大学におけるネットワーク利用の特徴抽出に関する研究

橋口 育弥[1]　齊藤 匠一郎[2]　岡村 耕二[3]

概要：ネットワークが利用目的ごとに分類可能であれば、ネットワークの分散によるトラフィックの緩和や新たなネットワークの利用目的の発見が可能であり、よりよいネットワーク環境を作り出すことが期待できる。分類するには、利用目的が明確であるネットワークの特徴が必要であるが、利用者層が広く、利用目的の全く異なる利用者が混在する大きなネットワークの通信は複雑であり特徴を見いだすことが困難である。そこで、大学では、学生、教員、職員と利用目的が明確であり、ネットワーク利用を分類することが可能であると考えられるため、特徴を抽出し教師データとして用いることで利用目的の異なる利用者が混在するネットワークを利用目的ごとに分類することが期待できる。本研究の目的は九州大学内の利用目的が異なり利用者が限定される 2 つのネットワークを対象としてデータフローから通信の時間的な特徴と空間的な特徴を抽出し定量的に示すことにある。ある一日におけるネットワークを利用する時間帯内の 10 分間隔におけるバイト数、TCP 通信の割合、バイト数の増加量の 3 つを時間的な特徴の特徴量として、各時間帯におけるパケット数およびバイト数と各時間帯の学内通信の割合を空間的な特徴の特徴量としてそれぞれグラフ化を行い、データセットから閾値を設定することで定量的に特徴を示した。その特徴を用いて、データセットにはない別の日のデータに対し判定を試み、特徴の評価を行った。

Study on feature extraction of network users in university

Ikuya HASHIGUCHI[1]　Shoichiro SAITO[2]　Koji OKAMURA[3]

1. はじめに

ネットワークは利用者の目的や用途により様々用意されている。しかしながら、利用者の層が広く、利用目的が複数存在する大きなネットワークも存在し、このようなネットワークの通信は複雑であり特徴を見いだすことが困難である。大学においても利用目的の異なる複数の立場の利用者が共有するネットワークが存在する。しかし、大学においては学内ネットワークに自由に接続でき、何時でも自由に自分のペースで学習できる環境を提供する「教育用ネットワーク」や事務作業や学外との通信を行う際に大学職員が使用する「事務用ネットワーク」といったようにネットワークを利用する立場や目的が明確であるためそれぞれのネットワーク利用を分類することが可能である。そこで大学内に目的や用途ごとに用意されているネットワークを「教師データ」とすることで大学内の利用目的の異なる複数の立場の利用者が共有するネットワークを利用目的ごとに分類することが可能であると考えた。つまり、利用者や利用目的が明確であるネットワークの場合、ネットワークの利用範囲が限られているためデータフローを解析することでデータフローに含まれる特徴量からネットワークの特徴を抽出することが可能であり、抽出した特徴を「教師データ」とすることで通信の傾向を定量的に示し、利用目的が複数存在し、通信が複雑であるネットワークの特徴を利用目的ごとに捉えることを可能にするということである。これにより大学において学生や教職員、申請を行えば一時的に学内のネットワークを利用できる学外者が共有する利用目的が複雑なネットワークにおいて、利用目的ごとにネットワークを分散しトラフィックを緩和することや「教師データ」により分類されなかったデータから新たなネットワークの利用目的を発見することができると考えた。

本研究の目的は九州大学内の利用目的が異なっており利用者が限定される 2 つのネットワーク (以下それぞれ「ネットワーク A」「ネットワーク B」とする。) を対象として主

[1] 九州大学大学院システム情報科学府
[2] 九州大学大学院統合新領域学府
[3] 九州大学情報基盤研究開発センター

に活動する時間帯のデータフローから通信の特徴を抽出し定量的に示すことにある。本研究では時間的な特徴と空間的な特徴の 2 つに分類し調査を行い、それぞれのデータフローより得られた特徴に対して閾値を設けることで数値化を図り適当なデータを用いて評価を行う。最終的に抽出した特徴によるネットワーク判定の妥当性や抽出に対する考察を行う。

2. ネットワーク利用者の空間的特徴に関する評価

2.1 調査を行った LAN の解析

データフローの解析は Elasticsearch および Kibana を用いた。Elasticsearch とは Elastic 社の開発した全文検索エンジンでありデータの蓄積、分析、ログ解析など様々な分析を可能としている。Kibana は Elasticsearch と連携して使用するデータ解析、可視化プラットフォームであり Elasticsearch で行う分析を視覚的に捉えることを可能にしている。本研究では、大学内に存在し利用目的の異なるネットワーク A とネットワーク B を対象に、活発に動く平日一日の 8:00 から 19:00 を 1 時間ごとに区分して調査を行った。調査には 2019 年 11 月 26 日のデータを扱った。解析からパケット数とバイト数の変化に注目し、横軸を 8:00 から 19:00 を 1 時間ごとに区分した時間帯、縦軸を各通信量として折れ線グラフをとった。(図 1)(図 2) さらにネットワーク A およびネットワーク B が学内と通信を行っている通信量を求め、図 1 や図 2 と同様に折れ線グラフをとった。(図 3)(図 4)

図 1、図 2 から各時間帯の通信量を見るとネットワーク A の方が常に多いことが分かった。次に図 3、図 4 から各時間帯の学内通信量を見ると、パケット数に特徴は確認できないがバイト数に関してはネットワーク A と比較してネットワーク B が多くの時間帯で上回っていることが確認できた。各ネットワークを通信相手先のアドレスによって

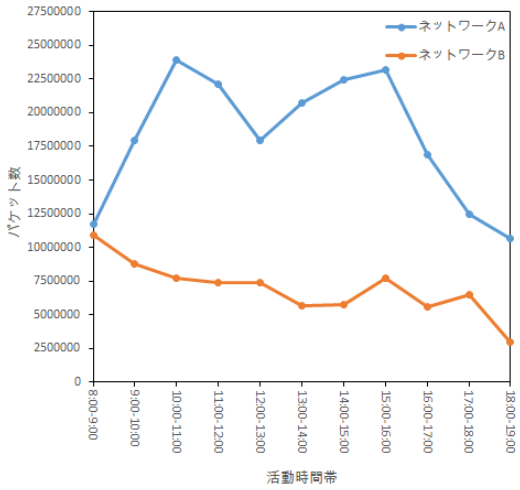

図 1 2019 年 11 月 26 日の活動時間における総パケット数の推移

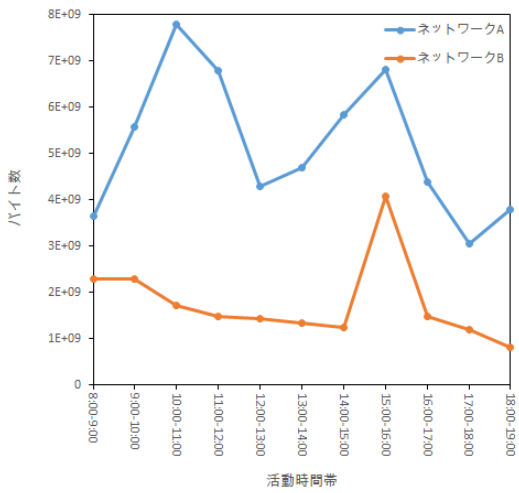

図 2 2019 年 11 月 26 日の活動時間における総バイト数の推移

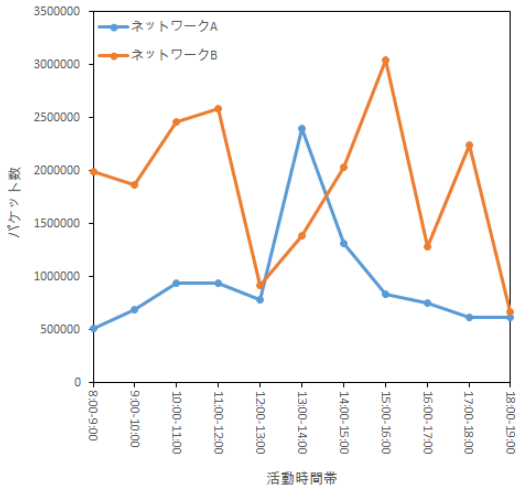

図 3 2019 年 11 月 26 日の活動時間における学内通信のパケット数の推移

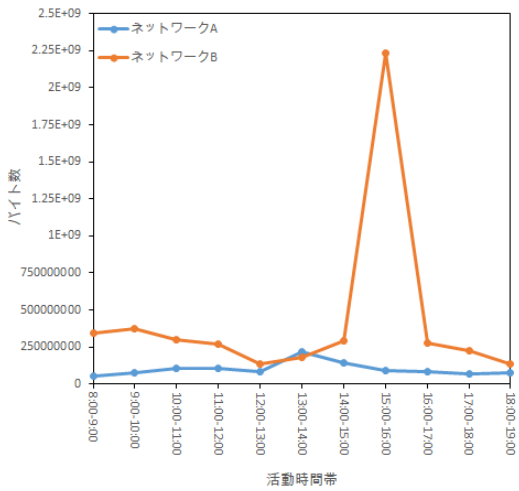

図 4 2019 年 11 月 26 日の活動時間における学内通信のバイト数の推移

表 1　2019 年 11 月 26 日の活動時間帯における各ネットワークの学内通信量の割合

時間帯	ネットワーク A		ネットワーク B	
	パケット数 [%]	バイト数 [%]	パケット数 [%]	バイト数 [%]
8:00-9:00	4.35	1.55	18.22	15.10
9:00-10:00	3.83	1.39	21.23	16.19
10:00-11:00	3.94	1.39	32.00	17.41
11:00-12:00	4.24	1.61	35.03	18.37
12:00-13:00	4.36	2.01	12.41	9.19
13:00-14:00	11.59	4.67	24.54	13.45
14:00-15:00	5.86	2.46	35.21	23.49
15:00-16:00	3.62	1.34	39.69	54.85
16:00-17:00	4.48	1.96	22.85	18.69
17:00-18:00	4.96	2.36	34.67	19.02
18:00-19:00	5.83	1.92	22.47	16.60

