의 유효성 검증을 위해 2014년 1월 한달간 더치트에 등록된 인터넷물품사기사건 총 2,144건에 대해 사건정보를 수집하여 시각화하였다. 등록된 사건 내 많은 정보 중 제2장에서 선정한 인터넷물품사기 주요수사단서인 계좌, 아이디, 전화번호에 대해서만 추출하였으며, 더치트에 등록된 게시물 번호를 하나의 사건(node)으로 표현하고, 사건 간 중복되는 수사단서가 있는 경우 연결(edge)하였다.

Fig. 6.(a)는 현재 경찰청 인터넷물품사기 이송규칙에 따라 동일한 계좌를 수사단서로 가진 사건을 시각화한 결과 그래프이다. 많은 수의 사건 중 몇 개의 군집들이 확인됨을 볼 수 있다. Fig.6.(b)는 피의자가 범행에 사용한 아이디가 동일한 사건 간 연결을 하여 시각화한 결과이다. 동일계좌정보로 연결한 그래프 보다 사건 간 군집이 약하게 형성되었지만, 아이디도 동일 범죄 판단의 주요수사단서로서 활용할 수 있음을 보여주고 있다. Fig.6.(c)는 범행에 사용된 전화번호가 동일한 사건 간 연결을 하여 시각화한 결과이다. 기존의 수사방법에서 중요치 않게 생각했던 수사단서인 전화번호에 대해 시각화 한 결과는 의

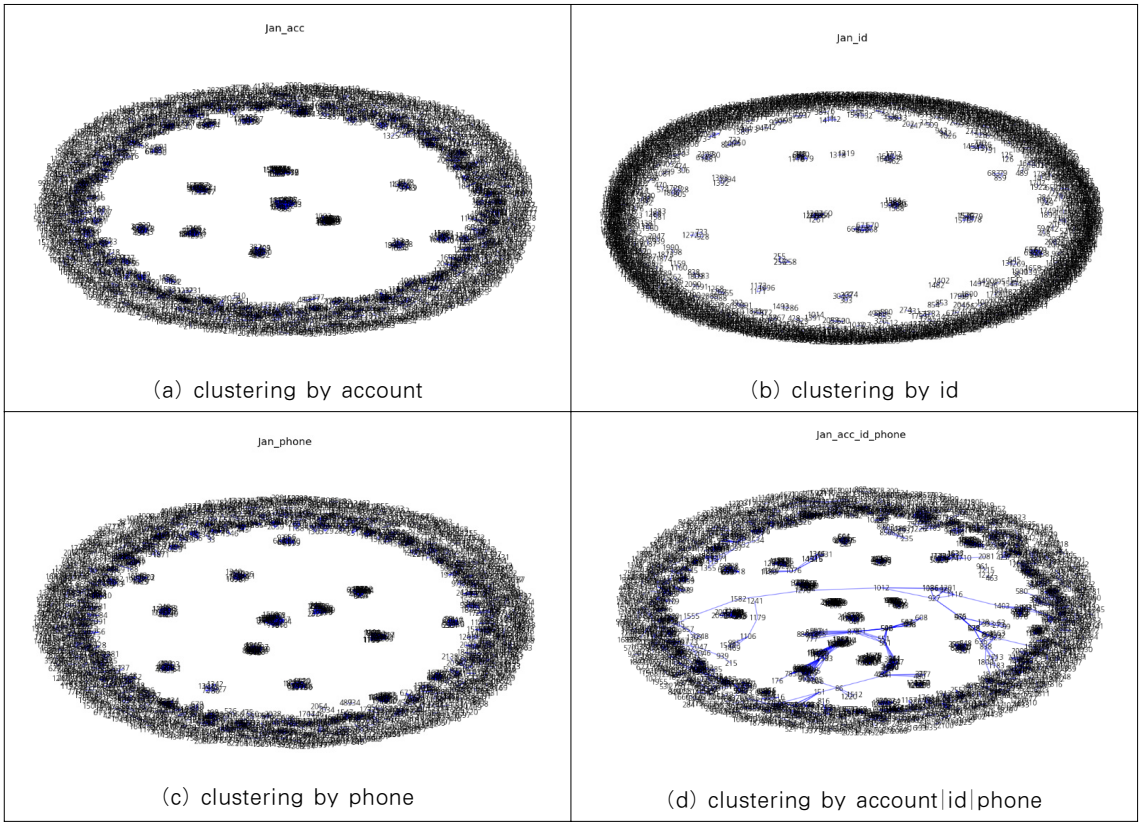

Fig. 6. Clustering cases of Thecheat in Jan 2014 by clues

외로 계좌정보를 이용한 시각화의 결과와 비슷할 정도의 사건 간 결속력을 보여주었다. Fig.6.(d)는 계좌번호, 아이디, 전화번호 중 중복 수사단서가 하나 이상 존재할 경우 선으로 연결한 결과이다. 그래프를 보면 (a), (b), (c)에서 보이지 않았던 군집들이 갑자기 출현된 것을 볼 수 있으며, 이는 사건 간 하나의 특정 수사단서가 아닌 여러 수사단서가 유기적으로 연관성을 가지고 있음을 설명해 준다.

