DNS TXT 레코드의 사용 현황 및 오남용 사례 분석

황은비 [1], 김현수 [2], 권태경 [3]
[1] 서울대학교 컴퓨터공학부 석사과정
[2] 서울대학교 컴퓨터공학부 박사과정
[3] 서울대학교 컴퓨터공학부 교수

eunbee.hwang@snu.ac.kr, wayles@snu.ac.kr, tkkwon@snu.ac.kr

Analysis of DNS TXT Record Usage and Misuse

Eunbee Hwang[1], Hyunsoo Kim[1], Taekyoung “Ted” Kwon[1]
[1]Dept. of Computer Science and Engineering, Seoul National University

요 약

현대 인터넷 인프라의 핵심 요소 중 하나인 DNS 는 도메인 이름을 IP 주소로 변환하여 사용자가 원하는 웹사이트에 접근할 수 있게 한다. 이 과정에서 다양한 레코드 형식이 사용되는데, 그중 TXT 레코드는 이메일 인증 및 도메인 소유권 검증과 같은 중요한 역할을 수행한다. 그러나 최근 TXT 레코드의 사용이 증가하면서 무분별한 남용 사례가 발생하고 있으며, 이는 성능 저하 및 보안 위협을 초래할 수 있다. 본 논문은 국내 인터넷 환경에서 DNS TXT 레코드의 사용 현황을 최초로 조사 및 분석한 연구로, 84,005 개 도메인에서 57,680 개의 TXT 레코드를 수집하여 TXT 레코드의 분포 및 오남용 사례를 파악하였다. 이를 통해 TXT 레코드 남용이 시스템 성능 및 보안에 미치는 영향을 규명하고, 향후 효과적인 관리 방안을 제시하고자 한다.

1. 서론

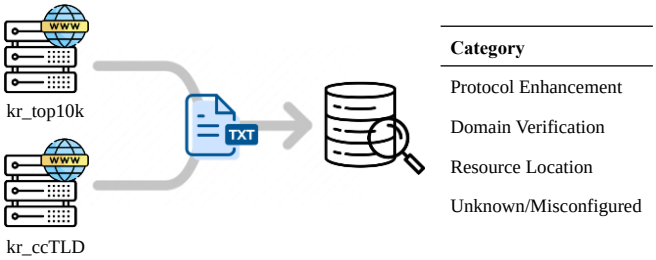
DNS(Domain Name System)는[1][2] 도메인 이름을 IP 주소로 변환하여 사용자가 원하는 웹사이트에 접근할 수 있게 하는 현대 인터넷 인프라의 핵심 요소이다. DNS 는 A, AAAA, CNAME, MX, TXT 등 다양한 레코드를 포함하고 있으며, 각 레코드 유형에 따라 서버와 관련된 여러 종류의 정보를 요청하고 받을 수 있다. A, AAAA, CNAME, MX 등의 레코드는 고정된 형식으로 정의되어 있지만, TXT 레코드는 자유 형식으로 도메인과 관련된 텍스트 정보를 저장하는 데 사용된다. 특히 TXT 레코드는 이메일 인증 및 도메인 소유권 검증을 위한 SPF(Sender Policy Framework)[3]나 ACME(Automated Certificate Management Environment)[4]와 같은 여러 표준에서 중요한 역할을 하고 있다.

그러나 최근 TXT 레코드의 사용 증가와 함께 무분별한 남용 사례가 나타나고 있다[5]. 이러한 남용은 실제로 TXT 레코드를 필요로 하는 프로토콜의 쿼리 처리 시 지연을 초래하고, UDP 데이터그램 사이즈의 증가로 인해 패킷 드롭의 위험을 높일 수 있다. 또한, DNS Resolver 의 캐싱 부담을 증가시키고, Amplification 공격에 악용될 가능성도 존재한다[6][7].

본 논문은 국내 인터넷 환경에서 DNS TXT 레코드의 사용 현황을 조사하고 분석하는 최초의 연구로, 관련 연구가 부족한 상황에서 DNS TXT 레코드의 오남용이 시스템의 성능 및 보안에 미치는 영향을 명확히 파악할 수 있는 기회를 제공한다. 이를 통해 향후 효과적인 관리 방안을 도출하는 것이 본 연구의 주요 목표이다.