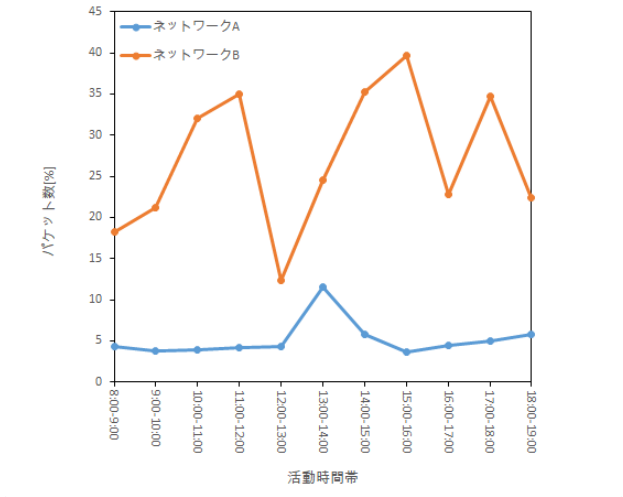

図 5　2019 年 11 月 26 日の活動時間における学内通信のパケット数の割合の推移

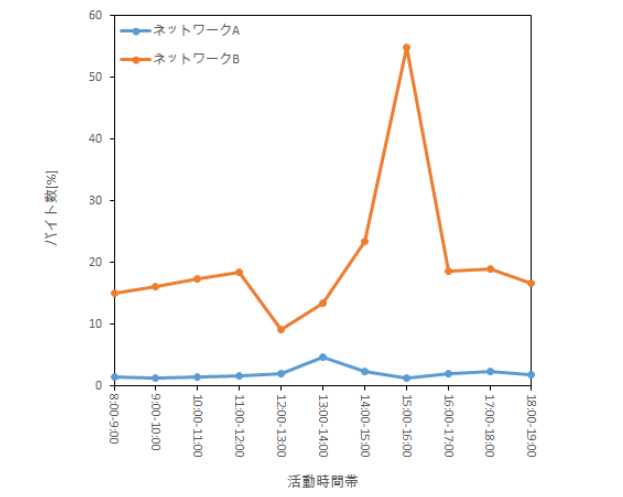

図 6　2019 年 11 月 26 日の活動時間における学内通信のバイト数の割合の推移

識別可能にするために、学内との通信に関してさらに分析を行った。各時間帯に学内との通信が行われている割合を算出し、横軸を 8:00 から 19:00 を 1 時間ごとに区分した時間帯、縦軸を学内との通信の割合として折れ線グラフをとった。(表 1)(図 5)(図 6)

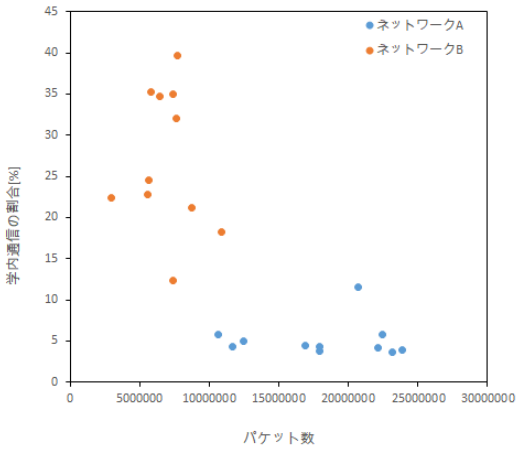

図 7　2019 年 11 月 26 日の各時間帯のパケット数と学内通信の割合の関係

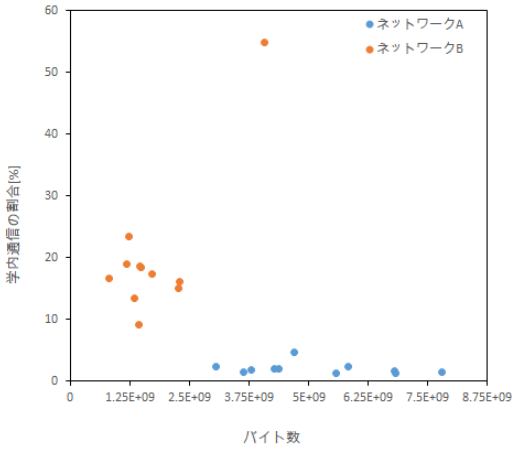

図 8　2019 年 11 月 26 日の各時間帯のパケット数と学内通信の割合の関係

図 5、図 6 からネットワーク B の方がネットワーク A よりも常に高い割合で学内通信を行っていることが確認できた。このことから各ネットワークの 1 時間ごとの総通信量のみでは使用するサービスの特徴を捉えてネットワークの識別を行うのが困難であったところを、ネットワークが学内と通信を行っている割合を参考にすることで通信相手先の特徴を掴み正確なネットワークの識別が可能になると考えた。したがって本研究において、各時間帯における総通信量と各時間帯における学内通信の割合を特徴量として識別を行うことにした。2 つの特徴量の関係を可視化するために、横軸を各時間帯の通信量、縦軸を各時間帯の通信量に対する学内通信の割合として各時間帯のデータを座標として捉えた。(図 7)(図 8)

2.2　閾値の決定

図 7、図 8 よりネットワーク A とネットワーク B がと各時間帯の総通信量と各時間帯の学内と通信を行う割合により区別が可能であると確認した。しかし調査した日と同様に他の日においても同じような結果が得られるとは限ら

ないため各ネットワークの傾向を数値で捉えて正確なネットワークの識別を行う必要があると考えた。本研究では閾値を用いることで数値的にネットワークの識別を正確に行うことが可能であると考え、閾値を設けることにした。まず、各ネットワークの傾向を掴むために 2020 年 1 月 6 日から 2020 年 1 月 10 日までの平日 5 日間のデータセットを解析し、図 7、図 8 と同様に座標として捉え横軸の通信量をネットワーク A のパケット数、ネットワーク A のバイト数、ネットワーク B のパケット数、ネットワーク B のバイト数と分類して 4 つの散布図を用意した。(図 9)(図 10)(図 11)(図 12) さらに散布図から最小二乗法により回帰直線を導出した。

最小二乗法とは、測定により得られた数値の組に対して想定される関数 $f(x)$ で近似するとき以下の式が最小となるように $f(x)$ を求めることである。

$$\sum_{i=1}^{n}\{y_i - f(x)\}^2 \tag{1}$$

$f(x)$ が 1 次関数のとき回帰直線となり、回帰直線を $y = ax + b$ とすると以下の条件を満たす。ここで、n は 2 変数 (x, y) の総数、x_i, y_i は各変数個々の数値、$\bar{x}, \bar{y}$ は各変数の平均値を表す。

$$a = \frac{\sum_{n=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{n=1}^{n}(x_i - \bar{x})^2} \tag{2}$$

$$b = \bar{y} - a\bar{x} \tag{3}$$

導出した回帰直線より切片が 30%大きい直線と 30%小さい直線を用意し、2 つの直線の間に出来る領域を閾値として捉えて領域内にデータが含まれているとき各ネットワークのデータであると識別することにした。直線の範囲は横軸においてはデータセットで通信量が最小である値から最大である値まで、縦軸においては学内通信の割合が最小である値から最大である値までにした。薄い青の直線が回帰直線であり、赤の直線が回帰直線より切片が 30%大きい直線、濃い赤の直線が回帰直線より切片が 30%小さい直線である。

2.3 閾値による評価

2.2 節で決定した閾値により 2019 年 11 月 26 日のデータの識別正確性を調べた。(表 2) ここで T はデータがネットワークに含まれていることを、F はデータがネットワークに含まれていないことを示している。各時間帯でパケット数、バイト数がともに T であればネットワークに含まれる正しいデータ、ともに F であればネットワークに含まれない誤ったデータ、一方が T で他方が F であればネットワーク識別不可のデータであると捉えてそれぞれの割合を算出した。例えば、一日の全ての時間帯でパケット数、バイト数がともに T であるならばネットワークに 100%含まれていると考えることが可能である。データセットの 2020 年 1 月 6 日から 9 日までを 1 つのデータフローとして見たとき、ネットワーク A と正確に識別されたデータは 80.00%、誤りであると識別されたデータは 7.27%と高い割合で判定が行われており誤りであると判定した割合も低いことから

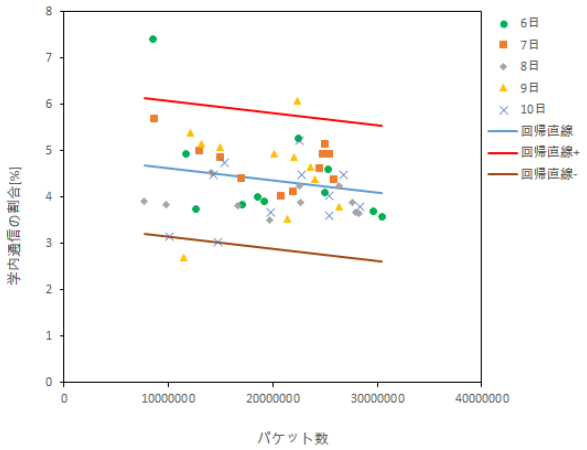

図 9 ネットワーク A における各時間帯のパケット数と学内通信の割合の関係

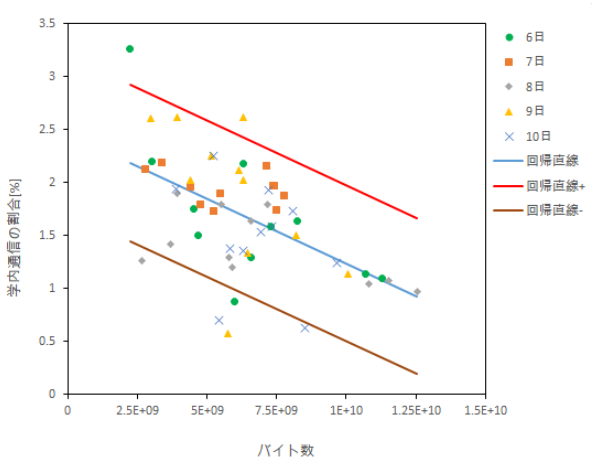

図 10 ネットワーク A における各時間帯のバイト数と学内通信の割合の関係

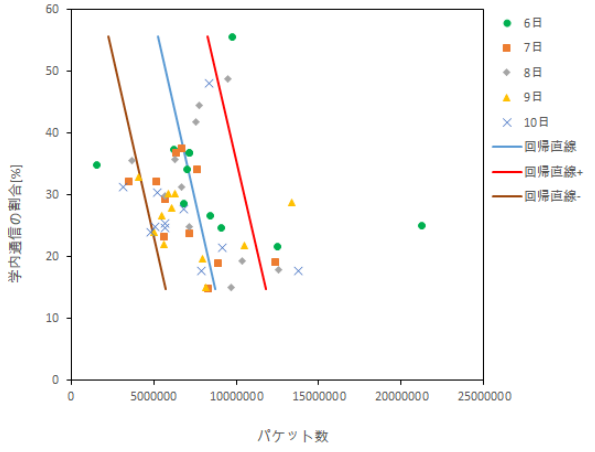

図 11 ネットワーク B における各時間帯のパケット数と学内通信の割合の関係

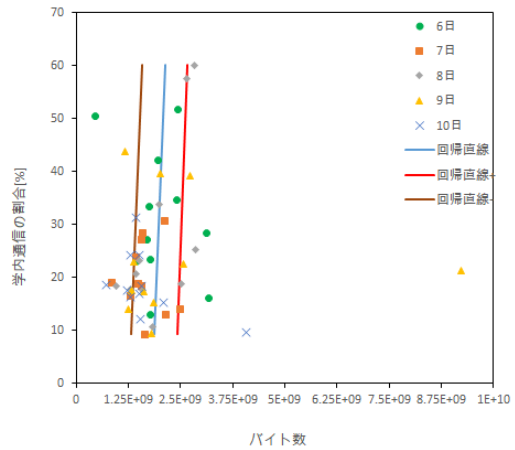

図 12 ネットワーク B における各時間帯のバイト数と学内通信の割合の関係

表 2 2019 年 11 月 26 日におけるネットワークの判定

時間帯	ネットワークA		ネットワークB	
	パケット数	バイト数	パケット数	バイト数
8:00-9:00	T	T	T	T
9:00-10:00	T	T	T	T
10:00-11:00	T	T	T	T
11:00-12:00	T	T	T	T
12:00-13:00	T	T	F	F
13:00-14:00	F	F	T	T
14:00-15:00	F	T	T	F
15:00-16:00	T	T	T	F
16:00-17:00	T	T	T	T
17:00-18:00	T	T	T	F
18:00-19:00	T	T	F	F

