「マルチメディア，分散，協調とモバイル
(DICOMO2019)シンポジウム」 令和元年7月

Kappa 指標による大学別 UDP リフレクタ数の分析

今村 弦[1] 岡村 耕二[†1]

概要： DDoS に悪用されかねない UDP リフレクタを逓減することは永続的な課題のひとつであり、そのためには、各組織にどの程度の UDP リフレクタが存在するかを定量的に把握し相互に比較するための指標が必要である。そのような指標としては既に国単位で、国別コードのドメインである ccTLD から、ある ccTLD に存在する UDP リフレクタ数と増幅率および経路の帯域幅かその ccTLD が送出できる最大の DDoS トラフィック流量を推定した DDoS ポテンシャルが提案されている。DDoS ポテンシャルは各 ccTLD の状況を知る上で有用であるが、絶対量による指標であるため往々にして割当済 IP アドレス数の大きな ccTLD のスコアが高くなる。UDP リフレクタの状況を比較するに当たっても正規化した指標である Kappa 指標を、インターネットセキュリティ全般について情報収集・事案対応と調整・国際および国内連携等を行う組織である JPCERT/CC が提案し、指標を用いた分析を行っている。これに対し本研究では、ccTLD 単位に比べてより細かい単位での分析に注目し、大学単位での分析を JPCERT/CC 提案の Kappa 指標を用いて行った。また、国単位での Kappa 指標による評価に比べデータ数が必然的に少ないためフリースケール性の確認を行い、対数を取ることで線形性が現れること、正規性の確認を行いサンプルが適正に抽出されていることの 2 点において JPCERT/CC が行った回帰分析と同様の回帰分析が行えることの確認を行い、大学別 UDP リフレクタ数の分析を行った。

Analysis of UDP Reflectors per University by Kappa Index

GEN IMAMURA[1] KOJI OKAMURA[†1]

1. はじめに

DDoS 攻撃は反射型 DDoS 攻撃の踏み台として用いられる DNS サーバや NTP サーバなど UDP リフレクタは UDP 通信を用いているため送信元の IP アドレスの偽装が容易であり 2006 年に観測されて以降これらの UDP リフレクタは攻撃の踏み台として利用されてきた。このような UDP リフレクタを用いた反射型 DDoS 攻撃は踏み台になりうるサーバの管理者が適切な設定を行うことで防ぐことのできる攻撃であるが適切な設定を行えていないサーバはネットワーク上に存在しており、そのようなサーバを多く保有している組織は反射型 DDoS 攻撃に加担するリスクが高い組織であると言える。反射型 DDoS 攻撃に加担するリスクを評価する際に絶対数を用いて評価を行った場合、ネットワーク規模が大きくなるにつれて UDP リフレクタの数も多くなり、ネットワーク規模が大きい組織はネットワーク規模の小さい組織に比べてリスクが高くなる傾向がある。

JPCERT/CC は反射型 DDoS 攻撃の踏み台になり攻撃に意図せず加担してしまう可能性のあるサーバが相対的に多く存在するネットワークを明らかにし、評価を行うためネットワークサイズに対して平均的に期待されるリスクを基準とした指標として国や地域単位で反射型 DDoS 攻撃の踏み台になるリスク評価を行ったインターネット雑草指標と呼ばれる Kappa 指標を考案しインターネット可視化サービスとして公開している。Kappa 指標は絶対数を用いた比較とは異なりネットワークサイズに対して平均的に期待される UDP リフレクタ数を用いているためネットワークサイズに依存しない比較が可能となり、ネットワークの規模は小さいが UDP リフレクタが多く存在するような対策を行っていない可能性の高い組織を見つけることが可能となっている。Kappa 指標ではネットワークサイズと UDP リフレクタ数をそのまま散布図にプロットしても相関が得られないため、ネットワークサイズと UDP リフレクタのフリースケール性の証明を行った上でそれぞれべき乗則に従うことを示し、両対数を用いて線形回帰分析を行っており回帰直線からの乖離より Kappa 指標の算出している。また算出した Kappa 指標が正規分布であることを言及しており乖離を用いた指標の有用性を示している。

JPCERT/CC 考案の Kappa 指標では国や地域単位での評価指標であるが本研究では新たにもっとミクロな大学組織単位で Kappa 指標による反射型 DDoS 攻撃の踏み台にされるリスク評価を大学別 Kappa 指標として国単位での Kappa 指標を基に線形回帰分析を行うための前段階としてフリースケール性の証明、その後検索エンジンの Shodan が提供しているスキャンデータベースである HighRedCenter いて抽出した各大学のネットワークサイズと DNS、NTP、RPC、SIP、SNMP、SSDP の各 UDP リフレクタについて線形回帰分析より大学別 Kappa 指標の算出と比較、サンプルに用い

1 九州大学大学院システム情報科学府
†1 九州大学情報基盤開発研究センター

たデータが適正なものであるかの検定として算出した Kappa 指標について正規性の検定を行った。

2. 背景

2.1 反射型 DDoS 攻撃

DDoS 攻撃は 2000 年にアメリカ合衆国で攻撃事例が観測されて以降現在に至るまで特定のサービスへの攻撃によるサービス停止や従量課金制のクラウドへの攻撃による経済被害など多くの被害が報告されており具体的な対策が提案されているが現在でもその被害は続いており 2018 年 12 月に IIJ が対応した DDoS 攻撃の累計は 396 件と少なくはなく、その中で一番大きな規模であった攻撃には DNS サーバを用いた反射型の DDoS 攻撃である。

これまでの DDoS 攻撃は脆弱性を持っている複数のデバイスにボットネットと呼ばれる指令に従い同時に命令を動作するマルウェアを利用してデバイスを乗っ取り大量の通信を送り付ける方法が中心だったが反射型の DDoS 攻撃は UDP リフレクタ（以下リフレクタ）と呼ばれる送信側パケットの大きさに比べて受信側パケットが非常に大きいプロトコルを増幅器として利用し大量の通信を送り付ける方法である（図 1）

