E-09

トラフィックから判断する WebDAV の利用検出手法の提案

The proposal of the use detection technique of WebDAV judged from traffic

剱本 倫章†　　川橋 裕‡
Michiaki Kenmoto　Yutaka Kawahashi

１．はじめに

インターネットは個人や企業，大学を問わず様々な社会システムに活用されるインフラストラクチャとなった．しかし，社会システムの基盤として高い利便性をもたらす一方で，システムに障害が発生した際には重大な影響をおよぼす危険性がある．また，インターネットの急激な拡大やネットワーク構成の複雑化にともない，障害を起こさない対策を実現することは極めて困難となっている．

上記に起因して，経済産業省から「障害は起こりうる」ことを前提とした，事故前提社会[1]の考え方が提唱されている．事故前提社会では，インシデントの予防や被害の最小化・局所化，回復力の高い仕組みのネットワーク構築が要求される．

すなわち，ネットワーク運用管理者（以下管理者）は障害に迅速に対応し，どのような経緯で障害が発生したかの原因を究明することが必要不可欠となっている．

事故前提社会において重要なことは，企業や公共機関などの組織において，組織内ネットワークと組織外ネットワークの間で，過去にどのような通信がおこなわれたか（以下通信記録）を把握することである．通信記録は，インターネットにおける通信の管理情報であり，「誰と誰が，いつ，どのくらいの量の情報をやりとりしていたか」の記録である．通信記録の解析によりネットワーク内の障害を把握することができる場合がある．例えば，ファイル共有をおこなった際にウイルスに感染し，情報が流出した場合，通信記録からウイルス感染した端末や感染経路を特定することが可能である．しかし，通信記録を管理する問題点として，管理者への負荷が大きいことが挙げられる．組織内ネットワークと組織外ネットワークの間では膨大な量の通信がされているため，全てを監視することが困難である．

管理者が障害を発見し，原因を究明する際に，現時点の通信記録だけではなく過去の通信記録を参照することも多い．しかし，和歌山大学の対外接続線を例に挙げると，1 日のうち最も通信が多い時間帯で，毎秒約 1 万パケットもの情報がやりとりされている．このような状況下において，管理者が障害発生前に不審な通信を発見することや，障害発生後すぐに原因を特定することは困難である．そこで，ウイルス感染の原因となりうる WebDAV のトラフィックを自動で検出することができれば，管理者の負担を軽減することにつながると考えられる．WebDAV とは，HTTP を拡張したプロトコルで実現された分散ファイルシステムで，IETF[2] によって RFC 2518[3] として定義されている．そのため，ファイアウォールで HTTP の利用を制限，禁止することにより，WebDAV の利用を制限，禁止することは可能である．しかし，HTTP を禁止することは Web ページの閲覧や公開を禁止することになるため，利便性が低下する．ファイアウォールの機種によっては，Web ページの閲覧や公開を禁止せずに WebDAV の利用を禁止することが可能なものもある．これは WebDAV と Web の閲覧を判別する際にペイロードの中を解析する必要があるので，プライバシを侵害する可能性がある．

本研究では，ペイロードの中を解析することなく，ヘッダ情報とペイロード長から WebDAV のトラフィックを自動で検出する手法を提案する．具体的には，WebDAV の利用を検出するシステムを組織内ネットワークと組織外ネットワークの境界に設置する．組織外ネットワークから組織内ネットワークへの通信，組織内ネットワークから組織外ネットワークへの通信をそれぞれ通信記録として保存する．通信記録として保存する情報は宛先・送信元 IP アドレス，宛先・送信元ポート番号，通信時間，ペイロード長である．上記の通信記録を基に，HTTP 通信における，Web ページへのアクセスの通信と WebDAV を用いたファイル共有の通信の挙動の違いを利用して識別する．さらに，ペイロードの中を解析せずに識別するため，ユーザのプライバシも守られる．

本論文では，第 2 章で従来の検出手法について述べる．第 3 章で研究目的について述べる．第 4 章で提案手法と提案システムについて述べる．5 章では本研究の評価と考察を述べる．

