

階層的ネットワーク計測における計測項目の相関分析

浅葉 祥吾¹ 北口 善明² 石原 知洋³ 高嶋 健人⁴ 阿部 博⁵ 篠田 陽一⁶

概要：ネットワーク管理者がネットワーク運用している時に、管理下のネットワークにおいてユーザからネットワークに繋がらないという漠然とした報告が上がることもある。「つながらない」という状況を把握するために、ユーザサイドからのネットワーク観測が有効な場合があるが、ユーザからの報告は「つながらない」という漠然としたものになることが多くネットワーク障害を的確に報告することは難しい。そこで、ネットワーク運用者は、センサデバイスを用いユーザサイドのネットワーク状態を情報収集する事を行った。ネットワーク状態の情報を複数の階層的レイヤに整理することで、どこの階層的レイヤに問題があるのかを把握することができる。本投稿では、大学の無線ネットワーク環境で、センサデバイスを用いて各階層的レイヤを作成して計測した。ユーザサイドのネットワークをクラスタリングして時系列に解析し、ユーザサイドのネットワークとインターネットの相関を調査した。また、非階層型クラスタリングのアルゴリズムの k-means を利用して ping と DNS の RTT の分類をおこなった。

Analysis of Multilayer Network Measurement

Shogo Asaba¹ Yoshiaki Kitaguchi² Tomohiro Ishihara³ Taketo Takashima⁴ Hiroshi Abe⁵
Yoichi Shinoda⁶

1. はじめに

本稿では、この階層的ネットワーク計測手法の計測データを用い計測項目の相関分析を実施した。特にユーザサイドで広く利用されている無線 LAN 環境を対象とし、無線 LAN 環境の計測項目間の相関や無線 LAN 環境がネットワーク通信に与える影響など複数の観点から相関分析を行った。相関分析手法にはクラスタリングなどの可視化手法を用い、階層的ネットワーク計測手法の有効性や課題点を考察した。

ネットワーク管理者がネットワークを運用している時に管理下のネットワークで、ユーザからネットワークにつながらないという障害報告を受けることがある。「つながらない」という障害報告には、ユーザ端末のアドレス問題や無線ネットワーク問題、DNS における名前解決の問題など様々な原因がある。このような「つながらない」という

状況を把握するために、ユーザサイドからのネットワーク観測が有効であるが、ユーザは「つながらない」という状況を正確に捉えることが困難であるので、ネットワーク運用者に的確にネットワーク状態を的確に報告することは難しい。

そこで、本研究のもととなる SINDAN プロジェクトでは、ユーザ環境においてネットワーク障害点を検出する手法を、それぞれのレイヤ毎に整理と確立を行い、迅速にネットワーク状態を把握できるシステムの研究をおこなっている。また、特に今後の利用がますます考えられる IPv4/IPv6 デュアルスタック環境を評価対象とし、複雑なネットワークにおいても十分利用可能なものを目標としている [7]。

本稿では、この階層的ネットワーク計測手法の計測データを用い計測項目の相関分析を実施した。特にユーザサイドで広く利用されている無線 LAN 環境を対象とし、無線 LAN 環境の計測項目間の相関や無線 LAN 環境がネットワーク通信に与える影響など複数の観点から相関分析を行った。相関分析手法にはクラスタリングなどの可視化手法を用い、階層的ネットワーク計測手法の有効性や課題点を考察した。

¹ 北陸先端科学技術大学院大学

² 東京工業大学 学術国際情報センター

³ 東京大学大学院 総合文化研究科

⁴ フリーランス

⁵ 北陸先端科学技術大学院大学, 株式会社レビダム

⁶ 北陸先端科学技術大学院大学 情報社会基盤研究センター

計測層	計測概要
データリンク層 datalink	隣接機器との接続性
インターフェース設定層 interface	IP アドレス設定
ローカルネットワーク層 localnet	同一セグメントにおける IP の到着性
グローバルネットワーク層 globalnet	組織外の外部サーバへの IP 的な到達性
名前解決層 dns	ドメイン名から IP アドレスの取得
ウェブアプリケーション層 web	ウェブアプリケーションに特化