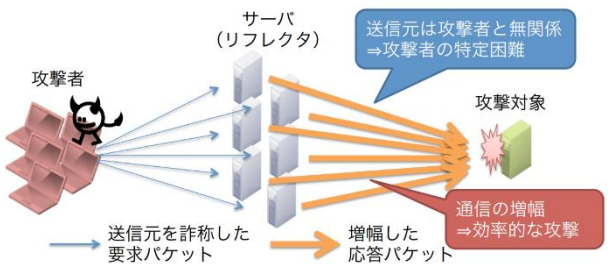

図 1：反射型 DDoS 攻撃のイメージ図

リフレクタには家庭用のルータや通信プロトコルが用いられ、通信プロトコルでは DNS（53 番ポート）や NTP（123 番ポート）が利用されることが多いがここ数年は SNMP（161 番ポート）や SSDP（1900 番ポート）がリフレクタとして利用されている例も報告されている。具体的な攻撃の手法として DNS サーバを用いた攻撃を例にするが、攻撃者はリフレクタに非常に大きなパケットの要求をあらかじめ埋め込んでおき、その後攻撃者は IP アドレスを被害者のものになりすましあらかじめ埋め込んだ要求を呼び出すことで大量の通信を送り付ける形になっている。反射型 DDoS 攻撃は以前多く見られたボットネットを用いた DDoS 攻撃に比べて乗っ取りを行うデバイスの数を少なくすることができ、UDP を用いているため攻撃者は被害者の IP アドレスに成りすましてリフレクタに対して要求を行うため攻撃者の特定を行うことが困難であると言った違いがある。

DNS サーバのようなプロトコルが外部に向けてオープンになっていると反射型 DDoS 攻撃の踏み台にされるのだが、DNS サーバを保有しているネットワークの管理者がキャッシュサーバとコンテンツサーバ分離し外部からキャッシュサーバに問い合わせができないようにするといった適切な設定を行うことで反射型 DDoS 攻撃の踏み台になることは避けることが可能であり反射型 DDoS 攻撃の踏み台は少しずつ減少してきてはいる。具体的な方法が示されている現在でも対策を行っていないサーバは存在しており、そのようなサーバは反射型 DDoS 攻撃に踏み台に利用されるリスクを持っているのが現状である。

2.2 Kappa 指標

反射型 DDoS 攻撃は 2006 年あたりから観測され具体的で効果的な対策の方法が提案されてきているが反射型 DDoS 攻撃の被害は無くなっていない。反射型 DDoS 攻撃に関する情報をまとめたサイトも多く存在し定量的な評価を行っている可視化サイトも存在するが、多くの可視化サイトがリフレクタの絶対数を用いたものであり、国単位での評価においては保有しているネットワークの規模が大きいアメリカや中国がいつも上位に頻出しており、ネットワーク規模が大きい国や組織ばかりが目についてネットワーク規模は小さいがリフレクタ放置しているようなネットワーク内で管理がしっかりとなされていない民度の低い国や組織を見つけることが難しい。

日本区内のセキュリティインシデントへの対応や情報連携を行っている JPCERT/CC が国や地域単位で反射型 DDoS 攻撃の踏み台にされるリフレクタの数からネットワークにサイズに対する反射型 DDoS 攻撃に加担するリスク示したインターネット雑草指標（Kappa 指標）という指標を考案し、ネットワーク規模に寄らない反射型 DDoS 攻撃に関する定量評価を行った。国や地域単位での反射型 DDoS 攻撃への加担リスク評価を行っている Kappa 指標は同サイト“インターネットリスク可視化サービスーMejiroー”で公開されている。

Kappa 指標はネットワークの規模が大きくなるに従って攻撃に利用されるリフレクタの数も増大するといった仮説の基、国や地域別にある一定規模のネットワークに対して平均的に期待されるリフレクタの数を回帰直線として導出し、平均的に期待されるリフレクタの数に比べて実際のリフレクタの数が多いか少ないかを偏差値を用いて算出している評価指標である。Kappa 指標を用いることで各ネットワークがネットワーク規模に対してどの程度偏差しているかが分かりネットワーク規模の違う国同士で比較が可能になるため、ベストプラクティスを学ぶ手がかりになり、同じ国の中で DNS や NTP などの各リフレクタについてどの項目がシビアな状況にあるかの比較も可能になるため対応の優先順位を決めることも可能になる。

また、Kappa 指標ではネットワークサイズとして対象のネットワークが保有している IP アドレス、UDP リフレク

タ数を Shodan のデータベースを用いて算出を行っているがネットワークサイズと UDP リフレクタ数をそのまま散布図にプロットしただけでは相関が出ないことが分かっている。これはネットワークの持つフリースケール性によるものであり、数の大きいものは少なく数の少ないものほど多くなるべき乗則に従う。フリースケールのようなべき乗則に従うデータは両対数を取ることでデータが線形性を持つので線形回帰分析が可能となるので Kappa 指標の散布図には両対数グラフが用いられている。

“インターネットリスク可視化サービスーMejiroー”で公開されている DNS についての UDP リフレクタとネットワーク規模の散布図を以下の図に示した。（図２）

縦軸に UDP のリフレクタ数、横軸に IP アドレスの総数を取っておりそれぞれ両対数を取った値を散布図にプロットして線形回帰分析を行っている。図２中の直線よりも上側の部分は平均的に期待されるリスク数より多く反射型 DDoS 攻撃に加担するリスクが高い民度の高い国、逆に直線より下側の部分は平均的に期待されるリスク数よりも少なく反射型 DDoS 攻撃に加担するリスクが低い民度の低い国であると予測される。

