大学キャンパス Wi-Fi アクセスポイント利用状況データセットの構築

升屋 正人, 根本 貴弘, 佐藤 亮介

東京農工大学 総合情報メディアセンター

masatom@go.tuat.ac.jp

Construction of a Dataset for University Campus Wi-Fi Access Point Usage

Masato Masuya, Takahiro Nemoto, Ryosuke Sato

Information Media Center, Tokyo University of Agriculture and Technology

概要

大学キャンパスにおいて，Wi-Fi アクセスポイントは学習や研究に不可欠なインフラとなっている．多くの大学が多数の Wi-Fi アクセスポイントを学内のさまざまな場所に設置しているが，その利用状況データはセキュリティ確保とプライバシー保護のため，一部の限られたスタッフしか参照することができない．このため，混雑状況の可視化やデータサイエンス教育など，さまざまな応用が期待されるにも関わらず，データを十分に活用することができていない．そこでわれわれは，オープンデータとして公開が可能な，セキュリティとプライバシーを確保した Wi-Fi 利用状況データセットを構築することにした．これにより，学内外の研究者による研究やデータサイエンス教育，大学 DX 推進に寄与することができる．

1 はじめに

大学キャンパスにおいてインターネット接続を提供するネットワークインフラの一つである Wi-Fi アクセスポイントは，学習，研究，そして，日常のコミュニケーションを行う上で，学生と教職員の双方にとって必要不可欠なものとなっている．学生は，授業中にわからないことがあった場合にその場でインターネットにアクセスして情報を補足できるほか，教室や図書館，カフェテリアやオープンスペースなど大学構内のどこからでもオンライン教材や講義資料にアクセスできる．研究者は，研究室のほか実験室やフィールドにおいても論文や研究プロジェクトの共有リソースにアクセスでき，データ収集やデータ共有を迅速に行うことができる．学生同士がグループワークを行う際や，教員と学生がオンラインで連絡を取り合う場合にも Wi-Fi が有用であるほか，大学内で開催される学会や研究会の参加者の利用に供されることも多い．

このため，多くの大学でキャンパス内のさまざまな場所に多数の Wi-Fi アクセスポイントが設置されている．大学の許可を得て，あるいは，許可を得ずに教員が研究室や実験室に設置しているものや，上流に携帯電話回線を用いるモバイルルータを教職員や学生が利用していることもあるが，大半はキャンパス情報ネットワークを運用管理する情報系センターなどが整備している Wi-Fi アクセスポイントである．

情報系センターが大学全体に Wi-Fi アクセスポイントを整備するには，5 年から 7 年おきなどのタイミングで行われるキャンパス情報ネットワークの更新に合わせてその一部として整備する方法，臨時に予算措置を行なってまとまった台数を整備する方法，そして，学内からのリクエストに応じて少数台を随時整備する方法などがある．いずれの方法で整備する場合も，九州工業大学の例 [1] のように，学内からの要望を聴取するほか既設の Wi-Fi アクセスポイントの利用状況を調べることが一般的である．

一方，Wi-Fi アクセスポイントの利用状況のデータは，キャンパス内の混雑状況の可視化 [2] や都市の群衆密度の視覚化 [3] など，群衆分布の分析にも有効である．COVID-19 のパンデミックに際して混雑した場所の特定に応用された例 [4] もある．

このような Wi-Fi アクセスポイントの利用状況の把握に際して欠かせないのは，機器ごとの接続クライアント数や接続ユーザ数の情報である．接続数の数ヶ月から 1 年の長期的な推移を見ることで増設や撤去を判断できるほか，時間帯による多寡を見ることで Wi-Fi アクセスポイント設置場所が混雑している時間帯を知ることができる．マサチューセッツ工科大学（MIT）キャンパスにおける Wi-Fi 利用状況をリアルタイムに可視化した iSPOTS プロジェクト [5] のように，Wi-Fi

アクセスポイントの利用状況データそのものを公開や可視化の対象とすることができれば，施設の開放時間の検討，緊急事態に際して必要となる建物内の人数推定，食堂の混雑状況の把握などへの応用が可能となる．