2. 본문

본 논문은 국내 인터넷 환경에서 DNS TXT 레코드가 어떤 형태로 사용되는지 알아보기 (그림 1)에서와 같이 두 종류의 도메인 데이터셋으로부터 TXT 레코드들을 수집한 다음 분석을 진행한다.



(그림 1) 데이터셋 확보 및 분석

2.1. 도메인 데이터셋

(1) kr_top10k

해당 데이터셋은 SIMILARWEBS[8]에 의해 수집되었으며, 우리나라 이용자들의 트래픽이 가장 많이 발생된 10,000 개의 도메인을 모아 놓은 데이터셋이다. 2024 년 6 월부터 8 월까지의 트래픽양을 기준으로 만들어졌으며, 해외(yotube.com)와 국내(naver.com) 도메인이 모두 포함되어 있다. kr_top10k 로 국내 웹서비스 제공자들의 TXT 레코드 사용 현황을 파악하기는 어려우나, 활발하게 운영중인 도메인들의 TXT 사용 현황을 파악하는데 용이할 것으로 판단된다.

(2) kr_ccTLD

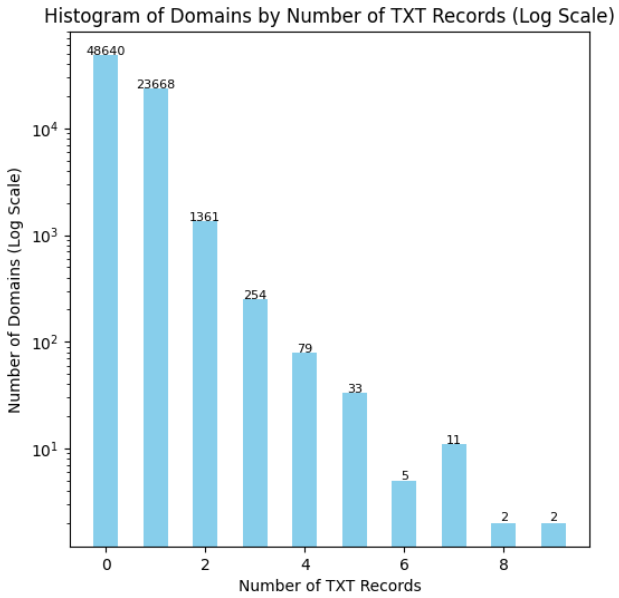
해당 데이터셋은 “.kr” ccTLD(Country code top-level domain)에 속한 도메인 74,005 개로 구성된 데이터셋으로, 여기에는 *.co.kr, *.ac.kr, *.go.kr 등이 포함된다. 현재 “.kr”에 속한 도메인을 파악하기 위하여, 본 논문에서는 CT(Certificate Transparency)[9] 모니터 가운데 하나인 Censys[10]를 활용한다. 2024 년 9 월 기준, CT 로그와 인터넷 스캔을 통해 파악된 “.kr”로 끝나는 도메인의 수는 1,400 만개 이상이나, 이 가운데 현재 유효하면서 chain-of-trust 로 검증 가능한 TLS(Transport Layer Security) 인증서를 보유하고 있는 도메인은 74,005 개로 확인되었다. 국내 도메인들 가운데 “.com”, “.net”과 gTLD 에 등록된 도메인들도 많기 때문에 해당 데이터셋은 국내의 모든 도메인을 보유하지 않으며, 더 나아가 “.kr”에 속하지만 해외에서 운영되는 도메인들도 충분히 존재할 수 있다. 그럼에도 앞서 언급한 kr_top10k 데이터셋과 국내 인터넷 환경에서의 TXT 사용 현황을 파악함에 있어 두 데이터셋이 유의미한 결과를 제공해줄 것으로 기대된다.

2.2. TXT 레코드 수집

다음으로 도메인 데이터셋들을 활용하여 각 도메인으로부터 TXT 레코드를 수집한다. 데이터 수집은 리눅스 환경의 수집 서버에서 약 1 분당 100-150 개의 도메인을 쿼리하는 속도로 진행된다. UDP 로 전송되는 DNS 특성상 중간에 패킷이 유실될 수 있어, 응답을 받지 못한 경우에 대한 재전송 로직을 필요로 한다. 결과적으로 kr_top10k 데이터셋에 해당하는 10,000 개의 도메인으로부터 도합 29,912 개의 TXT 레코드가 수집됐고, kr_ccTLD 의 74,005 도메인들로부터는 도합 27,768 개의 TXT 레코드가 수집됐다.