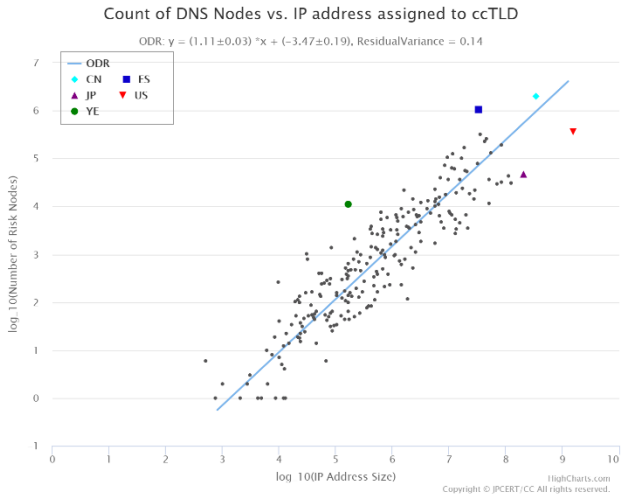

図 ２：DNS についての UDP リフレクタ数とネットワーク規模の散布図

図２中の直線は回帰直線を表し、プロットは全体的に回帰直線に沿っていて相関がある事がうかがえる。中国（CN）、日本（JP）、イエメン（YE）、スペイン（ES）、アメリカ（US）の５国を例に Kappa 指標の特徴について、まず絶対数での比較を行った場合スペインと中国、アメリカの３国は保有 UDP リフレクタ絶対数が多く絶対数では上位に入る国であり中国が最も多くの UDP リフレクタを保有している。その３国に比べて日本やイエメンは UDP リフレクタの絶対数は少ない値となっている。また IP アドレスの絶対数を見るとイエメン以外の国は多く、アメリカが最も多くの IP アドレスを保有している。これら５か国の Kappa 指標を取ると中国は+0.549、日本は-1.889、イエメンは+3.076、スペインは+2.044、アメリカは-2.071 となっている。

Kappa 指標では正の値は回帰直線の上側にプロットされていることを示しており平均的に期待されるリスク数よりも実際のリスク数が多く値が大きくなるほどハイリスクであり、逆に負の値では回帰直線の下側にプロットされていることを示しており平均的に期待されるリスク数よりも実際のリスク数よりも少なく値が小さくなるほどローリスクである。先程の各国の Kappa 指標の値を見ると最も Kappa 指標の値が高い国はイエメンの+3.076 であり、絶対数ではイエメンよりも多く UDP リフレクタを保有しているアメリカや中国と比べても非常に高い値となっている。これはネットワーク規模に対して UDP リフレクタを保有しすぎていると言え、逆にネットワーク規模の大きく UDP リフレクタの数も他の国に比べて多いアメリカや日本はそれぞれ負の値を取り、ネットワーク規模に対して平均的に期待されるリスク数よりも少なくネットワークの民度は高いと言える。最も多くの UDP リフレクタ数を保有している中国も+0.549 と平均的に期待されるリスク数と比べると多いがイエメンやスペインと比べると非常に良い値であると言える。

Kappa 指標での結果を見ると絶対数だけでは中国やアメリカのような国ばかりが目立ち、絶対数がさほど多くないため見つかりにくいイエメンのような相対的にネットワークの民度が低い国をネットワークサイズに依存しない比較方法で見つけることができるため優先的に対策を行うと言った対策の優先順位を決める点で生かすことができる。また、Mejiro では一つの国や地域についての各 UDP リフレクタの Kappa 指標の値をレーダーチャートにするサービスもあり UDP リフレクタ同士で Kappa 指標の比較を行うとどの UDP リフレクタの対策を優先的に行うべきかの指針にもなる。

2.3 研究目的

日本国内の組織についての反射型 DDoS 攻撃への加担リスクについて、大学内にも反射型 DDoS 攻撃の踏み台になりうるポートが存在しており比較を行う上で企業と大学では、保有している UDP リフレクタには差があり大学単位でも同様に保有している UDP リフレクタには差がある。Kappa 指標での評価において、細かい単位での評価を行った方がより正確にリスク評価可能である。大学単位で反射型 DDoS 攻撃の踏み台にならないよう対策や啓発活動を行う際に各ネットワーク管理者が対策優先度や対策の有無を決めるために利用できるミクロな指標が必要であると考え、大学単位での Kappa 指標を算出し評価と比較を行うことを本研究の目的とする。国や地域単位での Kappa 指標の導出方法を参考に線形回帰分析を行うための前提としてフリースケール性の検証を行いネットワークサイズと各 UDP リフレクタ数より大学別 Kappa 指標の算出、また本研究で利

用したサンプル数が少ないため算出した大学単位での Kappa 指標の正規性の検定によってサンプルデータが適正であるかの確認を行う。

3. Kappa 指標を用いた大学単位での評価

3.1 回帰分析と決定係数

大学単位での Kappa 指標の算出を行う上で用いたデータは HighRedCenter と言うデータ提供サービスを利用した。このサービスは 2009 年に開設されたネットワーク上に接続されているオープンな機器やサーバを検索することが可能な検索エンジンである Shodan.io が提供しているサービスであり、Shodan.io のスキャンデータを用いてネットワーク上のオープンポートに関する検索を可能としたデータベースで直近約 200 日分のスキャンデータを保有している。この HighRedCenter のデータベース内から今回調査対象とした 46 大学について、日本ネットワークインフォメーションサービス（JPNIC）が提供している Whois である“JPNIC Whois Gateway”を利用して各調査対象の大学のセグメントを検索しそれぞれのセグメントに対して Shodan の検索キーより各 UDP リフレクタの計数を行った。また反射型 DDoS 攻撃の踏み台にされる可能性のある UDP ポートとしてオープンである DNS（53/UDP）、NTP（123/UDP）、RPC（111/UDP）、SIP（5060/UDP）、SNMP（161/UDP）、SSDP（1900/UDP）の 6 つのポートを調査対象とした。UDP リフレクタを用いた反射型 DDoS 攻撃のリスク評価であるので調査対象の 6 ポート全て UDP ポートを調査対象としている。