妥当な閾値判定が可能であると考えられた。一方で、ネットワーク B と正確に判断されたのは 56.36%、誤りであると識別されたデータは 20.00%と、扱った半分程度のデータは正しく判定可能であるが誤りであると判定する割合も目立つことから妥当な閾値判定を行うことが困難であると考えられる。2019 年 11 月 26 日のデータに閾値を適用することによりネットワーク A と識別されたのは 81.82%、ネットワーク B と識別されたのは 54.55%であり、ネットワーク A の判定は妥当であるがネットワーク B の判定は妥当ではないと考えられた。また閾値により識別不可とされたデータはデータセットのネットワーク A において 10.91%、ネットワーク B において 23.64%、2019 年 11 月 26 日のデータのネットワーク A において 9.09%、ネットワーク B において 27.27%であった。ネットワーク A は誤った識別も少なく正確な判定を行えていると言えたが、ネットワーク B においても正確な判定を行うためには誤ったデータであると識別する割合を小さくするだけで無く、識別不可とされたデータの割合も小さくする必要があると考えた。そこで各時間帯の総通信量と各時間帯の学内通信の割合の 2 つを特徴量として扱っていたが、各時間帯の総通信量を通信プロトコルで分類することで特徴量を増やしてネットワークの判定を行い評価の改善を図った。

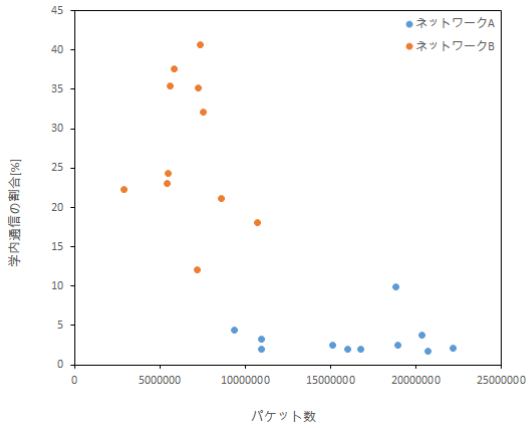

図 13 2019 年 11 月 26 日の TCP における各時間帯のパケット数と学内通信の割合の関係

2.4 通信プロコトルにより分類を行った LAN の解析

本研究では通信プロトコルを TCP と UDP の 2 種類に分類することにした。TCP は受信側がデータを正常に受信すると送信側に受け取ったことを報告し、一定時間経過して受信側から報告が無い場合は再度送信側がデータを送り直すことで確実性を保証する通信プロトコルであり、UDP は TCP と異なり受信側がデータを受け取ったことを報告は行わずパケットロスや到着順序の入れ替わりが起きても正しい順に並べ替えない確実性よりも速やかにデータが送信されることを重視する通信プロトコルである。

2019 年 11 月 26 日のデータを TCP と UDP に分類し、TCP における各時間帯の通信量と各時間帯の学内通信の割合の関係、UDP における各時間帯の通信量と各時間帯の学内通信の割合の関係を 2.1 と同様に散布図で示した。(図 13)(図 14)(図 15)(図 16)

図 13、図 14 より TCP における各時間帯の通信量と学内通信の関係は通信プロトコルによって分類する前の各時間帯の総通信量と学内通信の関係と大きな変化は見られないが、図 15、図 16 より UDP における各時間帯の通信量と学内通信の割合の関係は全く異なる分布をしており通信プロトコルによるデータの分類で傾向に違いが生じることが確認できた。

2.5 閾値の再決定

2.2 節で閾値を決定する際に用いた 2020 年 1 月 6 日から 1 月 10 日までの 5 日間のデータセットを TCP と UDP に分類した上で再度傾向を読み取るために、それぞれのネットワークの各通信プロトコルにおける通信量と学内通信の割合の関係を散布図で示した。(図 17)(図 18)(図 19)(図 20)(図 21)(図 22)(図 23)(図 24) そこから回帰直線を新たに導出して 2.2 節と同様に導出した回帰直線より切片が 30%大きい直線と 30%小さい直線を用意し、2 つの直線の間に出来る領域を閾値として再度決定させた。こちらも直線の範囲は各軸においてデータセットの最小値となる点から

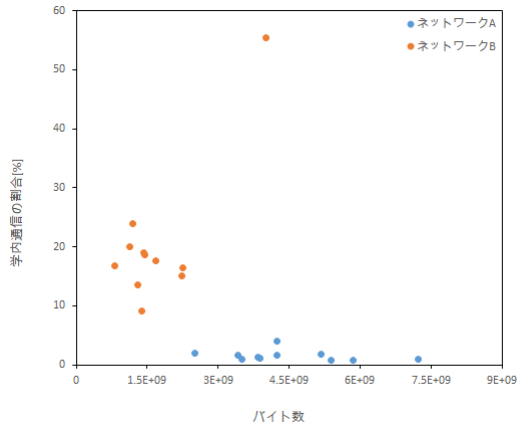

図 14 2019 年 11 月 26 日の TCP における各時間帯のバイト数と学内通信の割合の関係

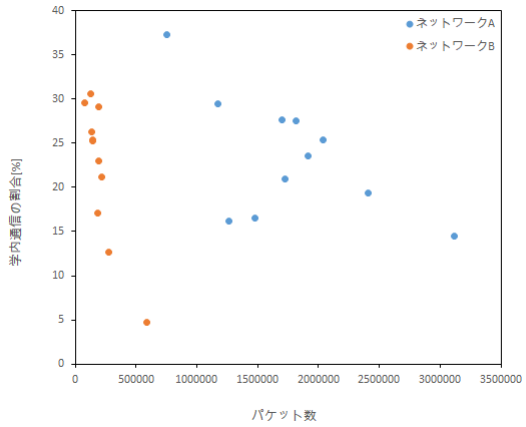

図 15 2019 年 11 月 26 日の UDP における各時間帯のパケット数と学内通信の割合の関係

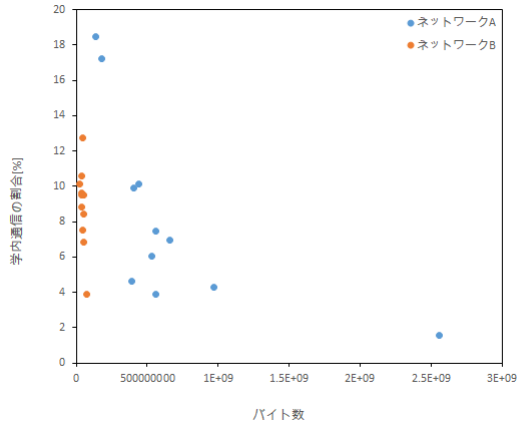

図 16 2019 年 11 月 26 日の UDP における各時間帯のバイト数と学内通信の割合の関係

最大値となる点までにした。なお、薄い青の直線が回帰直線であり、赤の直線が回帰直線より切片が 30%大きい直線、濃い赤の直線が回帰直線より切片が 30%小さい直線である。

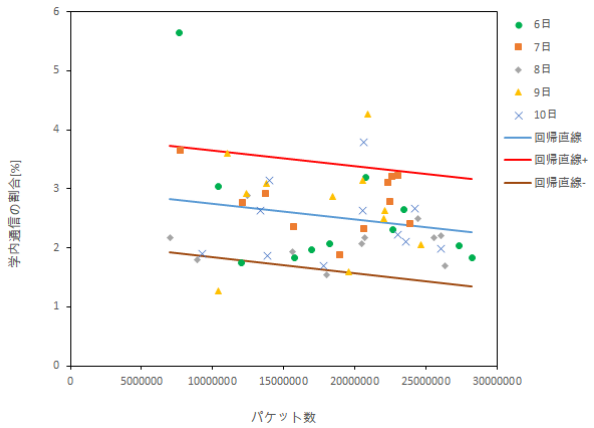

図 17 TCP のネットワーク A における各時間帯のパケット数と学内通信の割合の関係

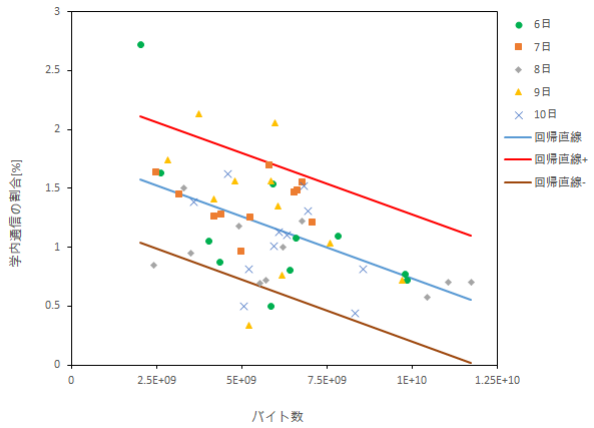

図 18 TCP のネットワーク A における各時間帯のバイト数と学内通信の割合の関係

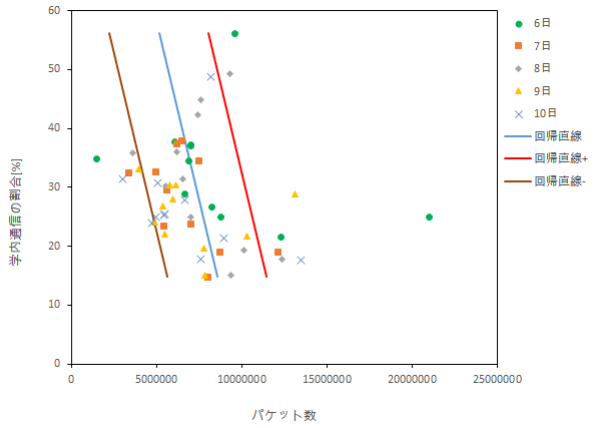

図 19 TCP のネットワーク B における各時間帯のパケット数と学内通信の割合の関係

2.6 閾値による再評価

2.5 節で再決定した閾値による評価を 2019 年 11 月 26 日のデータで行った。(表 3) 通信プロトコルで分類したことで各時間帯の判定がパケット数とバイト数の 2 つから TCP のパケット数、バイト数、UDP のパケット数、バイト数の 4 つとなり分類前に比べ厳密な判定を行うことを可能にし

