침해지표 기반의 사이버 위협수준 분석

조혜선, 이슬기, 김낙현, 김병익, 유동영, 김문현*
한국인터넷진흥원
*성균관대학교 정보통신대학원
e-mail: {hscho, sglee, knh, kbi1983, ydy}@kisa.or.kr, mhkim@skku.edu

Analysis of Cyber Threat Level based on Indicator of Compromise

Hyeisun Cho, Seulgi Lee, Nakhyun Kim, Byungik Kim, Dongyoung Yoo, Moon-Hyun Kim*
Korea Internet & Security Agency
*Dept.of Information Security, Sungkyunkwan University

요 약

최근 국내에서 신·변종 공격이 대량으로 발생함에 따라, 한정적인 보안전문 인력과 기존의 장비로 분석 및 대응하는데 어려움이 있다. 본 논문에서는, 대량으로 발생하는 침해사고에 대해 분석 우선순위를 확인하고자, 침해사고에 활용된 침해지표들의 위협을 분석하고 이를 정량적인 값인 침해지표 위협수준(TL_IoC, Threat Level of IoC)로 도출하는 방안을 제안한다. 이를 통해, 침해지표의 위협수준을 직관적으로 확인함으로써 침해사고의 대응수준을 신속하게 판단하고, 위협수준이 높은 침해사고에 대해 능동적으로 빠르게 분석함으로써 대량의 침해사고를 효율적으로 대응할 수 있다.

1. 서론

최근 국내에서 신/변종 공격이 대량으로 발생됨에 따라, 침해사고 대응에 대한 패러다임 변화가 요구되고 있다.[1] 실제로, 한국의 APT공격 노출률은 38%로 전 세계 평균(15%)에 비해 2배 이상 높은 수치이며 미국(13%)과 비교하면 3배 가까운 수치이다.[2]

이와 같이, 공격자들은 특정 타겟에 대한 공격목표를 달성하기 위해 기존의 보안장비를 우회하는 신·변종 공격을 대량으로 제작하고 있으나[3], 기존장비의 한계 및 한정된 분석인력으로 대량의 침해사고 분석 및 대응에 어려움을 겪고 있다. 한정된 자원으로 최근 발생하는 대량의 지능형 사이버 공격을 효율적으로 방어하기 위해서는, 침해지표를 분석하여 위협수준이 높은 침해사고에 대해 능동적으로 빠르게 분석 및 대응하는 새로운 전략이 필요하다[4].

본 논문에서는, 대량으로 발생하는 침해사고에 대해 효율적으로 분석하고자, 침해사고에 활용된 각각의 침해지표에 대해 위협을 분석하고, 이를 정량적인 값으로 도출하는 방안을 제안한다.

2. 침해지표 기반의 사이버 위협분석

2.1. 침해지표에 따른 위협분석 요구사항

침해지표(IoC, Indicator Of Compromise)란, 특정 침해사고의 증거 및 흔적들을 의미하는 데이터이며, 본래 디지털 포렌식 및 사고대응 분야에서 사용되었지만 현재는 보안장비/솔루션에서도 침해사고 이력을 확인하는 용도로 점차 확장되어 쓰이고 있다. 침해지표는 침해사고 공격에 활용된 인프라(유포지, 경유지, C&C 등으로 활용된 서버 및 도메인), 공격발생 일시, 악성코드 등을 포함한 공격 및 이와 연관된 정보들을 의미한다.[5]

본 연구에서 활용된 침해지표는 과거에 발생한 침해사고에 대한 분석결과 및 공개된 채널에서 획득할 수 있는 공격정보를 수집하여 연구를 수행한다. 본 연구에서 수집되는 침해지표는 아래 <표 1>과 같다.

<표 1> 사이버 위협정보 수집

구분		수집내용	총 수집량 ('15.10.~ '16.12.)
위협정보	침해사고 탐지정보	C&CIP	96건
		좀비IP	146,122건
		경유지	5,131건
		유포지	4,042건
	평판정보	RBL_IP	31,622,940건
		RBL_Domain	69,690건
	신/변종 악성코드	악성코드 - 신종	372,117건
		악성코드 - 변종	
합 계			32,220,138건

침해지표의 위협을 분석하기 위해서는 "위협의 정도" 결정할 수 있는 기준(Factor)이 필요하며, 이를 위해 위협분석 요구사항을 추출 할 수 있다. 공격 탐지/대응관점에서 침해지표의 탐지방법, 탐지시간, 평판정보, 행위정보 등을 고려한 위협분석 요구사항은 <표 2>와 같다.