また、ネットワークサイズに関して国単位での Kappa 指標では IP アドレスの総数をネットワークサイズとして利用しているが大学別の比較を行うに際し IP アドレスの総数を参照すると多くの大学が同じクラスのネットワークを持っているため IP アドレスの総数で算出を行った Kappa 指標と UDP リフレクタ数の比較が同じになってしまう、そのため大学別 Kappa 指標ではネットワークサイズに観測されたオープンポートの総数と日数と累計をネットワークサイズとして利用する。また、Shodan では 1 日に 1 度スキャンを行うため本研究で利用している UDP リフレクタ数とネットワークサイズは観測されたポート数と日数の累計となっている。

HighRedCenter のデータから調査対象大学、対象の UDP リフレクタをそれぞれ抽出したのち横軸にネットワークサイズ、縦軸に UDP リフレクタ数のそれぞれ常用対数を取って散布図の作成、回帰直線の導出を行う。大学別 Kappa 指標は国や地域単位での Kappa 指標での導出方法を基本としてネットワークサイズと対象の UDP リフレクタ数との相関を調べる為、ある一定規模のネットワークサイズに対して平均的に期待される対象 UDP リフレクタの数を回帰直線として導出、対象大学と同サイズのネットワーク規模で平均的に期待される UDP リフレクタの数と実際の対象大学に存在する UDP リフレクタの数との差を直交回帰の乖離度として算出、偏差値の導出を行っている。Kappa 指標の導出に用いる回帰直線に導出について y を UDP リフレクタの数、x をネットワークサイズのパラメータとして以下の式により回帰直線の導出を行っている。

$$y = ax + b \tag{1}$$

$$a = r \times \frac{\sigma_y}{\sigma_x}、b = \bar{y} - a\bar{x} \tag{2}$$

$$r = \frac{s_{xy}}{\sigma_x \times \sigma_y} \tag{3}$$

y（縦軸）を UDP リフレクタの数、x（横軸）をネットワークサイズとして回帰直線での回帰係数を a、回帰直線の切片を b、相関係数を r、x と y の標準偏差をそれぞれσ_xとσ_y、x と y の共分散をs_{xy}で表記している。回帰直線はネットワークサイズに対して平均的に期待されるリスク数を表しネットワークサイズの対する予測値である。散布図中の回帰直線を基準として上側の部分は平均よりリスクの多いハイリスクなネットワーク、下側の部分は平均よりリスクの少ないローリスクなネットワークであると言える。また以下の式により回帰直線からの乖離度 d の導出、Kappa 指標 κ の導出を行っている。

$$d = \frac{(ax_i + b) - y_i}{\sqrt{(-a^2) + 1}} \tag{4}$$

$$\kappa = \frac{d - \bar{d}}{\sigma_d} \tag{5}$$

回帰直線を$y = ax + b$として$\bar{d}$は乖離度の平均、σ_dは d の標準偏差を表している。Kappa 指標の値が正であり数値が大きい程ネットワークサイズに対して平均的に期待されるリスクに比べ反射型 DDoS 攻撃に加担するリスクが高くなり、逆に負であり数値が小さい程攻撃に加担するリスクは低くなると言える。また、回帰直線の導出と共にモデルの当てはまりの良さの数値である決定係数の導出も行った。決定係数の導出式は以下の通りである。

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{6}$$

y_iは実際の UDP リフレクタ数、$\hat{y}_i$は回帰直線上の平均的に期待される UDP リフレクタ数の予測値を$\bar{y}$は UDP リフレクタ数の平均を表している。決定係数は 0 から 1 の値をとり、1 に近いほど極端な値が少なくモデルの当てはまりよく、データが回帰直線に近い形で分布をしていると言える。

3.2 フリースケール性

線形回帰分析を行う上でフリースケール性の検証が必要であるため、以下フリースケール性の検証を行う。

対象大学ごとに HighRedCenter で算出を行った各 UDP リフレクタに数とオープンネットワーク数を表にまとめた。以下 JPCERT/CC での呼称に則り各 UDP リフレクタ

をリスクノードと呼称する。データオープンポートの累計であるネットワークサイズと DNS リフレクタの数について散布図にプロットを行ったのが以下の図である。（図３）

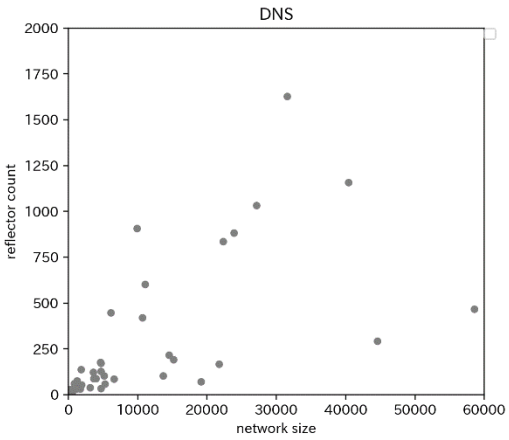

図 3：ネットワークサイズと DNS リフレクタ数の散布図

図３より散布図の左下に密集するように分布しており、逆にネットワークサイズや DNS リフレクタ数が大きい大学は少ない傾向にある。ネットワークサイズでは上位３大学が全体の 66.9%を DSN リフレクタ数では上位３大学が全体の 61.5%を占めている。また、図３の散布図内の四角で囲まれた部分を抜き出したものが図４である。（図４）