この利用状況データである接続クライアントや接続ユーザの情報は，Wi-Fi アクセスポイントの管理画面にアクセスしたり，ログを参照することで得ることができる．しかし，セキュリティを確保しプライバシーを保護するため，Wi-Fi アクセスポイントの管理画面へのアクセスやログの参照が許されるのは，保守業務を委託された業者や障害対応を行う製造メーカーのほか，情報系センターや情報システム部門の限られたスタッフのみとされることが一般的である．MIT の iSPOTS プロジェクトのように研究組織と情報システム部門が連携できる大学は少なく，人員が限られている情報系センターや情報システム部門が自ら行う研究開発には限界がある．Wi-Fi アクセスポイントの利用状況データは，さまざまな応用が考えられる有用なデータであり，その活用や分析により新たなサービスの創出や新たな分野への展開も期待できるにも関わらず，ほとんど活用できていない．

そこでわれわれは，情報系センターや情報システム部門以外の学生や研究者が利用できるオープンデータとして Wi-Fi アクセスポイントの利用状況データセットを構築することにした．情報系センターのスタッフであり Wi-Fi アクセスポイントの管理画面にアクセスして情報を参照できる立場にあるわれわれが，接続クライアントや接続ユーザの情報を入手して，匿名化など必要な対策を講じて公開する．この方法であればセキュリティやプライバシーの問題は生じない．Wi-Fi アクセスポイントの利用状況データをオープンデータとして公開することで，学内外の研究者による研究に資することができるほか，データサイエンス教育への応用や多くの学生と教職員の参画による大学 DX の推進を実現できるようになる．

2　Wi-Fi アクセスポイント利用状況データセット

本研究では，Wi-Fi アクセスポイントの利用状況を示すデータとして，対象とするすべての Wi-Fi アクセスポイントに接続しているクライアントとユーザの情報を用いることにした．情報は 10 分おきに収集し，データ収集時点で接続しているクライアントについて，以下の各項目をデータベースに格納する．

- 時刻
- Wi-Fi アクセスポイント名
- Wi-Fi アクセスポイント設置場所
- SSID
- MAC アドレス
- ユーザ ID（IEEE802.1X 認証の場合）

それぞれの項目に基づく集計や，クロス集計を行うことにより，利用の多い時刻や利用の多い Wi-Fi アクセスポイント，さらには，Wi-Fi アクセスポイントごとの利用の推移や特定の機器やユーザの利用動向，特定の Wi-Fi アクセスポイント周辺の混雑状況の把握など，さまざまに活用できる．また，10 分おきにデータを収集するので，例えば 1 時間のデータであれば 6 回分のデータから，1 日のデータであれば 144 回分のデータから重複を省くことで求めることができる．

なお，10 分未満で移動したクライアントや接続を終了したクライアントなど，10 分おきのデータ収集時に接続していないクライアントの情報は得ることができない可能性がある．しかし，10 分を超えるセッションタイムアウト時間が設定されていれば，クライアントが明示的に切断した場合を除いて Wi-Fi アクセスポイントからは接続中と見なされる．本研究で対象としたシステムではセッションタイムアウト時間が 30 分（1,800 秒）となっており，補足できないクライアントはほぼ存在しないと考えている．

2.1　セキュリティの確保

データセットの各項目のうち，Wi-Fi アクセスポイント名，設置建物，そして，SSID について，公開することで攻撃の対象になるなど，情報セキュリティを確保する上で懸念が持たれる可能性がある．しかしこれらの情報は，リモートからの攻撃に悪用される IP アドレスとは関連しておらず，また，設置場所で攻撃するのであればその場で知り得る情報であるので，公開によりセキュリティ上のリスクが高くなることは無い．

図 1　小金井キャンパス内の建物の天井に設置されている Wi-Fi アクセスポイントの例（Wi-Fi アクセスポイント名：AP-1108）．

例えば Wi-Fi アクセスポイント名は，図 1 のように装置に貼られており，設置場所においては容易に視認できるため機密情報とは言えない．また，学生の利用に供するため設置場所の多くが自由に出入りできる場所であり，第三者が Wi-Fi アクセスポイント名と設置場所の情報の紐付けを行うことも可能である．このことから，Wi-Fi アクセスポイント名及びその設置場所を公開することについてセキュリティ上の問題は無いと判断した．なお，建物内の設置位置については，盗難や破壊などに対する物理的セキュリティを確保する観点から，学内向けのサービス等にデータを活用する場合に限って機密保持契約等を締結した上でのみ開示することにし，公開用データには含まない．