<표 2> 침해지표 위협분석 요구사항

① **침해지표 탐지경로(출처) 관점**: 침해지표의 특성을 결정하는(C&C, 좀비, 경유지, 유포지 등) 탐지시스템의 출처 확인을 통해, 침해지표의 위협을 파악할 수 있다.
② **탐지시간 관점**: 최근 탐지된 침해지표일수록 최신/현재시점에 근접한 위협일 가능성이 크므로, 탐지시간에 따라 위협을 파악할 수 있다.
③ **Blacklist 등록여부 관점**: Blacklist로 등록된 위협정보는 다수의 평판정보를 반영하기 때문에, Blacklist등록여부에 따라 침해지표의 위협을 파악 할 수 있다.
④ **DNS변경이력 관점**: DNS변경이력이 많을수록, 공격추적을 회피하기 위한 공격자의 전략으로 해석될 수 있으므로, DNS변경이력에 따라 위협을 파악할 수 있다.
⑤ **탐지된 악성URL 관점**: 공격정보와 연관된 악성URL의 개수에 따라 침해지표의 위협을 파악할 수 있다.
⑥ **변종 악성코드 관점**: 연관된 변종 악성코드의 개수를 통해 공격그룹의 규모 및 파급도를 추측할 수 있으므로, 변종 악성코드의 개수에 따라 침해지표(악성코드)의 위협을 파악할 수 있다.
⑦ **악성코드 유포 관점**: 침해지표(유포지)에서 유포한 악성코드가 많을수록 공격이 활발히 진행되고 있음을 파악할 수 있으므로, 침해지표의 위협을 파악할 수 있다.
⑧ **유포지/경유지 활용 관점**: 특정 도메인이 유포지/경유지로 활용이 많을수록 공격자가 주로 활용하는 인프라로 해석될 수 있으므로, 유포지/경유지 활용이력에 따라 침해지표의 위협을 파악 할 수 있다.
⑨ **웹 변조이력 관점**: 웹 변조 여부에 따라 웹페이지 해킹여부를 확인할 수 있으므로, 웹 변조 여부에 따라 침해지표의 위협을 파악 할 수 있다.
⑩ **Drop악성코드 여부 관점**: Drop되는 악성코드의 유무를 통해 공격 복잡도를 파악할 수 있으므로, 침해지표(악성코드)의 위협을 파악할 수 있다.

2.2. 침해지표 위협분석 및 정량화

2.2.1. 위협수준 분석기준에 따른 가중치 설정

2.1절에서 언급한 침해지표 위협분석 요구사항을 분석해 보았다. 위 요구사항을 통해서, 침해지표 분석을 통해 Semantic한 침해사고 분석이 가능하나, 앞서 말한 대량의 침해사고를 각각 분석하여 의미를 파악하기에는 시간이 많이 소요된다. 따라서, 2.2절에서는 침해지표 위협분석 요구사항에 따라 위협을 분석하여 정량적으로 표현하는 방안에 대해서 제시한다.

침해지표의 위협을 정량적으로 도출하기 위해서는 <침해지표 위협 정량화 Matrix>를 구성한다(2.2.2절).[6] Matrix 구성하기 이전에 침해지표 위협분석 Factor별 중요도를 위협수준에 반영시키기 위해 "가중치"를 설정한다. "가중치"는 휴리스틱 기법에 의해 각각의 Factor가 위협분석에 미치는 영향을 제어하기 위한 장치로 활용한다.<표3>

<표 3> 침해지표 분석기준에 따른 상관관계 및 가중치 설정

침해지표 분석기준 (Factor)	위협수준 상관관계	가중치*
① 탐지경로(출처) 관점	위협수준 ∝ 탐지경로1(C2,유포지, 악성코드) >탐지경로2(공격활용, 경유지) >탐지경로3(RBL) >탐지경로4(악성코드 감염호스트)	10
② 탐지시간 관점	위협수준 ∝ 탐지된 시간(최근)	1
③Blacklist 등록여부 관점	위협수준 ∝ Blacklist등록여부	5
④ DNS변경이력 관점	위협수준 ∝ DNS이력(count)	5
⑤ 탐지된 악성URL 관점	위협수준 ∝ 악성URL활용 이력(count)	10
⑥ 변종 악성코드 관점	위협수준 ∝ 변종 악성코드(count)	15
⑦ 악성코드 유포 관점	위협수준 ∝ 유포한 악성코드(count)	15
⑧유포지/경유지 활용 관점	위협수준 ∝ 유포/경유지 활용이력(count)	9
⑨ 웹 변조이력 관점	위협수준 ∝ 웹 변조 여부	15
⑩Drop악성코드 여부 관점	위협수준 ∝ Drop악성코드 여부	15
합 계		**100**

* "가중치"는 침해사고 분석 기관 및 적용환경에 따라 가변적으로 조정이 가능한 변수임(variable).

