Vol.2014-IOT-27 No.15
Vol.2014-SPT-11 No.15
2014/10/10

学内LAN利用ログの分析と応用

宮下 健輔[1,a)]

概要：著者の所属する大学では昨年度から無線 LAN 利用ログを蓄積しているが，これは単にログをそのままの形で蓄積するだけのものであった．今回，収集対象を有線 LAN にも広げて学内 LAN 全体とし，蓄積されたログを取得しやすくすることを念頭に蓄積方法について検討を行ったので報告する．さらにその分析結果の応用について，可能性も含めて議論する．

A Study of Usage Logs in Campus Networks

Kensuke Miyashita[1,a)]

Abstract: The university to where the author belongs has had a storage of wi-fi usage logs since last year and the logs have been stored in raw format. Now the logs of wired LAN usage are also stored in the proposed data warehouse. The author reports a study of the way for storing the logs into the data warehouse that is intended to be easy to retrieve contents and analyze them. In addition, the author discusses the application of analysis.

1. はじめに

ラップトップPCと携帯電話に始まったインターネット接続可能な可搬型情報機器は四半世紀に亙って進化と普及を続け，今や多くの人々がスマートフォンやタブレット型PC，ラップトップPC等を持ち歩くようになっている．BYOD（Bring Your Own Device）という言葉が定義されたのは2004年のこと [1] であり，一般企業等にそのような方針が浸透しはじめたのは最近のことである．しかし著者の経験から，多くの大学では1990年代からBYODが始まり，教員や学生が主にラップトップPCを学内ネットワークに接続して研究や教育，学習に利用してきたと考えられる．

海外における大学の学内無線LANの利用動向調査に文献 [2] や [3] 等がある．これらは学内無線LANについてそれぞれの手法で利用動向を調査し，前者では以前に比べて利用されるアプリケーションや通信形態に大きな変化があることが発見され，後者は可搬型情報機器と据置型情報機器での通信の差異についてまとめている．また，文献 [4] では251の大学の約160万人の学生に対して情報機器に関する調査を行っている．この調査の主要な結果として，学生が自分の可搬型情報機器を学習のために利用したいと考えていることや，教育機関に対してそのような機会を設けてもらいたいと考えていること等が挙げられている．これらの結果から，学内LANにおける無線LANの利用は今後ますます盛んになっていくことが予想できる．

一方，近年の国内大学における無線LANの利用動向について，文献 [5]，[6]，[7] 等での調査が行われている．これらはそれぞれ学習院大学，筑波大学，京都女子大学での学内無線LANの利用動向をそれぞれの手法で調査したものである．このような調査の手法には海外と国内の区別なく，まずは無線LAN機器の出力するログ（記録または履歴）を分析することから始まっている．ネット接続可能な可搬型情報機器が大量に学内へ持ち込まれることに対応するためにはこのような利用動向の調査が必要不可欠であり，そのような情報を大学間で共有することが望まれるところである．

本稿では，まず一般的な学内LAN利用ログの性格や取

[1] 京都女子大学現代社会学部
Faculty for the Study of Contemporary Society, Kyoto Women's University
[a)] miyasita@cs.kyoto-wu.ac.jp

り扱い方について考えを述べた後，著者の所属する大学で実際に取得しているログを紹介し，最後にこれをもっと利用しやすくするシステムについて提案する．

2. 学内 LAN 利用ログ

株式会社日本レジストリサービス（JPRS）の統計情報[*1]によれば，大学など高等教育機関による AC.JP ドメインの登録数は 3538 件（2014 年 9 月 1 日）と国内の高等教育機関数を遙かに超えている．大学数の増加と 18 歳人口の減少への対策として多くの大学が WWW や SNS 等を利用した広報に力を入れている昨今，インターネットに接続せずドメイン名も登録していない大学は皆無ではないかと思われる．

インターネットに接続している大学では，たとえそれが広報戦略のみの目的であったとしてもごく小規模な学内 LAN が敷設されると思われる．現在では多くの大学が学生サービスの一環として学内の情報設備やインターネット接続環境の充実を謳っており，学生の利用できる学内 LAN を持たない大学が存在するとは考え難い．

そのネットワークをどんな機関が運用しているかに関わらず，ネットワークを利用すれば何かしらのログが残るものである．ユーザが機器を接続するメディアが無線であれば無線アクセスポイント（AP）または AP を統括しているコントローラに，有線メディアであればエッジスイッチにポート利用のログが残る．その接続機器が DHCP を利用して IP アドレスを取得すれば DHCP サーバにログが残り，さらに利用開始に何らかの認証が必要であれば認証サーバにログが残る．その後も利用終了まで，WWW やメールに代表されるサービスを利用する度にそれらのサーバやファイアウォール等にログが残る．本稿では大学の設備に残されるこれらのログをまとめて学内 LAN 利用ログまたは単にログと呼ぶ．