２．従来の検出手法

本章では，トラフィック解析やファイル共有検出に対する既存技術，既存研究について述べる．その後，これらの問題点に述べる．

2.1 トラフィック解析手法の分類

組織内外ネットワークの境界でおこなうトラフィック解析手法は，大きく分けて以下の 2 つに分類することができる．

●ゲートウェイ型

組織内ネットワークと組織外ネットワークの境界にパケット解析システムを設置してトラフィックを解析する

●トラフィックモニタ型

組織内ネットワークと組織外ネットワークの境界を流れるパケットを解析システムへ転送し，解析する

以下の節でそれぞれの技術について説明する．

2.2 ゲートウェイ型

ゲートウェイ型は，組織内ネットワークと組織外ネットワークの境界にゲートウェイとしてシステムを設置することで，組織内ネットワークを出入りするトラフィックを解析する手法である．

†和歌山大学, Wakayama University

‡ 和歌山大学システム情報学センター, Center for Information Science, Wakayama University

ゲートウェイ型は，異常なトラフィックを検出した場合，即座にトラフィックの遮断が可能である．しかし，ゲートウェイ端末でトラフィックの遅延が発生すると組織内ネットワーク全てに影響を与えてしまうため，ゲートウェイに高速なトラフィック転送性能が求められる．ゲートウェイ端末にはパケットのペイロードを検閲するものが多く，ユーザのプライバシを侵害することになる．事前に定義したルールを基にパケットのペイロードを照合することで異常なトラフィックを検出するため，新しいソフトウェアやウイルスなどに対応できない．

以下で既存のゲートウェイ型の検出手法について述べる．

2.2.1 Cisco ASA 5500 シリーズ

Cisco 社製のファイアウォール Cisco ASA 5500 シリーズ[4]は，堅牢なファイアウォールに加えて，侵入防御，VPN，コンテンツセキュリティなどを監視できる製品である．トラフィックを解析することにより，HTTP アクセスを安全な HTTP メソッド，安全ではない HTTP メソッド，WebDAV メソッドに分類し，WebDAV の利用を禁止することが可能である．トラフィック解析にはペイロードの中も解析する．

2.3 トラフィックモニタ型

トラフィックモニタ型は，組織内ネットワークと組織外ネットワークの境界を流れるトラフィックをキャプチャ，解析する手法である．ゲートウェイ型と違って，トラフィックを即座に遮断することはできないが，トラフィックパケットのペイロードを検閲せずにトラフィックを解析することが可能である．

以下で既存のトラフィックモニタ型の検出手法について述べる．

2.3.1 SVM を用いた P2P トラフィック判別手法の提案

本研究の先行研究である阪上らの手法[5]は，P2P ファイル共有ソフトウェアを用いた通信の特徴を訓練データとし，機械学習させ，P2P トラフィックを判別する手法である．P2P ファイル共有ソフトウェアを用いた通信の特徴として，クライアント端末の通信先ホスト数とペイロード長が挙げられている．SVM(Support Vector Machine) で学習させるパラメータは，通信先ホスト数とペイロード長である．通信先ホスト数を用いて一般トラフィックと P2P トラフィックを分離し，ペイロード長によって各ソフトウェアを判別することが可能である．さらに，P2P トラフィックの特徴として，P2P トラフィックは一般トラフィックに比べ，well-known ポートを使用しない通信が多いことが挙げられている．

実験環境は，和歌山大学内ネットワークと大学外ネットワークの境界に設置されたレイヤ 2 スイッチからポートミラーリングで転送されたパケットをキャプチャするようになっている．実験では，和歌山大学ネットワークでトラフィック判別実験をおこない，P2P ソフトウェア 9 種類と Skype の計 10 種類の P2P トラフィックを検出することに成功した．

しかし，データのインデックス情報をインデックスサーバが監視するハイブリッド型の P2P 通信では，検索リンクの確立やファイル検索はインデックスサーバとの通信のみで実現できるため，通信先ホスト数は少ない傾向にある．そのため，通信先ホスト数が少ないハイブリッド P2P を利用したファイル共有ソフトウェアは，先行研究では判別できない可能性がある．

2.4 問題点

ゲートウェイ型のトラフィック解析では，ペイロードの中も解析するため，ユーザのプライバシを侵害する．さらに，ゲートウェイ端末に異常が発生した場合，組織内ネットワーク全体に影響を与えるという問題がある．

