Vol.2013-IOT-23 No.7
2013/9/27

学内無線LANシステムにおける利用統計データの分析とその課題

杉木 章義[1,a)]　佐藤 聡[1]　和田 耕一[1]

概要：教育や研究環境の充実を目的として，多くの大学で学内無線 LAN システムが導入されている．大規模データ解析の手法を学内無線 LAN の統計データに応用すれば，従来の手法では見つけることのできなかった潜在的な問題を発見し，また戦略的なネットワークの配備計画が可能になることから，利用者から見た学内ネットワークの品質を大幅に向上させることができると期待される．本論文では，学内無線 LAN システムに関する利用統計データを蓄積し，対話的に解析を行うためのシステムについて提案する．また，いくつかの解析事例について紹介し，その際に課題となる利用者のプライバシーやセキュリティに関する配慮などについても議論する．

1. はじめに

近年，研究や教育環境の充実を目的として，数多くの大学で学内無線 LAN システムが導入されている．この学内無線 LAN システムは，e ラーニングを通じた学習や，電子ジャーナルを活用した文献調査など，さまざまな目的に利用されている．さらに最近では，従来からのノート PC に限らず，スマートフォンやタブレットなどのさまざまな端末が学内無線 LAN システムに接続され，利用されるようになっており，大学における学内無線 LAN システムの重要性がますます高まっている．

一方で，ネットワーク分野だけに限らず，大規模データ解析によって得られる知識や情報を活用して，生活環境を豊かにしていくことが広く模索されている．大規模データ解析では，大量のデータを取得し，解析することによる知識の獲得と，その知識を活用した応用の 2 つの段階からなっている．学内無線 LAN システムに対しても大規模データ解析を適用してくことで，利用者から見た品質を飛躍的に高め，情報環境を中心としたキャンパス環境を充実していくことができる可能性がある．

本論文では，学内無線 LAN システムの利用統計データを活用したネットワークの潜在的な問題の発見や，整備方針決定の可能性について探索する．最初の試みとして，本論文では，データ解析のためのプロトタイプシステムを構築し，その結果得られた事例について紹介する．

[1] 筑波大学システム情報系（学術情報メディアセンター）
University of Tsukuba, Tsukuba, Ibaraki 305–8573, Japan
[a)] sugiki@cc.tsukuba.ac.jp

運用における利用統計データの活用に関しては，利用者のプライバシーやセキュリティに十分配慮し，各種ガイドラインや国内法との関係についても慎重に検討を進めていく必要がある．その点について，論文の後半で議論を行う．

本論文の構成を次に示す．まず，第 2 章で本研究の動機と従来からの統計データ管理方法との違いについて説明する．次に，第 3 章で本研究で作成したプロトタイプシステムについて紹介する．第 4 章で，活用事例をいくつか紹介し，第 5 章で関連研究について述べ，最後に第 6 章で本論文をまとめる．

2. 従来の管理方法の問題点

大学キャンパスにおける情報系センターの運用では，従来からログを取得し，運用に用いることが広く行われており，その点では変わらないと言える．大きく異なるのは，解析しやすい形でデータを蓄積し，対話的に問い合わせを行うための仕組みにあると言える．

本章では，従来の運用方法における問題点を指摘することで，本研究の目指すところを明らかにする．

- **ログの点在：** 従来よりネットワーク機器が出力するログは一定期間保存され，運用状況のモニタリングや障害対応などに活用されていたが，それぞれの機器が出力するログのフォーマットは異なっており，また複数のストレージに分散して格納されることも多かった．多くの機器は syslog や SNMP（Simple Network Management Protocol）などの標準規格での出力が可能であるが，その詳細なフォーマットは機器によって異なっている．そのため，統計的データが容易に活用

