# 統計データを基にした サブドメイン名と公開ポートの関係性の推定

坂田 拓美（福井大学大学院工学研究科）
・小高 知宏　黒岩 丈介（福井大学大学院工学研究科）
・諏訪 いずみ（仁愛女子短期大学）・白井 治彦（福井大学工学部）

## 1　はじめに

昨今では，インターネット上の様々な情報が収集され，多種多様な目的に利用されている．公開された情報をもとにして，目的とする情報を収集・分析する手法を OSINT (Open Source Intelligence) と呼ぶ [1].

ドメイン名に関する情報も OSINT として利用できる場合がある．現在 DNS における OSINT としては，サブドメインをリストアップすることによる稼働マシンのリストアップや，whois 情報が一般的となっている [2]．しかし，これら以外にも利用方法がある可能性が考えられる．

そこで新たな情報として，サブドメイン名から対象サーバーで稼働中のサービスを推定し，攻撃手法の選別に用いられてしまう可能性を考える．本研究の目的は，サブドメイン名とサービスに関連の深い開放ポートの間に，推定可能な関係性が存在するか統計データから検証することである．

## 2　サブドメイン名の統計的特徴

Rapid7 が Project Sonar の一環として公開している Open Data[3] の一つである，Forward DNS (FDNS) 2021-04-23-1619136719-fdns_a.json をデータとして用いた．

サブドメインの命名に固有のパターンが存在するのかを知るため，同一のサブドメインが命名されることが多いのかを確認した．図 1 は，横軸にデータ内で同一のサブドメインが何回出現してきたかで分けた組を，縦軸に各組に該当するサブドメインの種類数を記している．重複がなく，1 回しか出現しないサブドメインが圧倒的に多いことが分かる．これによって，同一のサブドメインでのデータをもとに開放ポートを推測するといった手法は，対象となるサブドメインと同一のものが存在する確率が低くなり，適用することは難しいと言える．

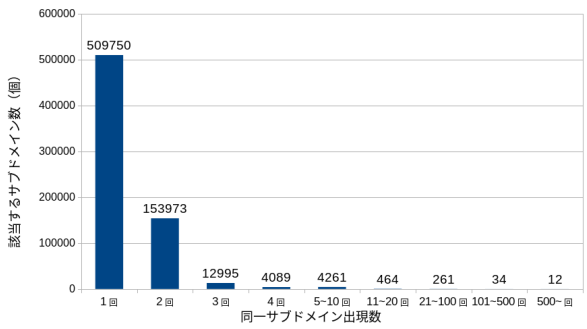


図 1: 各出現回数ごとのサブドメイン数

## 3　サブドメイン名と開放ポートの統計的特徴

対象サーバー上で稼働中のサービスをサブドメインから推測するため，サービスと関係が深い開放されたポート番号を用いる．そのためまずは，サブドメイン名と開放ポート間に関係性があるか確認を行う．

統計上，もっとも多い開放ポートは HTTP の 80 番ポートとなっており，99%近くの割合でドメインが割り当てられたサーバーでは開放されていることが確認できた．反対に，それ以外のポートは次点で 7%程度の 53 番と，大きな開きがある．

## 4　考察とまとめ

今回の検証の結果，80，443 番ポートについては多くのサブドメインで現れたが，それ以外のポートとサブドメイン名の間に関係性を見出すことはできなかった．統計データのみでサブドメイン名から開放ポートの推測を行うことは難しいと言える．

しかし，サブドメイン名は主に人間が命名していることから，文字の並びと開放ポートの間には何らかの関係性が見出される可能性が考えられる．課題として，サブドメイン名について自然言語処理を適用して検証を行っていく．

## 参考文献

[1] Michael Glassman and Min Ju Kang. Intelligence in the internet age: The emergence and evolution of open source intelligence (osint). *Computers in Human Behavior*, Vol. 28, No. 2, pp. 673–682, 2012.

[2] Media Sonar. Osint techniques for domain pois. `https://opendata.rapid7.com/`, 2020. [Online; accessed 7-July-2021].

[3] Rapid7. Rapid7 labs - open data. `https://opendata.rapid7.com/`, 2021. [Online; accessed 7-July-2021].