2.2.2. 침해지표 위협수준 판단 Matrix 구성

각각의 침해지표에 대한 분석결과를 정량적으로 변환하기 위해서 <침해지표 위협 수준 판단 Matrix>를 구성한다.<표 4>

<침해지표 위협수준 판단 Matrix>에서는 침해지표의 분석 기준(Factor)별로 분석결과를 반영할 수 있는 "위협속성"을 설정하고, 해당 "위협속성"이 얼마나 위협적인지를 정량적으로 나타내는 위협등급을 부여한다. "위협등급"은, 위협속성의 특성에 따라 0~5로 정의하고, 0은 '정보 없음'을 의미하며 5에 가까울수록 위협수준이 높은 것으로 판단한다.[7][8]

<표 4> 침해지표 위협수준 판단 Matrix

침해지표 분석기준 (Factor)	위협속성	위협 등급	위협 지수*	최대위협 지수**
① 탐지경로(출처)	탐지경로1(C2,유포지)	5	50	50
	탐지경로2(공격활용, 경유지)	4	40	50
	탐지경로3(RBL)	3	30	50
	탐지경로4(악성코드 감염호스트)	1	10	50
② 탐지시간	~1Month	5	5	5
	1~3Month	4	4	5
	3~6Month	3	3	5
	6~12Month	2	2	5
	12Month ~	1	1	5
③ RBL등록여부	Existence	3	15	15
	Non-existence	0	0	15

④DNS변경이력	41~	5	25	25
	31~40	4	20	25
	21~30	3	15	25
	11~20	2	10	25
	1~10	1	5	25
	0	0	0	25
⑤탐지된 악성URL	41~	5	50	50
	31~40	4	40	50
	21~30	3	30	50
	11~20	2	20	50
	1~10	1	10	50
	0	0	0	50
⑥변종 악성코드	41~	5	75	75
	31~40	4	60	75
	21~30	3	45	75
	11~20	2	30	75
	1~10	1	15	75
	0	0	0	75
⑦악성코드 유포	41~	5	75	75
	31~40	4	60	75
	21~30	3	45	75
	11~20	2	30	75
	1~10	1	15	75
	0	0	0	75
⑧유포지/경유지 활용	41~	5	45	45
	31~40	4	36	45
	21~30	3	27	45
	11~20	2	18	45
	1~10	1	9	45
	0	0	0	45
⑨웹 변조이력	Existence	5	75	75
	Non-existence	0	0	75
⑩Drop악성코드 여부	Existence	5	75	75
	Non-existence	0	0	75

* (위협지수) = (위협등급) x (가중치)
** (최대위협지수) = (지표별 최대위협등급) x (가중치)

2.2.3. 침해지표 위협 정량화

침해지표 위협분석에 따른 정량화는 2.2.2절에서 정의한 <침해지표 위협수준 판단 Matrix>를 기반으로 계산한다. 침해지표에 대한 위협을 분석하여 정량화 한 값은 백분율로 표시하고, 이를 침해지표 위협수준(TL_IoC, Threat Level of IoC)이라고 정의한다.

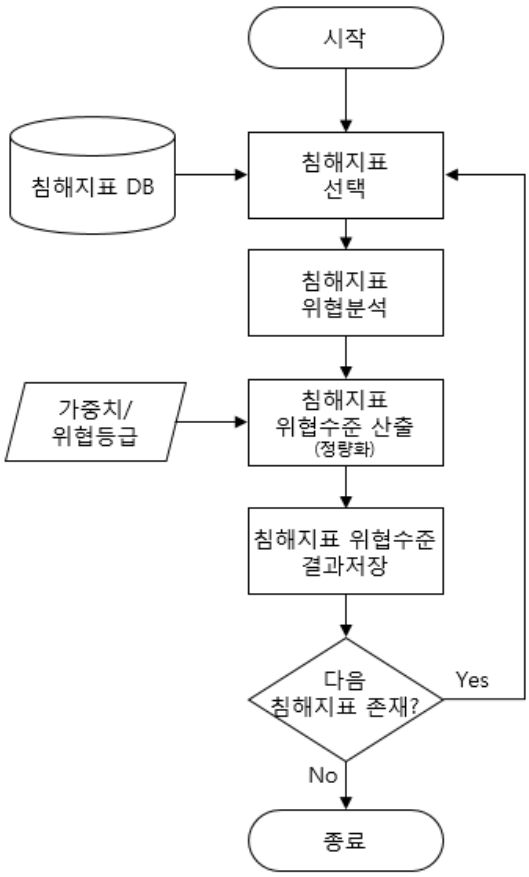
침해지표 위협분석 기준으로 정의한 10가지 Factor중, TL_IoC도출에 활용된 Factor가 n개라고 할 때, 다음의 식 (1)을 통해 침해지표 위협수준을 도출할 수 있다.