ⓒ 2013 Information Processing Society of Japan

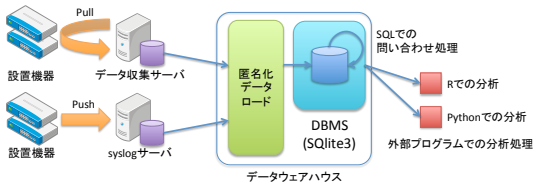


図 1 システム構成
Fig. 1 System Overview

できない状態になっていたと言える．

- **アドホックな解析：** 上記のデータは，運用のモニタリングに必要なものを除き，障害などの問題の発生時などに初めて参照されることが多かった．多くのログは人手によって解析が行われ，またその場で記述した簡単なスクリプトが利用されることも多く，解析手法が限定されていたと言える．
- **有機的なログの結合：** 多くの作業は単一のログの解析で完結しており，また解析の困難さから，複数のログを参照して問題を特定する場面も非常に限られていた．典型的な例として，認証ログを参照しながら，その利用者が行った通信をファイアウォールログで分析するなどの場面があるが，これらのログの結合作業は主に人手で行われており，困難さが伴っていた．

以上の問題を改善することによって，利用者からの障害やサービス品質に関する問い合わせに迅速に回答し，また，事前に潜在的な問題を排除してくことが可能となることから，運用の品質を大幅に向上させることができると期待される．

3. データウェアハウス化

3.1 概要

本論文のシステム構成を**図 1** に示す．本システムでは各機器が出力するログを整形し，共通の場所にデータを蓄積することで，容易な分析を可能とする．また，蓄積されたデータに対して対話的な問い合わせを可能とすることで，繰り返し分析を可能にする．本システムは，データベースの分野におけるデータウェアハウスそのものであり，著者らもそれを意識して構築している．

3.2 プロトタイプ

本節では，本論文のプロトタイプシステムについて説明する．このシステムは恒久的なものではなく，今後も蓄積するデータの量や運用上の必要に応じて，変更していく予定である．

データの蓄積にあたっては，一旦，共通のフォーマットとして CSV 形式に加工した後，SQLite にデータをロードする．現在，主に蓄積しているのは端末と AP（Access Point）間のコネクション品質に関するデータであり，データベースの大きさが約 2.72 GB と SQLite で十分対応できるデータ量である．



地図データ ©2013 Google. ZENRIN

図 2 筑波キャンパスにおける AP 配置
Fig. 2 Wireless APs in Tsukuba Campus.

分析に関しては，SQL による問い合わせで十分対応できるケース，SQL でデータを出力し，Python で分析を行うケース，Python で中間加工を実施し，R を使用して分析を行うケースの 3 つがある．

データのロードにあたっては，慎重に匿名化を実施している．基本的には，一方向ハッシュ関数を使用して実施する．匿名化を実施しているのは，ユーザ ID と MAC アドレスである．ユーザ ID に関しては，現在では学生や教職員などの属性情報も省略し，匿名化を実施している．一方の MAC アドレスに関しては，品質の分析上，必要なベンダー情報のみを記録し，MAC アドレス自身は匿名化している．ハッシュ関数としては，現在は SHA1 を使用している．

4. 蓄積データ

4.1 データの取得環境

データ取得環境となる筑波大学では，学群に 9,853 人，大学院に 6,687 人の学生が在籍している．一方，教職員は 4,508 人である（2013 年 5 月 1 日現在）．このほぼ全てが，学内無線 LAN の利用対象者である．筑波大学では，キャンパスとして，つくば市内にある約 256 ha の筑波キャンパスと，約 1.2 ha の東京キャンパス文京校舎がある．特に，筑波キャンパスは南北 4 km，東西 1 km にわたって分散しており，それに対応するように学内無線 LAN の AP も配置されている（**図 2**）．AP は 6 月 15 日の時点で 475 台である．データの蓄積は継続的に実施しているが，本論文の分析では，このうち 3 月 31 日から 8 月 31 日の約 5 ヶ月分のデータを分析に使用した．

4.2 データの取得機器

本研究では，Cisco 5500 シリーズ（バージョン 7.0）3 台から，無線 LAN のコネクション品質に関するデータを取得して，実験を行った．

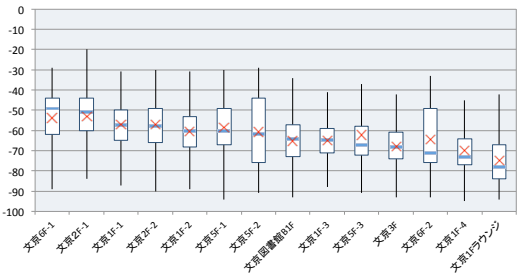


図 3 AP ごとの RSSI 品質統計データ
Fig. 3 RSSI distribution collected from APs.