表 1 各階層と計測内容

データリンク層の計測項目	計測内容
ssid	無線ネットワークの SSID
bssid	無線ネットワークで接続している AP の BSSID のアドレス
チャンネル channel	利用している無線ネットワークのチャンネル
受信信号強度 RSSI	利用している無線ネットワークの電波信号強度
ノイズ信号強度 noise	受信しているノイズの信号強度
データレート datarate	利用している無線ネットワークのデータレート

表 2 データリンクにおける計測項目

2. 既存研究

ネットワーク状態やサービス状態を監視するアプリケーションとして、Nagios[1] や zabbix[2] などがある。それらのモニタリングアプリケーションは、サーバやネットワークのリソース状態の管理や TCP/IP におけるレイヤ 2 やレイヤ 3 の到達性の確認が主となっているため、ユーザサイドで発生しているネットワーク障害を特定することが難しい。また、構成しているネットワークやシステムが複雑になるにつれて管理が困難になり、ユーザサイドで発生しているネットワーク障害の発生検出が難しい。

また、ネットワークの状態計測手法としては、Apache HTTP server benchmarking tool[3] や Chaos Monkey[4] などがあるが、これらはネットワークや Web サーバの負荷に対する性能計測や障害復旧システムが正常に動作しているのか確かめるものであり、また、RIPE NCC の RIPE Atlas[5] を利用し、ISP(Internet Service Provider) 間のネットワーク状態の計測 [6] あるが、いずれの手法においてもユーザサイドのネットワーク計測は行っていない。

3. 階層的ネットワーク状態計測の各項目

計測レイヤ設計としては、データリンク層 (datalink)、インターフェース設定層 (interface)、ローカルネットワーク層 (localnet)、グローバルネットワーク層 (globalnet)、名前解決層 (dns)、ウェブアプリケーション層 (web) がある [9]。表 1 に、各計測レイヤで確認している概要を示す。

今回の実験では、ユーザサイドのネットワーク状態とグローバルなインターネット状態が目的であるため、データリンク層とローカルネットワーク層、グローバルネットワーク層、名前解決層の計測項目を利用した。

3.1 階層的ネットワーク状態計測のデータリンク層

表 2 に、階層的ネットワーク状態計測のデータリンク層の計測項目を示す。

階層的ネットワーク状態計測のデータリンク層は、TCP/IP の階層におけるネットワークインターフェイス層と同じ階層構造をしており、隣接機器との接続性を確認するための計測層である。ネットワークインターフェイスの down 状態から up を行いリンクアップできるかまで

を確認している。以下は、無線ネットワーク環境の計測項目である。

ssid は、Service Set Identifier のことで利用している無線ネットワークの SSID 情報である。bssid は、Basic Service Set Identifier のことで利用している無線ネットワークで接続している AP(Wireless LAN Access Point) の BSSID のアドレスである。無線環境を特定するために用いる。

channel は、利用している無線ネットワークのチャンネル情報である。2.4GHz 帯は 1 から 14CH であり、5GHz 帯は 36 から 54CH と 100 から 140CH である。

rsssi は、利用している無線ネットワークの電波信号強度 RSSI(Received Signal Strength Indication) 情報である。単位は dBm であり、負の値である。センサデバイスがアクセスポイントからどの程度の信号を受信しているのかを計ったものである。受信信号強度は、チップメーカーがそれぞれの 0 から -255 の範囲の値と定義している。

noise は、受信しているノイズの信号強度情報である。単位は dBm であり、負の値である。無線環境におけるバックグラウンドノイズの量である。ノイズレベルが高い場合、無線信号強度が低下する。信号強度とノイズレベルで、SN 比 (信号対雑音比) を計算することができる。

データレートは、利用している無線ネットワークのデータレート情報である。単位は Mbps である。

3.2 階層的ネットワーク状態計測の

ローカルネットワーク層とグローバルネットワーク層階層的ネットワーク状態計測のローカルネットワーク層とグローバルネットワーク層は、TCP/IP 階層モデルにおけるインターネット層の組織内、組織外のサーバへの IP 的な到達性を確認する層である。Ping コマンドにより RTT(Round Trip Time) とパケットロス率、traceroute コマンドによるパス計測の到着性の確認とパス MTU を計測する。