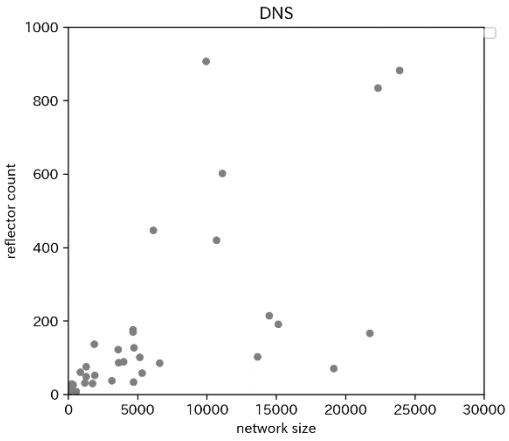

図 4：図３の軸範囲を縮小したネットワークサイズと DNS リフレクタ数の散布図

図３と同様に散布図の左下に密集しサイズの大きなサンプルは少ない傾向が出た。これはべき乗則を持つフリースケールに近い傾向が見られるためフリースケール性が期待できる。フリースケール性を持つデータでは両辺対数を取ることで線形の相関が出る為、線形回帰分析が可能になる。JPCERT/CC が行っている定量評価である Kappa 指標ではフリースケール性を示し線形回帰分析を用いて指標の算出を行っている。図３、図４より大学の各データにおいてもフリースケール性が期待され国や地域単位での指標算出の時と同様に両対数を取ることでデータが線形性を持つことも期待でき線形回帰分析を行うことが可能であると望めるため以下フリースケール性の確認を行った。

ネットワークサイズと DNS リフレクタ数について10^nの指数関数状に範囲を設定し相対密度をヒストグラムにしたものがそれぞれ図５と図６である。相対密度とは積分した時に１となる様な各範囲の発生確率のことであり、それぞれの範囲内に存在する大学数を範囲の広さで平均した値を発生確率の算出値として利用している。指数関数状に範囲を広げているため（図５）（図６）

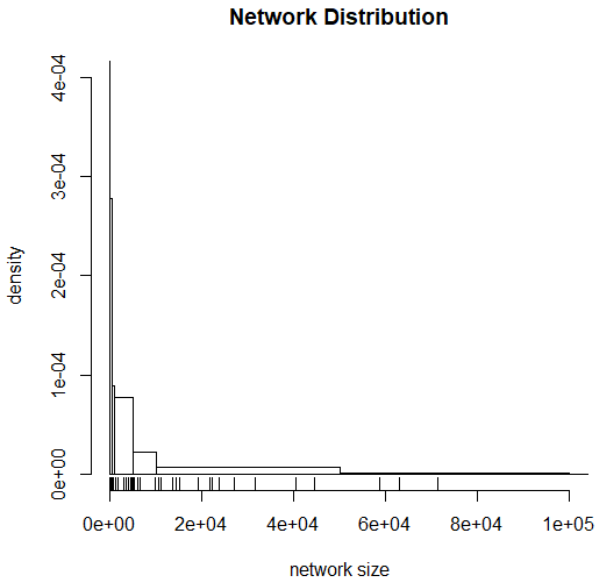

図 5：ネットワークサイズの相対密度

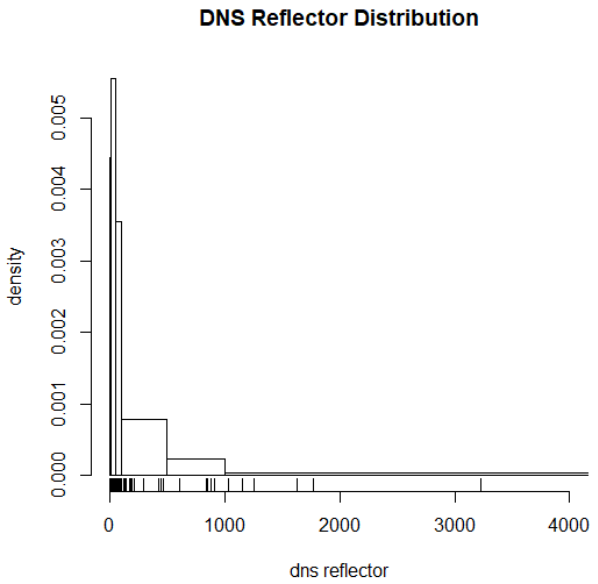

図 6：DNS リフレクタ数の相対密度

図５、図６よりネットワークサイズ、DNS リフレクタに関してフリースケール性の特徴であるロングテールが見られるためネットワークサイズ、DNS リフレクタ共にフリースケール性を持ちべき乗則に従うとする。NTP リスクノードにおいてもネットワークサイズ、DNS リフレクタと同様のロングテールが確認でき、Kappa 指標の算出が可能となった。

3.3 大学単位での Kappa 指標

3-2 で大学のネットワークサイズ、UDP リフレクタ数においてものフリースケール性が確認できたため両対数を用いた線形回帰分析を行う。DNS と NTP 以外の UDP リフレクタに関しては十分なデータが収集できず Kappa 指標を算出しても評価指標として信頼性が低いと考えられるため、本研究では DNS と NTP、２つの UDP リフレクタを線形回帰分析と Kappa 指標の算出対象とした。

まず DNS について縦軸を DNS リスクノード数、横軸をネットワークサイズとして散布図でプロットし回帰直線を導出それぞれの標準偏差 σ を回帰直線に足し引きしたものをそれぞれ図にまとめた。青色の実線は回帰直線を表し、橙色と緑色の実線はそれぞれ回帰直線に標準偏差足した直線と引いた直線を表している。（図７）