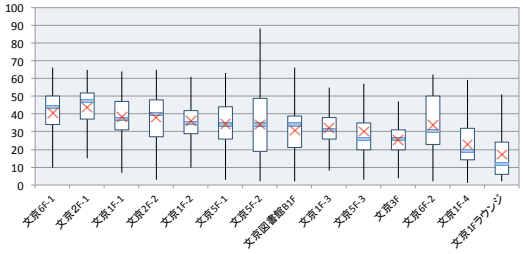


図 4 AP ごとの SNR 品質統計データ
Fig. 4 SNR distribution collected from APs.

本研究で，蓄積しているログは主に下記の 2 つである．

- **clients ログ**
  匿名化を実施した無線 LAN に接続した端末 MAC アドレス，接続中の AP，SSID などの概要を記録したログである．10 分おきに取得し，期間全体の総取得回数は 22,196 回，総エントリ数は 11,237,605 件である．
- **detail ログ**
  匿名化を実施した無線 LAN に接続したユーザ，認証や暗号化の方式，RSSI（Received Signal Strength Indicator）や SNR（Signal-to-Noise Ratio）などの電波状況などの詳細な情報を記録したログである．各端末ごとに HTTP 要求を生成する必要があるため，負荷軽減から 30 分おきに取得し，総取得回数は 7,399 回，総エントリ数は 3,781,376 件である．

上記のサンプリング間隔は，経験的に決定し，運用に影響を与えないよう保守的な値に設定している．

4.3 RSSI/SNR データの活用

東京キャンパスにおいて，複数の利用者から頻繁にコネクションが切断されるという報告が度々寄せられていた．無線 LAN では，外来波や利用者の端末の状況などさまざまな要素がコネクション品質に影響することから，有線 LAN に比べて原因の特定が難しく，一般的なトラブルシュートの説明に留まっていた．

東京キャンパスの RSSI のグラフを**図 3**，SNR のグラフを**図 4** に示す．両者のグラフは箱ひげ図で分布を表しており，×印は平均値を表している．**図 5** の大学全体の品質の分布と比較すると，東京キャンパスの各 AP はむしろ品質の高い方に位置していると言える．図 5 は中央値の分布のみを示しており，RSSI と SNR で個別に AP のランク付けを行い，ソートした結果をプロットしている．また，データの再送や重複に関しても分析を行ったが，こちらもキャンパス平均以上の品質であった．以上から，RSSI や SNR に関連した影響を棄却することができた．

次に，別の観点から AP ごとの同時接続数について調査を行った．同時接続数は直接記録されていなかったが，AP が付与する端末番号の最大値を見ることで間接的に知ることができた．その結果，東京キャンパスの図書館の地下 1 階で度々，上限の 200 台まで払い出していることが確認された．状況について確認したところ，このフロアでは，定期的にセミナーを開催しており，時間帯により接続しにくいことが確認された．

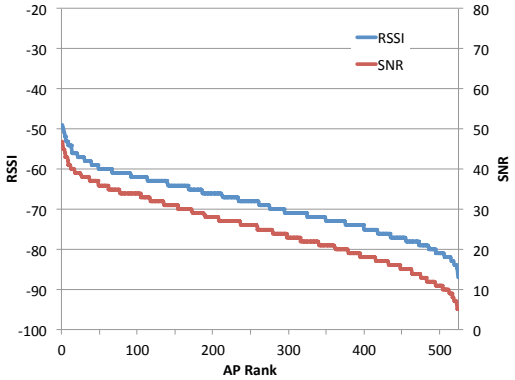


図 5 AP ごとの品質統計データ
Fig. 5 Sorted RSSI and SNR distribution for all APs.

今回の分析が従来のものと異なるのが，収集しているデータの規模と，問い合わせの繰り返しによる問題の分析である．収集しているデータに関しては，各端末が AP に報告する RSSI や SNR を直接収集し，蓄積しているため，一般的なサイトサーベイによる方法以上に実態に近いデータが収集できていることが期待される．また，データの分析に関しては，SQL を何度も記述しながら分析を進めることで，徐々に分析の精度を高めている．以上から，従来以上のデータ分析結果をもとに利用者からの問い合わせに回答できていることが期待される．

