オープン利用時の履歴情報統計システムの試作

佐藤悦教*・中村庸郎**・松尾征夫***

Trial development of log analyzer for student-users during free use

Yoshinori SATO, Tsuneo NAKAMURA and Yukio MATSUO

要　旨

情報処理センター施設を使用する上で問題となりつつある，様々なアプリケーションからの乱雑な印刷，ブラウザを用いた不適当なWebページの閲覧や書き込み等の情報を収集，オープン利用時に何らかのトラブルが発生した場合の調査を行うためPCを使用していたユーザを特定するプログラムを作成した。この統計情報の一部を公開することで不適切な印刷やWebページへのアクセスを各ユーザ個人が考え抑制することで施設の使用マナーの向上を期待する。

Abstract

Recently unsuitable printings and web accesses by student-users during free use have remarkably increased in number. Therefore we trially developped the programs to count printings, to rank accessed sites during free use, and to identify student-user who used the specific computer immediately. We think that student-users will take better way in free use, if those results are open to student-users.

1．緒　　言

平成10年4月の教育用計算機の更新に伴い，DOSからWINDOWSへとクライアントOSが移行され，NT Serverによりドメインを構築し（STドメイン）クライアントの集中管理を行い使用環境を向上させている。それと同時期にネチケット（ネットワークにおけるエチケット）に関する講習およびPCを使用した実習を受けることを前提にWebページの閲覧を全学生に対して利用可能としている。

便利になる一方，WINDOWSやブラウザの操作に慣れるに従って重複印刷，情報処理系の課題とは一切関係の無い個人の趣味として印刷されるWebページ等の印刷が多くなり，本来の目的である授業の課題やプログラムの印刷枚数を上回るほどである。多い時で一日2000枚程度の印刷が行われることもあり用紙だけではなくトナーの消耗も激しく一年ももたない状況である。プリンタは授業や情報処理系授業の課題の提出などで使用するため，常に稼動させる必要がある。そこでプリンタの乱用を減少させるため何らかの対策を講じる必要性が生じてきた。

また，Webページの閲覧についても問題点として，不適当なページへのアクセス，掲示板やチャットを使った書き込み等がある。実際にトラブルも発生しているためサーバー側である程度までは対応しているが通常使用しているユーザにまで影響を及ぼしているため改善が望まれている。また，何らかのトラブルがオープン利用時に発生したことを考えトラブルの発生した時間に特定のPCを使用していたユーザを短時間で割り出す必要性もある。

今回，統計情報の収集を行うのは学生が授業又はオープン利用で使用できるドメイン内のネットワークである（5年生については研究室も可）。ドメインの構成は図1の通りである。

これらの問題を解決するために市販されている各種ツールの導入も考えられるが，まずユーザの意識やモラルを高めることが重要であると考え，各ユーザ毎のプリンタの使用頻度，Webページへのアクセス回数を出力するプログラムを作成し統計情報を公開することですこしでも各自が情報処理センターの利用状況について考えてくれるこ

＊　技　官　情報処理センター
＊＊　助教授　情報工学科
＊＊＊　教　授　機械工学科

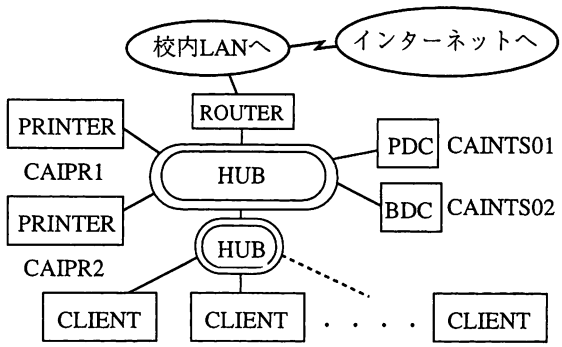

図1　ネットワーク図

とを期待して統計情報の収集プログラムを作成した。

2．プリンタ使用統計

2－1　プリンタの使用に関するログ

プリンタに関するログはPDC (Primary Domain Controller)であるホスト caints01のシステムログへと記録される。

システムログにはプリンタに関する記録以外にも ServicePack のインストールの結果や Eventlog の開始等システムに関する事柄が全て記録される。システムログは通常アーカイブファイルとなっているため NT Server のアプリケーションの一つであるイベントビューアを用いて内容を確認する。プログラムを作成するにあたりシステムログをアーカイブファイルからテキストファイルへ変換する必要があるため，イベントビューアを用いてテキストファイルとしてシステムログの保存を行う。