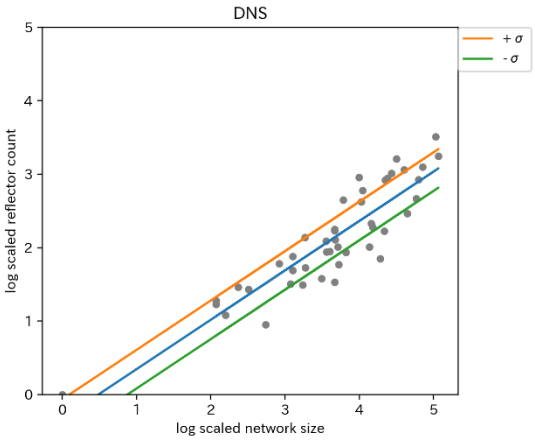

図 7：DNS リスクノードとネットワークサイズの比較

DNS についての直交回帰直線と決定係数R^2は以下の式で表される。

$$y = 0.672x - 0.326 \quad (7)$$

$$R^2 = 0.802 \quad (8)$$

また NTP についても DNS の時と同様に縦軸を NTP リスクノード数、横軸をネットワーク規模として散布図に回帰直線と共にまとめた。（図８）

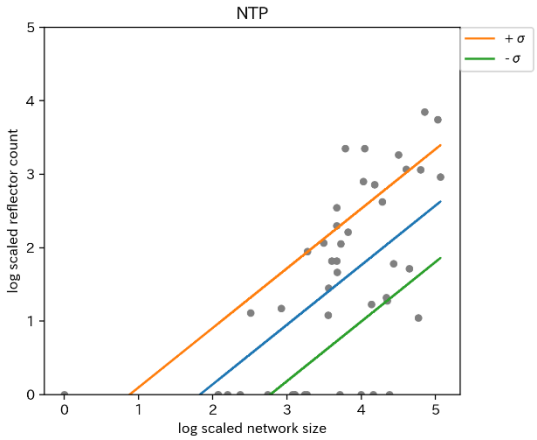

図 8：NTP リスクノードとネットワークサイズの比較

NTP リスクノードの回帰式と決定係数は以下の通りである。x はネットワーク規模を、y は NTP リスクノードを表している。

$$y = 0.811x - 1.481 \quad (9)$$

$$R^2 = 0.377 \quad (10)$$

3.4 大学別 Kappa 指標の正規性

Kappa 指標は中央値を基準とした指標であるため基準に近いほど数が多くなり、逆に極端にリスク数が多い大学や極端にリスク数が少ない大学は少なくなる正規分布に近くなることが予想される。また、調査大学が正規母集団から抽出されているかの証明として以下大学単位での Kappa 指標の正規性についての検証を行う。

DNS、NTP の各リスクノードの大学別 Kappa 指標について 0.5 ずつの区間でヒストグラムにした確率密度分布がそれぞれ図９、図１０である。縦軸は確率密度を表しており、ヒストグラム内の曲線はそれぞれ DNS、NTP リスクノードにおける大学単位での Kappa 指標の平均値と分散値を用いた理想的な正規分布の確率密度関数を表している。（図９）（図１０）

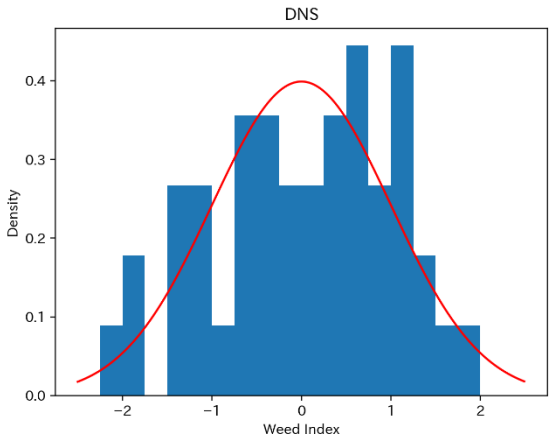

図 9：DNS リスクノードにおける大学別 Kappa 指標の確率密度分布

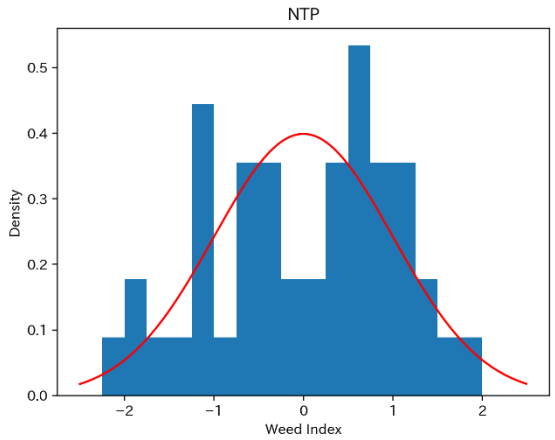

図 10：NTP リスクノードにおける大学別 Kappa 指標の確率密度分布

また各リスクノードについて各大学の Kappa 指標の相対的な出やすさである確率密度を合計していった関数である累積密度関数を取ったものが図１１、図１２である。確率密度分布の累計を取っているため範囲は 0 から 1 となる。また、図９、図１０と同様に曲線は各リスクノードの平均値と分散値を用いた理想的な正規分布の累積密度関数を表している。（図１１）（図１２）