次に潜在的な問題を把握するため，各端末から見た場合の RSSI と SNR の中央値の分析を行った．その結果を，**図 6** に示す．これらの数値が低いと，どの AP を利用しても利用者が繋がりにくいと感じるようになり，大きく満足度が低下することが予想される．図 6 を見ると全般に良好な品質を維持しているが，RSSI が-80 db 以下，SNR が 20 以下の端末については注意が必要である．この原因として，相対的に品質が低い AP ばかりを利用しているか，端末自身に何らかの問題を抱えていることが予想される．

ⓒ 2013 Information Processing Society of Japan

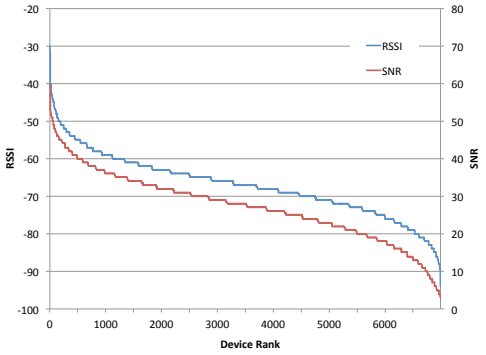


図 6 端末から見た RSSI 品質統計データ
Fig. 6 RSSI and SNR distribution seen from each device.

4.4 共通して利用されている AP の抽出

よりデータマイニングらしい手法として，各端末が利用した AP の組み合わせに対してアソシエーション分析を行った．本手法を適用すると，一体的に共通して利用されている AP のグループを抽出することができ，今後の AP の配備計画に役立てることができる．R の arules パッケージに含まれている Apriori アルゴリズムを使用し，support を 0.10 以上，confidence を 0.90 以上，ルールの最大長を 3 に設定し，分析を行った．また，訪問した AP に対して直接アソシエーション分析を行うと，ルールが複雑になりすぎることから，建物ごとにまとめて分析することとした．その結果，568 個のルールが生成された．

ルールを分析すると，全般に中央図書館を含むルールが多く，ルールの左側（LHS）に含むものは 23 個，ルールの右側（RHS）に含むものは 242 個ある．以上から，中央図書館に設置された AP がよく利用されていることが分かる．学生向け施設であるステューデントプラザの AP もよく活用されており，LHS にステューデントプラザを含むものが 60 個ある．興味深い例として，大学会館の AP はよく利用されているが，下記のように体育関連施設と結びついたルールがよく見られる．

```
LHS            RHS            supp conf lift
{合宿所} => {大学会館}    0.10 0.90 2.22
{体育風洞} => {大学会館} 0.13 0.90 2.22
```

これは体育芸術エリアと大学会館が近接しているため，体育専門学群の学生が利用していることも考えられるが，体育の授業のオリエンテーション会場が大学会館となっており，一般の学生も授業の前後で利用していることが推測される．

その他，全般には各専門分野のエリアごとにルールが生成されいる．筑波大学では，専門分野ごとに大まかに第 1 エリア，第 2 エリア，第 3 エリアなどのように分かれているが，エリア内の建物で完結するルールが多い．ただし，食堂に設置された AP に関しては，多少エリアをまたいで利用されている．

5. 関連研究

スマートフォンやタブレット型端末などの持ち込み端末の普及により，大学キャンパスのネットワークを取り巻く状況は無線 LAN システムを中心に大きく変化している．

近年の海外における大学キャンパスの無線 LAN の状況については，Chen ら [1] や Gember ら [2] によって報告されている．Chen らはコネチカット大学における Wi-Fi の利用状況を 3 日間，パケットキャプチャにより取得し，TCP 層および HTTP 層での解析を行っている．Gember らは，2 つのキャンパスを対象に HTTP 層を中心により上位の層に着目し，比較を行っている．

日本国内の大学キャンパスの事情については，佐藤ら [3] による学習院大学での事例報告があり，本研究の先行研究 [4] においても，筑波大学の事例を報告している．スマートフォンやタブレットなどの急速な普及に対応して，ネットワークの品質を適切に維持してくために，今後も調査や情報共有を進めていく必要がある．

一般的なスマートフォンの振る舞いについては，Falaki ら [5] や，Maier ら [6]，Afanasyev ら [7] の研究が知られている．スマートフォンの普及はまだまだ始まった段階にあり，今後も継続的に研究を進めていく必要がある．