2－2　ログ内容および PAD

イベントビューアによりテキストとして保存されたシステムログは以下のフォーマットの羅列によって各事柄が表される。

日付 時刻 ソース 種類 分類 EVENTID ユーザ
　サーバ　その他の情報（事柄により異なる）
…… (1)

プリンタの使用統計を採取するためプリンタに関する記録以外は必要としない。システムログを確認することで，

EVENTID ＝10

であれば，その行がドキュメントの印刷に関する記録であるということがわかる。

プログラムの最終出力をユーザ毎の印刷回数，全印刷枚数および全印刷サイズ(byte)および総印刷枚数，総印刷サイズ(byte)とすると，プログラムを作成する上で必要なデータとなるのは(1)の内容から

・EVENTID
・ユーザ名（ユーザより）
・一回の印刷枚数（その他の情報より）
・一回の印刷サイズ（その他の情報より）

となる。これらのデータをユーザ単位でまとめ印刷の多い順（降順）にソートを行いプリンタの使用状況の統計を採取する。

以下に作成するプログラムの PAD を図2として示す。

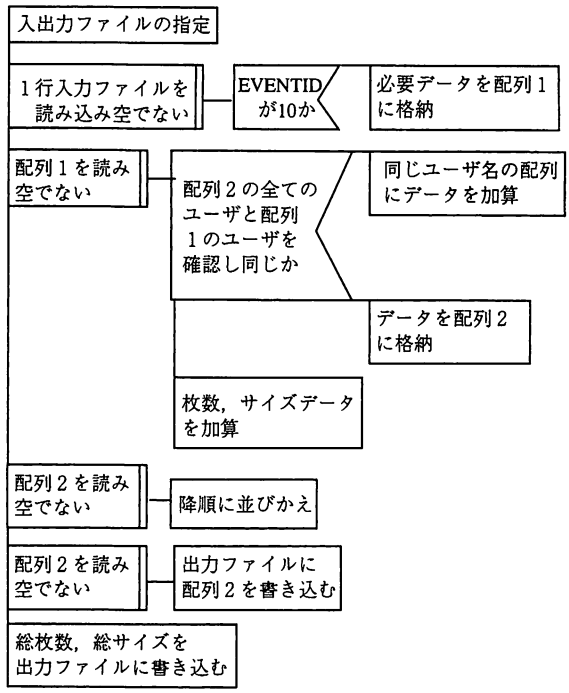

図2　プリンタ集計プログラムの PAD

上記，PAD に従いプログラムを作成した。

2－3　実行結果

入力ファイルをある一ヶ月のシステムログとし，作成したプログラムにより実行結果を求めた結果以下のような出力を得た。尚，以下の実行結果ではユーザ名は user1，user2…..となっているが実際には個人が特定できるユーザ名となっている。

Owner : user1 Bytes = 2149815 Page = 90
Owner : user2 Bytes = 1349939 Page = 87
Owner : user3 Bytes = 1979337 Page = 76
Owner : user4 Bytes = 1589954 Page = 57

：
：
Owner : user10 Bytes = 27808 Page = 1
Total _bytes = 194253015 total _print = 3846

実行結果より，各ユーザの出力回数，全印刷バイト数，全印刷枚数および総印刷枚数，総印刷バイト数が確認することができる。

3．アクセス統計

3－1　Webページの閲覧に関するログ

Webページの閲覧を行うためにはクライアントからのリクエストに応じてBDC (Backup Domain Controller)であるホストcaints02のProxy Serverが当該ファイルを取得しクライアントへ返すことで閲覧が可能となる。Proxy Serverはクライアントのリクエストに答える以前にアクセスの可否，URLのフィルタリング等を行いProxy Server上で起こった事象全てについてログへと記録する。このログはクライアントのアクセス毎に逐次追加され通常はProxy Serverのツールである View Access Logを用いて確認することができる。

データの内容はテキスト形式により追加されるためログが記録されているファイルはWindowsのアプリケーションであるメモ帳やワードパッド等でも確認することは可能である。ただし，全てのアクセスに関して記録を行うためログサイズは大変大きく学生用のProxy Serverであるcaints02でも1週間で10Mbyte以上のデータ量となりWindowsのアプリケーションでは確認できない場合もある。

3－2　ログ内容およびPAD

保存されているログはProxy Serverの管理者が定める任意のフォーマットの羅列により表される。学生が使用できるドメイン内のProxy Serverでは以下のフォーマットの羅列により事象の記録を行う。

