사이버 침해정보 연관 그래프 구축 및 활용방안 연구

이슬기, 조혜선, 김병익, 신영상, 이태진
한국인터넷진흥원
e-mail : {sglee, hscho, kbi1983, ysshin, tjlee}@kisa.or.kr

A Study on Building a Cyber Incidents Information based Relational Graph and Using Plan

Seulgi Lee, Hyeisun Cho, Byungik Kim, Youngsang Shin, Taijin Lee
Korea Internet & Security Agency

요 약

사이버 침해사고 정보를 공유하는 체계가 전 세계적으로 확산되고 있는 추세이다. 상호 네트워크 통신을 위하여 필요한 인터넷기반정보와 사이버 침해사고 관련 정보를 획득하기 위한 채널은 다양하게 존재하고 공공의 이익을 목적으로 공유되고 있으며 침해정보에 대한 세부적인 분석정보 또한 오픈소스 프로젝트를 통해 손쉽게 획득할 수 있다. 한국인터넷진흥원에서는 공인된 사용자 혹은 기관을 대상으로 침해사고에 활용된 악성정보를 공유하고 있다. 본 논문은 이러한 인터넷기반정보와 침해사고와 관련된 연관정보를 활용한 사이버 침해정보 연관 그래프 구축방안에 대하여 논하며 그 활용방안이 어떠한 것이 있는지 제안한다.

1. 서론

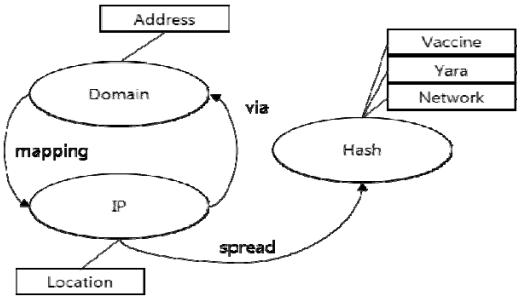
오늘날, 사이버 침해사고는 악성코드 자동생성 툴과 같이 침해사고 발생이 용이한 환경이 확보되고 있다. 자동화 도구의 유행과 더불어 해킹사고는 점차 증가하는 양상을 보이고 있으며 그 대표적인 도구인 메타스플로잇은 실제 공격에도 사용될 수 있을 정도로 고도화된 툴이다[1]. 이 도구는 공개된 취약점을 이용하여 해킹공격을 손쉽게 수행할 수 있도록 지원한다. 스크립트 키디가 일으키는 공격의 확대와 동시에 크래커들의 공격 또한 점차 지능화되고 있다. 이렇게 대량으로 발생하는 침해사고에 대한 모든 공격을 분석할 수 있는 인력과 자원을 확보하기 제한되기 때문에 자동화된 침입탐지/방지시스템 고도화 또는 분석가의 분석지원을 통해 이를 해결할 수 있다. 본 논문에서는 기존에 발생한 침해사고에 대하여 연관된 정보를 조회하는 기능과 해당 침해사고와 동일/유사한 침해정보가 활용된 침해사고를 조회할 수 있는 침해사고 연관 그래프 구축 및 활용방안에 대하여 제안한다.

2. 사이버 침해정보 연관 그래프 구성요소

본 논문에서 사이버 침해정보는 사이버 침해사고를 수행하는데 필요한 IP, Domain, 악성코드 등의 객체 및 분석정보, 그리고 인터넷기반정보는 IP, Domain 에 대한 분석정보로서 정의한다. 예를 들어, IP 는 특정 시점에 맵핑되는 Domain 이 존재하며 Domain 은 도메인 등록대행업체를 통하여 등록하기 때문에 등록자이메일 등의 정보가 후이즈 서비스를 통해 수집할 수 있다. 악성코드 분석을 수행할 수 있는 Cuckoo Sandbox 와 같은 도구를 통해 악성코드의 행위분석정보를 수집할 수 있다. 사이버 침해정보 연관 그래프는 위에서 정의한 정보를 이용하여 구성 가능하고 최초로 수집된 정보로부터 계속적인 정보수집 및 분석을 통해 방대한 량의 데이터를 구축할 수 있다.

3. 사이버 침해정보 연관 그래프 구축

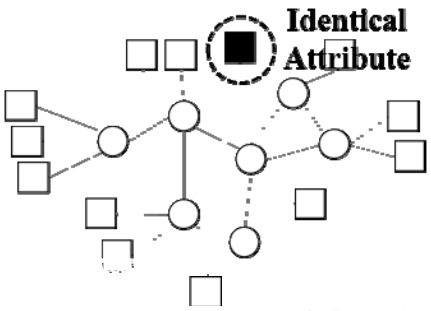
사이버 침해정보 연관 그래프는 위의 구성요소들이 결합되어 구성된다. 데이터는 조회할 수 있는 단위인 객체(IP, Domain, Hash, Email)와 지리정보, Yara Signatures 와 같이 더 이상 조회할 수 없는 정보인 속성으로 분류한다. 또한 데이터를 연계하여 정보를 수집/분석하고 산출된 정보에서 또 다시 수집/분석하는 행위를 반복하기 때문에 입력 및 출력 간 연결관계를 맺을 수 있다. 단순하게 데이터 간 연결관계를 맺는 것뿐 아니라 데이터가 산출된 근거가 존재하기 때문에 산출근거 기반 연결관계 특성화는 개선사항으로서 고려할 사항이 될 수 있다.