外部サーバの Google のパブリック DNS[10] と WIDE プロジェクトサーバへの Ping の IPv4 往復遅延の平均値、最小値、最大値を Ping を使い計測している。単位は msec である。Ping により、ネットワークの帯域の混雑さを計測している。

3.3 階層的ネットワーク状態計測の名前解決層

階層的ネットワーク状態計測の名前解決層は、TCP/IP 階層モデルにおけるアプリケーション層の DNS(Domain Name System) による名前解決の確認を行う計測層である。アプリケーションを利用する際に必須となる機能として、ドメイン名から IP アドレスを取得する名前解決があり、この層が、名前解決の計測層である。OS の resolver API 毎に挙動が異なることが想定されるので、DHCP/DHCPv6 等で得られたネームサーバと Google のパブリック DNS サーバに対して名前解決の問い合わせをする。

IPv4 ネームサーバに対して、v4only.sindan-net.com の FQDN に対して、A レコードの応答遅延時間を計測している。IPv4 ネームサーバは、ローカルにあるサーバと google のパブリック DNS サーバである。単位は msec である。

4. 実験結果

ユーザサイドのネットワークで広く利用されている無線 LAN 環境の計測項目を時系列に解析し、重要な要素である計測項目毎にクラスタリングをおこなった。また、ユーザサイドのネットワークとインターネットの相関を調査した。

東京工業大学の無線 LAN 環境である eduroam JP[8] のネットワーク環境を対象とした。eduroam JP とは、大学等教育期間の間でのキャンパス無線 LAN の相互利用を実現している国立情報学研究所 (NII) のサービスである。国際無線 LAN ローミング基盤 eduroam は、業界標準の IEEE802.1X に基づいており、安全で利便性の高い無線 LAN 環境を提供している。東京工業大学の無線 LAN 環境である eduroam は、IEEE802.11n の規格を採用しており、Cisco Systems, Inc の AP を利用している。

eduroam に、Raspberry Pi[11] に USB の無線 LAN アダプタをつけて、2.4GHz 帯と 5GHz 帯を計測できるようにしたセンサデバイスを利用した。図 1 に、階層的ネットワーク状態計測の実験環境を示す。

東京工業大学の無線 LAN 環境である eduroam は、IEEE802.11n の規格を採用しており、Cisco Systems, Inc の AP を利用している。2018 年 3 月における 1 ヶ月間の階層的ネットワーク状態計測のデータリンク層とグローバルネットワーク層、名前解決層の一部の計測項目を利用した。データは、IPv4 のネットワーク環境で 5 分に一回計測したものである。

4.1 データリンク層における相関分析

階層的ネットワーク状態計測のデータリンク層で計測している ssid と bssid, channel, rssi, noise, rate の値を使い調査した。各項目の外れ値は、channel は 0CH 以上、rssi は -200 から 0, noise は -200 から 0, RATE は 10 以上とした。