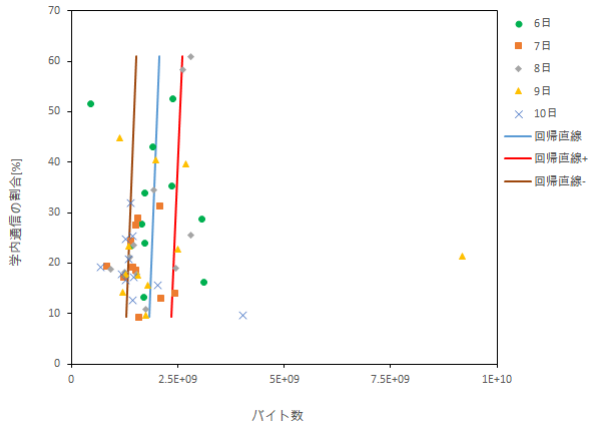

図 20 TCP のネットワーク B における各時間帯のバイト数と学内通信の割合の関係

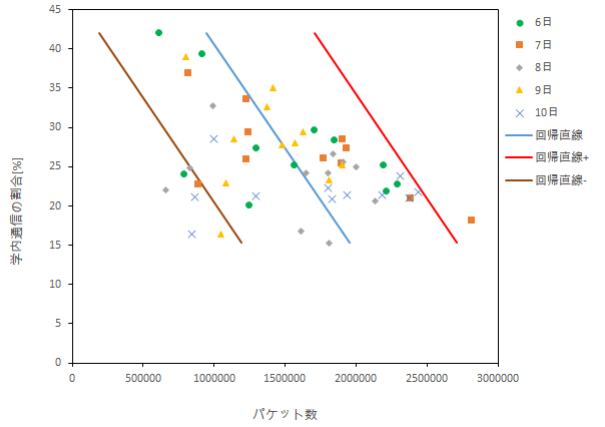

図 21 UDP のネットワーク A における各時間帯のパケット数と学内通信の割合の関係

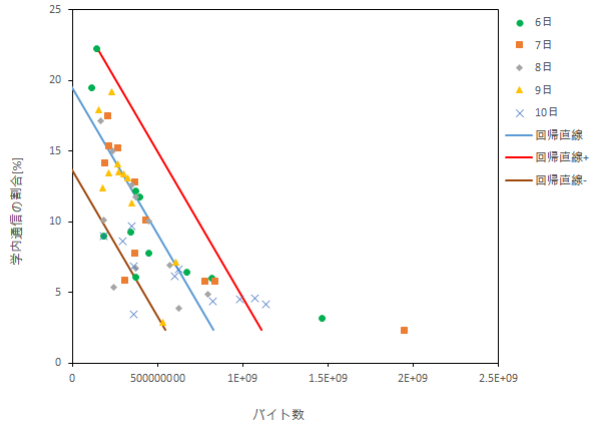

図 22 UDP のネットワーク A における各時間帯のバイト数と学内通信の割合の関係

た。評価は各時間帯において T が 3 つ以上であるならばその時間帯のデータは正しく識別が行われているとし、F が 3 つ以上であるならばその時間帯のデータは誤りであると識別されている、T と F が 2 つずつであればその時間帯のデータは識別不可であるとした。データセットの 2020 年 1 月 6 日から 9 日までを 1 つのデータフローとして見たと

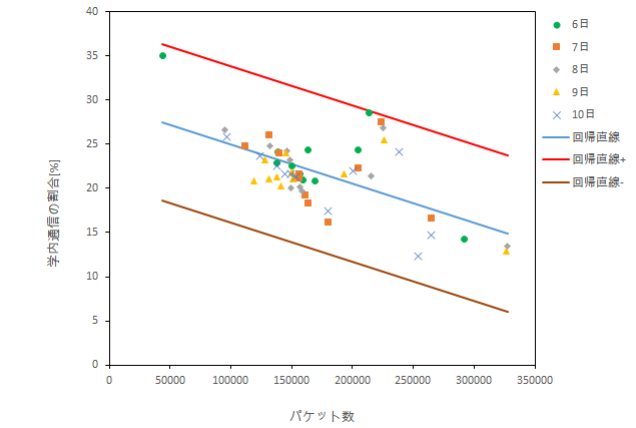

図 23 UDP のネットワーク B における各時間帯のパケット数と学内通信の割合の関係

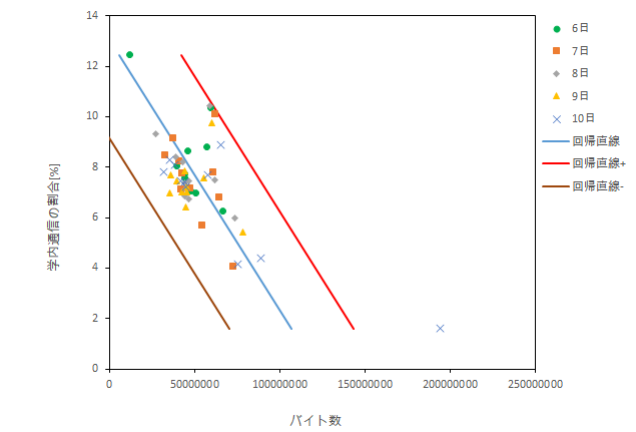

図 24 UDP のネットワーク B における各時間帯のバイト数と学内通信の割合の関係

表 3 分類後の 2019 年 11 月 26 日におけるネットワークの判定

時間帯	ネットワークA				ネットワークB			
	TCP		UDP		TCP		UDP	
	パケット数	バイト数	パケット数	バイト数	パケット数	バイト数	パケット数	バイト数
8:00-9:00	T	F	T	T	T	T	T	F
9:00-10:00	T	T	T	T	T	T	T	T
10:00-11:00	T	T	T	T	T	T	T	T
11:00-12:00	T	T	F	F	T	T	T	T
12:00-13:00	T	T	T	T	F	F	T	T
13:00-14:00	F	F	T	T	T	T	T	T
14:00-15:00	F	F	T	T	T	F	T	T
15:00-16:00	T	T	T	T	T	F	T	T
16:00-17:00	T	T	T	T	T	T	T	T
17:00-18:00	T	T	T	T	T	F	F	T
18:00-19:00	F	T	T	F	F	F	T	T

き、ネットワーク A と正確に識別されたデータは 72.36%、誤りであると識別されたデータは 9.09%、ネットワーク B と正確に識別されたのは 81.82%、誤りであると識別されたデータは 5.45%でありどちらのネットワークも閾値による正確な判定を行うことが可能であると考えられた。2019 年 11 月 26 日のデータを見ると閾値によりネットワーク A と識別されたのは 63.64%、ネットワーク B と識別されたのは 72.73%、誤りであると識別されたデータはどちらのネットワークにおいても 0%であり誤りであると判定したデータの少なさとネットワークに正しく識別されたデータの割合から閾値により正しく判定が行われたと言える。閾

値により識別不可とされたデータについて、データセットにおいてはネットワーク A で 14.55%、ネットワーク B で 12.73%、2019 年 11 月 26 日においてはネットワーク A で 36.36%、ネットワーク B で 27.27%であった。

3. ネットワーク利用者の時間的な特徴の抽出に関する評価

3.1 データの条件設定

データを取得するにあたり調査範囲として、主に通信が行われている時間帯の 8:00 から 21:00 の範囲でデータを取得することとした。ネットワーク A、ネットワーク B ともにデータフローについてグラフ化を行った際の軸について説明を行う。グラフの横軸は時刻とし、単位を 10 分とした。縦軸はバイト数、バイト数の増加量、全通信に対する TCP 通信のバイト数の割合の 3 種を設定した。バイト数の増加量、全通信に対する TCP 通信についての導出式は以下のとおりである。

$$\text{増加量} = (10\text{ 分間のバイト数}) - (\text{前 }10\text{ 分間のバイト数}) \quad (4)$$

$$\text{TCP 通信の割合} = \frac{\text{TCP 通信のバイト数}}{\text{全通信のバイト数}} \times 100 \quad (5)$$

3.2 定量的な特徴の抽出の流れ

定量的な特徴の抽出の過程について概要の説明を行う。まず、2020 年 1 月 6 日から 1 月 10 日までの 5 日間のデータに対し「時間変化に対するバイト数の変化」、「時間変化に対するバイト数の増加量」、「時間変化に対する TCP 通信の割合の変化」のグラフを作成する。グラフ作成後、グラフをもとに特徴の予測を行う。次に、特徴の予測をもとに特徴を抽出するための閾値を設定する。閾値を設定した後、2020 年 1 月 14 日のデータを用いて閾値の評価を行う。

3.3 グラフ作成、特徴の予測

1 月 6 日から 1 月 10 日までの 5 日間のデータに対し、時間変化に対するバイト数の変化についてのグラフ化を行った。ネットワーク A が図 25、ネットワーク B の結果が図 26 である。

図 25 よりネットワーク A では九州大学の講義時間内のときバイト数が大きくなる。ここでの講義時間とは 8:40 から 16:20 までの間に 90 分ごとに行われるものであり、講義間には 20 分間の休憩ならびに 12:00 から 13:00 の間は 1 時間の昼休憩を設けている。以降講義時間については前文で述べたとおりとする。図 26 よりネットワーク B では 8:00 からバイト数が大きくなりその後 21:00 まで緩やかに小さくなっていく。

続いて、1 月 6 日から 1 月 10 日までの 5 日間のデータに対し、時間変化に対するバイト数の増加量についてのグラフ化を行った。ネットワーク A が図 27、ネットワーク B の結果が図 28 である。

図 27 よりネットワーク A では講義開始時刻 20 分間か講義終了時刻前 30 分間に大きく増減する。図 28 より大きく増減する箇所が数カ所存在しているが、全体を通して増減量は少ない。