SSID についても，隠蔽設定はされておらず，ビーコンの到達範囲の無線クライアントであれば容易に知ることができる情報である．キャンパス敷地外の道路からでも確認でき，秘匿性は無い．特に，利用者向けの tuatnet, eduroam, guestnet の 3 つの SSID については，接続方法を含む詳細な情報をホームページで公開している [6] ことから，そのままの情報をデータセットに含むことについてセキュリティ上の問題は無いと判断した．一方，管理用など一般の利用者が接続できない SSID については，対応ルールは定めずに 1 対 1 で対応させた misc + 数字 2 桁 (misc01, misc02, ...) に置換して公開することにした．

2.2 プライバシーの保護

データセットの各項目のうち MAC アドレスとユーザ ID については，接続クライアントを識別しての集計や利用者の動向を把握する上で不可欠の情報である．しかし，プライバシーの保護の観点から，そのまま公開することは適切ではない．

MAC アドレスは，物理アドレスやハードウェアアドレスとも呼ばれ，ネットワークに接続するインターフェース一つ一つに付された 12 桁の 16 進数で表記される 48 ビットのアドレスで，データの宛先を特定するために用いられる．OUI (Organizationally Unique Identifier) と呼ばれる前半の 6 桁が IEEE により機器メーカーに割り当てられ，後半の 6 桁をメーカーが一意に割り当てている．MAC アドレスを知ることでネットワークインターフェースのメーカを知ることができるため，機器のメーカーについてもある程度類推できる場合が多い．また，MAC アドレスが特定の個人と紐付けられてしまえば，行動追跡が可能となり，場合によっては犯罪に悪用される．このことから，MAC アドレスは秘匿すべき個人情報と言える．

ただし，MAC アドレスは利用者が変更することもできる．自ら任意のアドレスを設定することができるほか，スマートフォンでは自動的にランダムな MAC アドレスを設定する仕組みが標準となっている（iPhone ではプライベート Wi-Fi アドレス，Android ではランダム MAC と呼ぶ）．ランダムに MAC アドレスを設定する場合，MAC アドレスの最初のオクテットの 7 ビット目が 1 の LAA (Locally Administrated MAC Address) の中から設定することになっており，16 進数表記の場合，2 桁目が `2,6,A,E` のいずれかの MAC アドレスがそれに該当する．しかし，すべての利用者が MAC アドレスのランダム化を適切に利用しているとは限らない．スマートフォンであっても，設定をオフにすればランダム化機能は無効になる．このため，本研究では LAA であるかどうかに関わらず，MAC アドレスを匿名化することにした．

また，ユーザ ID は，言うまでもなく個人と密接に関連する情報である．様々なサービスで利用されているほか，メールアドレスの一部となっている場合も多い．たとえ氏名，住所，生年月日などの個人情報と結びつく ID 体系でなかったとしても，情報セキュリティを確保する観点からも，公開することは極めて不適切である．

なお，MAC アドレスとユーザ ID の匿名化にあたっては，集計のため元の情報と 1 対 1 で対応している必要がある．匿名化した値を元に戻す必要はなく，また，桁数を保持する必要はないため，安全性が高く広く用いられている一方向ハッシュ関数 SHA-256 を用いたハッシュ化により MAC アドレスとユーザ ID を匿名化することにした．SHA-256 によるハッシュ化は不可逆であり，総当りによる攻撃は不可能ではないものの，必要な時間を考えれば現実的には解読の可能性は無い．また，異なる入力値から同一のハッシュ値が得られる “衝突” についてもその可能性は極めて低く，事実上あり得ないこととされている．

ただし，SHA-256 でハッシュ化していることを公開するので，知人のユーザ ID や機器の MAC アドレスを知り得た者が自らハッシュ値を求め，求めたハッシュ値に基づいて個人を追跡できてしまう可能性がある．これを防ぐため，任意の不定長の秘密の文字列を MAC アドレスとユーザ ID に付加した後にハッシュ値を求めることにした．ユーザ ID や機器の MAC アドレスを知ったとしても，付加されたこの任意の文字列をハッシュ値から求めることは極めて困難であるため，プライバシーの問題が生じることはない．