2.3. TXT 레코드 분석

kr_ccTLD 데이터셋을 활용하여 TXT 레코드의 분포를 분석하였다. (그림 2)에서 확인할 수 있듯이, 가장 많은 TXT 레코드를 보유한 도메인은 9 개의 TXT 레코드를 갖고 있었으며, 두 개의 도메인이 발견되었다. 이 데이터셋에서 65.73%의 도메인은 0 개의 TXT 레코드를 보유하고 있었고, 2 개 이상의 TXT 레코드를



(그림 2) kr_ccTLD 의 TXT 레코드 개수에 대한 히스토그램

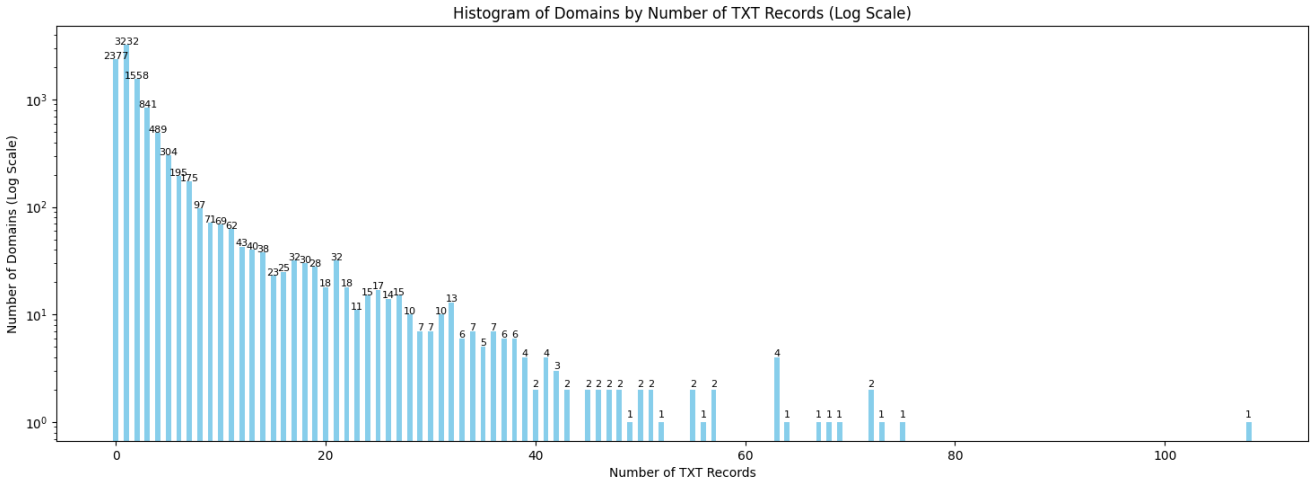
활용하는 도메인은 겨우 2.36%에 불과하였다. 결과적으로, 하나의 도메인은 평균적으로 0.38 개의 TXT 레코드를 보유하고 있어, 대다수의 도메인이 TXT 레코드를 거의 사용하지 않음을 시사한다.

반면, kr_top10k 데이터셋을 활용한 분석에서는 (그림 3)에서 확인할 수 있듯이, 가장 많은 TXT 레코드를 보유한 도메인이 108 개의 TXT 레코드를 갖고 있었으며, 하나의 도메인은 평균적으로 2.99 개의 TXT 레코드를 보유하고 있었다. 이러한 결과는 “.kr” 도메인보다 TXT 레코드의 사용이 상대적으로 더 활발함을 나타낸다. 이는 트래픽이 더 많은 도메인에서 TXT 레코드의 활용도가 높다는 점과, 도메인의 운영 목적이나 활용 방식이 상이할 수 있음을 보여준다.