最後に、1 月 6 日から 1 月 10 日までの 5 日間のデータに対し、時間変化に TCP 通信の割合の変化についてのグラフ化を行った。ネットワーク A が図 29、ネットワーク B の結果が図 30 である。

図 29 より、ネットワーク A では TCP の割合が 80

グラフ化をしたことによって閾値を用いて以下のように特徴を求めることができると予測した。

(1) ネットワーク A は講義時間内のバイト数が大きくなる。
(2) ネットワーク B は活動時間中常に TCP 通信の割合が高い。

3.4 バイト数の特徴

ネットワーク A を赤、ネットワーク B を青で時間変化に対するバイト数の変化をグラフ化したところ（図 31)、突発的な点を除いて 12:00〜13:00、16:20〜18:30 の時間帯以外ではネットワーク A の方がバイト数が明らかに大きくなっていた。

このことから、講義時間内でバイト数を閾値として設定し、閾値以上をネットワーク A、閾値未満をネットワーク B とする方針で閾値の決定を行う。図 31 より、仮の閾値を 500MBytes とし、10 分間ごとに正誤判定を行い、精度を求めた。(表 4)(表 5)

ネットワーク A の精度を可能な限り下げず、ネットワーク B の精度を上げなければならないため閾値を上げていき 5 日分の精度の平均が高いときの閾値を求める。(表 6)

表 6 より、調査した中で閾値が 600M バイトのとき一番精度の平均が高かった。このことから講義時間内で閾値 600M バイト以上でネットワーク A、600M バイト未満でネットワーク B という特徴を抽出した。

3.5 TCP 通信の割合の特徴

バイト数の時と同様に講義時間内で、ネットワーク A を赤、ネットワーク B を青で時間変化に対する TCP 通信の変化についてグラフ化したところ (図 32)、ネットワーク B は常に高い TCP 通信の割合を維持しているが、ネットワーク A は比較的低くなっていた。

このことから、TCP 通信の割合を閾値として設定し、閾値以上ならネットワーク B、閾値以下ならネットワーク A とする方針で閾値の決定を行う。図 32 より、仮の閾値を 95 %とし 10 分ごとに正誤を判定し全体に対する割合を求

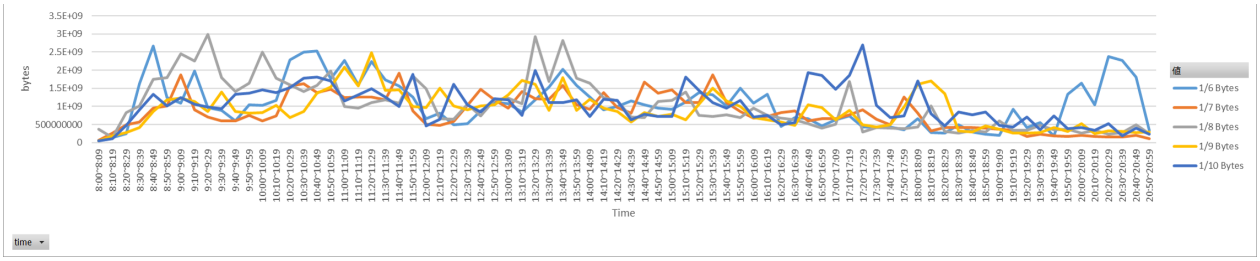

図 25　ネットワーク A の時間経過に対するバイト数の変化

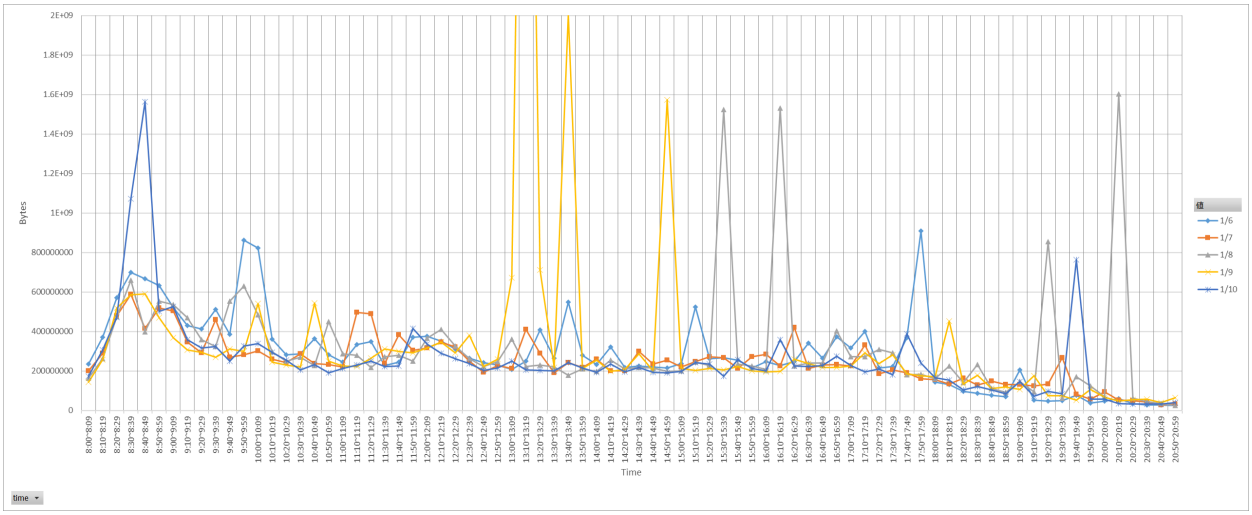

図 26　ネットワーク B の時間経過に対するバイト数の変化

めた。(表 7)(表 8)

ネットワーク B の判定精度を可能な限り下げず、ネットワーク A の判定精度を上げなければならないため閾値を上げていき 5 日分の精度の平均が高いときの閾値を求める。(表 9)

表 9 より、調査した中で閾値が 96 %のとき一番精度の平均が高かった。このことから講義時間内で TCP の割合が閾値 96 %以上でネットワーク A、600M バイト未満でネットワーク B という特徴を抽出した。

表 4　閾値 500MBytes のときの正誤判定 (ネットワーク A)

time	1月6日	判定	1月7日	判定	1月8日	判定	1月9日	判定	1月10日	判定
8:40~8:49	2665765124	T	948328142	T	1748471981	T	862896609	T	1332027990	T
8:50~8:59	1246283907	T	1018193668	T	1795643824	T	1185590010	T	1015540544	T
9:00~9:09	1085916877	T	1877000893	T	2447652049	T	1227432635	T	1241264261	T
9:10~9:19	1983398077	T	903267814	T	2260374689	T	1143394992	T	1058155176	T
9:20~9:29	971528834	T	713539050	T	2984778608	T	853213181	T	978077242	T
9:30~9:39	889387593	T	600941522	T	1791511139	T	1393407405	T	941630963	T
9:40~9:49	592935571	T	602431734	T	1412757034	T	862227689	T	1330149823	T
9:50~9:59	1048211550	T	758949034	T	1639446374	T	819603111	T	1365589205	T
10:00~10:09	1023907338	T	593203244	T	2499997250	T	830553492	T	1454299107	T
10:30~10:39	2491095398	T	1623920576	T	1413532191	T	867527069	T	1776791755	T
10:40~10:49	2532460213	T	1387618244	T	1576846188	T	1371590833	T	1816538988	T
10:50~10:59	1770271150	T	1471143476	T	1972031247	T	1528821917	T	1699973539	T
11:00~11:09	2267140560	T	1249304054	T	991875076	T	2082069266	T	1146263402	T
11:10~11:19	1599565697	T	1257808007	T	959164099	T	1565766487	T	1326319443	T
11:20~11:29	2239754548	T	1252290001	T	1099457351	T	2488342136	T	1484836184	T
11:30~11:39	1739448908	T	1186814768	T	1177168870	T	1434505818	T	1255832077	T
11:40~11:49	1569671288	T	1918256507	T	1091357309	T	1464423162	T	990702412	T
11:50~11:59	1266023745	T	883490267	T	1842387170	T	1004935077	T	1879048228	T
13:00~13:09	1071178479	T	955812791	T	1231110769	T	1327075904	T	1176190244	T
13:10~13:19	841015446	T	1414104926	T	1072971827	T	1714258596	T	746047369	T
13:20~13:29	1182916120	T	1217137082	T	2931556625	T	1604640467	T	1994409324	T
13:30~13:39	1541237431	T	1202030106	T	1679935880	T	887983565	T	1104952134	T
13:40~13:49	2016512528	T	1573050290	T	2823634115	T	1790542676	T	1107748329	T
13:50~13:59	1574609417	T	1038213132	T	1771585106	T	888497284	T	1185770874	T
14:00~14:09	1254844945	T	924808388	T	1641704008	T	1195818114	T	721062682	T
14:10~14:19	900915260	T	1384596330	T	1240901044	T	958230233	T	1201362073	T
14:20~14:29	992443385	T	965396182	T	1151825441	T	864068080	T	1166800384	T
14:50~14:59	948632560	T	1369416246	T	1132375369	T	726384554	T	716400196	T
15:00~15:09	927530882	T	1460038343	T	1168071010	T	782242625	T	719862056	T
15:10~15:19	1117336519	T	1099230839	T	1399379617	T	634786533	T	1816137728	T
15:20~15:29	1385334522	T	1106357076	T	748244835	T	1058552090	T	1448697804	T
15:30~15:39	1315845709	T	1869773251	T	721125154	T	1509163236	T	1101864532	T
15:40~15:49	1015978811	T	1108523745	T	767667674	T	1152087612	T	944911944	T
15:50~15:59	1509971482	T	853650066	T	686829342	T	999091138	T	1168580005	T
16:00~16:09	1089950931	T	682858254	T	952016078	T	695442160	T	711498572	T
16:10~16:19	1341831069	T	729603685	T	760525017	T	633864233	T	759399832	T
	全項目	36	全項目	36	全項目	36	全項目	36	全項目	36
	T	36	T	36	T	36	T	36	T	36
	精度	100	精度	100	精度	100	精度	100	精度	100