トラフィックモニタ型のトラフィック解析では，ペイロードの中を解析しない為，ユーザのプライバシは守られていると考える．しかし，先行研究である阪上らの手法では，P2P ファイル共有ソフトウェアのトラフィックの検出には成功しているが，クライアント・サーバ方式のファイル共有には対処ができない．加えて，対象となっているポートは Ephemeral ポートのみであるため，well-known ポートでの通信は検出することが不可能である．

３．研究目的

本研究の目的は，ユーザの利便性を保ちつつ，セキュリティを向上させることである．そのために，80 番ポート宛ての通信である Web アクセスと WebDAV のファイル共有をトラフィックから判別し，分離，検出する．WebDAV は HTTP の拡張機能であり，Web アクセスと同じ 80 番ポートを使用する．そのため，ファイアウォールで 80 番ポートの通信を禁止することにより，WebDAV でおこなうファイル共有も防ぐことが可能である．しかし，80 番ポートでの通信を禁止することは Web アクセスを禁止することになるため，Web ページの閲覧や公開ができなくなる．WebDAV を利用し，ファイル共有をおこなうことは，ウイルス感染などの原因になるが，Web ページの閲覧や公開を禁止することはユーザの利便性を損なうことになるのである．

ユーザの利便性を追求すれば，セキュリティそのものがなくなり，セキュリティを追求すれば利便性がなくなる．ネットワークを管理する際，ネットワークのセキュリティとユーザの利便性は二律背反すると言える．本研究は，Web アクセスと WebDAV のファイル共有をトラフィックから判別，分離，検出するため，ファイアウォールで 80 番ポートの通信を禁止しなくてもよい．さらに，障害発生時には，WebDAV を利用したファイル共有が過去におこなわれたか調べることができるため，障害の原因を究明する際，WebDAV が原因であるかどうかの切り分けに利用できる．つまり，ユーザの利便性を保ちつつ，セキュリティを向上させることが可能である．

４．提案手法

本章では，和歌山大学のネットワーク構成を説明し，WebDAV の利用によるファイル共有をトラフィックから検出する手法を提案する．

4.1 和歌山大学のネットワーク構成

現在の和歌山大学のネットワーク構成は図 1 のようになっている．学外ネットワークから見ると，通信パケットはまず初めにファイアウォールを通過し，その後レイヤ 3 スイッチを介して学内ネットワークへと入る．しかし，ファイアウォールでは，宛先，送信元ポートともに 80 番ポー

トは禁止しておらず，80 番ポート宛の通信は通過することが可能である．本研究では，学外ネットワークと学内ネットワークの境界部分に設置されているレイヤ３スイッチからポートミラーリングしたパケットを解析することを想定する．

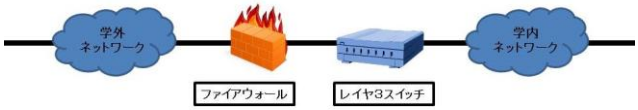

図 1　和歌山大学のネットワーク構成

4.2 検出手法

本節では，Web アクセスと WebDAV を使用したファイル共有によるトラフィックについての特徴を示し，その後検出手法について説明する．

4.2.1 Webアクセスによるトラフィックの特徴

Web アクセスによる通信は，図 2 に示すように，クライアント端末が Web サーバに対して HTTP リクエストを送り，Web サーバは html ファイルやイメージファイルなどを返す．html ファイルとは Web サーバがインターネット上に公開しているホームページ用のファイルで，イメージファイルとはホームページで公開している画像のことである．

Web アクセスでは，クライアント端末から Web サーバに対する通信量と Web サーバからクライアント端末に対する通信量を比較すると，後者の方が多くなる．これは，クライアント端末は Web サーバに対して，HTTP リクエストを送るだけだが，Web サーバは HTTP リクエストに対し HTTP レスポンスを返すだけでなく，html ファイルやイメージを送るため，どうしても通信量が多くなる．