Fig.7.은 Fig.6.(d)에서 확인된 군집들 중 확연히 군집을 이루고 있는 A군집, B군집에 대해 선택하여 확대한 그림이다. 해당 그래프 내 각 사건들은 강한 결속력을 가지고 있는 다른 군집들과는 달리 군집 내에서도 또 다시 서로 다른 수사단서에 의해 하위레벨의 군집을 이루고 있다.

Fig.7.의 2개의 군집 중 B군집의 계좌정보와 전화번호를 통해 확인한 결과 해당 군집은 실제로 모두 동일 피의자 수법으로 발생된 인터넷물품사기로 확인되었다. 이 사건의 특징은 피의자들이 휴대폰 대리점을 사칭하며 최신형 스마트폰을 판매할 것처럼 피해자들을 기망한 것으로, 1~2일 사이에 하나의 계좌당 최대 30명 이상의 피해자를 발생시킨 사건이다. 특히 거래물품의 금액이 고가로 피해자 1인당 피해금액이 상당하였으며, 그로 인하여 하나의 계좌로 인해 발생된 피해금액의 총합이 평균 8백만원 이상에 달하는 것으로 확인되었다. 피의자들은 이러한 수법으로 몇 년에 걸쳐 수 십 개의 대포통장을 이용하여 범행을 저질렀으나, 각 계좌의 개설지점이 전국적으로 흩어져 있는 이유로, 경찰청 인터넷 사기 이송규칙에 의해 각기 다른 경찰관서에서 수사가 진행되었을 것으로 추정된다. 이러한 이유로 피의자들이 다시 새로운 대포통장을 이용하여 범행에 착수한다 하더라도 해당 사건을 접수한 경찰서에서는 피의자들의 범죄 상습성을 알아채기 어려울 것이다. 하지만 Fig.6.(d)와 같이 주요수사단서를 활용한 사건 간 연관성을 분석하면 숨겨져 있던 여죄를 파악할 수 있다.