表 5　閾値 500MBytes のときの正誤判定 (ネットワーク B)

time	1月6日	判定	1月7日	判定	1月8日	判定	1月9日	判定	1月10日	判定
8:40~8:49	666022011	F	414839376	T	398676211	T	589581036	F	1563837532	F
8:50~8:59	631344272	F	520037669	F	552818768	F	468992119	T	502101228	F
9:00~9:09	520682084	F	504950416	F	535727290	F	368504352	T	526015067	F
9:10~9:19	428910880	T	345465364	T	468993468	T	306305069	T	358814830	T
9:20~9:29	413396838	T	292075567	T	358450509	T	294737876	T	317249775	T
9:30~9:39	518455560	F	459361628	T	322783399	T	269074260	T	322814716	T
9:40~9:49	385261107	T	269611932	T	554264245	F	311377449	T	250315698	T
9:50~9:59	861948473	F	282672212	T	630276691	F	299184425	T	326899060	T
10:00~10:09	823059287	F	302014298	T	485384299	T	540366506	F	338674460	T
10:30~10:39	287903268	T	285643938	T	270583550	T	220522491	T	204178340	T
10:40~10:49	363538559	T	237506488	T	227648107	T	542689755	F	232530925	T
10:50~10:59	281920796	T	233342894	T	448996146	T	250597592	T	190366008	T
11:00~11:09	245134720	T	221215679	T	287549863	T	225039024	T	213379062	T
11:10~11:19	334725208	T	497926825	T	280072487	T	222876812	T	228889132	T
11:20~11:29	347533234	T	489667433	T	217801313	T	263808824	T	249946840	T
11:30~11:39	227235126	T	237603591	T	273258332	T	310682902	T	223330482	T
11:40~11:49	244943612	T	383504176	T	279533770	T	299335730	T	225532626	T
11:50~11:59	370731151	T	304236195	T	248868893	T	291906117	T	415901398	T
13:00~13:09	208645408	T	211602329	T	361836985	T	671370920	F	250565758	T
13:10~13:19	249728501	T	409225660	T	223421607	T	5410810139	F	206048078	T
13:20~13:29	408839969	T	289197480	T	230385221	T	710818899	F	202040823	T
13:30~13:39	264570117	T	190043365	T	220319743	T	217272738	T	199908207	T
13:40~13:49	549765347	F	242693362	T	178656892	T	2005939847	F	243345357	T
13:50~13:59	279521512	T	218640542	T	210097136	T	212931321	T	216430939	T
14:00~14:09	232502679	T	260720681	T	196989652	T	254801843	T	192417723	T
14:10~14:19	321259377	T	200661937	T	253630972	T	198350427	T	234988912	T
14:20~14:29	218585635	T	200284042	T	210537046	T	207478270	T	196556084	T
14:50~14:59	215205459	T	254877058	T	198867853	T	1574005587	F	190213599	T
15:00~15:09	238628570	T	220298978	T	200449205	T	212360996	T	198058716	T
15:10~15:19	523282016	F	247284823	T	244818145	T	203119108	T	241892438	T
15:20~15:29	262226909	T	272572737	T	227000927	T	213578883	T	234080044	T
15:30~15:39	268102623	T	267051799	T	1523097865	F	206579494	T	174006757	T
15:40~15:49	255921641	T	212125283	T	236175665	T	223444906	T	258391824	T
15:50~15:59	215834649	T	271679122	T	226003531	T	200394318	T	212261802	T
16:00~16:09	246954958	T	284743767	T	207950196	T	194392286	T	200060570	T
16:10~16:19	228249205	T	225156346	T	1530256156	F	198374847	T	359177615	T
	全項目	36	全項目	36	全項目	36	全項目	36	全項目	36
	T	28	T	34	T	30	T	28	T	33
	精度	77.78	精度	94.44	精度	83.33	精度	77.78	精度	91.67

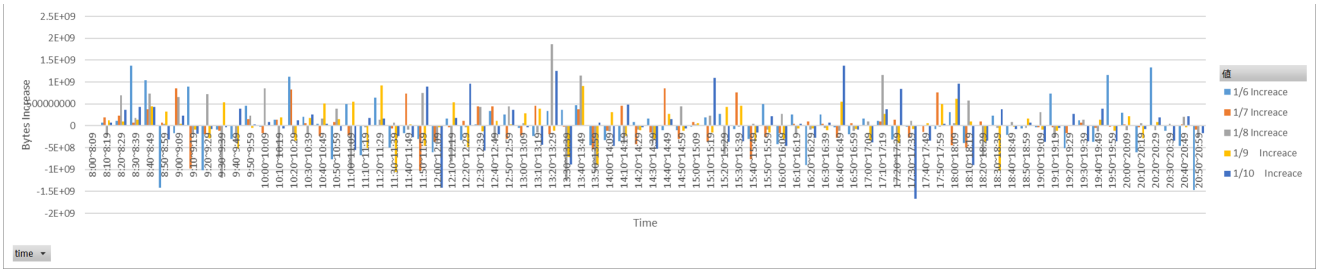

図 27　ネットワーク A の時間経過に対するバイト数の増加量

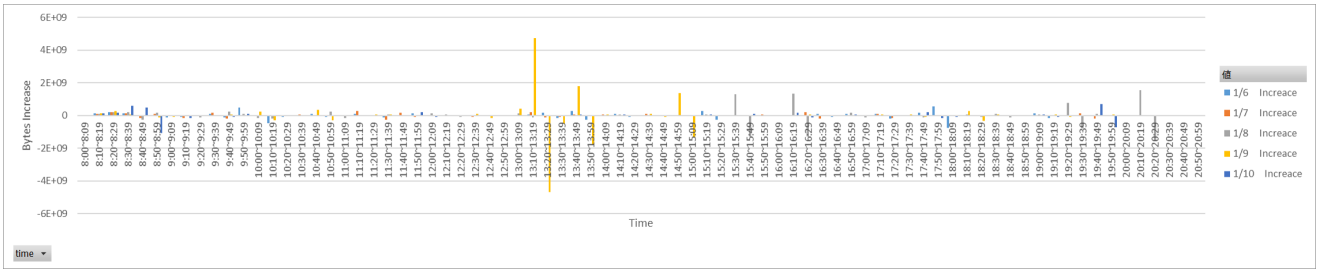

図 28　ネットワーク B の時間経過に対するバイト数の増加量

3.6　閾値の検証

3.5 節より、講義時間内でバイト数が 600M バイト以上で TCP 通信の割合が 96 %以下のとき教育用ネットワークであるという特徴を抽出した。ここで閾値の正当性を確かめるために、新しく 1 月 14 日のデータを用いて閾値の評価を行う。1 月 14 日のデータを表 10 、用意した閾値で判定したときの判定結果を表 11 に示す。

また、14 日のデータも併せてもう一度閾値ごとの精度を調べたところ表 12 と表 13 のように精度の平均が変化した。

14 日のデータを加えたことで、バイト数の方はどの閾値でも精度が上がったが、TCP 通信の割合の方では精度が上下していた。また、精度が一番高くなる閾値は変わらず、600M バイトと TCP の割合が 96 %のときであったため閾値は正当なものであった。

4.　考察

空間的特徴の評価では、通信プロトコルの分類後の評価で両ネットワークともに正しく識別されたデータの割合は高く、誤りであると識別されたデータの割合は低かった。しかし識別不可とされたデータに関しては再評価によりデータセットにおいてネットワーク A では 3.64 %上昇、ネットワーク B では 10.91 %下降しており、2019 年 11 月 26 日のデータにおいてネットワーク A では 27.27 %上昇、ネットワーク B では変化は無く、識別不可とされたデータがネットワーク B では減少したがネットワーク A では増加したことが確認できた。この識別不可のデータの判定結果により評価が変わるため再評価により識別不可のデータの割合を減らすことが必要であったが結果としてネットワーク B は改善されたがネットワーク A は精度が落ちたと言える。このことからネットワークの判定において評価するデータを増加させることで誤りであると識別するデータの割合を減らし正しい識別を行うことを可能にするが、ネットワークによっては判定が複雑になり識別が困難になるデータが増加すると考えられた。また、誤りであると識別したデータや識別不可であるデータが存在する原因として、閾値を決定する際に求めた回帰直線が扱った日数が 5 日間という少ないデータセットから導出したため正確な傾向を掴めなかったことや本研究で扱ったデータの中に普段行うことの無い量の通信が行われていた可能性が挙げられる。

時間的特徴の評価では、グラフより特徴を予測し閾値による判定を高い精度で行うことができた。しかし、精度の高い閾値を探すうえであまり複雑な細かい数字を閾値として取り扱わなかったことが問題としてあげられる。取り扱わなかった理由は本研究ではわかりやすく特徴を評価できるよう手動で調査を行っているためである。しかし、最終的な目標としては手動ではなく自動での閾値の設定を考えており、自動化が可能になるとさらに厳密な閾値を設定可能になると推測している。また、取り扱うデータが少なかったため閾値が厳密に設定出来ず良い検証を行うことが出来なかったと考えている。今回グラフ化の際に縦軸をバイト数、TCP 通信の割合を取り上げたが、それ以外のパケット数や id などに着目するとまた違う特徴を予測する余地があるため、今後は今回の研究とは異なるパラメータで特徴を抽出することも検討している。

以上のことからネットワークの判定を正確に行うためには空間的特徴、時間的特徴ともにデータセットを組む際に 1 ヶ月や 2 ヶ月と多くのデータを取り扱うことが必要であると考えられた。