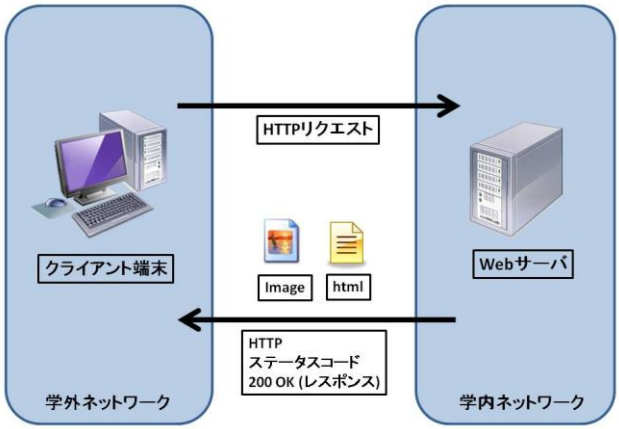

図 2　Webアクセス

4.2.2 WebDAV を使用したファイル共有によるトラフィックの特徴

WebDAV を利用したデータのアップロードによる通信は，図 3 に示すように，クライアント端末から Web サーバに対してデータを送信し，Web サーバはレスポンスを返す．この場合，クライアント端末から Web サーバに対するデータ転送量は，Web サーバからのレスポンスによる通信量よりも多い．

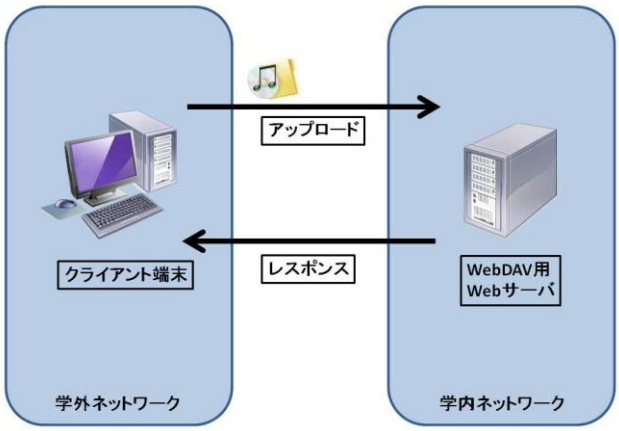

図 3　WebDAVを利用したデータのアップロード

4.2.3 検出手法

4.2.1，4.2.2 から，Web アクセスと WebDAV を利用したデータのアップロードによるトラフィックを比較すると下記のことがわかる．

●Web アクセスでは，Web サーバからクライアント端末へ流れる通信量の方が多い

●WebDAV を利用したファイルのアップロード時は，クライアント端末から Web サーバへ流れる通信量の方が多い

この特徴を利用し，Web サーバへのアクセスと WebDAV の利用が混在するトラフィックから WebDAV の利用を検出する．

4.3 提案システム

本節では，提案システムのハードウェア構成について説明し，その後システム構成について説明する．

4.3.1 ハードウェア構成

提案システムのハードウェア構成を以下に示す．提案システムを動作させるには十分な性能である．

●OS : CentOS6.5
●CPU : Intel(R) Celeron(R) CPU G540 @ 2.50 GHz
●HDD : 1.00 TB
●Memory : 4.00 GB
●NIC : Intel(R) 82579V PRO/1000 Gigabit Ethernet

4.3.2 ソフトウェア構成

提案システムのソフトウェア構成を以下に示す．Apache[6]，MySQL ともにオープンソースソフトウェアで，いずれも広く利用されており，安定性が高い．提案システムは Apache や MySQL に加え，php とデータベース接続クライアントツールである phpMyAdmin を構築している．

●MySQL : 5.1.71
●Apache : 2.2.15

4.3.3 システム構成

本項では，本システムの構成について述べる．本システムの構成図を図 4 に示す．組織内ネットワークと組織外ネットワークの境界に設置されているレイヤ 3 スイッチから，

ポートミラーリングを利用してパケットを本システムへミラーリングする．パケットキャプチャ部では，キャプチャしたパケットを整形し，必要なデータのみをデータベース部に格納する．格納するデータベースには MySQL を利用する．トラフィック判別部では，格納しているデータベース部の通信記録を基に，80 番ポートを利用した通信を Web アクセスと WebDAV に分けて WebDAV を検出する．トラフィック判別部で検出した WebDAV の利用に関する通信記録を WebDAV.log というログファイルに格納する．