表 1　東京農工大学総合情報メディアセンターが運用管理する Wi-Fi アクセスポイントの一覧．すべて Cisco Systems 社製．

型番	数量	アンテナ	規格	最大通信速度	デュアル 5GHz
AIR-CAP2702I-Q-K9	158	内蔵	802.11a/g/n/ac (Wave 1)	1.3 Gbps	
AIR-AP2802I-Q-K9	84	内蔵	802.11a/g/n/ac (Wave 2)	4.68 Gbps	✓
C9105AXI-Q	48	内蔵	802.11a/g/n/ac/ax	1.488 Gbps	
C9115AXE-Q	6	外付	802.11a/g/n/ac/ax	5.38 Gbps	
C9115AXI-Q	55	内蔵	802.11a/g/n/ac/ax	5.38 Gbps	
C9120AXI-9	111	内蔵	802.11a/g/n/ac/ax	5.38 Gbps	✓

3　対象とした Wi-Fi アクセスポイント

Wi-Fi 利用状況データセットは，東京農工大学総合情報メディアセンターが 2 つのキャンパスと 5 箇所の学外拠点に設置して運用管理している Wi-Fi アクセスポイント 462 台を対象として構築した．設置場所ごとの台数は，小金井キャンパスが 248 台，府中キャンパスが 206 台，FM 唐沢山，FM 大谷山，FM 本町が各 2 台，FM 津久井と FM 多摩丘陵が各 1 台である[*1]．

これらの Wi-Fi アクセスポイントの仕様を表 1 に示す．すべての Wi-Fi アクセスポイントに送受信機が 2 組内蔵されており，2.4GHz 帯と 5GHz 帯にそれぞれ対応した送受信機を内蔵している機種と，2.4GHz 帯と 5GHz 帯の双方に対応した送受信機と 5GHz 帯に対応した送受信機を内蔵している機種がある．送受信機が 2 組とも 5GHz に対応している後者の機器を表 1 において “デュアル 5GHz” とした．これらすべての Wi-Fi アクセスポイントを Cisco Catalyst 9800-40 ワイヤレスコントローラ（Cisco IOS XE Software,Version 17.3.8a）を用いて集中管理している．

Cisco Catalyst 9800-40 は，管理できる Wi-Fi アクセスポイントの最大数が 2,000 台，最大クライアント数が 32,000，最大スループットが 40Gbps の Wi-Fi アクセスポイント集中管理装置である．管理下のすべての Wi-Fi アクセスポイントを制御できるほか，すべての Wi-Fi アクセスポイント及び接続クライアントの記録を参照できる．本研究では，この装置にアクセスすることですべての Wi-Fi アクセスポイントに接続しているクライアントの情報を一括して得た．

4　接続クライアント情報の取得

Wi-Fi アクセスポイントを集中管理している場合，管理装置であるワイヤレスコントローラーにアクセスすれば，すべての接続クライアントの情報を得ることができる．しかし，管理装置にアクセスできればすべての情報の参照と設定の変更が可能となるため，管理者以外の第三者によるアクセスは当然許されない．このため，Wi-Fi アクセスポイント利用状況データがこれまで十分に活用されていなかった．本研究では，参照できる権限を持つわれわれが情報を取得し，セキュリティとプライバシーを確保してオープンデータとすることができるデータセットを構築する．

Cisco Catalyst 9800-40 の GUI 管理画面では，Wi-Fi アクセスポイントごとの接続クライアント数上位 10 位と下位 10 位，または SSID ごとの接続クライアント数上位 10 位と下位 10 位をダッシュボード上で確認できる．しかし，それ以外の詳細な情報を知るには，Monitoring > Wireless > Clients と進み，接続クライアントの一覧表示画面にアクセスする必要がある．この画面では，Client MAC Address, IPv4 Address, IPv6 Address, AP Name, SSID, WLAN ID, State, Protocol, User Name, Device Type, Role の情報を参照でき，本研究で構築するデータセットに必要なすべてのデータが含まれている．しかし，ここからデータを抽出するには Excel ファイルに情報をエクスポートした上で，それを加工しなければならない．これを手作業で行うのは現実的ではなく，自動化は簡単ではない．

一方，Cisco Catalyst 9800-40 に Linux や BSD などの UNIX 系 OS が稼働しているマシンや仮想マシンから SSH を用いてアクセスすれば，以下のコマンドで管理下の Wi-Fi アクセスポイントに接続しているすべてのクライアントの情報を表示できる．

```
# show wireless client summary detail
```

このコマンドで得られるクライアントのデータは，MAC Address, SSID, AP Name, State, IP Address, Device-type, VLAN, BSSID, Auth Method, Created, Connected, Protocol, Channel, Width, SGI, NSS, Rate, CAP, Username, Rx packets, Tx packets, Rx bytes, Tx