階層的ネットワーク状態計測

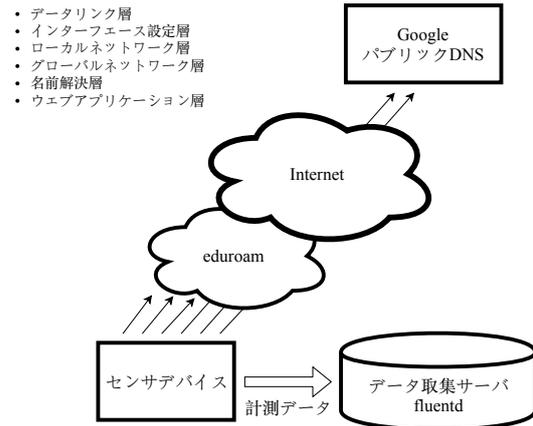


図 1 階層的ネットワーク状態計測の実験環境

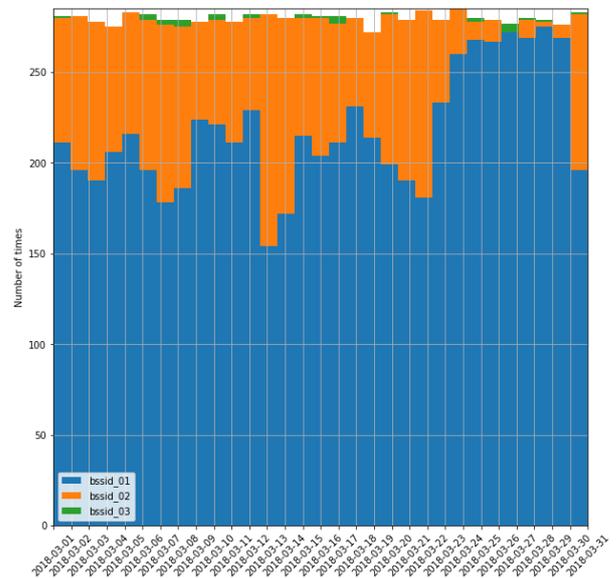


図 2 各計測で接続した BSSID の日別利用頻度

無線 LAN 環境は、AP の場所や距離、電波強度によって変化するので、時間帯によってどこの AP に繋がっているのか調べるために、BSSID を利用した。期間でどの BSSID を利用しているのかを調べた。図 2 に、時間帯の BSSID を利用している積み上げヒストグラムを示す。期間中に観測した BSSID は 3 つであり、それぞれを BSSID01, BSSID02, BSSID03 とした。

BSSID は、BSSID01 が多く次に BSSID02 が利用されており、BSSID03 が少ない。1 日の中でも BSSID の変化していることがある。BSSID が各計測データに影響を与えているのかを調べるため、図 3 に、期間での BSSID で色分けした channel と rssi, noise, rate の時系列グラフを示す。

次に、各計測項目ごとの相関を調べるために、BSSID ごとに色分けした CHANNEL と RSSI, NOISE, RATE の相関図を作成した。図 4 に、CHANNEL ごとに色分けした RSSI と NOISE, RATE の相関を示す。