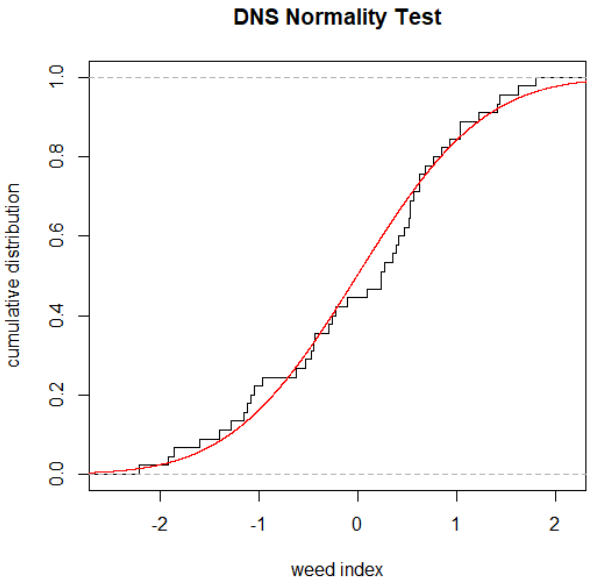

図 11：DNS リスクノードにおける大学別 Kappa 指標の累積密度関数

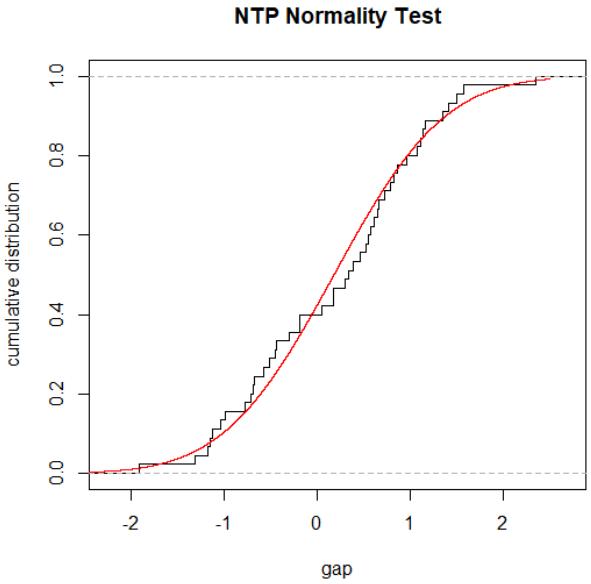

図 12：NTP リスクノードにおける大学別 Kappa 指標の累積密度分布

また、具体的な正規性検定の方法としてコルモゴロフースミルノフ検定（KS 検定）とシャピローウィルク検定（SW 検定）を大学単位での Kappa 指標の正規性の検定に用いた。コルモゴロフースミルノフ検定では“検定対象の累積密度分布関数が他の累積密度分布関数に一致すること”ここでは各リスクノードにおいて“大学別 Kappa 指標の累積密度関数が正規分布の累積密度関数に一致すること”を、シャピローウィルク検定では“大学単位での Kappa 指標のサンプルデータが正規母集団から抽出されていること”を帰無仮説として検定を行っている。コルモゴロフースミルノフ検定の統計量 D の導出式については以下の通りである。

$$D = \max_{-\infty<x<\infty} |S_n(x) - F(x)| \quad (11)$$

$$F(x) = \int_{-\infty}^{x} f(y)dy \quad (12)$$

$S_n(x)$は検定対象の累積密度分布を F(x)は確率密度関数 f(x)の累積密度分布を表しており、正規性の検定であるので f(x)は今回理想正規分布の確率密度関数を表している。有意水準を α とすると統計量 D、サンプル数 n を用いて以下の式で表される。

$$\alpha = 2\sum_{j=1}^{\infty}(-1)^{j-1}exp\{-2j^2(D\sqrt{n})^2\} \quad (13)$$

またシャピロ―ウィルク検定の統計量 W の導出式については以下の通りである。

$$W = \frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (14)$$

x は標本データを表し、$x_{(i)}$は i 番目に小さい標本である i 番目の順序統計量を、$\bar{x}$は x の平均を表している。またa_iについては標準正規分布からサンプリングされた独立同分

布の確率変数の順序統計量の期待値である m の転置行列とこの順序行列の分散共分散行列の逆行列を用いて以下の式で与えられる。

$$(a_1, \dots, a_n) = \frac{{}^t mV^{-1}}{\sqrt{{}^t mV^{-1}V^{-1}m}} \quad (15)$$

$$\mathrm{m} = {}^t(m_1, \dots, m_n) \quad (16)$$

有意水準をαとすると統計量 W とサンプル数 n を用いて以下の式で表される。

$$\alpha = \frac{\log(1-W) - \mu}{\sigma} \quad (17)$$

$$\mu = -1.2725 + 1.0521 \times (\log\frac{\log n}{n}) \quad (18)$$

$$\sigma = 1.0308 - 0.26758 \times \{\log(\log n) + \frac{2}{\log n}\} \quad (19)$$

シャピローウィルク検定では 95%の確率で正規分布に一致するとした W の値が 0.95 以上のものを正規分布とする方法もあるが、本研究では各検定において有意水準を 5%としてα > 0.05のとき帰無仮説を保留し正規分布に従うものとした。DNS リスクノード、NTP リスクノードの各検定を行った数値を表にまとめた。（表 1）

		DNS	NTP
KS	D	0.075	0.207
	p−value	0.96	0.0356
SW	W	0.978	0.914
	p−value	0.53	0.00264

表 1：各リスクノードにおける KS 検定と SW 検定の統計量と p 値

4. 考察

4.1 フリースケール性

図５、図６よりネットワークサイズ、DNS リフレクタ共に数が少ないほど相対密度が高くなり逆に数が大きくなるほど相対密度が大きく下がっており、ロングテールの形になっていることが確認できた。フリースケール性の証明においてネットワークサイズや各 UDP リフレクタの対数と取ったものをヒストグラムにすると線形性を示すため直線上になる。しかし今回のフリースケール性の証明ではきれいな直線は見受けられなかった。相対分布を用いた方法では範囲で一度潰すためサンプル数に影響されにくくロングテールの形は確認ができるが、フリースケールによる線形性の確認に関しては現状のサンプル数では見受けられなかった。サンプル数を増やすことで綺麗な直線になる可能性があるためサンプル数を増やして確認を行う必要がある。

また、国や地域単位での Kappa 指標では累積有意性モデルを仮定し実際の値と比較を行うことでフリースケール性の検討を行っている。フリースケール性の証明として現状ではフリースケール性を言いきれない部分があるためサンプル数を増やし、国別 Kappa 指標と同様に線形性の確認や累積有意性モデルを利用した比較等を行う必要があると思われる。

4.2 大学単位での Kappa 指標

図７、図８より決定係数を見ると DNS リスクノードは 0.802、NTP リスクノードでは 0.377 と大きく差が出た。DNS リスクノードは全体的に良い相関が出ていると言えるが、NTP リスクノードに関しては相関が無いと言ってもよい。原因としては NTP リスクノードを保有していない組織がネットワークサイズに寄らず半数近くを占めていること、ネットワークサイズが 3.5 から 4.0 の中規模のネットワークサイズの NTP リスクノードが広く散布しており極端な値を取っている可能性が高い部分が影響していると思われる。もしくは国や地域単位での評価の時とは違い、大学単位で評価を行った場合では NTP とネットワーク規模には相関が元々ない可能性も存在する。