クライアント名　日付　リクエストサイト名
　ステータス　その他の情報
　（移動時間やサイズ等の様々な数値データ）
…… (2)

Webページの統計については，フルリクエスト名で行った場合では数値が分散され，実行結果も膨大となってしまうことから出力データとしてはサイト名（www.tomakomai-ct.ac.jp 等のドメイン名）までとする。また，実際にWebページを閲覧することができたリクエストについてのみ集計を行うためステータス（Status returned to client）を確認する。

Status returned to client ＝200

であれば，リクエストに対して要求どおりの返答があったことを示す。その他の値の場合はエラー（Not Found，Internal Server Error 等）があり閲覧できなかったったことになるため出力データから除外する必要がある。

プログラムの最終出力をサイト毎のアクセス数とするとプログラムを作成する上で必要なデータとなるのは(2)の内容から

・接続ホスト（クライアント名より）
・接続先サイト（リクエストサイト名より）
・閲覧の可否（ステータスより）

となる。これらのデータをサイト単位でアクセス回数をまとめアクセス回数順にソートを行うことでサイト毎のアクセス数を採取することが可能である。

以下に作成するプログラムのPADを図3として示す。

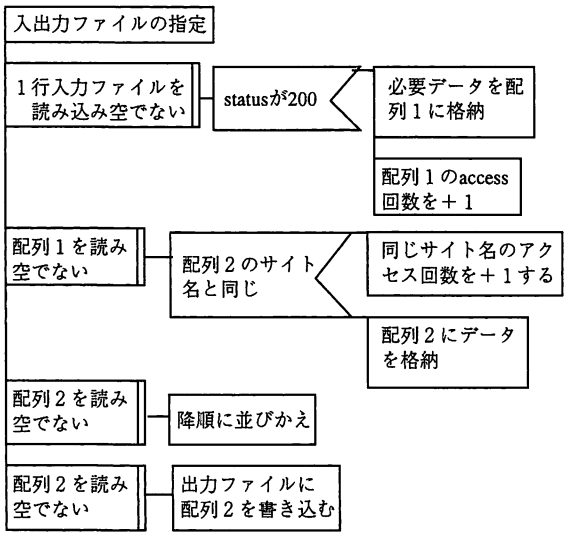

図3　アクセス統計のPAD

3－3　実行結果

入力ファイルをある一ヶ月のProxy Serverのアクセスログとしてプログラムにより実行結果を求めた結果以下のような出力が得られた。

11635-www.yahoo.co.jp
7939-www.geocities.co.jp
6256-home.netscape.com

3431-www.cyberclick.net
3361-village.infoweb.ne.jp
2817-www.geocities.com
2673-www.interq.or.jp
2592-www.infoseek.co.jp
2541-search.yahoo.co.jp
1809-www.goo.ne.jp
1471-www2.t-network.ne.jp
：
：

実行結果より，各サイト毎のアクセス回数が確認することができる。

4．ユーザの特定

4－1　ユーザ認証に関するログ

学生が通常使用できるネットワークはSTドメイン内のみである。学生がドメインのネットワーク資源を使用するためには必ずWindows NT Serverによるユーザ認証を行わなければならない（学生は必ずドメインユーザであるため)。結果ログオン，ログアウトどちらについてもNT Serverのセキュリティログに記録されることになるためトラブルの情報からセキュリティログを確認することでトラブルを起こしたユーザを特定することができる。例えば単独では通常ユーザ認証を行わないブラウザにおいてもアクセスの経歴はProxy Serverのログとして記録されるためホストや時刻等からアクセスしたユーザを特定することが可能となる。しかし，セキュリティログについても全学生の記録が保存されているため多きなデータサイズとなりユーザを特定するだけでも大きく時間を消費してしまうため効率化を考えたい。

セキュリティログにはユーザのログオン，ログアウトの記録以外にもアカウントに関する設定，オブジェクトアクセス等についての事柄が記録される。セキュリティログはシステムログと同様にアーカイブファイルとして記録されているため通常イベントビューアを使用してログの内容の確認を行う。本報ではテキストファイルを扱ってプログラムを作成するため2－1と同様にイベントビューアを用いてテキストファイルとしてセキュリティログの保存を行った上でプログラムの作成を行う。

4－2　ログ内容およびPAD

テキストとして保存されたセキュリティログは以下のフォーマットの羅列により表される。

日付 時刻 ソース 種類 分類
　EVENTID ユーザ コンピュータ
　　その他の情報（事柄により異なる）
…… (3)

