.KR 최상위 도메인 DNS 질의로그의 악성행위 탐지 관점에서의 분석*

이기룡,[a] 이제현,[a] 권종훈,[a] 이희조,[a]† 박해룡[b]

고려대학교 컴퓨터학과,[a]

한국인터넷진흥원[b]

Analysis on .KR TLD DNS Query Log for Malicious Behavior Detection

Kiryong Lee,[a] Jehyun Lee,[a] Jonghoon Kwon,[a] Heejo Lee,[a] and Haeryong Park[b]

Dept. of Computer Science and Engineering, Korea University[a]

Korea Internet & Security Agency[b]

요 약

도메인이름체계(DNS)는 인터넷의 핵심요소로써 인터넷사용자 뿐 아니라 악성코드에 의해서도 활발히 이용되고 있다. 이 연관성을 이용하여 DNS 트래픽이나 질의로그를 분석하여 악성코드 및 연관 악성도메인의 활동과 감염상태를 탐지하려는 연구가 국내외에서 진행되어오고 있다. 그러나 대다수의 연구들은 DNS 계층구조 중 캐시서버와 같이 하위수준 DNS 서버의 로그의 분석에 그 초점이 맞춰져 있었으며, 상위수준 도메인서버의 로그에 대한 연구는 소수의 해외사례만이 존재하였다. 본 연구에서는 대한민국 국가코드최상위도메인(ccTLD)의 DNS 질의로그를 악성행위 탐지의 관점에서 분석하여 .KR 도메인을 사용하는 악성행위의 특징과 국내사정에 적합한 ccTLD기반 악성행위 탐지기법의 방향을 제시한다.

I. 서론

1.1 연구목적

도메인이름체계(DNS)는 중앙 집중 된 위치에서 넓은 범위의 네트워크를 관찰할 수 있는 인터넷의 핵심 요소다. 특히 DNS의 질의로그는 적은 정보로 많은 영역의 악성행위 탐지를 할 수 있어 많은 네트워크기반의 악성행위 탐지연구[1, 2, 3, 4, 5, 6]들에 활용되고 있다.

하나의 DNS 서버(군)이 관할하는 네트워크의 범위는 DNS 계층구조상에서의 위치에 따라 달라지는데, 개별사용자 컴퓨터에 가까운 하위수준의 재귀적 도메인이름서버(Recursive DNS Server)는 이를 이용하는 모든 사용자의 질의를 일차적으로 관찰할 수 있어 인터넷서비스제공자(Internet Service Provider)와 같이 특정 네트워크 내의 악성코드 활동을 분석하고자 하는 목적에 적합하지만, 그 범위가 제한적이다. 반면 일반 최상위도메인(generic top-level domain: gTLD)서버와 국가코드 최상위도메인(country code top-level domain: ccTLD)서버와 같은 상위수준의 서버에서는 관할 차상위 또는 최상위 도메인을 질의하는 인터넷 전체의 질의를 관찰할 수 있다. 특정 도메인구역(Domain Zone) 내의 악성도메인 탐지와 광범위한 악성행위 관찰에 유용하다. 그러나 상위수준 DNS서버의 질의로그는 그 접근의 어려움과 그 방대한 데이터 크기로 인해 연구사례가 많지

* 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 정보통신●방송 연구개발사업의 일환으로 수행하였음. [14-824-06-001, 사이버 공격의 사전 사후 대응을 위한 사이버 블랙박스 및 통합 사이버보안 상황분석 기술 개발]

† 교신저자. heejo@korea.ac.kr

않으며 상당수가 해외도메인에 집중되어있다.

1.2 관련 연구사례

TLD에 대한 연구는 TLD의 계층위치적 특징을 분석하고 이용한 연구와 ccTLD의 지역적 특징을 분석한 연구로 분류할 수 있다. TLD가 RDNS와 다르게 도메인의 정보를 보유하고 있다는 점을 이용한 연구로는 Spring와 Felegyhazi, Hao의 연구[1, 2, 3]가 있다. 이 연구들은 TLD와 대형 DNS 데이터로부터 획득한 DNS로그를 기반으로, 정상 도메인과 악성 도메인간의 신규도메인 등록 후 최초 응답되기까지의 경과일의 시간차를 분석하여 그 이용 패턴에 차이가 있음을 분석하였다. 이 연구에서 밝혀낸 도메인 사용패턴은 도메인의 평판을 이용하여 악성도메인을 탐지하는 연구들에서 중요한 평가근거로 사용될 수 있음을 제안하고 있고, 실제 TLD를 대상으로 한 연구사례 [4]에서 그 판단척도로써의 가치가 입증된 바 있다.