回帰直線と Kappa 指標の導出から、ある程度回帰直線に沿う形でリスクノードが分散していた。国や地域単位での Kappa 指標と同様に絶対数を見ただけでは分からないリスクの高さや低さや潜在的に存在するリスクの高い組織を大学単位での評価においても見つけることができ、ネットワークサイズに依存しない比較が行えたと言える。

また、今回調査対象とした大学以外の大学の Kappa 指標を求めたい場合、評価したいネットワークのオープンネットワーク数と各リスクノード数を調査、DNS リスクノードと NTP リスクノードの回帰式にネットワークサイズを代入すると評価したい大学のネットワーク規模において平均的に期待される各リスクノード数が分かるため平均的に期待されるリスクノード数と評価大学に存在する実際のリスクノード数との乖離度から Kappa 指標を導出することが可能となっている。

4.3 大学単位での Kappa 指標の正規性

図９、図１０より、DNS リスクノードに関して確率密度分布より多少のずれはあるものの中央値が多く極端な値は少なくなっている。サンプル数が少ないため極端な値が目立ってしまうがサンプル数を増やすとより理想的な正規分布に近い分布をしていく事も予想される。逆に NTP リスクノードに関してはリスクノードを保有していない大学とリスクノードを保有している大学との差が大きく、中央値が少ない形となり中央付近が少なくなっている。また図１１、図１２より累積分布関数を見るとほとんど理想的な正規分布に沿っており、累積分布関数を見ると正規分布に沿っていると言える。正規分布の特徴である−σから+σの範囲内にサンプル総数の約 68%が収まると言う特徴で

ある 1σ 区間については DNS リスクノード、NTP リスクノード共に 60%近い大学が 1σ 区間に分布しており、ある程度は正規分布の 1σ 区間に近いと言える。

また表 1 より各検定の p 値が有意水準である 0.05 を DNS、NTP リスクノード共に上回っておりコルモゴロフースミルノフ検定の帰無仮説である“検定対象の累積密度分布関数が他の累積密度分布関数に一致すること”とシャピローウィルク検定の帰無仮説である“大学単位での Kappa 指標のサンプルが正規母集団から抽出されていること”は DNS リスクノードに関しては棄却されず保留された。帰無仮説を保留することは直接正規分布であることを証明する訳ではないが今回はシャピローウィルク検定の統計量である W が DNS リスクノードでは 0.978 と 0.95 を上回っており 95%以上で正規分布であると言える。また NTP リスクノードに関してはコルモゴロフースミルノフ検定、シャピローウィルク検定の 2 種類の検定においてどちらも帰無仮説が棄却されて正規分布でない可能性が高いという結果となった。

5. まとめ

本研究では JPCERT/CC 考案のネットワーク規模と UDP リフレクタ数から平均的に期待されるリスクを回帰分析によりネットワーク規模に寄らない比較を可能にした Kappa 指標を基に大学単位で Kappa 指標の算出を行い比較するミクロな評価を行った。大学単位での Kappa 指標の算出するための回帰分析を行うための前提条件として必要なフリースケール性の確認を行い、大学単位での Kappa 指標の算出を行った。また、大学単位での評価では国単位での評価に比べデータ数が少ないため、サンプルに用いたデータ群が適正に抽出されているかの確認のため正規性の検定を行った。

HighRedCenter のデータベースより各大学のネットワークサイズと UDP リフレクタの絶対数をプロットした結果と指数関数状に相対密度のヒストグラムを取った結果、大学別のネットワークでもフリースケール性がある程度確認した上で、Kappa 指標の導出に必要な線形回帰分析を行った。DNS、NTP の各リスクノードについて両対数での線形回帰分析の結果より平均的に期待されるリスク数である回帰直線と回帰直線からの乖離度からの偏差値を用いて Kappa 指標と決定係数の導出を行い、DNS リスクノードについてある程度の相関が見られた。また、国や地域単位での Kappa 指標の時と同様にネットワーク規模に依存しない比較を行うことで絶対数では見つけることのできないハイリスクなネットワークを見つけることができた。また大学単位での Kappa 指標の正規性の証明において DNS リスクノードについては正規性が見られたが、データの少なかった NTP リスクノードに関しては正規性を確認することが出来なかった。

本研究の改善点として国や地域単位と大学単位での Kappa 指標を比較した時、国や地域単位での UDP リフレクタ数とネットワーク規模に強い相関が出ているのに比べ大学単位での NTP リスクノードにおける回帰分析ではデータのばらつきが大きくリフレクタ数とネットワーク規模に相関が得られなかったこと、DNS リスクノードと NTP リスクノード以外のリスクノードに関して十分なデータが集まらなかったこと、NTP リスクノードに関してコルモゴロフースミルノフ検定とシャピローウィルク検定において正規分布が見られなかったことの点の 3 点が改善点として挙げられる。

また、今後の研究の方針として大学に所属している学生数のような大学特有のパラメータと Kappa 指標との相関を調べること、国立大学や私立大学のような大学の種別で Kappa 指標との相関の有無を調べ各大学のリスク分析や大学の種類単位で傾向分析をすることができないか。また、企業などの大学以外の組織単位での Kappa 指標の導出と大学単位での Kappa 指標との比較を今後の研究とする予定である。

6. 謝辞

また本研究の先行研究である“ccTLD 別 UDP リフレクタ数の指標化と分析”を手掛けられた一般社団法人 JPCERT コーディネーションセンター経営企画室兼早期警戒グループ担当部門長の洞田慎一博士、国際部サイバーメトリクスライン情報セキュリティアナリスト川崎基夫氏には本研究で利用した HighRedCenter のデータを提供していただくと共に研究を進める上で有益なご助言を頂きました、ここに両氏に対して感謝申し上げます。

参考文献

[1] インターネットリスク可視化サービスーMejiroー (https://www.jpcert.or.jp/research/mejiro.html)
[2] ccTLD 別 UDP リフレクタ数の指標化と分析
[3] Defending against DNS reflection amplification attacks
[4] ユーザ標的型 Web サイト改ざんに対する検索エンジン検知手法における提案