以下で各部の詳細について述べる．

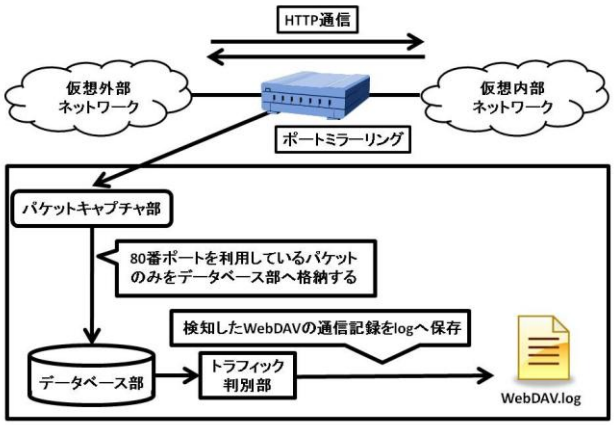

図 4　システム構成

パケットキャプチャ部

本システムは，組織内ネットワークと組織外ネットワークの境界に設置されているレイヤ 3 スイッチから，ポートミラーリングを利用してパケットを本システムへミラーリングする．ミラーリングにより流れてきたパケットをパケットキャプチャ部でキャプチャし，TCP 通信のヘッダ情報を保存する．保存するヘッダ情報を以下に示す．

●宛先 IP アドレス
●送信元 IP アドレス
●宛先ポート番号
●送信元ポート番号
●時間
●ペイロード長

データベース部

パケットキャプチャ部では，すべての通信のヘッダ情報を保存している．しかし，本システムは 80 番ポートを利用した通信についてトラフィック判別をおこなうため，送信元，または宛先ポート番号が 80 番ポート以外の通信記録は必要ない．

そこで，データベース部へ格納する通信記録を送信元，または宛先ポート番号が 80 番ポートのパケットのみにすることにより，サーバの負荷を軽減する．

トラフィック判別部

トラフィック判別部では，格納しているデータベース部の通信記録を基に，80 番ポートを利用した通信を Web アクセスと WebDAV に分けて WebDAV を検出する．検出には，4.3 節で述べた特徴を利用する．検出後は，ログファイルに送信元，宛先 IP アドレスと通信時間を保存する．

５．　評価

本章では，提案システムの実験の結果を示し，既存技術や既存研究と提案システムとの比較評価，本研究の今後の課題について述べる．

5.1 実験

本節では，4.3 節で述べた提案システムを用いた実験について述べる．

5.1.1 実験概要

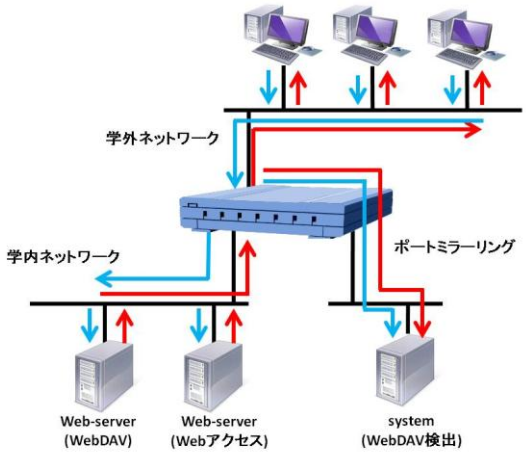

図 5　実験環境

本項では，実験概要について述べる．実験は，図 5 で示すように，仮想的な組織内ネットワークと組織外ネットワークを構築し，組織外ネットワークから組織内ネットワークへ WebDAV を利用してファイルのアップロードをおこない，検出する実験をおこなった．1 バイト，1K バイト，1M バイト，10M バイト，100M バイト，1G バイトの計 6 種類のファイルを用意し，各ファイルを 10 回アップロードした．

WebDAV を利用したファイルのアップロードの他に，別に用意した Web サーバへ Web アクセスもおこない，Web アクセスと WebDAV の利用をそれぞれ検出することができるかを検証した．Web アクセスする対象の Web ページは，phpinfo 関数(PHP の設定情報を出力する PHP の関数)のみを記述した PHP ファイルと下記のトップページを用いた．