1998 年に米国で制定されたデジタル・ミレニアム著作権法（DMCA）を受け，我が国では 2002 年に「特定電気通信役務提供者の損害賠償責任の制限及び発信者情報の開示に関する法律」（いわゆるプロバイダ責任制限法）が施行された．この法によりインターネット上で発生した主に権利や社会的法益の侵害による賠償請求等に対して大学がもつ責任が明確化され，保全すべきログの種類や期間もこれを拠り所として検討された．また，文献 [8] で言及されているようなデジタル・フォレンジック確立のためにもログの保全がまず基本となる．さらに，大学が従来から保有する学生原簿や成績，就職先情報等の電子ファイル化が進み，これらは「個人情報の保護に関する法律」（2003 年に一部施行の後 2005 年に全面施行となった，いわゆる個人情報保護法）により個人情報データベースと呼ばれ，時代に即してより適切に取り扱われるべき対象となった．

著者の所属する大学（以下，本学という）では学生番号や氏名から容易に推測できる文字列をユーザ名として用いており，学内 LAN 利用ログにもそれが残るようになっている．すなわち，学内 LAN 利用ログを大学が保有する他の個人情報データベースと突合したり他のログと突合したりすることにより個人を特定できることになるので，これらログも慎重に取り扱わねばならない．

前述の文献にも見られるように，学内 LAN 利用ログを統計的に分析することによりネットワークの利用動向を明らかにしたり，運用改善の用に供したりすることは一般的に行われている．例えば利用の集中している箇所を特定してネットワーク機器を増設したり，ユーザからの障害やサービス品質の問い合せに対して素早く対応したりするためにログが利用されている．このときログから個人を特定できるような状況のまま分析を行う必要はなく[*2]，事前に匿名化処理をすることが適切である．しかし完全に匿名化してしまうと分析に支障が生じることとなる．例えば同一のユーザに関する情報を複数のログ間で突合することが不能となれば，学内でのそのユーザの利用状況をネットワーク機器を越えて追跡すること等が不可能となってしまう．ただし，「そのユーザが誰であるか」という情報を保存する必要はなく，例えばログ A のエントリ a とログ B のエントリ b は同一ユーザのものであるということが保証できればよい．そのため，このような目的のために学内 LAN 利用ログを活用する場合は，これを識別非特定情報に変換してから扱うのがよかろう．この「識別非特定情報」とは文献 [9] で定義されている概念で，「一人ひとりは識別されるが、個人が特定されない状態の情報（それが誰か一人の情報であることがわかるが、その一人が誰であるかまではわからない情報）」（[9] より引用）のことであり，この変換のことを「非特定化」と呼ぶ．つまり，学内 LAN 利用ログにより利用動向等を分析する際には，前もって非特定化処理を行うべきである．

3. 実際の学内 LAN 利用ログ

本学の学内 LAN には有線 LAN と無線 LAN が混在しており，それぞれについてログの集積場所や収集方法が異なる箇所とそれぞれに共通する箇所がある．

[*1] http://jprs.jp/about/stats/registered/

[*2] もちろん，特定のユーザの訴える不具合等に対処する場合にはログからそのユーザに該当する箇所だけを抽出する必要があろう．しかし学内 LAN の全体傾向としての利用動向調査や局所的な通信の集中等を発見するためには個人を特定することは不要である．ただし場合によってはユーザの属性（学生かどうか等）が必要となる場合がある．

3.1 有線 LAN

有線 LAN はエッジスイッチに Apresia 社製品を利用している．このスイッチの各ポートと壁や床の情報コンセント（RJ-45 モジュラジャック）が接続されており，ユーザはこれら情報コンセントに直接各自の情報機器を接続したり，情報コンセントに接続されたスイッチや無線 AP に情報機器を接続したりする．このスイッチや無線 AP は大学が設置したものではなく，例えばユーザの所属する研究室やユーザ自身が設置したものである．すなわち，LAN 管理者側からは有線 LAN に接続したかに見える機器が実際には無線 LAN に接続されていることがあるので分析の際は注意が必要となる．

学内 LAN に接続された機器にはまず DHCP で IP アドレスが振られ，その後エッジスイッチによって認証（RADIUS による）が実施される．その際はまず MAC アドレス認証を行い，これに失敗した（接続された機器の MAC アドレスが認証サーバに登録されていなかった）ときには WWW を利用したユーザ認証を行う．MAC アドレスが登録されているのは WWW によるユーザ認証が不可能な機器（例えばプリンタや無線 AP）等ごく一部であるので，大抵の場合はユーザ認証となる．