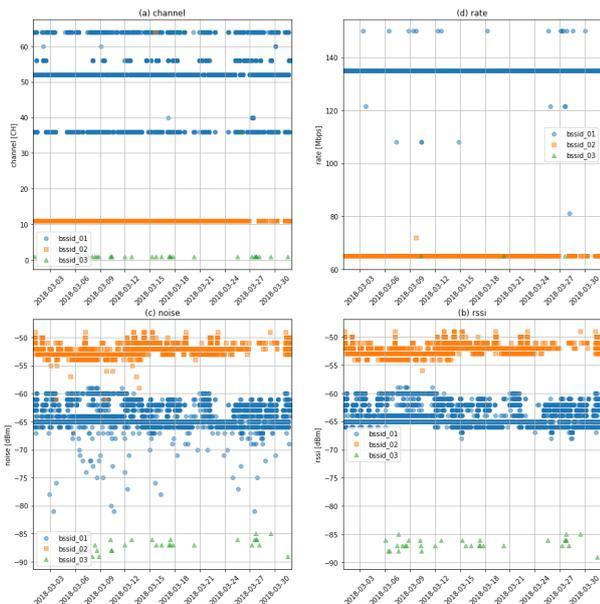


図 3 BSSID 毎の CHANNEL, RSSI, NOISE, RATE の時系列変化

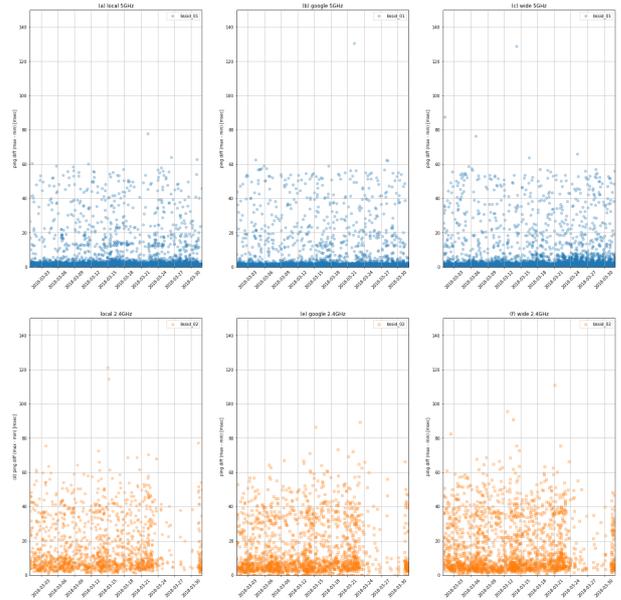


図 5 利用無線 LAN 周波数帯毎の通信遅延変動幅の時系列変化

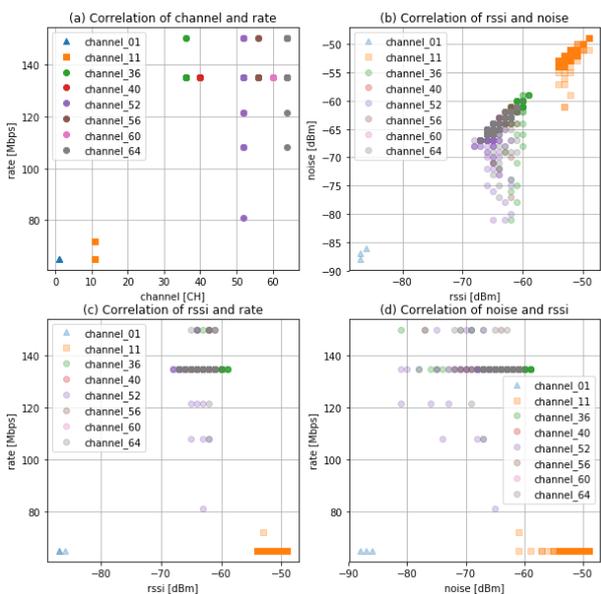


図 4 CHANNEL 毎の無線 LAN パラメータ間の相関図

図 2 の (a) より, BSSID01 は 5GHz 帯, BSSID02 と BSSID03 は 2.4GHz 帯を多く利用している。

BSSID01 と BSSID02 は BSSID アドレスが近く, 5GHz 帯と 2.4GHz 帯で別れているので, 同じ AP であると考えられる。BSSID02 が 65CH や BSSID03 が 36CH などを使って計測をしようとしているが, 計測した値は外れ値である。同じ BSSID では, 2.4GHz の channel が変化していないが, 5GHz 帯の channel が変化している。また, rssi と noise は, 5GHz 帯が共に中間にあり, 2.4GHz 帯が共に高いことがわかるが, 5GHz 帯を多く選択している。

期間中, 無線 LAN 環境の各計測の値は, 変動率が高いので, センサデバイスを固定して計測をしていたが, AP

の接続数や電波の干渉などが起きていると考えられる。

今回の実験では, 一台のセンサデバイスで一つ SSID だったがセンサデバイスを様々な SSID を複数台で計測すると, 無線 LAN 環境の詳細なネットワーク状態が把握できる可能性がある。

4.2 データリンク層とネットワーク層の相関分析

ユーザ側のネットワークとグローバルネットワークの比較を行うために, センサデバイスから Google パブリック DNS と WIDE プロジェクトのサーバに Ping コマンドの 10 回分の RTT の最大値と最小値の差分を利用した。この 2 つのターゲットに Ping で RTT の差分を計測することで, 2 つとも差分が僅差であるならグローバルネットワークがどの程度安定していることがわかって考えられる。図 5 に, BSSID アドレスごとに Ping コマンドの RTT の差分を示す。

次に, 図 6 に, BSSID アドレスごとに Ping コマンドの RTT の差分のヒストグラムを示す。差分を 0 から 60msec までの間のヒストグラムに表した。

BSSID ごとに比較すると Google パブリック DNS と WIDE プロジェクトの Ping の RTT の差分で見ると, 5GHz 帯と 2.4GHz 帯は, どちらも近い形をしているので, グローバルネットワークは, 安定していると考えられる。5GHz 帯は 0 に近い所にピークがあるのに対して, 2.4GHz 帯は 0 から少し離れたところにピークがあることがわかる。このピークの差により, 5GHz 帯は, 2.4GHz 帯より無線 LAN 環境が安定していると考えられる。