Fig.8.는 위 Fig.7.의 B군집에서 확인된 35건의 사건 간 어떤 수사단서가 연관성이 있는지에 대해 시

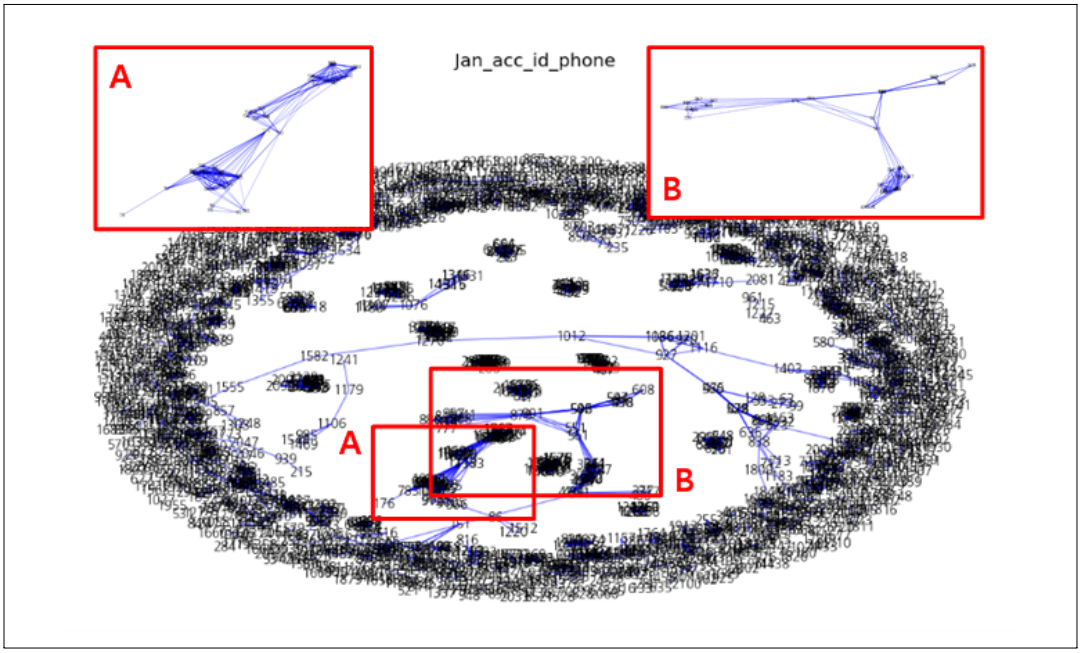

Fig. 7. Extract specific clusters from graph

각화한 결과이다. 각 사건은 노드(node)로 표현하고, 사건 간 동일한 값을 가지는 수사단서는 각 속성에 따라 다양한 선(edge)의 종류로 표현하였다. B 군집에 속해있는 사건들의 수사단서 값을 확인한 결과 총 9개의 금융계좌, 13개의 아이디, 7개의 전화번호가 사용된 것으로 확인되었다. 또한 노드 내 숫자가 높을수록 나중에 발생된 사건으로, 노드 내 숫자를 통해 시간요소를 표현하였으며, 실제 각 사건의 발생일자는 Fig.9.와 같다.

Fig.8.에서 형성된 그래프 내 노드를 보면 오른쪽의 낮은 숫자에서 왼쪽의 높은 숫자로 이어지고 있으며, 노드 간 연결선(edge)의 종류를 통해 사건 간 각자 다른 수사단서 중복에 의해 연결되어 있음을 알 수 있다. A 영역의 사건들은 모두 동일 계좌 값을 가지고 있으며 B 영역에 들어서면 서로 같은 ID, phone 값을 가진 사건들이 군집으로 나타난다. 그러나 B 영역을 기점으로 새로운 계좌정보가 출현하기 시작한다. 사건 361, 374, 344, 464, 427은 새로운 계좌정보를 가지며 이들은 다시 C 영역에서 사건 551, 561과 동일한 계좌정보를 가지는 군집을 이룬다. 사건 551, 561부터는 또 다시 새로운 ID 정보가 출현하고 있으며, D, E, F를 경계로 나뉘는 사건들은 같은 아이디를 수사단서로 가지고 있으나 각자 다른 계좌정보를 가지고 있다.

종합적으로 보면 사건들은 최초 계좌(점선으로 연결)로 연결되었다가 다시 아이디와 전화번호로 연결(실선으로 연결)이 이어지다가 재차 계좌로 연결되는 흐름을 반복하고 있다. 이와 같은 데이터 마이닝 과정을 통해 다량의 대포통장을 사용하는 피의자들이 새로운 대포통장으로 범행도구를 변경하더라도, 각 수사단서의 변경시점은 서로 다른 패턴을 보이므로 계좌 외의 수사단서를 활용하면 피의자들의 범행을 놓치지 않고 추적할 수 있다.

Fig.9.는 사건을 발생일자별로 나열한 그래프로, 전체 범행기간 중 몇 일간 범행을 하지 않은 시점(2014. 1. 10 ~11)이 확인되는데, 이 기간은 주말이어서 피의자들이 범행을 잠시 멈추었다는 시간 별 행동패턴을 파악할 수 있다. 이처럼 데이터에서 시간에 따라 반복 측정되는 값이나 사건의 순차에 따라 데이터마이닝을 하는 방법을 시계열 데이터마이닝이라고 하는데, 본 연구에서 선정한 주요수사단서 외에도 사건정보와 같은 다양한 정보들이 범죄패턴을 파악하는 단서가 될 수 있음을 알 수 있다.

본 사건처럼 단기간에 대량의 피해가 발생되는 경우 피의자 검거도 중요하지만, 추가 피해를 차단하는 것 역시 필요하다. 현재 경찰에서는 상습사기에 사용된 것으로 확인된 금융계좌에 대해서는 금융기관을 상대로 지급정지를 요청할 수 있다. 만약 사건을 접