[*1] FM は Field Museum の略で，演習林や農場の総称．

bytes の情報であり，本研究で構築するデータセットに必要な情報がすべて含まれている．

コマンドの出力は固定長であるため，カラムを指定して本研究で用いる MAC Address, SSID, AP Name, State, Username の値を cut コマンドにより抽出できる．このうち，State はクライアントの状態を示す．得られるデータには接続途中のクライアントも含まれるが，接続が完了したクライアントはこの値が “Run” となるため，State の値を用いることで接続しているクライアントの情報のみを抽出できる．また，IPv4 アドレスと IPv6 アドレスが割り当てられている場合など，クライアントの情報が 2 段で出力される場合もあるが，本研究のデータセットでは IP アドレスの情報は用いないため，1 段目のみを用いることにした．それには，Wi-Fi アクセスポイント名に含まれる文字列「AP-」が含まれる行のみを抽出すればよい．

本研究では, Debian 12.12 をインストールした Linux マシンを管理用ネットワークに接続し，SSH を用いてワイヤレスコントローラに接続し，CUI コマンドによりデータを得て，それを加工してデータベースに格納する．この処理をスクリプト化し，cron を用いて 10 分おきにスクリプトを実行した．用いたスクリプトを右に示す．スクリプト中の username にはコマンド実行権限を有するユーザ名，password にはパスワード，xxx.xxx.xxx.xxx にはワイヤレスコントローラの IP アドレスの実際のものを設定している．salt 変数にはハッシュ化の前に付加する任意の文字列を指定した．この文字列をランダムで長いものにすることで，総当たり攻撃に対する耐性を高めることができる．

スクリプトではまず, expect コマンドを用いて SSH アクセスを行なった後にクライアント情報を得るコマンドを実行する．得られた出力から必要な項目を cut コマンドにより抽出した後，1 段目のみを対象として awk コマンドを用いてハッシュ化を行なって CSV ファイルに出力し，sqlite3 コマンドで SQLite データベースファイル wlan.db に格納する．なお，IEEE802.1X 認証を用いていない SSID に接続するクライアントについてはユーザ名の情報が含まれないため，ユーザ名が含まれる場合のみハッシュ値を求めるようにした．また，データを格納する SQLite データベースファイル wlan.db は以下の SQL 文を用いて空のテーブルを作成してから用いた．

SQL 文

```
CREATE TABLE data (date,time,apname,ssid,mac,user);
```

一般の利用者が接続できない SSID は数が少ないため，sub() 関数を用いて置換する．以下のスクリプトでは, 例として “xxxx”, “yyyy”, “zzzz” の 3 つの SSID をそれぞれ”misc01”, “misc02”, “misc03” に置換する場合を示した．

Wi-Fi 利用状況データセット生成スクリプト

```
#!/bin/sh
USERNAME="username"
PASSWORD="password"
IPADDR="xxx.xxx.xxx.xxx"
date="`date +%Y%m%d`"
hour="`date +%H`"
min="`date +%M`"
CSVFILE=$date$hour$min.csv
salt="D5GYQrOLpUkEzQNoZm7FviQmm5pmtl+zy2Z3FhAs"

expect -c "
set timeout -1
spawn ssh ${USERNAME}@${IPADDR}
expect \"assword:\"
send \"${PASSWORD}\n\"
expect \"#\"
send \"terminal length 0\n\"
expect \"#\"
send \"show wireless client summary detail\n\"
expect \"#\"
send \"exit\n\"
" | \
tr -d '\r' | \
cut -c1-15,16-48,49-81,82-95,304-368 | \
grep "AP-" | \
awk -v date=${date} -v time=${hour}${min} \
    -v salt=${salt} \
 'IF $4 == "Run" { \
  apname=$3; ssid=$2; mac=$1; user=$5; \
  sub("/^xxxx$/","misc01",ssid);\
  sub("/^yyyy$/","misc02",ssid);\
  sub("/^zzzz$/","misc03",ssid);\
  cmd1="echo -n " mac salt "|sha256sum"; \
  cmd1 | getline sha256_1; close(cmd1); \
  split(sha256_1,hash1," "); mac=hash1[1]; \
  if (user != "") { \
   cmd2="echo -n " user salt "|sha256sum"; \
   cmd2 | getline sha256_2; close(cmd2); \
   split(sha256_2,hash2," "); user=hash2[1];\
  } \
  print date "," time "," \
        apname "," ssid  "," mac "," user \
  }'>> $CSVFILE
sqlite3 -separator , wlan.db \
    ".import $CSVFILE data" > /dev/null 2>&1 \
rm -f $CSVFILE
```