4.3 ネットワーク層と名前解決層の相関分析

Google のパブリック DNS サーバがターゲットとなって

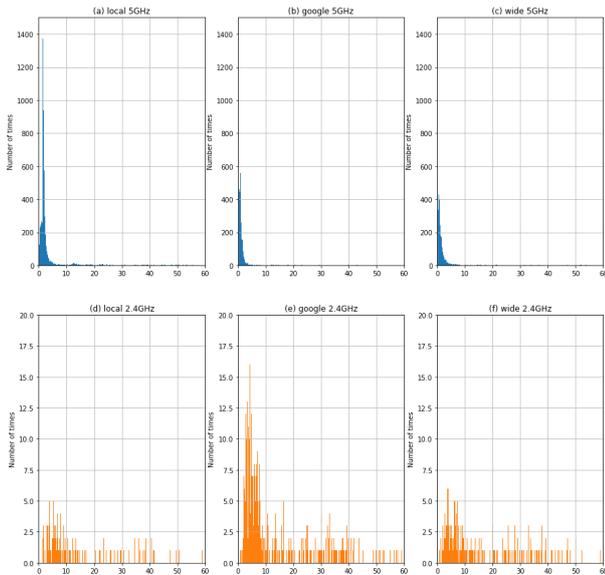


図 6 BSSID アドレスごとに Ping コマンドの RTT の差のヒストグラム

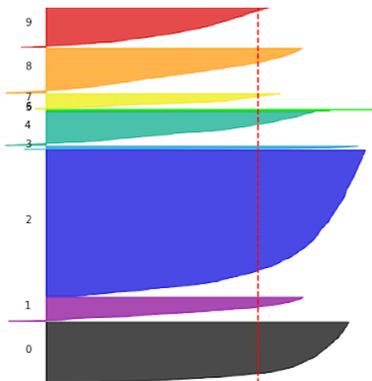


図 7 ping rtt と dns rtt の相関分析のためのシルエット分析図

いるグローバルネットワーク層の往復遅延の平均値と名前解決層の応答遅延の相関関係を調べた。それぞれの外れ値を無くすために四分範囲を決めて、同じ時刻帯で計測した各データを対にした。往復遅延の平均値を Ping RTT, 名前解決層の応答遅延を dns rtt とした。それぞれの遂にしたデータから、ネットワーク状態のどのような分類ができるのか調査するために k-means(k-means clustering) を利用した。k-means は、非階層型クラスタリングアルゴリズムであり、クラスタの平均を用いてに分類する方法であり、scikit-learn[12] で python のライブラリが提供されている。今回は、k-means の初期値の選択を改良した k-means++ を利用した。k-means で正確なクラスタリングを行うためにそれぞれの値を標準化して、K-means の妥当な分類数を調べるためにシルエット分析をした。図 7 に、k-means のクラスタ数が 10 の時のシルエット分析の図を示す。

シルエット分析 (Silhouette analysis) は、クラスタ内のサンプルがどの程度の密度のクラスに属しているのかを調べるものである。また、一つのクラスが隣接するクラ

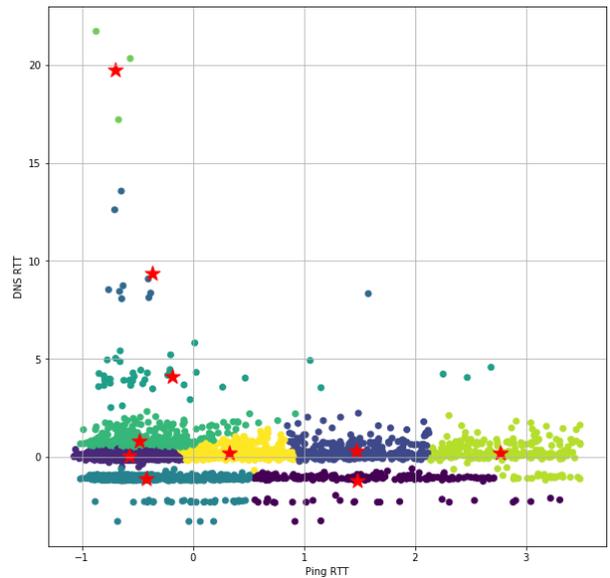


図 8 k-means 法でクラスタリングした ping rtt と dns rtt の相関図

スタとどれくらい離れているのかを調べられる。クラスタが正確にシルエット図のグラフが均一な面積になる。図 7 は、クラスタ数が 10 の時であり 0 から 9 までの面積が均等に近いので、シルエット分析によりクラスタ数は 10 が妥当であると考えられる。次に、クラスタと時系列の関係を調べるために、クラスタをそれぞれのグローバルネットワーク層の往復遅延の平均値と名前解決層の応答遅延に反映させた。図 8 に k-means でクラスタリングした Ping の RTT と DNS の RTT の相関図を示す。各軸は正規化した値である。