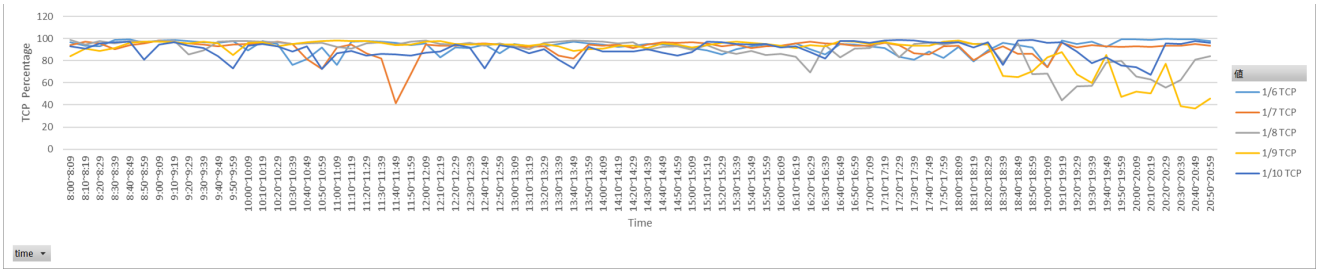

図 29　ネットワーク A の時間経過に対する TCP 通信の割合の変化

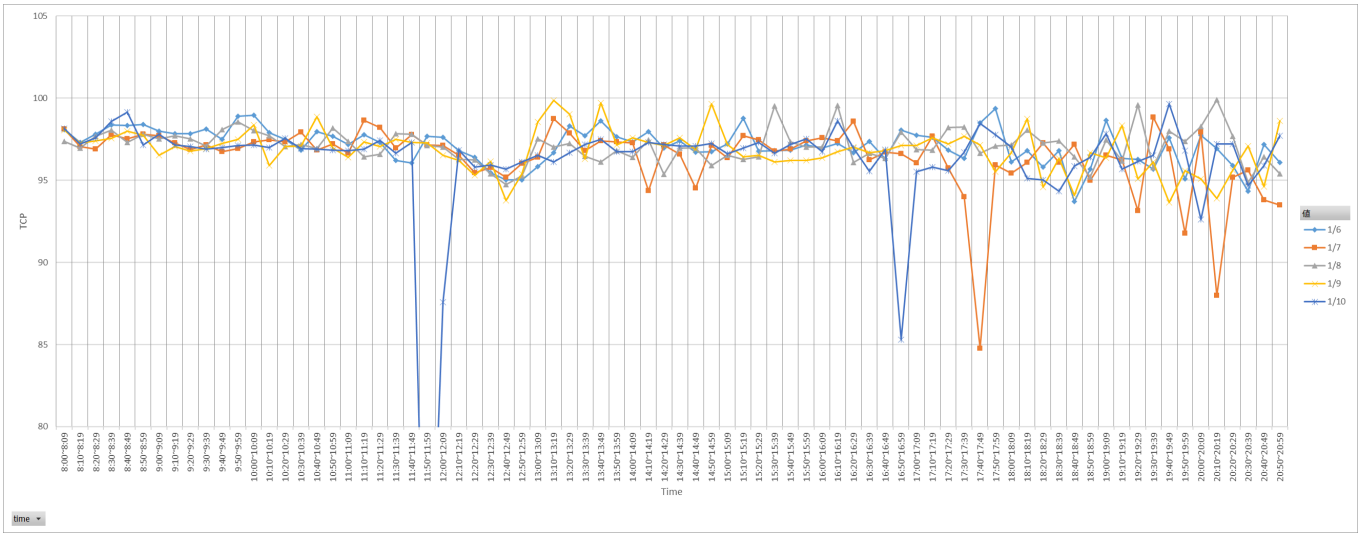

図 30　ネットワーク B の時間経過に対する TCP 通信の割合の変化

5.　おわりに

本研究ではデータフローから大学内における利用目的の異なる 2 つのネットワークの特徴を空間的および時間的な観点から定量的に示して抽出することを目的として研究を行った。

空間的特徴の研究では、調査対象としたネットワークが活発に動く時間帯において 1 時間ごとのパケット数およびバイト数と 1 時間ごとの学内通信の割合から使用するサービスの違いを数値で捉えることでネットワークの識別を行うことが可能であることを示し、データセットを用いて閾値を決定することにより正確なネットワークの識別の割合がデータセットと同様に評価可能であることをデータセットに含まれない別日に適用して確認した。ネットワーク A においては高い割合で識別を可能にしたがネットワーク B においては 50 %程度の割合で識別を行うことから両ネットワークともに高い割合で識別を行うことは困難であった。評価を改善するために 1 時間ごとのパケット数およびバイト数と 1 時間ごとの学内通信の割合を通信プロトコルの TCP と UDP に分類することでネットワークの傾向を多くの面から捉えて評価基準を増やし、データセットを用いて再度閾値を決定することでネットワークの識別の再評価を可能にした。結果として、両ネットワークともに誤りであると識別したデータの割合が低く正しく識別したデータの割合が高かったことから通信プロトコルによりデータを分類して識別を行うことで評価を改善可能であることが確認できた。したがって通信の傾向をデータフローから数値化して捉え閾値を設けることでネットワークを利用する目的により通信は異なっていることが示せた。

時間的特徴の研究では、ネットワークのデータフローから時間変化におけるバイト数の変化、TCP 通信の割合の変化に着目しグラフ化を行い、閾値を用いてネットワーク特徴を定量的に示すことでネットワーク利用者の特徴の抽出を試み、バイト数と TCP 通信の割合にそれぞれ閾値を設定することで特徴を抽出できた。また、抽出した特徴が正当であるかを確認するために実際に特徴を抽出したときに使用したデータとは別のデータに対し抽出した特徴が表れるか評価を行った。その結果、判定精度が高かったため閾値が妥当なものであると示した。改善可能なこととして閾値を設定するうえでデータサンプル数が少なかったこと、本研究ではわかりやすく特徴を評価できるように手動で作業を行っていたため時間がかかり効率が悪いことがあげられた。以上の点をふまえて厳密な閾値を導くためにより多くのサンプルデータから閾値を設定すること、また多くのデータを取り扱おうとするにあたり手動で行っていた閾値を吟味する作業の高速化のために自動化することが必須と

なる。

今後の課題としてはネットワークの判定をより改善するために閾値を決める際に用いたデータセットで扱う日数を増やし、より多くのデータから傾向を掴むことが必要である。また、学内において利用者が複雑であるネットワークが本研究で扱った閾値によって利用目的ごとに識別が可能であることの確認と通信相手先を変更することで各時間帯の通信量と相手先との通信の割合に似た関係を調査することで本研究のネットワーク判定の妥当性を確認する必要がある。最終的には、本研究で得られた結果をもとに特徴を捉えたネットワークを「教師データ」として扱い、様々な利用者が混在するネットワークの特徴の抽出を機械学習的に行えるよう研究する予定である。

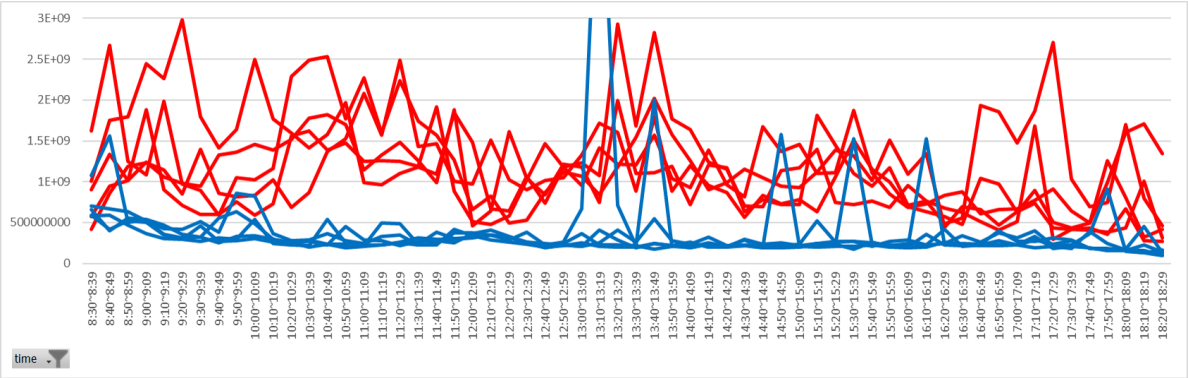

図 31　ネットワーク A とネットワーク B の時間経過に対するバイト数の変化の比較

表 6　閾値を変化させたときの精度

500MBytes			550MBytes			600MBytes			625MBytes			650MBytes			700MBytes		
全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度
36	36	100	36	36	100	36	35	97.22	36	35	97.22	36	35	97.22	36	35	97.22
全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度
36	36	100	36	36	100	36	35	97.22	36	33	91.67	36	33	91.67	36	32	88.89
全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度
36	36	100	36	36	100	36	36	100	36	36	100	36	36	100	36	35	97.22
全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度
36	36	100	36	36	100	36	36	100	36	36	100	36	34	94.44	36	33	91.67
全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度
36	36	100	36	36	100	36	36	100	36	36	100	36	36	100	36	36	100
全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度
36	28	77.78	36	32	88.89	36	32	88.89	36	32	88.89	36	33	91.67	36	34	94.44
全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度
36	34	94.44	36	36	100	36	36	100	36	36	100	36	36	100	36	36	100
全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度
36	30	83.33	36	31	86.11	36	33	91.67	36	33	91.67	36	34	94.44	36	34	94.44
全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度
36	28	77.78	36	30	83.33	36	31	86.11	36	31	86.11	36	31	86.11	36	32	88.89
全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度
36	33	91.67	36	35	97.22	36	35	97.22	36	35	97.22	36	35	97.22	36	35	97.22
平均精度			平均精度			平均精度			平均精度			平均精度			平均精度		
92.50			95.56			95.83			95.28			95.28			95.00		

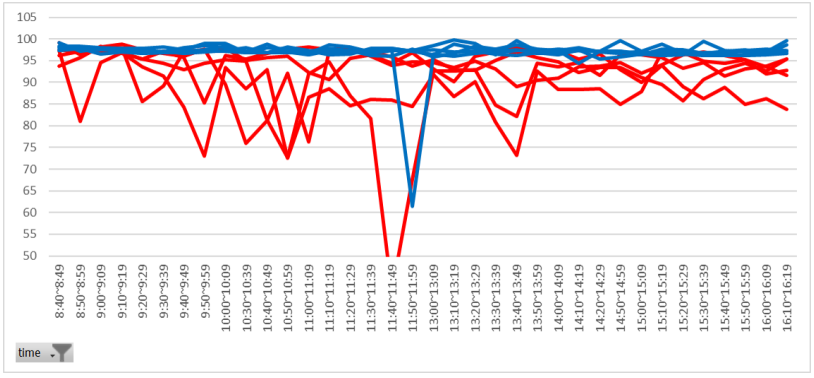

図 32　ネットワーク A とネットワーク B の時間経過に対する TCP 通信の割合の変化の比較

表 7　閾値 95 %のときの正誤判定 (ネットワーク A)