●和歌山大学(www.wakayama-u.ac.jp/)
● amazon(www.amazon.co.jp/)
● google(www.google.co.jp/)
● nicovideo(www.nicovideo.jp/)
● wikipedia(www.wikipedia.org/)
● yahoo(www.yahoo.co.jp/)
● youtube(www.youtube.com/)

Web アクセスは，アクセス先のキャッシュがある状態，キャッシュがない状態で，Web アクセスをそれぞれ 10 回おこなった．

5.1.2 実験結果

検出実験の結果を表 1，2，3 で示す．

表 1，2，3 より，下記のことがわかる．

●ファイルサイズが一定以上の場合のみアップロードの検出可能

●ブラウザにキャッシュがない状態での Web アクセスは全て検出可能

●ブラウザにキャッシュがある場合での Web アクセスを検出できない場合がある

表 1　WebDAV利用検出実験の結果

ファイルサイズ	検出回数 / 実験回数
1 バイト	0 / 10
1K バイト	7 / 10
1M バイト	10 / 10
10M バイト	10 / 10
100M バイト	10 / 10
1G バイト	10 / 10

表 2　Webアクセス検出実験の結果(キャッシュなし)

アクセス先	検出回数 / 実験回数
和歌山大学	0 / 10
amazon	0 / 10
google	0 / 10
nicovideo	0 / 10
wikipedia	0 / 10
yahoo	0 / 10
youtube	0 / 10
phpinfo	0 / 10

表 3　Webアクセス検出実験の結果(キャッシュあり)

アクセス先	検出回数 / 実験回数
和歌山大学	10 / 10
amazon	10 / 10
google	0 / 10
nicovideo	10 / 10
wikipedia	10 / 10
yahoo	10 / 10
youtube	10 / 10
phpinfo	10 / 10

5.1.3 実験結果の考察

WebDAV の利用実験の結果から，ファイルをアップロードする場合，1K バイトのファイル検出率が 7 割であり，1 バイトのファイルは検出できなかった．考えられる要因として下記の理由が考えられる．

●基本的なクライアント端末とサーバ間の通信は，サーバからクライアント端末への通信量の方が多い

●不定期にクライアント端末とサーバ間で通信がおこなわれ，この通信はサーバからクライアント端末への通信量の方が多い

WebDAV を利用してファイルをアップロードする場合，基本的なクライアント端末とサーバ間の通信はクライアント端末からサーバへの通信量の方が多い．さらに，WebDAV の仕様で，不定期にクライアント端末とサーバの間では通信がおこなわれる．この通信はサーバからクライアント端末に流れるパケット量の方が多い．このため，データサイズが少ないファイルをアップロードした場合はデータサイズでは検出できないと考えられる．

Web アクセスでは，まずブラウザにキャッシュが残っていないかを確認する．キャッシュが残っていなければ全てのデータを取得するために Web サーバへアクセスする．ブラウザにキャッシュが残っていれば Web サーバに対して GET リクエストを発行しない．キャッシュを使用するかどうか判断できない場合には，サーバにリクエストを発行するが，特にコンテンツが更新されていなかった場合は「304 Not Modied」というレスポンスが返ってくる．この HTTP ステータスコードはリクエストしたリソースが更新されていないことを示す．結果として，キャッシュによりサーバからクライアント端末に対する通信量が変化するため誤検出する．

5.2 評価

本節では，提案システムの評価をゲートウェイ型，阪上らの手法と比較し述べる．

5.2.1 ゲートウェイ型との比較

2.2 節で挙げたゲートウェイ型でのトラフィック解析手法では，WebDAV を利用したファイル共有を検出，禁止することが可能である．しかし，既存技術として挙げたファイアウォールでは，トラフィックを解析する際にペイロードの中を解析する必要があるので，ユーザのプライバシを侵害する．

提案システムは，ペイロードを見ることなく WebDAV を検出するのでユーザのプライバシは守られる．

5.2.2 阪上らの手法との比較

2.3.1 項で挙げた阪上らの手法では，トラフィック解析をおこなう対象のポート番号は Ephemeral ポートのみで，well-known ポートでの通信は検出することができない．さらに，宛先ホスト数でファイル共有と通常トラフィックを判別するため，P2P ファイル共有ソフトウェアを検出することはできるが，クライアント・サーバ方式のファイル共有は検出することができない．