また，大学キャンパスにおける無線 LAN の研究は，大学への普及が始まった 2000 年代前半に集中的に実施されており，Henderson ら [8]，Shwab ら [9]，Hernández-Campos ら [10]，Balazinska ら [11] の報告がある．当時と無線 LAN を取り巻く状況は大きく変化しており，改めて研究を実施していく必要がある．

6. まとめと今後の課題

本論文では，大学の情報系センターにおけるデータウェアハウス導入の可能性について議論した．従来より，情報系センターでは必要なログが一定期間保存され，適切に管理されてきたと言えるが，その情報の活用は，障害時等の問題解決など一部の場面に限られてきた．また，その場で記述したスクリプトによる解析など，アドホックな利用であることも多く，また解析作業の困難さから，複数のログの情報を結びつけ，有機的に活用してきたとは言えなかった．データベース分野におけるいわゆるデータウェアハウスのようなシステムを導入することで，運用水準の向上に大いに役立てることができる．本研究では，そのためのプロトタイプシステムを構築し，運用水準の向上におけるいくつかの事例を紹介した．今後，OLAP（OnLine Analytical Processing）などのより本格的なシステムを導入することで，利用者の期待により応えることができると期待される．

一方で，無線 LAN システムの利用状況に関するデータは利用者の振る舞いに近く，利用者のプライバシーやセキュ

ⓒ 2013 Information Processing Society of Japan

リティに十分配慮する必要がある．また，諸外国と比較して，日本は通信の保護に関する法律やガイドラインが整備されており，その観点からも活用方法に関して慎重な検討が必要である．単にデータを匿名化するだけではなく，学内における各種ガイドラインの策定や利用規約の整備も含めて今後，研究を進めていく予定である．

謝辞および免責事項

本研究の一部は，筑波大学学術情報メディアセンターの研究開発経費で実施した．また，日々の運用を支えているセンターのスタッフに感謝する．さらには，本研究の成果は，研究者としての著者らの個人的な見解を述べたものであり，所属機関を代表する意見ではないことに注意が必要である．

参考文献

[1] Chen, X., Jin, R., Suh, K., Wang, B. and Wei, W.: Network Performance of Smart Mobile Handhelds in a University Campus WiFi Network, *IMC'12*, pp. 315–328 (2012).
[2] Gember, A., Anand, A. and Akella, A.: A Comparative Study of Handheld and Non-handheld Traffic in Campus Wi-Fi Networks, *PAM'11*, pp. 173–183 (2011).
[3] 佐藤　真，村上登志男，磯上貞雄，城所弘泰，久保山哲二：キャンパス内の無線 LAN 利用動向分析，情報処理学会研究報告（IOT），2013-IOT-22(3)，pp. 1–5 (2013).
[4] 杉木章義，佐藤　聡，和田耕一：Mobile Devices in the Wild (Campus), 情報処理学会研究報告（OS），2013-OS-126(21)，pp. 1–8 (2013).
[5] Falaki, H., Mahajan, R., Kandula, S., Lymberopoulos, D., Govindan, R. and Estrin, D.: Diversity in Smartphone Usage, *MobiSys'10*, pp. 179–194 (2010).
[6] Maier, G., Schneider, F. and Feldmann, A.: A First Look at Mobile Hand-Held Device Traffic, *PAM'10*, pp. 161–170 (2010).
[7] Afanasyev, M., Chen, T., Voelker, G. M. and Snoeren, A. C.: Usage Patterns in an Urban WiFi Network, *IEEE/ACM Trans. on Networking*, Vol. 18, No. 5, pp. 1359–1372 (2010).
[8] Henderson, T., Kotz, D. and Abyzov, I.: The Changing Usage of a Mature Campus-Wide Wireless Network, *Computer Networks*, Vol. 52, No. 14, pp. 2690–2712 (2008).
[9] Schwab, D. and Bunt, R.: Characterising the Use of a Campus Wireless Network, *INFOCOM'04*, pp. 862–870 (vol. 2) (2004).
[10] Hernández-Campos, F. and Papadopouli, M.: A Comparative Measurement Study of the Workload of Wireless Access Points in Campus Networks, *PIMRC'05*, pp. 1776–1780 (2005).
[11] Balazinska, M. and Castro, P.: Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network, *MobiSys'03*, pp. 303–316 (2003).

ⓒ 2013 Information Processing Society of Japan