time	1月6日	判定	1月7日	判定	1月8日	判定	1月9日	判定	1月10日	判定
8:40~8:49	99.16	F	93.81	T	97.55	F	96.26	F	97.37	F
8:50~8:59	96.40	F	95.63	F	97.03	F	97.23	F	80.96	T
9:00~9:09	98.09	F	98.34	F	97.87	F	97.38	F	94.52	T
9:10~9:19	98.84	F	96.73	F	98.07	F	96.89	F	96.82	F
9:20~9:29	97.48	F	95.44	F	85.61	T	95.40	F	93.53	T
9:30~9:39	96.74	F	94.45	T	89.23	T	97.25	F	91.39	T
9:40~9:49	96.05	F	92.91	T	97.16	F	95.73	F	84.21	T
9:50~9:59	97.64	F	94.47	T	97.55	F	85.28	T	72.98	T
10:00~10:09	89.55	T	95.18	F	97.89	F	96.16	F	93.36	T
10:30~10:39	75.96	T	94.90	T	95.02	F	95.33	F	88.50	T
10:40~10:49	81.21	T	81.36	T	95.75	F	96.70	F	92.97	T
10:50~10:59	92.15	T	72.77	T	95.98	F	97.62	F	72.50	T
11:00~11:09	76.34	T	92.03	T	92.25	T	98.15	F	86.64	T
11:10~11:19	97.11	F	94.66	T	90.71	T	97.46	F	88.57	T
11:20~11:29	97.67	F	86.84	T	95.59	F	97.93	F	84.63	T
11:30~11:39	97.24	F	81.74	T	96.43	F	96.10	F	86.08	T
11:40~11:49	95.95	F	41.67	T	94.45	T	93.90	T	85.84	T
11:50~11:59	93.82	T	67.64	T	96.93	F	94.73	T	84.43	T
13:00~13:09	95.20	F	92.63	T	93.11	T	94.57	T	91.82	T
13:10~13:19	92.87	T	92.63	T	90.17	T	93.35	T	86.75	T
13:20~13:29	92.76	T	92.99	T	96.04	F	94.96	T	90.16	T
13:30~13:39	95.03	F	84.84	T	97.09	F	93.06	T	80.86	T
13:40~13:49	96.99	F	82.12	T	98.15	F	89.05	T	73.23	T
13:50~13:59	95.75	F	94.34	T	97.54	F	90.54	T	92.59	T
14:00~14:09	94.66	T	93.51	T	97.13	F	90.99	T	88.32	T
14:10~14:19	92.35	T	94.60	T	95.38	F	93.60	T	88.44	T
14:20~14:29	93.48	T	91.62	T	96.45	F	93.78	T	88.45	T
14:50~14:59	93.50	T	96.35	F	92.96	T	94.51	T	84.86	T
15:00~15:09	91.20	T	96.43	F	89.97	T	92.16	T	87.85	T
15:10~15:19	89.44	T	95.69	F	93.94	T	94.04	T	96.98	F
15:20~15:29	85.78	T	93.21	T	89.02	T	96.32	F	96.80	F
15:30~15:39	90.58	T	94.50	T	86.19	T	97.05	F	94.90	T
15:40~15:49	93.09	T	91.42	T	88.80	T	96.24	F	94.43	T
15:50~15:59	94.36	T	93.03	T	84.99	T	95.17	F	95.20	F
16:00~16:09	92.65	T	93.78	T	86.32	T	93.55	T	91.90	T
16:10~16:19	95.40	F	95.44	F	83.75	T	91.63	T	92.79	T
	全項目	36	全項目	36	全項目	36	全項目	36	全項目	36
	T	19	T	27	T	16	T	17	T	31
	精度	52.78	精度	75.00	精度	44.44	精度	47.22	精度	86.11

表 8　閾値 95 %のときの正誤判定 (ネットワーク B)

time	1月6日	判定	1月7日	判定	1月8日	判定	1月9日	判定	1月10日	判定
8:40~8:49	98.33	T	97.53	T	97.32	T	98.00	T	99.16	T
8:50~8:59	98.39	T	97.81	T	97.80	T	97.75	T	97.16	T
9:00~9:09	97.99	T	97.71	T	97.54	T	96.52	T	97.81	T
9:10~9:19	97.83	T	97.27	T	97.71	T	97.05	T	97.16	T
9:20~9:29	97.82	T	96.88	T	97.51	T	96.76	T	97.06	T
9:30~9:39	98.12	T	97.15	T	97.09	T	96.92	T	96.91	T
9:40~9:49	97.49	T	96.74	T	98.07	T	97.24	T	96.98	T
9:50~9:59	98.89	T	96.94	T	98.55	T	97.50	T	97.11	T
10:00~10:09	98.96	T	97.34	T	98.03	T	98.37	T	97.15	T
10:30~10:39	96.83	T	97.93	T	97.18	T	97.09	T	96.93	T
10:40~10:49	97.95	T	96.83	T	96.92	T	98.87	T	96.89	T
10:50~10:59	97.67	T	97.21	T	98.19	T	96.97	T	96.83	T
11:00~11:09	97.18	T	96.66	T	97.36	T	96.39	T	96.81	T
11:10~11:19	97.78	T	98.66	T	96.42	T	97.32	T	96.89	T
11:20~11:29	97.32	T	98.21	T	96.60	T	97.06	T	97.42	T
11:30~11:39	96.22	T	96.95	T	97.83	T	97.49	T	96.66	T
11:40~11:49	96.04	T	97.78	T	97.80	T	97.29	T	97.31	T
11:50~11:59	97.68	T	97.16	T	97.10	T	97.28	T	61.42	F
13:00~13:09	95.83	T	96.38	T	97.52	T	98.55	T	96.51	T
13:10~13:19	96.69	T	98.73	T	97.01	T	99.85	T	96.12	T
13:20~13:29	98.30	T	97.88	T	97.25	T	99.03	T	96.67	T
13:30~13:39	97.70	T	96.80	T	96.46	T	96.31	T	97.17	T
13:40~13:49	98.62	T	97.38	T	96.13	T	99.70	T	97.50	T
13:50~13:59	97.66	T	97.34	T	96.80	T	97.14	T	96.73	T
14:00~14:09	97.32	T	97.27	T	96.38	T	97.59	T	96.74	T
14:10~14:19	97.96	T	94.37	F	97.45	T	97.30	T	97.30	T
14:20~14:29	96.97	T	97.15	T	95.36	T	97.12	T	97.18	T
14:50~14:59	96.74	T	97.11	T	95.91	T	99.66	T	97.24	T
15:00~15:09	97.21	T	96.36	T	96.51	T	97.16	T	96.63	T
15:10~15:19	98.78	T	97.70	T	96.27	T	96.43	T	96.95	T
15:20~15:29	96.76	T	97.46	T	96.44	T	96.53	T	97.32	T
15:30~15:39	96.79	T	96.76	T	99.54	T	96.13	T	96.65	T
15:40~15:49	96.85	T	96.90	T	97.35	T	96.20	T	97.18	T
15:50~15:59	97.15	T	97.41	T	97.01	T	96.20	T	97.52	T
16:00~16:09	96.96	T	97.59	T	97.01	T	96.38	T	96.73	T
16:10~16:19	97.23	T	97.38	T	99.56	T	96.73	T	98.62	T
	全項目	36	全項目	36	全項目	36	全項目	36	全項目	36
	T	36	T	35	T	36	T	36	T	35
	精度	100	精度	97.22	精度	100	精度	100	精度	97.22

表 9　閾値を変化させたときの精度

95%			95.50%			96%			96.50%			97%		
全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度
36	19	52.78	36	22	61.11	36	24	66.67	36	26	72.22	36	28	77.78
全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度
36	27	75.00	36	30	83.33	36	32	88.89	36	34	94.44	36	35	97.22
全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度
36	16	44.44	36	18	50.00	36	21	58.33	36	24	66.67	36	25	69.44
全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度
36	17	47.22	36	20	55.56	36	21	58.33	36	26	72.22	36	28	77.78
全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度
36	31	86.11	36	32	88.89	36	32	88.89	36	32	88.89	36	35	97.22
全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度
36	36	100	36	36	100	36	35	97.22	36	33	91.67	36	26	72.22
全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度
36	35	97.22	36	35	97.22	36	35	97.22	36	33	91.67	36	24	66.67
全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度
36	36	100	36	35	97.22	36	34	94.44	36	28	77.78	36	24	66.67
全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度
36	36	100	36	36	100	36	36	100	36	29	80.56	36	23	63.89
全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度	全項目	T	精度
36	35	97.22	36	35	97.22	36	35	97.22	36	34	94.44	36	18	50.00
平均精度			平均精度			平均精度			平均精度			平均精度		
80.00			83.06			84.72			83.06			73.89		

表 10　1 月 14 日のデータ表

time	ネットワークA		ネットワークB	
	バイト数	TCP通信の割合	バイト数	TCP通信の割合
8:40~8:49	1123999306	96.53	753882121	98.59
8:50~8:59	555469807	95.44	1416906179	99.30
9:00~9:09	702525652	96.76	565395554	98.18
9:10~9:19	1536000717	98.26	505289883	98.08
9:20~9:29	904960976	97.25	436996526	98.02
9:30~9:39	783425231	96.50	324743907	96.85
9:40~9:49	1064260971	97.10	312992642	97.10
9:50~9:59	715818919	95.39	348515528	97.48
10:00~10:09	921758225	80.22	345382153	98.04
10:30~10:39	2404679546	90.11	241839038	96.76
10:40~10:49	1232828827	96.51	241931927	96.87
10:50~10:59	1038193843	94.25	283773789	97.12
11:00~11:09	1046546242	95.25	245125786	96.62
11:10~11:19	1212743614	95.23	215536341	96.38
11:20~11:29	890656030	93.15	234636721	96.37
11:30~11:39	933353237	95.24	233697801	97.11
11:40~11:49	953535858	95.84	239569600	97.14
11:50~11:59	862026630	94.46	320084704	97.78
13:00~13:09	1152330602	95.26	175251243	95.69
13:10~13:19	1029994823	94.23	172683137	96.04
13:20~13:29	1662880046	94.08	196950399	96.25
13:30~13:39	1384168850	93.58	286056547	97.15
13:40~13:49	1062743047	87.52	218414140	96.62
13:50~13:59	1098812337	86.69	280495085	97.24
14:00~14:09	1253418695	88.40	337640369	97.48
14:10~14:19	1220475290	83.73	304031302	97.88
14:20~14:29	1305717070	86.60	242357918	96.90
14:50~14:59	911490008	78.55	161377018	94.94
15:00~15:09	1468704485	83.31	183630492	96.19
15:10~15:19	1862998162	70.63	198968762	96.86
15:20~15:29	1332878294	73.97	191817425	96.61
15:30~15:39	1044486901	74.72	179837815	96.44
15:40~15:49	1150469952	65.85	208060799	96.82
15:50~15:59	1222127292	85.48	214588503	96.97
16:00~16:09	1618041028	79.81	191213379	96.99
16:10~16:19	799427324	84.69	225286313	97.82

表 11　1 月 14 日の判定結果

ネットワークA				ネットワークB			
バイト数		TCP通信の割合		バイト数		TCP通信の割合	
全体	36	全体	36	全体	36	全体	36
T	35	T	29	T	34	T	34
精度	97.22	精度	80.56	精度	94.44	精度	94.44

表 12　14 日のデータを加えたことによる精度の変化 (バイト数)

閾値候補	平均精度	
	1月6日から10日	1月6日から10日+1月14日
500MBytes	92.50	92.82
550MBytes	95.56	95.60
600MBytes	95.83	95.83
625MBytes	95.28	95.37
650MBytes	95.28	95.37
700MBytes	95.00	95.14

表 13　14 日のデータを加えたことによる精度の変化 (TCP 通信の割合)

閾値候補	平均精度	
	1月6日から10日	1月6日から10日+1月14日
95.00	80.00	79.86
95.50	83.06	83.80
96.00	84.72	85.19
96.50	83.06	82.64
97.00	73.89	73.15