또한, 수집된 kr_top10k 와 kr_ccTLD TXT 레코드를 각각 4 가지 카테고리 Protocol Enhancement, Domain Verification, Resource Location, Unknown/Misconfigured 로 분류하여[11] 분석한 결과를 <표 1>과 같이 확인 할 수 있다. 여기서 Protocol Enhancement 는 SPF, DMARC, DKIM 과 같이 타 프로토콜 보안성 증대를 위해 TXT 레코드가 활용되는 경우이며, Domain Verification 은

Category (kr_top10k)	**Use Cases**	**Count (%)**
Protocol Enhancement	SPF, DMARC, DKIM, …	6897 (23.06%)
Domain Verification	N/A	21981 (73.49%)
Resource Location	Email, URL, …	455 (1.52%)
Unknown/Misconfigured	N/A	579 (1.94%)
Category (kr_ccTLD)	**Use Cases**	**Count (%)**
Protocol Enhancement	SPF, DMARC, DKIM, …	23453 (84.46%)
Domain Verification	N/A	4121 (14.84%)
Resource Location	Email, URL, …	88 (0.32%)
Unknown/Misconfigured	N/A	106 (0.38%)

<표 1> kr_top10k 와 kr_ccTLD TXT 레코드 분류 결과



(그림 3) kr_top10k 의 TXT 레코드 개수에 대한 히스토그램

SaaS 솔루션 활용을 위해 도메인 소유권 증명이 필요할 때 사용되는 TXT 레코드들이다. Resource Location 은 TXT 레코드를 통해 다른 URL 이나 Email 주소를 포인팅하고 있는 경우에 해당하며, 이외 알 수 없는 활용 형태(unclassified) 및 잘못된 양식으로 사용되거나 오남용으로 판단되는 경우 Unknown/Mis-configured 로 분류하였다.

kr_top10k 데이터셋에서 TXT 레코드의 활용도를 살펴본 결과, Domain Verification 이 전체의 73.49%로 가장 큰 비중을 차지했다. 그 다음으로는 Protocol Enhancement 가 23.06%, Resource Location 이 1.52%로 뒤를 이었다. 반면, kr_ccTLD 데이터셋에서는 Protocol Enhancement 가 84.46%로 가장 많이 사용되었으며, 이어서 Domain Verification 이 14.84%, Unknown/Mis-configured 가 0.38%를 차지했다. 이를 통해 kr_top10k 와 같이 실제 인터넷 사용자들이 트래픽이 많이 발생하는 웹서비스의 경우 도메인에 대한 소유권 검증을 다양한 서비스 제공자(예: Google, Facebook, Microsoft, …)에게 수행하는 것이 부각되는 반면, 단순하게 “.kr” 로 끝나는 도메인을 갖고 있는 측면에서는 SPF 와 같이 이메일 보안과 연계된 TXT 레코드의 비중이 늘어나는 것을 확인할 수 있었다.

2.4. TXT 레코드 오남용

수집된 57,680 개의 TXT 레코드 가운데 정해진 표준 규칙을 따르지 못하는 레코드들과 오남용 되는 레코드들을 확인할 수 있었다. 대표적으로는 <표 2>의 첫번째 사례와 같이 kr_ccTLD 도메인 가운데 10,517 도메인이 모두 단 한개의 SPF TXT 레코드를 갖고 있다. IPv6 주소 fd1b:212c:a5f9::/48 는 ULA(Unique Local Address)로, IPv4 의 192.168.x.x 나 10.x.x.x 와 유사하여 인터넷에서는 라우팅이 불가능하다. 따라서, 이 주소를 도메인의 SPF 레코드에 포함하는 것에는 호스팅 네임서버상의 이슈가 있었고, 이것이 10,517 개 도메인에 공통 적용된 것으로 예상된다.

이외에도 DKIM, DMARC 의 TXT 레코드가 표준대로 입력이 안 된 사례가 전체 TXT 레코드 가운데 79 개를 확인할 수 있었는데, 레코드를 봤을 때 DNS 네임서버 프로그램에 TXT 레코드를 추가하는 과정에서 입력 오류가 있었던 것으로 보여진다. 마지막으로는 단순 영어문장, 자기소개, “ ”(공백), 1, ~과 같은 single character 레코드들도 총 17 개를 발견했다.