各クラスタは色を固定してグラフに反映している。クラスタリングを、それぞれの時系列に反映させたところ、各グラフから 4 クラスは、分類がわかりやすく明示されているが、残りの 6 クラスが様々なところに散っている。

図 7 と図 8 より、グローバルネットワーク層の往復遅延の平均値と名前解決層の応答遅延で、有益な相関関係がないように考えられる。また、ネットワーク状態を k-means でおこなったクラスタリングを時系列にクラスタリングを反映させてみたが、クラスタリングと時系列の相関関係がないことが判明した。

5. おわりに

本投稿では、ネットワーク状態の評価を、階層的ネットワーク計測の計測データの相関に焦点をあて、ネットワーク状態を的確に評価できるかを検討した。今回の相関分析により、無線 LAN 環境を利用したネットワーク通信では、利用する周波数帯による影響が大きい傾向が読み取れた。したがって、無線 LAN ネットワーク環境が、インターネット全体の通信に与える影響が大きいと言え、無線 LAN 環境の評価は重要と考えられる。また、今回の解析

では、BSSID ごとに比較をおこない、同一 BSSID の基地局でも使用周波数がばらつくことと、定点で計測をしていた場合でも BSSID が変わる、すなわち異なる基地局にアソシエーションすることが観測された。これは、AP とセンサデバイスまでの間で何らかの遮蔽が発生したことによる信号強度が変わったことに起因して発生した、ないしはより負荷が低い AP へ振り分けるように AP 側から指示された、などの要因が考えられるが、今回おこなった計測ではそれらの原因について明確にすることはできなかった。そのため、今後の計測において無線のコントロールフレームのモニタなどを行うことで、より詳しく調査する必要がある。また、今回と同様の実験を様々なネットワーク環境で実験を行うことで、ネットワーク計測の値の正確性を調査することができると思う。

参考文献

- [1] Nagios - The Industry Standard In IT Infrastructure Monitoring, <https://www.nagios.org/> (2018 年 5 月 10 日参照)
- [2] Zabbix オフィシャル日本語サイト :: エンタープライズクラスの分散監視オープンソースソリューション, <https://www.zabbix.com/jp/> (2018 年 5 月 10 日参照)
- [3] ab - Apache HTTP server benchmarking tool, <http://httpd.apache.org/docs/current/programs/ab.html> (2018 年 5 月 10 日参照)
- [4] Chaos Monkey, <https://netflix.github.io/chaosmonkey/> (2018 年 5 月 10 日参照)
- [5] RIPE Atlas, RIPE, Atlas, <https://atlas.ripe.net/> (2018 年 5 月 10 日参照)
- [6] Romain Fontugne, Emile Aben, Cristel Pelsser, Randy Bush, Pinpointing Delay and Forwarding Anomalies Using Large-Scale Traceroute Measurements. <https://conferences.sigcomm.org/imc/2017/papers/imc17-final106.pdf> (2018 年 5 月 10 日参照)
- [7] SINDAN Project Simple Integrated Network Diagnosis And Notification, <https://www.sindan-net.com> (2018 年 5 月 10 日参照)
- [8] eduroam JP, <https://www.eduroam.jp> (2018 年 5 月 10 日参照)
- [9] 北口 善明, 石原 知洋, 高嶋 健人: センサデバイスを利用したネットワーク状態計測手法の評価, 情報処理学会 マルチメディア・分散・協調とモバイル (DICOMO) シンポジウム 2017 論文集, pp.1348-1353, June 2017.
- [10] Google Public DNS, <https://developers.google.com/speed/public-dns/> (2018 年 5 月 10 日参照)
- [11] The Raspberry Pi is a tiny and affordable computer that you can use to learn programming through fun, practical projects. Join the global Raspberry Pi Community. <https://www.raspberrypi.org/> (2018 年 5 月 10 日参照)
- [12] scikit-learn: machine learning in Python, <http://scikit-learn.org/stable/index.html> (2018 年 5 月 10 日参照)