TLD 중 ccTLD의 국가별, 지역별로 특화된 질의특징을 분석한 연구로는 중국의 .CN TLD 서버를 대상으로 한 연구[5]가 있으며, 지역별, 서버별 특징들을 분석하였다. 또, 캐나다의 .CA TLD 서버의 정보를 이용한 연구[6]에서는 DNS 서버에서 나타나는 질의자의 다양성, 동일 IP에 할당된 도메인들의 평판, BGP, AS 등의 네트워크 그룹 별 블랙리스트 등재 IP 주소 수 등의 특징을 이용하여 새로운 도메인들의 악성 가능성을 추정하는 시스템을 제안하였다.

이러한 종래의 연구결과를 근거로 볼 때, TLD에서 관찰되는 악성도메인의 활동양상과 분포, 시간대별 특징, ISP, RDNS 별 특징은 새롭게 등장하는 악성도메인을 탐지하거나 악성코드 활동에 최적화된 탐지기법을 연구하는데 필수적으로 고려되어야 할 요소라고 할 수 있으며, 기존 연구들에서 이러한 정보가 실제로 악성도메인의 탐지 정확성에 기여하고 있음이 확인되었다.

본 연구에서는 특히 TLD 서버에서의 악성행위탐지에서 고려해야할 질의양상과 활동패턴을 정상질의활동과 비교하고, 악성도메인 질의행위 및 이를 발생시키는 악성코드 탐지의 관점에서 각 현상의 의미와 응용가능성에 대해 분석한다.

II. 본론

2.1 .KR TLD 로그수집과 악성 DNS질의 선별

이 연구에서는 .KR TLD 서버에 질의되는 악성도메인의 행위를 분석하기 위하여, 2014년 8월 19일~25일의 7일간의 .KR TLD 서버 로그를 수집하고, 총 15개의 서버 중 국내에 위치한 b.dns.kr 서버의 로그를 분석대상으로 하였다.

총 236개국, 20,869개 ISP에 일평균 약 22.2GB의 데이터를 수집하였다. 질의 도메인은 일평균 약 850만 개 이고, IP는 일평균 약 56만 8천개였다.

이로부터, 통계적으로 관찰 가능한 악성도메인들의 활동양상을 분석하기 위해 수집된 질의로그의 질의도메인을 Virustotal [7]에 조회하여 알려진 악성도메인들에 대한 질의행위만을 선별하였다.

2.2 질의량 통계 요약

악성도메인에 대한 질의 발생량 통계에서 7일간에 걸쳐 공통적으로 나타난 특징을 요약하면 다음과 같다.

- **A와 AAAA 유형의 질의가 전체 질의량의 약 95%**
- **질의량 상위 10개 국가에서 발생한 질의가 전체 질의량의 약 92%**
- **질의량 상위 10개 도메인에 대한 질의가 전체질의량의 약 40%**
- **국내발생 질의는 주말이 주중의 약 70%의 질의량을 보이나, 시간대별 증감추이는 유사하게 반복**
- **국외발생 질의의 시간대별 증감추이는 국내와 유사하나, 각 국가의 현지시각에 따름**

2.3 전체 DNS 질의 대비 악성 DNS 질의 특징

전체도메인에 대한 DNS 질의량 통계와 악성 도메인에 대한 질의량 통계비교를 통해 다음과 같은 차이점을 도출하였다.

DNS 질의유형 별 분포에서 전체도메인에 대한 DNS 질의 중 도메인으로부터 IP 주소를

획득하는 목적의 A와 AAAA유형의 질의는 분석대상로그 7일 전체에 걸쳐 공통적으로 질의량의 약 90%를 차지하였다. 그러나 악성도메인에 대한 질의만을 집계한 통계에서는 약 95%의 비율을 보여, 악성도메인의 대다수가 A와 AAAA 유형으로 질의된다는 것을 확인하였다.

도메인 별 질의량의 측면에서, 전체 도메인에 대한 질의 중 질의량 상위 10개의 악성도메인에 대한 질의가 차지하는 비율은 약 5%이지만, 악성도메인들에 대한 질의량 중에서는 약 40%의 비중을 차지하는 것으로 나타났다. 이로부터 볼 때, 소수의 주요 악성도메인을 탐지하고 차단함으로써 상당수의 악성 DNS 질의행위를 감소를 기대할 수 있다.