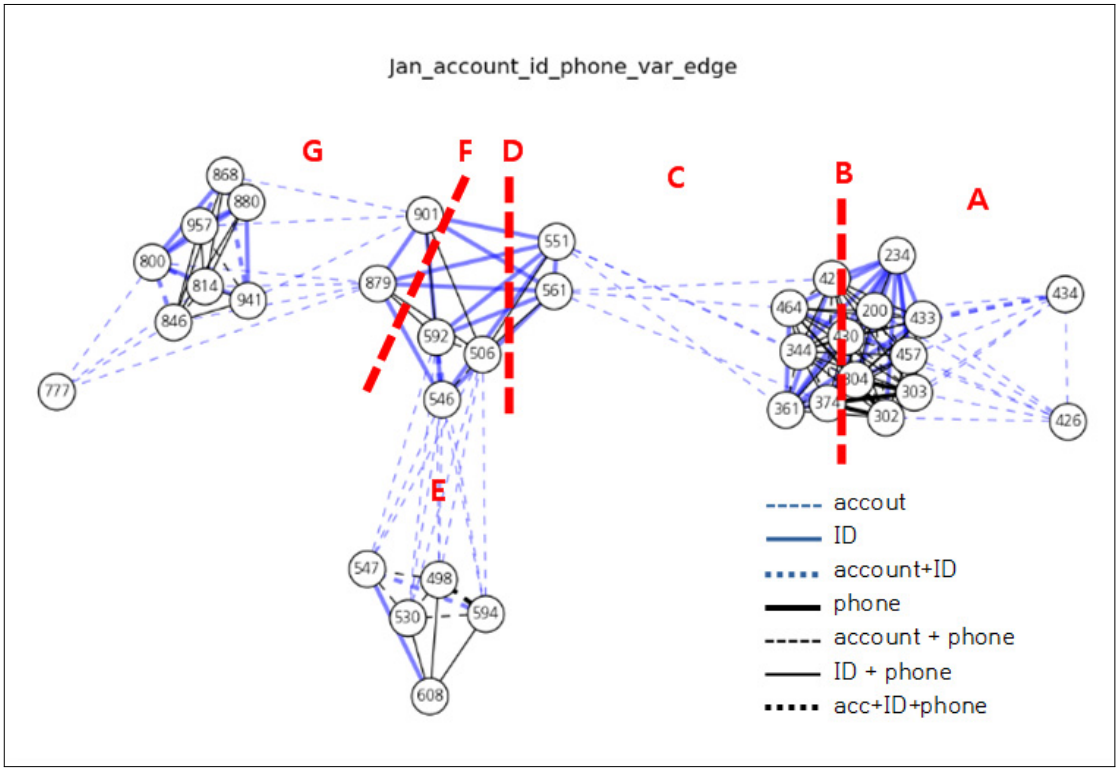

Fig. 8. Visualization of clues between cases in cluster B

수한 수사관이 범행에 사용된 아이디 또는 전화번호와 같이 주요 수사단서에 대한 연관성을 조회할 수 있는 시스템을 구축하여 활용할 수 있다면, 해당 수사단서가 상습사기에 사용되었음을 확인하는 등 접수된 사건단서를 대상으로 동일 범죄 조직에 의해 저질러졌음을 알 수 있고 이와 같은 대량피해의 확산을 조기에 차단할 수 있을 것이다.

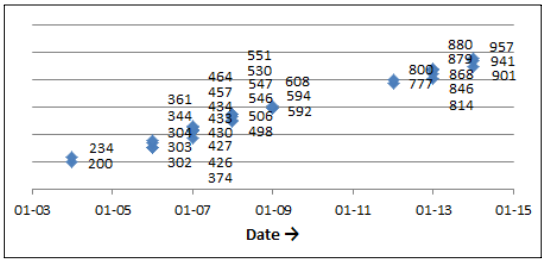

Fig. 9. The date of case occurred in Fig.8.

Ⅳ. 결 론

사소한 단서가 사건을 해결하는 열쇠가 되는 것처럼 사이버범죄 역시 사소한 디지털 정보 하나라도 의미 있는 수사단서가 될 수 있다. 사이버 사건에서 무수히 발생되는 수사단서 중 계좌정보와 같이 하나의 수사단서만이 유일한 분류기준이 되기는 어렵다. 의미 있는 수사단서를 활용하게 되면 동일 피의자의 여죄 추적, 중복 수사방지, 피의자 공범관계 확인, 사건발생 조기 차단 등의 효과를 거둘 수 있다.

본 논문에서는 주요 수사단서 선정 및 네트워크 시각화 기법을 이용하여 인터넷 사기 및 신종금융범죄의 주요단서의 중복확인을 통해 사건 간 연결고리를 추적하였다. 추적결과 서로 다른 계좌를 이용한 사건에서 계좌정보를 제외한 수사단서 간 연관성을 확인할 수 있었으며 실제로도 해당 사건들은 동일 피의자의 범행임을 확인할 수 있었다.