ユーザの特定を行うプログラムを作成するためセキュリティログからログオン，ログアウトに関する記録を必要とする。セキュリティログを確認することでログオンに関する記録であれば
　EVENTID ＝528
ログアウトに関する記録であれば
　EVENTID ＝538
となる。

ブラウザを使用してトラブルが発生した場合，何よりユーザの特定を行う必要があり，これをプログラムの最終出力と考える。またユーザを特定するための最低限の情報として時刻，ホスト名(IPアドレス）が必要となるためこの2項目（このデータについては他のログから抽出する）を入力必須とした場合，プログラムを作成する上で必要なデータとなるのは(3)の内容から

・日付
・時刻
・EVENTID
・ユーザ名（ユーザより）
・ログオンID（その他の情報より）
・ワークステーション名（その他の情報より）

以上のデータがホストを使用しているユーザを特定するために必要となる。

以下に作成するプログラムのPADを図4として示す。

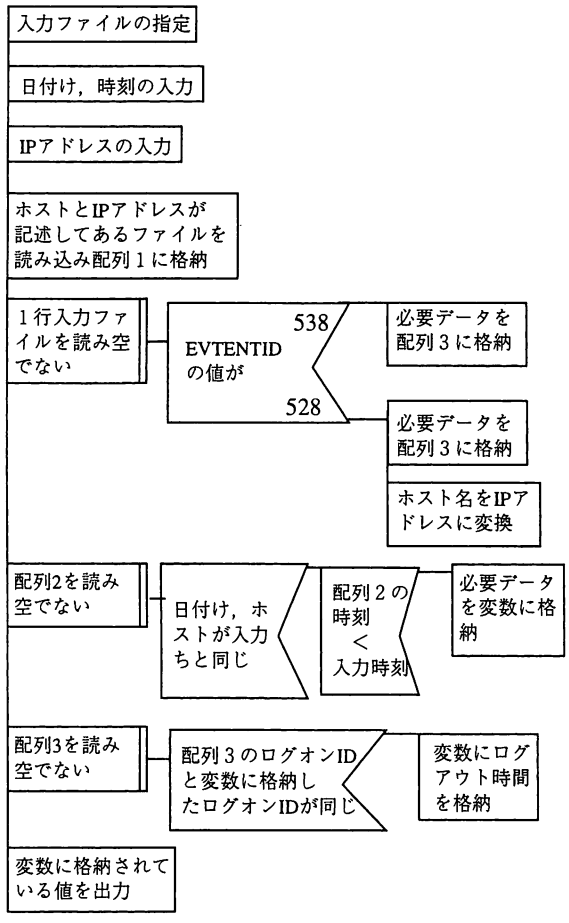

図4　ユーザー特定のPAD

4－3　実行結果

入力ファイルをセキュリティログ，入力項目を時刻，IPアドレスとし作成したプログラムにより実行結果を求めた結果以下のような出力が得られた。

［入　力］
Act Date (yymmdd) = 991014
Act Time (hhmmss) = 161023
IP address (www.xxx.yyy.zzz) = 123.45.67.89

［出　力］
99/10/14 IN-16:05:26 OUT-16:22:35
HOST:caicl25 IP:123.45.67.89
User : testuser

実行結果より，ワークステーションを使用していたユーザを時刻，ホストから特定することができる。

5　結　　言

プリンタに関する統計，Webページへのアクセスに関する統計，ユーザを特定するためのプログラムを作成することができた。本報には記述していないがユーザ毎のログオン時間を出力するプログラムも作成した。また，今後の拡張性を考え本報では全く使用していないデータもプログラム内に格納しており，必要に応じて新しい統計用，管理用プログラムを作成することも可能である。

しかし，作成した統計情報をどういった形でどこまで学生に公開するか問題として残っている。統計情報の内容などを確認し，センター内で問題が無いと判断した上で統計情報の掲示，ブラウザによる公開を順次行う予定である。

今後はVisual BasicやVisual C等も用いてよりグラフィカルな情報が提供できるようなものを作成したいと考えている。

参考文献

1）Digital, DEC C Language Reference Manual,March 1996
2）Microsoft Corp, Windows NT4.0 Server リソースガイド，March 1997
3）SUITESPOT, Netscape Proxy Server Administrator's Guide,July 1997
4）JACK PURDUM, C Programming Guide 丸善株式会社(1987)

（平成11年11月30日受理）