SQLite データベースファイルに蓄積されたデータは，SQL 文で処理が可能であるほか，CSV 形式へのエクスポートも容易である．CSV ファイルにエクスポートするには，例えば以下のようにすればよい．

SQLite コマンド +SQL 文

```
$ sqlite3 wlan.db
sqlite> .mode csv
sqlite> .output output.csv
sqlite> select * from data;
```

4.1 設置場所データ

Wi-Fi アクセスポイント設置場所については，キャンパス及び建物までのデータと，建物内の設置位置までを含むデータを，対象者を分けて提供することにしている．このため，接続クライアントのデータとは別のテーブルに格納することにし，公開用のデータについては以下の SQL 文で作成したテーブルに格納した．

SQL 文

```
CREATE TABLE aplist (apname,building);
```

別の SQLite データベースファイルにテーブルが格納されている場合には attach 文で抽出や結合を行うことができるほか，insert 文でデータをコピーして用いることができる．

4.2 データの例

日付と時刻を指定して，データを表示した例を以下に示す．MAC アドレスとユーザ名は 256 ビットのハッシュ値となるため 16 進数 64 桁であるが，ここでは上位 4 桁以外は省略している．

SQL 文

```
sqlite> select * from data
  ...> where date="20250717" and time="1710";
```

出力結果（先頭 6 行のみ示す）

```
20250717|1710|AP-1074|eduroam|2549...|8af0...
20250717|1710|AP-1215|misc01|4d4c...|
20250717|1710|AP-0274|tuatnet|17eb...|825b...
20250717|1710|AP-0146|tuatnet|ffa6...|992c...
20250717|1710|AP-0424|tuatnet|8a95...|5e17...
20250717|1710|AP-0453|tuatnet|849f...|2488...
:
```

2025 年 7 月の一月分のデータを蓄積した設置場所データを含まない SQLite データベースファイルには，5,321,689 件のデータが含まれ，サイズは 925,749,248 バイト（882.86 メガバイト）であった．長期の休暇のある期間ではないため，本学においては 1 ヶ月あたりのサイズとしてはこの程度を想定しておけばよいことになる．xz コマンドで-9 オプションを指定して最大圧縮率で圧縮すると 29,147,392 バイト（27.80 メガバイト）となった．なお，CSV 形式にエクスポートするとサイズは 850,651,531 バイト（811.24 メガバイト）となり，最大圧縮率オプションの xz コマンドで 18,597,476 バイト（17.74 メガバイト）に圧縮できた．データそのものを公開する場合は圧縮した CSV ファイルも有効であると考えられる．

5 データ活用の例

構築した Wi-Fi アクセスポイント利用状況データセットには様々な応用が考えられるが，例として Wi-Fi アクセスポイントの増設の要不要の検討のため，2025 年 7 月の全体の利用状況と Wi-Fi アクセスポイントごとの利用状況を調べたものを以下に示す．

ここでは，2025 年 7 月 1 日 0 時 0 分から 2025 年 7 月 31 日 23 時 50 分まで 10 分おきに取得した接続クライアント情報に基づいて作成されたデータセットを用いた．また，接続クライアントテーブル data と設置場所テーブル aplist は同じ SQLite データベースファイルに格納して用いた．

5.1 全体の利用状況

まず，全体の傾向を確認した．SSID ごとの接続クライアント数及び接続ユーザ数は以下の SQL 文で求めることができる．

SQL 文

```
sqlite> select ssid,count(distinct mac),
   ...> count(distinct user) from data
   ...> group by ssid;
```

出力結果を集計したものを表 2 に示す．

表 2　2025 年 7 月の SSID ごとの接続クライアント数と接続ユーザ数．

SSID	接続クライアント数	接続ユーザ数
tuatnet	19,320	6,461
eduroam	4,264	2,260
guestnet	44	34
misc.. の合計	126	-

接続ユーザ数は IEEE802.1X 認証を用いる tuatnet，eduroam，guestnet の 3 つの SSID の場合のみ有効な項目である．これらの SSID では，クライアント数がユーザ数の 1.3 倍から 3.0 倍となっている．これは，一人が複数台の機器を使用している場合のほか，ラン