有線 LAN の利用ログが残るのは下記の箇所となる．

- DHCP サーバ：
  - 日時
  - 接続機器の MAC アドレス
  - 発行した IP アドレス
- RADIUS サーバ：
  - 日時
  - ユーザ名
  - 接続機器の MAC アドレス
  - スイッチのホスト名
- 各エッジスイッチ：
  - 日時
  - 接続機器の MAC アドレス
  - 接続機器に割り振られた IP アドレス
  - ユーザ名（ユーザ認証の場合のみ）
  - ポートの番号

ただし，各エッジスイッチのログは syslog により syslog サーバに集約されている．

3.2 無線 LAN

無線 LAN は本学では 2011 年度より本格的導入が始まり [7]，建物の免震工事や新築・改築等の度に規模を拡大し今では 7 つの校舎に 144 台の AP が設置されている．本学には講義室や演習室をもつ校舎が 10 あるが，そのうちの 5 つに無線 LAN が設備されている（残りは研究所棟と学生会館）．

本学の無線 LAN は主に Aruba 社製品で構成されており[*3]，AP での通信はすべてコントローラに集約される構成になっている．そのため，無線 LAN 利用ログも主にこのコントローラに集約されており，SNMP を利用してそのログを取得している．各接続機器について下記の情報を取得している．

- IP アドレス
- MAC アドレス
- ユーザ名
- 接続時間
- 接続 AP 名

この情報を取得した時刻をこれらに加えて 1 レコードとなる．SNMP を利用した上記情報の取得は，運用の妨げにならないよう考慮して 10 分おきとしている．

また，前述の DHCP サーバと RADIUS サーバのログは無線 LAN への機器接続でも有線と同様のものが記録されるので，無線 LAN の場合も下記のログが収集される（実際には有線・無線の別なくそれぞれのサーバから収集される）．

- DHCP サーバ：
  - 日時
  - 接続機器の MAC アドレス
  - 発行した IP アドレス
- RADIUS サーバ：
  - 日時
  - ユーザ名
  - 接続機器の MAC アドレス
  - 無線コントローラのホスト名

文献 [6] では無線 LAN に接続していない機器の情報も収集できていると聞く．すなわちユーザの携帯している情報機器が無線 AP を探索する電波も収集対象としている．これが可能であればユーザの意識的な操作（無線 LAN に機器を接続する）なしにその機器の情報を収集できるが，本学で利用している無線 AP では無線 LAN に接続した機器についてのみ上記の情報を取得できる．著者は昨年，無線 LAN 利用ログを分析することで学内のユーザの時刻ごとの滞留箇所を明らかにし大学の避難計画策定に応用することを文献 [10] で試みたが，これは無線 LAN に接続している機器のユーザのみを対象としたものに限られてしまっていた．無線 LAN に接続していない機器の情報まで収集することが可能ならそのような実現可能性が増すものと考えられる．

4. 提案システム

上述したようなログを一定の形式に整えて蓄積し，対話的な問い合せに応答できるシステムをここに提案する．このシステムは Mac OS X 10.9 の動作している Mac mini

[*3] 学生会館のみ Aerohive 社製品（3 台）が設置されている．

上で構築し，後述するデータ形式の変換等テキスト処理は sed や awk 等を利用したシェルスクリプトで行い，DBMS として SQLite3 を利用している．

4.1 データ形式

このシステムでは上述した各サーバおよびスイッチ，無線コントローラのログに対応したテーブルを用意し，それぞれテキスト形式のカラムを用意してデータを蓄積している．蓄積するタイミングは，syslog 形式のログの場合はそれが syslog により出力された時点，SNMP で収集しているログの場合は上述の通り 10 分おきである．

上記のログは過去のものが 1 年分それぞれのサーバに保存されていたので，初回データ投入時にはそれらも変換して蓄積している．それでもこの原稿執筆時点で総量は数 GB なので，SQLite3 で十分対応できるデータ量である．

蓄積されるデータ量は増加する一方であり，翻ってシステムの応答時間は一定の水準を保たなければならないので，今後は必要に応じてより高性能なハードウェア上に移設することや，DBMS を変更して高速化すること等を検討すべきであろう．

4.2 非特定化

このシステムにデータを蓄積する際には，前述した「非特定化」処理を実施している．具体的には openssl コマンドを利用して SHA-1 ハッシュ関数でユーザ名と MAC アドレスを変換し，データ全体を非特定化している．ただし，ユーザ名には「学生かそれ以外か」および学生の場合「入学年度」と「所属する学部学科」という属性情報が内包されており，これらは統計処理の際に利用したいので保持することとした．また同じ理由で，MAC アドレスの上位 24 ビット（ベンダー ID 部）も変換せず保持している．

