Web メールフィルタリングシステムの機能拡張および性能評価

浦川順平†　鈴木健二†
電気通信大学大学院 電気通信学研究科†

1 はじめに

近年，企業での情報漏洩が問題になっている．原因の 1 つに Web メール，掲示板，ブログ，SNS 等の Web アプリケーションの利用があるが，その中でも Web メールは最も利用頻度が高く，情報漏洩となる危険性が極めて高いと考えられている．筆者等は先に，世の中で広く普及している Web メールの特徴調査に基づき，HTTP 通信の Web メールに対して，サーバから送信される Web ページの解析とクライアントから送信される POST リクエストの解析の 2 つを組み合わせたフィルタリング手法を提案し，その有効性を示した[1][2]．今回は，利用されている Web メール手順を踏まえ，対象を HTTPS 通信にまで拡大した．また，情報漏洩の原因が Web ページ取得でなく，ユーザからの情報送信であるという点を考慮して，規制実施タイミングの改善を図った．本稿では，システムの機能拡張の概要と性能評価について報告する．

2 HTTPS 通信解析のための拡張

(1) HTTPS 通信の Web メールへの対応

近年，セキュリティ上の理由から HTTPS 通信を利用した Web メールが注目されている．先の調査[2]によると Web メールにおける HTTPS 通信の利用はフリーWeb メールでは 7% (5/67) とまだ比較的少ないものの，ISP がその契約者に対して提供する Web メールでは 100% (6/6) の割合で利用されている．そのため今後の Web メールの利用規制では HTTPS 通信への対応が重要となる．

(2) HTTPS 通信解析機能の実現

HTTPS 通信を解析するために，プロキシにクライアント，サーバそれぞれの共通鍵を取得する仕組み（図 1）を導入した．これにより，サーバ（クライアント）から取得したデータをサーバ（クライアント）との共通鍵を使って復号化した後にフィルタリングし，そのデータをクライアント（サーバ）との共通鍵を使って暗号化してクライアント（サーバ）へ送信することが可能になる．

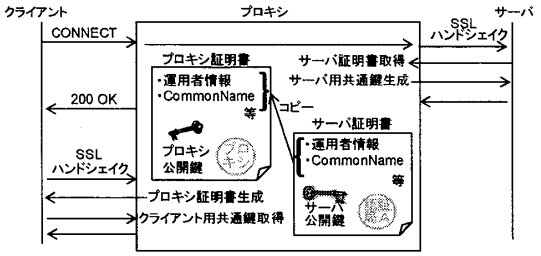

図1　共通鍵を取得するまでの流れ

Performance Evaluations of Web-Based Email Filtering System with New Functional Enhancement
†Junpei Urakawa and Kenji Suzuki
†Graduate School of Electro-Communications, The University of Electro-Communications

3 Web メールの規制実施タイミングの改良

(1) プロトタイプシステムの規制実施タイミング

プロトタイプシステムはプロキシとして動作し，以下に示すサーバから送信される Web ページ（メール作成ページ）の解析（方式 1）とクライアントから送信される POST データ（Web メール）の解析（方式 2）の 2 つで Web メールを判定するため，方式 1 で規制できない JavaScript で記述された動的な Web ページを方式 2 で規制でき，方式 2 で規制できないキーワードの出現頻度の低いものを方式 1 で規制することにより全体として高い規制率を実現することができる．

(方式 1) サーバからクライアントに対する Web ページ内で POST メソッドが指定された form タグ内にメール特有のキーワード（To，宛先，Bcc 等）が出現していた場合に規制する．各キーワードにはあらかじめ重みづけを行っておき，値の合計が閾値を超えた場合のみ規制する．

(方式 2) クライアントからサーバへの POST リクエスト内で Content-Type 固有の文字列（例. application/x-www-urlencoded の場合，"&"）とメール特有のキーワードの組が出現していた場合に規制する．方式 1 と同じく検出したキーワードの値が閾値を超えた場合のみ規制する．

プロトタイプシステムは各方式による判定直後に規制を実施する（図 2）．このため，図 3 で示すように実際は Web メールでないが方式 1 により Web メールと判断された場合，そのページは取得前に規制されるため，ユーザはそのページを閲覧することができない．

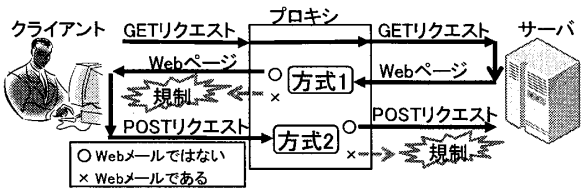

図2　メール送信までのプロトタイプシステムの動作

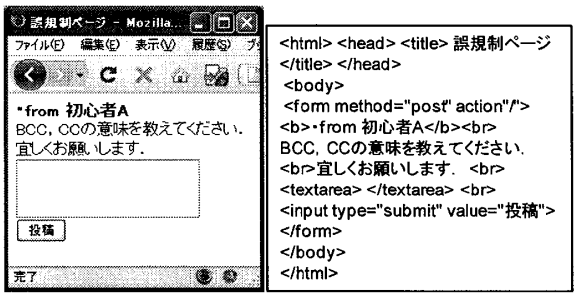

図3　規制により取得することができない Web ページ

(2) 適切な規制実施タイミングの実現

プロトタイプシステムでは方式 1 判定直後で規制を実施していたが，情報漏洩が発生するのは Web メール作成ページ取得時でなく，Web メール送信時である点を考慮すると，そのタイミングは図 4 に示すようにプロキシが POST リクエストを受理した後が適切であると考えられる．