ダムに MAC アドレスを割り当てるスマートフォンなど，同じ機器が集計した期間内に異なる MAC アドレスを持つ場合があったためと考えられる．

1 日あたりの接続クライアント数は，重複を除いて MAC アドレスを集計すればよいので以下の SQL 文で求めることができる．

SQL 文

```
sqlite> select date,count(distinct mac)
   ...> from data group by date;
```

この出力に基づいて作成したグラフを図 2 に示す．

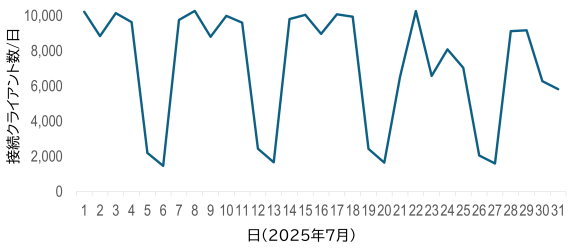

図 2　2025 年 7 月の 1 日あたり接続クライアント数の推移．

図 2 を見ると，授業の開講日程に応じて接続数が増減していることがわかる．例えば，2025 年 7 月 21 日（月）は海の日（祝日）だが授業実施日とされたものの，すべての授業が開講されたわけではない．このため他の月曜日より接続数が少ない．また，2025 年 7 月 23 日（水）以降は 15 回の授業に加えて実施される 16 回目の授業回の期間となり，開講される授業が減少したことから接続数が減少したことがわかる．

5.2　Wi-Fi アクセスポイントごとの利用状況

次に，Wi-Fi アクセスポイントごとの利用状況を明らかにするため，一日あたりの接続クライアント数，及び，ピーク時刻とその時の接続クライアント数を調べた．

一日あたりの接続クライアント数が上位 10 位までの Wi-Fi アクセスポイントは，以下の SQL 文で求めることができる．

SQL 文

```
sqlite> select data.apname,aplist.campus,
   ...> aplist.building,count(*) from data
   ...> inner join aplist
   ...> on data.apname=aplist.apname
   ...> where data.date="20250710"
   ...> group by data.apname
   ...> order by count(*) desc limit 10;
```

出力結果

```
AP-0091|小金井|13 号館|3723
AP-1057|小金井|12 号館|2895
AP-0093|小金井|13 号館|2802
AP-1080|小金井|小金井図書館|2524
AP-1081|小金井|小金井図書館|2484
AP-0229|小金井|総合会館|2460
AP-1009|府中|新 4 号館|2295
AP-1202|府中|邂逅館|2273
AP-0443|小金井|講義棟|2197
AP-1077|小金井|小金井図書館|2115
```

授業が開講されている大教室の Wi-Fi アクセスポイントと図書館の Wi-Fi アクセスポイントに多くクライアントが接続していることがわかる．その他の平日も調べたところ，同様の傾向であった．多くのクライアントが接続している場所に Wi-Fi アクセスポイントを追加すれば，利用が平準化されユーザ体感が向上する．このため，大教室で多くの利用がある小金井 13 号館と小金井 12 号館に Wi-Fi アクセスポイントを追加することが望ましいと言える．

一方，ピーク時刻とその時の接続クライアント数は，以下の SQL 文で求めることができる．

SQL 文

```
sqlite> select * from (select data.apname,
   ...> data.time,aplist.campus,
   ...> aplist.building,count(*),row_number()
   ...> over (partition by data.apname
   ...> order by count(*) desc) as rank
   ...> from data inner join aplist
   ...> on data.apname=aplist.apname
   ...> where date="20250710"
   ...> group by data.time,data.apname
   ...> order by count(*) desc) as ranking
   ...> where ranking.rank=1 limit 10;
```

出力結果

```
AP-0155|1210|府中|福利厚生センター|152|1
AP-0229|1210|小金井|総合会館|128|1
AP-0231|1210|小金井|総合会館|122|1
AP-0095|1620|小金井|【図書館前】|121|1
AP-1057|1510|小金井|12 号館|119|1
AP-0093|1420|小金井|13 号館|116|1
AP-0226|1220|小金井|総合会館|113|1
AP-0443|1320|小金井|講義棟|109|1
AP-0091|1020|小金井|13 号館|108|1
AP-0227|1220|小金井|総合会館|101|1
```