これらの処理で本当に非特定化が実現できているかどうかは，その時点で突き合わせられる他のログや資料にどんなものが存在するか等の状況に依存するので，随時検証することが求められる．例えば，特定の年度に特定の学科に入学した学生がただ 1 人であった場合，上記のように学生の属性情報を残したままでは個人が特定されてしまうことになる．

4.3 応用

このシステムのように，大量のデータを時系列順に一定の形式で蓄積し，ユーザが一定の手順で容易にデータを取得できるようにしたものをデータウェアハウスと呼ぶ [6]，[11]．このシステムでは DBMS に SQLite3 を利用しており，これはアプリケーションに組み込める DBMS ライブラリとしても広く普及しているので，蓄積されたデータの取得方法として例えば以下のようなものが考えられる．

- コマンドラインから直接 SQL を利用して取得する
- シェルスクリプトや Python，Ruby 等の言語を利用してプログラム中でデータを取得し加工する
- Ruby on Rails 等の Web アプリケーションフレームワークを利用して WWW 上にインタフェースを構築する

5. おわりに

本稿では学内 LAN 利用ログについて，その位置付けや取扱いについての考えを述べ，本学で取得しているログを紹介し，統計的にログを分析するためのシステム（データウェアハウス）を提案した．

学内 LAN 発足時よりこのようなログが蓄積され，プロバイダ責任制限法や個人情報保護法等を念頭に置きながら，サーバの二次記憶容量等の都合により取捨が選択されてきた．これまでこのログは，ネットワーク上の障害発生時に過去へ遡って原因特定を実施したり，日常の運用状態を大雑把に把握するために管理者が眺めるのに利用されたりとごく部分的な活用に留まっていた感がある．また，それらの場合の具体的な利用方法として，その場限りのスクリプト等による一時的な処理が行われていたり，複数の場所にあるログを横断的に検索するような大規模な利用が行われなかったりしていた．

本稿で提案したシステムでは，学内 LAN 利用ログの主要なものが有線と無線との区別なく 1 ヶ所に集約されており，一定の手順で問い合せを行うことが可能となっている．また，蓄積されたデータはプライバシーに配慮し非特定化されている．さらに DBMS としてアプリケーションに組み込むことの容易な SQLite3 を利用しているので，このシステムと対話しながら実行するアプリケーション等の開発も容易であると考えられる．

今後はこのシステムを利用して学内 LAN の利用動向を明らかにしたり，このシステムと連携する分析アプリケーションを開発したりすることを考えている．

参考文献

[1] Rafael Ballagas, Michael Rohs, Jennifer G. Sheridan, Jan Borchers: BYOD: Bring your own device, Proceedings of the Workshop on Ubiquitous Display Environments at Ubicomp 2004 (2004).
[2] Tristan Henderson, David Kotz, Ilya Abyzov: The changing usage of a mature campus-wide wireless network, Computer Networks, 52(14), pp.2690–2712 (2008).
[3] Aaron Gember, Ashok Anand, Aditya Akella: A comparative study of handheld and non-handheld traffic in campus Wi-Fi networks, Passive and Active Measurement, Springer Berlin Heidelberg, pp.173–183 (2011).
[4] Eden Dahlstrom, J.D. Walker, Charles Dziuban: The

ECAR Study of Undergraduate Students and Information Technology, EDUCAUSE Center for Applied Research (2013).
[5] 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）.
[6] 杉木章義，佐藤聡，和田 耕一：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，情報処理学会研究報告（IOT），2013-IOT-23(7)，pp.1–5（2013）.
[7] 宮下健輔：京都女子大学における無線 LAN 利用動向調査，情報処理学会研究報告（IOT），2013-IOT-23(6)，pp.1–5（2013）.
[8] 林紘一郎, 他：情報証拠論（Information Forensics）確立のための基礎検討，電気通信普及財団研究調査報告書，21，pp.17–24（2006）.
[9] 技術検討ワーキンググループ報告書，第 5 回パーソナルデータに関する検討会配付資料，首相官邸 (2013).
[10] Kensuke Miyashita: A Wireless LAN Usage Trends Survey on Campus for Evacuation Planning, International Conference on Signal-Image Technology & Internet-Based Systems, pp.865–869 (2013).
[11] William H. Inmon, Chuck Kelley: Rdb-VMS: Developing a Data Warehouse, John Wiley & Sons, Inc. (1993).