또한 인터넷 사기 사례에 대해 수사단서 연관성 분석을 통해 동일 범죄군임을 확인할 수 있었으며, 수사단서간의 변동 패턴에 따라 피의자 조직의 범행 추이에 대해서도 확인할 수 있었다.

이와 같이 실제 사례 적용을 통해 수사단서 수집의 중요성을 확인할 수 있었으나 단기적으로 피해자

로부터 수집되는 수사단서로는 한계가 있다.

이러한 한계를 극복하기 위해서는 다양한 수사단서를 수집하고 이를 DB로 구축하여 적극 활용할 필요가 있다. 우선 사건 접수단계에서는 실시간 수사단서 수집이 중요하다. 사이버 사기사건의 특성상 단기간에 대량으로 발생되므로, 동일 분류기준을 가진 사건들이 접수에서부터 실제 관할서로 이송되기까지 처리시간이 그만큼 지체되게 되면 이미 추가적인 피해를 막기에는 한발 늦어지게 된다. 이를 해결하기 위해서는 사건 최초 접수단계부터 계좌정보 외의 수사단서에 대해서도 실시간 DB 구축을 하고 이를 조회할 수 있는 수사단서 연관성 분석 시스템을 도입하여야 한다.

사건 종결단계에서는 범행에 사용된 계좌의 출금지점코드, CCTV영상, 공범 피의자간 상세한 연관성분석 정보를 입력하여 관리하여야 한다. 많은 수의 사이버 사기 사건의 경우 총책 피의자의 접속 IP는 중국지역으로 확인되고, 인출책이 국내에 머물며 전문적으로 인출하는 것으로 확인된다. 이러한 인출책들은 항상 금융기관의 CCTV에 노출되는데, 만약 사건 종결할 때마다 검거하지 못한 인출책의 CCTV 사진을 출금금융기관 지점코드와 지리정보시스템을 결합한 시스템에 저장하는 형태로 데이터를 축적하면, 이후 해당지역을 순찰하는 지역경찰은 이 시스템을 활용하여 해당 금융지점 ATM 인근 지역에 대해 주요 인출책에 대한 탐지 및 식별이 가능해진다. 또한 인출책을 검거한 경우에도 축척된 CCTV 사진 DB에서 안면인식 기능을 연동하여 여죄 검색에도 활용할 수 있을 것이다.

이와 같이 각 수사단계에서 수집된 수사단서들을 축적하고, 이를 유기적으로 결합하여 본 논문에서 제시된 연관성 분석 및 시각화 기법을 활용하면 사이버 범죄 수사의 효율성을 높일 수 있을 것이다.

References

[1] Dong hwan Lee and Changwon Pyo, "A Systematic Plan to Collect and Analyze Criminal Intelligence by Police," Korea Institute of Criminology, pp. 1-137, Dec. 2005.

[2] Joon Woo Kim, Joong Kweon Sohn, and Sang Han Lee, "Usefulness of Data Mining in Criminal Investigation," Journal of Forensic and Investigative Science, 2(2), pp. 5-19, Dec. 2006.

[3] Keiwon Kim, Jinwan Seo, "The Study on the Typology of Cyber Crime," Korean Public Management Review, 23(4), pp.95-118, Dec. 2009.

[4] Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques," Morgan Kaufmann Pub, pp. 608, Nov. 2005.

[5] Jeonghwa Lee, Youngrok Lee and Chi-Hyuck Jun, "Biclustering method for time series data analysis," Industrial Engineeering & Management Systems, 9(2), pp. 131-140, Jun. 2010.

[6] J.A. Hartigan , "Direct Clustering of a Data Matrix," Journal of American Statistical Association, Vol. 67, No. 337, pp. 123-129, Mar. 1972.

[7] Cheng, Yizong, and George M. Church. "Biclustering of expression data." Ismb. Vol. 8. pp. 93 - 103, Aug. 2000.

[8] Tanay A, et al. "Discovering statistically significant biclusters in gene expression data," Bioinformatics Vol. 18. No. 1, pp. 136-144, Mar. 2002.

〈저자소개〉

김 주 희 (Ju Hee Kim) 정회원
2004년 2월: 계명대학교 컴퓨터공학과 졸업
2006년 5월~2010년 12월: 잉카인터넷 NSC 드라이버개발팀 연구원
2015년 2월: 고려대학교 정보보호대학원 디지털포렌식학과 석사
2011년 5월~현재: 경기지방경찰청 사이버수사대 근무
〈관심분야〉 정보보호, 컴퓨터포렌식