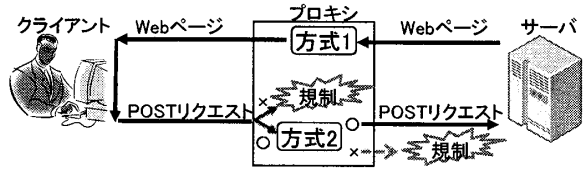

図 4　規制実施のタイミング

このため，本システムでは新たに POST 用ブラックリストを導入し，POST リクエスト先の URL がリスト内の URL と一致した場合に規制するよう拡張した．POST 用ブラックリストは図 5 に示すように各方式適用後に登録され，そこに登録する URL は GET リクエスト先の URL およびサーバから取得した Web ページ（HTML 文書）内の form における action フィールドを基に導出する．

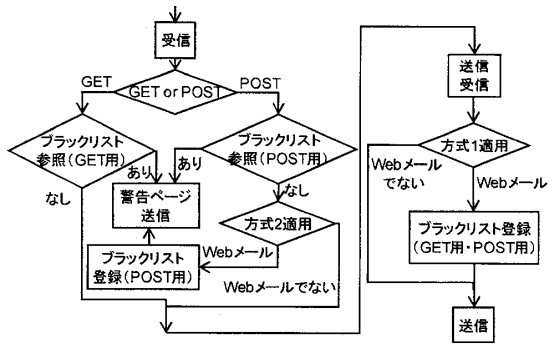

図 5　リクエストに対する本システムの動作

4 評価実験

機能拡張を実施後，本システムの性能評価を行った．測定項目は既存 Web メールの規制率およびクライアントからの同時リクエスト数に対する平均レスポンス時間であり，プロキシを通さない場合と既存プロキシである squid を通した場合を比較対象とした．本実験は LAN 内で実施し，3 台のクライアントとパフォーマンス測定ツールである JMeter[3]によりリクエストを生成した．HTTP 通信時および HTTPS 通信時の同時リクエスト数に対するレスポンス時間をそれぞれ図6，図 7 に示す．

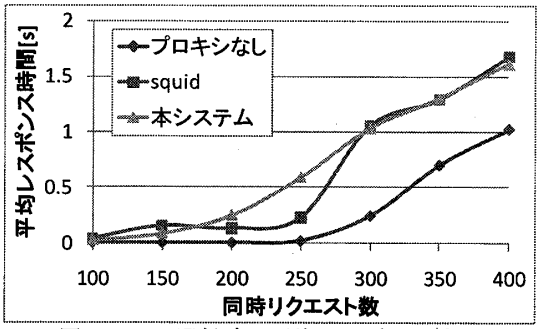

図 6　HTTP 通信時の平均レスポンス時間

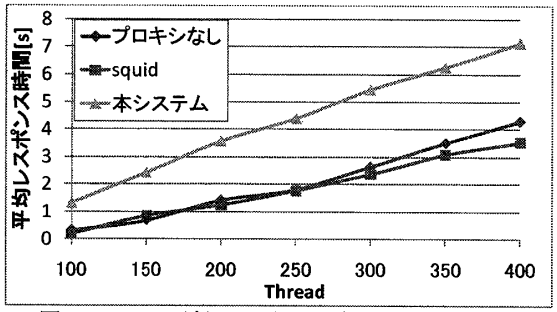

図 7　HTTPS 通信における平均レスポンス時間

本実験では 73 種類の Web メールにおいて 71 種類の Web メールを規制することに成功した（規制率=97%）．また，HTTP 通信時の本システムのレスポンス時間はプロキシなしに比べ大きいが，squid と比べるとその差は最大（同時リクエスト数=250）でも 0.4s であり測定したリクエスト数の範囲では同程度とみなすことができる．同様に HTTPS 通信の場合もプロキシなしおよび squid に比べ本システムのレスポンス時間は大きいが，これはクライアントおよびサーバからのデータに対する暗号化・復号化処理による時間と考えられる．

5 考察

(1) 本システムは HTTP 通信と共に HTTPS 通信を利用する Web メールに対応することでシステムの有用性をより高める結果となった．また，規制実施のタイミングを POST リクエスト受理後に変更することで図 3 のような Web ページであってもクライアントはそれを取得することができ，ユーザの作業効率を妨げる可能性を大幅に減らすことができる．

(2) 本システムのフィルタリング作業は，squid との比較からその処理時間は十分に小さく，全体のレスポンス時間から考えると無視できる程度であると考えられる．しかし，図 7 で示すように HTTPS 通信の場合は同時リクエスト数が増えるにつれ，暗号化・復号化処理によりレスポンス時間が増大する．そのため，評価実験と同程度の HTTPS リクエストを扱う環境では，SSL アクセラレータ等の高速な暗号化・復号化が可能なハードウェアで復号化を行った後，本システムでフィルタリングする等の検討が必要である．

(3) 機能拡張である Web メールの規制実施タイミングの改良は情報漏洩において必ず避けなければならない不正情報の通過（False Negative）を引き起こすことなく，ユーザにとって不適切な情報遮断（False Positive）を低減させることが可能であるため有効である．

6 おわりに

現在，本システムを大学内外の環境において実際に利用してもらう計画に進めている．

参考文献

[1] 浦川順平，鈴木信雄，鈴木健二： 情報漏洩を防ぐ Web メールフィルタリング手法の提案と設計， 情報処理学会全国大会第 71 回全国大会，No. 5E-1，pp. 331-332 (2009)

[2] 浦川順平，鈴木健二： Web メールの手順解析に基づくフィルタリング手法の 拡充提案，DICOMO2009 シンポジウム， 7C-2，pp. 1465-1472 (2009)

[3]The Apache Jakarta Project, "Apache JMeter" http://jakarta.apache.org/jmeter/