(그림 1) 그래프 구성요소 간 연결 예시

<표 1> 침해정보 연관그래프 정보

Category		Data Types
Data	Object	IP
		Domain
		Hash
		Email
	Attribute	Location
		Yara Signatures
		HTTP Requests
		Detection Name
		Registration Address
		Etc
Relationship	-	-



(그림 2) 연관 그래프 확장 예시

그림 1 은 수집한 구성요소를 결합하여 연관 그래프로 구축한 기본 예시이며 IP-Domain 의 관계에서 양방향성 연결관계를 가짐을 확인할 수 있다. 특히, 연결관계마다 수집/분석채널이 다르기 때문에 각각 연결선이 가지는 의미는 서로 상이하다. 단, 사이버 침해정보 연결 그래프 구축의 최초 단계에서 이를 고려하게 되면 복잡도가 증가하기 때문에 이를 제외하고 구성한 후 채널마다 가지는 중요도 및 신뢰성에 따라 가중치를 부여하는 개발이 용이하다.

표 1 은 침해정보 연관그래프를 구성하는 데이터 및 연결관계의 분류, 타입 별 대표적인 데이터유형을 정리한 자료이다. 객체는 앞에서 정의한 4 가지 데이터 유형으로 나누어지나 속성은 수집/분석 채널의 확보를 통해 추가적인 속성값을 적용할 수 있다.

그림 2 는 재귀적인 수집/분석 호출에 의하여 구축된 연결그래프 예시이다. 침해사고 연관 그래프가 해당 예시와 같이 확장되었을 때, 동일한 속성을 공유하는 두 개 이상의 침해자원이 존재한다면 두 개 침해자원 간 연관성이 존재함을 확인할 수 있다.

4. 사이버 침해정보 연관 그래프 활용

사이버 침해정보 연관관계 그래프가 그림 2 와 같이 구성되어 있을 때, 시스템이 제공하는 기본적인 기능은 사용자(혹은 분석가)가 객체/속성을 기준으로 침해사고 정보를 조회할 수 있다는 것이다. 시스템은 주기적으로 정보를 수집하고 이를 기반으로 다양한 정보 축적을 반복한다. 그렇기 때문에 사용자는 한 개 이상의 침해정보를 검색하면 그와 연관된 다양한 객체/속성 정보를 일순간에 확인이 가능하며 이는 수동으로 침해사고를 분석하는데 있어 효율성을 증대시킬 수 있을 것으로 기대하고 있다. 사이버 침해정보 연관관계 그래프는 수집/분석 채널 확보에 따라 획득한 정보에 근거하여 구성되기 때문에 수집/분석 채널에 의하여 연관그래프가 가지는 의미 및 가치가 변이할 수 있다. 어느 정도 분량의 데이터가 저장되는 지와 데이터 모델링을 통하여 저장관리체계가 어떻게 구성되어 있는지에 따라 검색 성능이 좌우될 수 있다.

사이버 침해정보 연관 그래프를 활용하는 다른 방안으로, 특정 객체로부터 다른 객체까지 얼마나 거리가 떨어져 있는 지와 그 사이에 존재하는 연관관계들의 시퀀스를 두 객체 간의 관계 규정에 사용하는 것이다. 두 객체 간의 거리는 몇 홉(hop)으로 이루어 졌고 연결관계가 도출된 근거의 신뢰성이 무엇인지에 따라 가중치를 부여하는 등의 작업을 통해, 특정 객체가 발생하면 다른 객체 혹은 속성이 발생할 확률을 구하는 등 보다 지능적인 보안관제를 위하여 사용할 수 있는 지표가 된다.

5. 결론

사이버 침해정보 연관 그래프 구축을 통해 침해사고를 일으킨 정보들의 연관관계를 손쉽게 파악할 수 있고 그로 인하여 어떻게 침해사고가 발생했는지 확인할 수 있다. 본 논문은 사이버 침해정보 연관 그래프를 어떻게 구축해야 하고 구축된 그래프에서 어떠한 방향으로 활용해야 하는지에 대하여 제안하였다. 연관 그래프를 구축하고 다수 침해사고 정보와의 연관분석, N 개 침해정보 간 어떠한 연결관계가 있는지, 그리고 발생확률을 산출해볼 수 있는 근거로서 활용 등이 기대된다. 그래프 구성이 용이하고 활용성 측면의 장점이 존재하기 때문에 사이버 침해사고 연관 그래프를 구축하고 추가적으로 연구할 가치가 있다고 제안한다. 본 논문에서 제안하는 사이버 침해정보 연관 그래프 구축은 시간정보를 제외한 데이터만을 이용하여 구성되어 있다. 그렇기 때문에 얼마나 자주 악용되었는지 산출할 수 있는 빈도정보와 객체/속성의 변화에 대하여 추적하고 이를 활용하는 방안이 제외되어 있다. 향후 이러한 정보들을 반영하여 보다 개선된 침해정보 연관 그래프를 구축하고 테스트하는 것이 필요하다.

Acknowledgment

이 논문은 2015 년도 정보(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.B0101-15-0300, 사이버 공격의 사전 사후 대응을 위한 사이버 블랙박스 및 통합 사이버보안 상황분석 기술 개발)

참고문헌

[1] 한국인터넷진흥원 “2014 년 10 월 인터넷침해사고 대응통계”

[2] 한인혜 “사이버 위협정보 분석 · 공유시스템(CTAS) 공유체계”