出力結果の右端のカラムは時刻別接続クライアント数の当該 Wi-Fi アクセスポイントにおける順位となっている．各 Wi-Fi アクセスポイントについて接続数

が 1 位の時刻を抽出し，接続クライアントが多い順に並べているためすべて “1” となった．SQL 文中の where ranking.rank=1 を削除すれば Wi-Fi アクセスポイントの重複を許して時刻別接続クライアントの多い順に集計することもできる．

学生食堂がある府中福利厚生センターと小金井総合会館の Wi-Fi アクセスポイントにおいて，昼食時間帯に接続クライアント数が最大となった．そのほか，4 時限終了後に図書館前の屋外に設置されている Wi-Fi アクセスポイントに多くのクライアントが接続し，大講義室では授業中に接続クライアント数が最大になっていることもわかった．大講義室については一日あたりの接続クライアント数の観点からも増設が望ましい．また，ピーク時の接続数を見ると，学生食堂がある建物への増設も検討した方がよい．これらを踏まえ，曜日による差異や数ヶ月間の動向をさらに確認した上で，増設の必要性を判断することにしている．

併せてわれわれは，接続クライアントが少ない Wi-Fi アクセスポイントを移設することによる配置の最適化についても検討している．このために，例えば以下の SQL 文を用いて接続クライアントが存在しない Wi-Fi アクセスポイントを確認している．

SQL 文

```
sqlite> select * from aplist where apname not in
   ...> (select apname from data
   ...> where data.date="20250710"
   ...> group by apname) order by apname;
```

6 まとめと今後の予定

東京農工大学総合情報メディアセンターが運用管理する Wi-Fi アクセスポイント 462 台の利用状況データセットを，公開可能な SQLite データベースファイルとして構築した．Wi-Fi アクセスポイントを集中管理するワイヤレスコントローラから 10 分おきに接続クライアントの情報を入手し，MAC アドレスと接続ユーザ ID については，SHA256 によりハッシュ化することで匿名化してプライバシーを確保している．このデータの取得と加工及び SQLite データベースファイルへの格納をシェルスクリプトと cron により自動化した．すでに東京農工大学では，構築したデータセットに基づいて Wi-Fi アクセスポイントの増設と移設の検討を行っており，有用な知見を得つつある．

今後は，構築した Wi-Fi アクセスポイント利用状況データセットの SQLite データベースファイル及び CSV ファイルを，設置場所データと合わせてホームページ上に公開する予定である．これにより，データサイエンス教育の教材となるほか，学生による新たなアプリケーション開発などにつながることを期待している．また，自らデータを活用することで，将来の利用がどのようになるのかの予測や混雑場所の可視化などを行いたいと考えている．さらには，データそのものの公開に止まらず，外部のアプリケーションから利用できる API を開発して公開し，より広く活用されることを目指したいと思う．

参考文献

[1] 福田 豊, 中村 豊, 佐藤 彰洋, 和田 数字郎, 岩崎 宣仁, “無線 LAN 利用状況調査に基づいて策定した改善策の検証”, 情報処理学会論文誌デジタルプラクティス, Vol.3, No.3, pp.1–9, 2022.

[2] A. Binthaisong, J. Srichan and S. Phithakkitnukoon, “Wi-Crowd: Sensing and visualizing crowd on campus using Wi-Fi access point data,” Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, pp.441–447, 2017.

[3] T. Wiangwiset, C. Surawanitkun, W.Wongsinlatam, T. Remsungnen, A. Siritaratiwat, C. Srichan, P. Thepparat, W. Bunsuk, A. Kaewchan and A. Namvong, “Design and Implementation of a Real-Time Crowd Monitoring System Based on Public Wi-Fi Infrastructure: A Case Study on the Sri Chiang Mai Smart City,” Smart Cities, Vol.6, No.2, pp.987–1008, 2023.

[4] M. Ribeiro, D. Teixeira, P. Barbosa and N. J. Nunes, “Using passive Wi-Fi for community crowd sensing during the COVID-19 pandemic,” Journal of Big Data, Vol.10, No.7, pp.1–22, 2023.

[5] A. Sevtsuk, S. Huang, F. Calabrese and C. Ratti, “Mapping the MIT campus in real time using WiFi,” Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City, pp.326–338, 2009.

[6] 東京農工大学総合情報メディアセンター, “無線 LAN 一覧”, https://www.imc.tuat.ac.jp/info-system0/campusnet/tuatnet.html, 2025 年 9 月 17 日閲覧.