2.4 악성 DNS 질의 중 국가별 통계분석

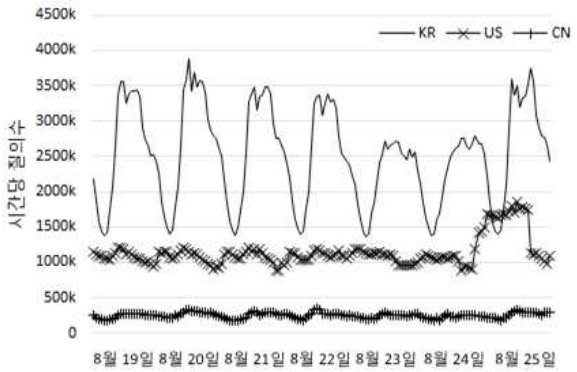
.KR TLD 서버의 7일간의 국가별 질의량 분포는 [그림 1]과 같이 국내에서 발생한 질의는 시간대별 질의량 변화가 극명하고, 주중간은 절대량의 차이와 증감추이의 변화가 크지 않았으며, 주말의 경우 그 증감추이는 유지되지만 절대량은 비율적으로 감소하는 것을 알 수 있다. 미국, 중국 등의 국외에서 발생하는 질의는 국내와의 시차만큼의 시간차이로 유사한 질의량 변화추이가 존재하였다. 예외적으로 8월 24일과 25일에 걸쳐 미국에서 발생한 대량의 질의는 ns11.whois.co.kr에 대한 질의가 미국 내 다수의 ISP로부터 발생한 특수한 경우로 확인되었다.

그러나 이를 세분화하여 [그림 2]와 같이 각 ISP별 질의량으로 살펴보면 개개 ISP에서 발생하는 질의량은 시간대별 증감폭이 크지 않고, 다수의 ISP의 질의량을 누적시켜 국가규모가 되었을 때, 그 시간대별 증감량이 부각되어 나타났다. 이는 악성행위 탐지의 측면에서 볼 때, 국외로부터 질의되는 악성 DNS질의의 경우, 특정 ISP에서 집중적으로 발생하는 질의는 시간대별 변화폭이 크지 않게 나타나지만, 여러 ISP에 걸쳐 감염되어있는 악성코드에서 발생하는 질의의 경우 사용자의 일과시간대에 종속적인 질의량 증감추이를 보이며, 해당 국가의 시간대에 종속되어 나타난다는 것을 알 수 있다.

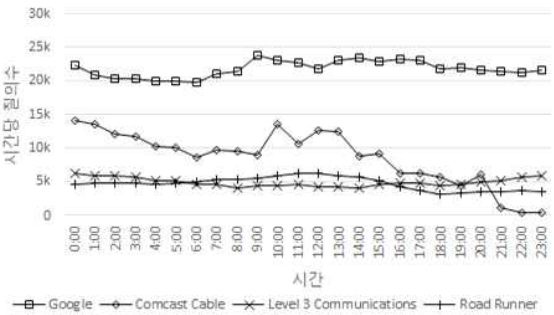
2.5 악성 DNS질의의 시간대별 증감추이 유형

악성도메인에 대한 시간대별 증감추이는 크게 3 가지 유형이 관찰되었다.

- ***유형1. 일과시간 종속 질의량 변화***
- ***유형2. 하루 종일 일정수준 유지***
- ***유형3. 특정시간대에 질의 집중발생***



[그림 1] 질의량 상위 3개 국가에서 발생한 시간대 별 악성도메인 질의량 추이

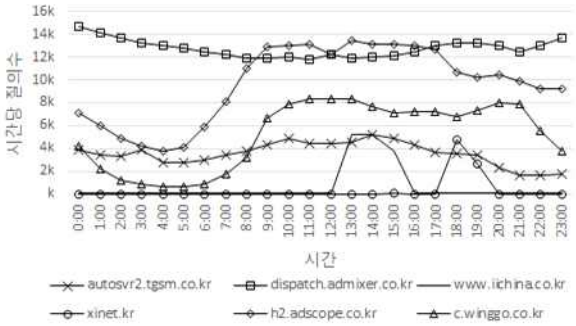


[그림 2] 미국에서 발생한 DNS 질의의 ISP별 질의량 분포 (8월 19일, 질의량 상위 4개 ISP)