提案手法は，阪上らの手法に比べ，Ephemeral ポートでのファイル共有は検出できないが，80 番ポートでファイル共有をする WebDAV を検出することができる．クライアント端末からサーバに対する通信のペイロード長で WebDAV を利用したファイルのアップロードを判別するため，クライアント・サーバ方式のファイル共有である WebDAV を検出することができる．

5.2.3 評価

5.2.1，5.2.2 より，提案手法はペイロードを見ることなく，80 番ポートでファイル共有をする WebDAV の利用を検出できるため，既存手法よりも有用性があると考えられる．

5.3 今後の課題

本節では，本研究の今後の課題について述べる．

5.3.1 ファイルサイズに影響しないで WebDAVを検出

実験結果から，提案手法では数 K バイトのファイルを共有した場合，検出できないことがあると判明した．これは，共有するファイルデータを除いた WebDAV の通信では，

クライアント端末からサーバへの通信量より，サーバからクライアント端末への通信量の方が多いためである．

そこで，ペイロード長に加え，WebDAV と Web アクセスのトラフィックを判別する新しい要因を発見，追加する必要がある．

5.3.2 監視する通信ポートの追加

本手法では，HTTP の WebDAV 利用を検出することに成功したが，セキュリティを考えるのであれば，WebDAV は HTTPS(HTTP over SSL/TLS) で通信するのが妥当である．そこで，HTTPS でも WebDAV を検出できるように拡張する必要がある．そのためにも，DropBox[7] などのオンラインストレージサービスの利用時のファイルのアップロードと，WebDAV を利用したファイルのアップロードを判別する必要がでてくる．

5.3.3 WebDAV検出後の動作

本手法では，WebDAV を検出しても log ファイルに残るだけで，管理者に対して通知する機能はない．今後，実運用するには検出した際に log ファイルに残すだけでなく，管理者に対して通知する機能も必要であると考えられる．log ファイルだけでは視覚的に判断し難いため，管理インタフェースを作成し，ユーザビリティを向上する必要がある．

5.3.4 実環境での実験

本実験は，仮想環境で WebDAV を利用したファイルのアップロードと Web アクセスをおこない検出した．今後，担当教官の監視のもと，実環境に導入し，WebDAV を検出することができるかを検証したいと考えている．

６． おわりに

現在は組織内の通信内容を閲覧し，HTTP メソッドから WebDAV を検出，禁止する手法が主流であるが，本研究では，通信の内容を見ることなく，WebDAV と Web アクセスをトラフィックの特徴から判別し，WebDAV を利用したファイルのアップロードを検出する手法を提案した．提案手法を実装したシステムを和歌山大学のネットワーク環境を模した仮想ネットワークに導入して検出実験をおこない，WebDAV を利用したファイルのアップロードの検出に成功した．今後は WebDAV 以外の手段を用いたファイルのアップロードを WebDAV の通信と誤検出しないかを検証していきたいと考えている．

参考文献

[1]経済産業省
"「情報セキュリティ総合戦略」の概要"
http://www.meti.go.jp/policy/netsecurity/downloadfiles/Strategy_Summary.pdf

[2]"IETF"
https://www.nic.ad.jp/ja/tech/rfc-jp.html

[3]Internet Engineering Task Force
"RFC2518"
http://www.ietf.org/rfc/rfc2518.txt

[4]CISCO SYSTEMS
"Cisco ASA 5500 シリーズ"
http://www.cisco.com/cisco/web/support/JP/docs/SEC/Multi-FunctionSecur/ASA5500AdaptiveSecurAppli/CG/004/webvpn.html?bid=0900e4b1825ae5e9

[5] 阪上竜太
"SVM を用いた P2P トラフィック判別手法の提案"
2012 年度修士論文和歌山大学大学院システム工学研究科

[6] The Apache Software Foundation
"Apache"
http://www.apache.org/

[7] DropBox Inc
"DropBox"
https://www.dropbox.com/