$$TL\_IoC(x) = \sum_{i=1}^{n}(t_i * w_i) / \sum_{i=1}^{n}(m_i * w_i) * 100 \quad (1)$$

$$T = \{t_1, t_2, \dots, t_n\}, M = \{m_1, m_2, \dots, m_n\}, W = \{w_1, w_2, \dots, w_n\}$$

여기서 T는 침해지표 위협분석 Factor별 위협속성을 의미하는 위협지수의 집합이고, M은 T가 가질 수 있는 최댓값의 집합이며, W는 T의 각 요소에 대응하는 가중치들의 집합을 의미한다.

침해지표 위협 정량화 시스템구성은 (그림 1)와 같으며, 1달에 1번씩 배치처리를 통해 침해지표 위협수준을 도출하여 침해지표 관리 DB에 저장한다.



(그림 1) 침해지표 위협분석 시스템 흐름도

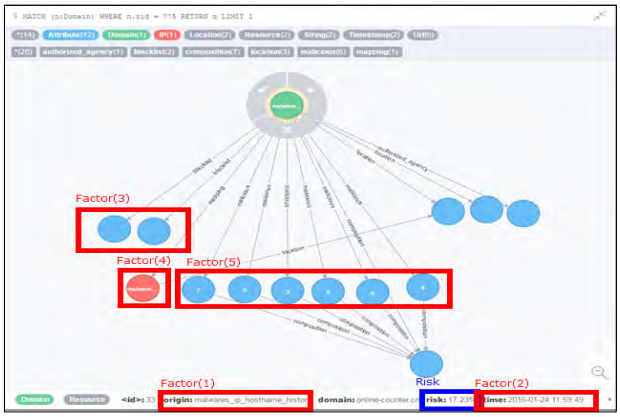
2.3. 침해지표 위협수준 도출 결과

위 침해지표 위협수준 도출 식 및 로직을 통해 실제 침해지표 관리 DB를 참조한 침해지표 위협수준 도출을 위한 모듈을 구현하였으며, 탐지시간 및 추가 Factor에 대한 정보수집 시 위협수준이 변경이 되므로 주기적인 배치처리를 통해 실행된다.

실제 침해사고에 활용된 악성 도메인에 대한 침해지표 위협 분석결과, 5가지 Factor를 통해 위협수준을 도출할 수 있었다.(그림 2)

3. 결론

본 논문에서는, 침해사고에 활용된 침해지표를 기반으로 위협을 분석하고 이를 정량적인 값으로 도출하는 방법을



(그림 2) 침해지표에 대한 위협수준 분석 결과

제시하였다. 도출된 침해지표 위협수준 적용을 통해, 실제 침해사고 발생 시 해당 위협수준을 직관적으로 파악하여 이에 대한 대응수준을 결정할 수 있고, 위협이 높은 침해사고를 선 대응함으로써 대용량 지능형 침해사고 공격에 효율적으로 대응할 수 있다.

향후, 침해지표 위협분석 기준 고도화를 통해 위협수준에 대한 신뢰성을 높이고 해당 알고리즘에 대한 상용 환경 검증을 수행할 예정이다.

ACKNOWLEDGMENT

이 논문은 2017년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.2017-0-00158, 국가 차원의 침해사고 대응을 위한 사이버 위협 인텔리전스 분석(CTI) 및 정보 공유 기술 개발)

참고문헌

[1] 한국인터넷진흥원, “2017년 7대 사이버 공격 전망(보고서)”, 2016.12.

[2] FireEye, “파이어아이, 국내 사이버 공격 현황 및 랜섬웨어 트렌드 발표”, 2016.04.

[3] Kaspersky, “What is known about the Lazarus Group: Sony hack, military espionage, attacks on Korean banks and other crimes”, 2016.

[4] 전자신문, “사이버 위협정보 '신뢰성 확보 시급'”, 2015.11.

[5] “Indicator of compromise”, 『Wikipedia』(2011)

[6] 유학용, 유동영, “사이버공격 대응 기본 매트릭스”, 한국인터넷진흥원, 2014.06

[7] Sallam, Hany. "Cyber Security Risk Assessment Using Multi Fuzzy Inference System." IJEIT 4.8 (2015): 13-19.

[8] 김태경. "악성코드 탐지 방법에 관한 연구." 보안공학연구논문지 9.5 (2012): 387-400.