앞선 절에서 언급한 바와 같이 전체 TXT 레코드에서 정확한 용도를 파악할 수 없어 미분류(unclassified)로 분류된 레코드들의 수도 꽤 많기 때문에 실제 오남용 되거나 잘못 설정되어 무의미한 TXT 레코드의 개수는 훨씬 많은 것으로 예상한다.

Examples	Category	Count
• v=spf1 ip6:fd1b:212c:a5f9::/48 -all	Misconfiguration	10517
• _dmarc.***.co.kr IN TXT \"v=DMARC1; p=none; rua=mailto:***@naver.com\" • default._domainkey IN TXT (\"v=DKIM1; h=sha256; k=rsa; \" \"p=MIIBIjANBCgKCA … Y5ffVQAB\") ; ----- DKIM key default for ***.co.kr	Incorrect record	79
• WHERE R U LOOKING AT? • enjoy hacking with dns? we search for curious minds! => ***.com <= • “ ”, “1”, “~”	Random string	17

<표 2> TXT 레코드 오남용 사례

3. 결론

본 논문은 국내 인터넷 환경에서 DNS TXT 레코드의 사용 현황과 오남용 사례를 파악하였다. 이를 위해 “.kr” ccTLD 에 속하는 도메인과 국내에서 트래픽이 가장 많이 발생한 도메인을 합쳐 총 84,005 개 도메인에서 57,680 개의 TXT 레코드를 수집하여 분석하였다. 이를 통해 최소 18.40%(10,613 개)의 TXT 레코드가 오남용되고 있음을 발견하였다.

본 논문에서 수행된 연구는 예비 분석(preliminary analysis)의 성격을 띠고 있으며, 보다 엄밀한 조사를 위해서는 “.kr” 도메인에 국한하지 않고 각 도메인의

IP 분석을 통해 국내에 서버를 두고 있는지 여부를 확인하는 작업이 필요할 것이다. 또한, 국내 호스팅 사업자들이 운영하는 네임서버의 방식과 사용하는 소프트웨어, 관리도구들에 대한 분석도 이루어져야 한다. 더 나아가 활용 방식을 알 수 없었던 TXT 레코드들을 파악하기 위해 네임서버 zone file 관리자와의 면담, TXT 레코드를 활용하는 논문/특허 등 기술 자료 조사가 추가적으로 필요하다. 향후 연구는 이러한 개선 사항을 반영하여, 더욱 효율적이고 안전한 DNS TXT 레코드 관리 방안을 제시할 수 있을 것으로 기대한다.

ACKNOWLEDGEMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터육성지원사업 및 클라우드-네이티브 이동통신 시스템 원천기술 개발 및 리더십 구축의 연구결과로 수행되었음 (IITP-2024-2021-0-02048, IITP-2024-RS-2024-00418784)

참고문헌

[1] Mockapetris, Paul V. "RFC1034: Domain names-concepts and facilities.", 1987.

[2] Mockapetris, Paul V. "RFC1035: Domain names-implementation and specification.", 1987.

[3] Kitterman, S. "RFC 7208: Sender Policy Framework (SPF) for Authorizing Use of Domains in Email.", 2014.

[4] Barnes, R., et al. "RFC 8555: Automatic Certificate Management Environment (ACME).", 2019.

[5] Van Der Toorn, Olivier, et al. "TXTing 101: finding security issues in the long tail of DNS TXT records." *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 2020.

[6] Kambourakis, Georgios, et al. "Detecting DNS amplification attacks." *Critical Information Infrastructures Security: Second International Workshop, CRITIS 2007*, 2007.

[7] Kovacs, Eduard. "Large DNS Text Records Used to Amplify DDoS Attacks: Akamai." *SecurityWeek*, 2014, www.securityweek.com/large-dns-text-records-used-amplify-ddos-attacks-akamai/.
Accessed 2024. 09. 20.

[8] Similarweb, www.similarweb.com.
Accessed 2024. 09. 20.

[9] Laurie, Ben. "Certificate transparency." *Communications of the ACM* 57.10, 2014.

[10] Durumeric, Zakir, et al. "A Search Engine Backed by Internet-Wide Scanning." *Proceedings of the 22nd ACM Conference on Computer and Communications Security*, Oct. 2015.

[11] Portier, Adam. “Security In Plain TXT Observing the Use of DNS TXT Records in the Wild.” MS thesis. Villanova University, 2018.