[그림 3]의 그래프는 각 증감추이 유형에 속하는 대표적 악성도메인들의 시간대별 질의량 변화를 나타낸 것이다. autosvr2.tgsm.co.kr와 dispatch.admixer.co.kr는 0시 ~ 7시에 적은 질의량을 보이고 7시 ~ 22시에 질의량이 증가하는 유형 1의 추이를 보이고, www.iichina.co.kr와 xinet.kr는 정적 질의량을 보이는 유형2, h2.adscope.co.kr와 c.winggo.co.kr는 특정시간대에만 높은 질의량을 발생시키고 나머지 시간동안 질의가 거의 발생하지 않는 유형3의 질의패턴을 가지는 도메인이다. 다수의 악성도메인의

질의패턴이 세 가지 유형에 속하는 것을 고려할 때, 각 유형에 적합한 악성행위 탐지전략을 택함으로써 효과적인 탐지를 달성할 수 있다.

유형을 고려한 탐지전략의 예로서, 유형1의 도메인은 그 질의량이나 빈도, 주기성 등의 특징이 악성도메인이나 악성코드의 고유속성이 아닌 발생지역의 시간대에 따라 크게 좌우되므로, 사용자 행위에 의한 질의량의 유동성이 적은 시간대의 질의로그를 대상으로 행위기반 탐지기법을 적용시켜 알려지지 않은 악성도메인을 탐지하고, 더 많은 악성질의가 발생하는 시간대에 종래 알려진 악성도메인이나 행위패턴을 이용하여 악성행위를 탐지하는 방법이 효과적이다.



[그림 3] 시간대별 증감추이 유형 별 질의량 변화추이 예시 (8월 19일)

유형 2의 악성도메인들은 일 중 고정적인 질의량을 보이므로 탐지정확도가 분석시점에 거의 영향 받지 않는다. 따라서 유형1의 악성도메인과 정상도메인의 질의가 적은 시간대를 택하여 탐지기법을 적용함으로써 탐지정확도를 높일 수 있다.

마지막으로 유형3의 악성도메인들에 대한 질의는 그 활동시간이 짧고, 질의량이 지속적이지 않으므로 행위기반의 탐지기법보다는 기존의 질의량 분포를 모델링하여 질의량의 급격한 증감을 탐지하는 방식의 기법이 효과적으로 사용될 수 있다. 또한 이 유형의 악성도메인들의 질의 IP 분포를 분석한 결과, 질의 IP주소가 서로 다른 ISP나 국가에 퍼져있지 않고 특정시간대의 소규모 지역에 집중되어있는 경우가 많은 것으로 나타났다.

III. 결론

.KR TLD DNS로그는 전 세계에서 .KR 도메인에 대해 발생하는 DNS 질의를 관찰 할 수 있어, 그 질의양상과 통계로부터 악성행위 탐지기법을 연구하고 전략을 수립하는데 유용한 정보를 획득할 수 있다. 우리는 .KR TLD DNS로그에서 악성도메인에 대한 질의만을 선별하여 시간대별, 질의국가별 질의양상 변화와 질의유형 분포를 도출함으로써 주목해야할 통계적 사실들과 서로 다른 탐지전략이 필요한 악성질의 패턴 유형에 대해 분석하였다. 분석 결과로부터 TLD에서는 RDNS에서 관찰되는 악성행위 고유의 행위특징에 더하여 질의자의 국가, ISP 별 분포정도에 따라 부가되는 특징을 고려하여야 한다는 것을 밝혔다.

[참고문헌]

[1] J.M.Spring, L.B.Metcalf, and E.Stoner, Correlating domain registrations and DNS first activity in general and for malware, Securing and Trusting Internet Names, 2011

[2] M.Felegyhazi, C.Kreibich, and V.Paxson, On the Potential of Proactive Domain Blacklisting, USENIX Workshop on Large-scale Exploits and Emergent Threats, 2010

[3] S.Hao, N.Feamster, and R.Pandrangi, Monitoring the initial DNS Behavior of Malicious Domains, ACM SIGCOMM, 2011

[4] L.Bilge, E.Kirda, C.Kruegel, and M.Balduzzi, EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis, NDSS, 2011

[5] Y.Xuebiao, W.Xin, L.Xiaodong, and Y.Baoping, DNS Measurements at the .CN TLD Servers, Fuzzy Systems and Knowledge Discovery, 2009

[6] M.Antonakakis, R.Perdisci, W.Lee, N.Vasiloglou II, and D.Dagon, Detecting Malware Domains at the Upper DNS Hierarchy, USENIX Security Symp., 2011

[7] H.Sistemas, "Virus total." URL: http://www. virustotal. com