横浜国立大学におけるウェブホスティングサービス

A Web Hosting Service in Yokohama National University

志村俊也 †， 吉岡克成 ‡， 徐浩源 †， 牧田大佑 ‡， 星澤裕二 ‡ *， 松本勉 ‡
Toshiya SHIMURA †， Katsunari YOSHIOKA ‡， Haoyuan XU †，
Daisuke MAKITA ‡， Yuji HOSHIZAWA ‡ *， Tsutomu MATSUMOTO ‡

tshimura@ynu.ac.jp， yoshioka@ynu.ac.jp， haoyuan@ynu.ac.jp
makita-daisuke-jk@ynu.ac.jp ， yuji_hoshizawa@securebrain.co.jp， tsutomu@ynu.ac.jp

† 横浜国立大学　情報基盤センター
‡ 横浜国立大学　大学院環境情報研究院
* 株式会社セキュアブレイン

† Information Technology Service Center， Yokohama National University
‡ Graduate School of Environment and Information Science， Yokohama National University
* SecureBrain Corporation

概要

横浜国立大学情報基盤センターでは、利用者向けサービスとしてウェブホスティングサービスを提供している。また、同大学「情報・物理セキュリティ研究拠点」及びその共同研究者である(株)セキュアブレイン社の協力の下、ホスティングしているウェブサイトのコンテンツ検査を定期的に実施している。本稿では、ウェブホスティングサービスの概要と運用状況、およびウェブコンテンツ検査について紹介する。

キーワード

ウェブホスティング， クライアントハニーポットシステム， 法人向け gred セキュリティサービス

1. はじめに

横浜国立大学情報基盤センターでは、学内利用者向けにウェブホスティングサービスを提供している。ホスティングサイト総数は、2012 年 6 月 30 日時点で 171 サイトであり、その内訳は、表 1 に示す通りである。ここで、本学公式ウェブサイト www.ynu.ac.jp （以後、全学ウェブと呼ぶ） は、同じコンテンツを学内からのアクセス用と学外からのアクセス用の 2 つに負荷分散してホスティングしているので 2 サイトとしてカウントしている。負荷分散は、学外向け DNS サーバと学内向け DNS サーバに対して異なる A レコードの値を登録することで実現している。一方、利用者自身（研究室等）でウェブサーバを運用しているサイトは 43 サイトであるので、本学の

全ウェブサイトの約 80%が、当センターのウェブホスティングサービスを利用して公開していることになる。

	ウェブサイトの種類	サイト数
1	横浜国立大学公式ウェブサイト (www.ynu.ac.jp， 全学ウェブ)	2
2	部局・専攻・学科等の学内組織	71
3	委員会や部局横断的プロジェクト	12
4	GCOE、科研費等の政府の支援を受けているプロジェクト	6
5	研究室・学内研究グループ	76
6	本学主催の学会・研究会	4

表 1 ホスティングサイトの内訳

当センターでは、単にウェブホスティングサービスを提供しているだけでなく、本学環境情報研究院の研究者を中心とする本学の「情報・物理セキュリティ研究拠点」、及びその共同研究者である（株）セキュアブレイン社[1]の協力の下、ホスティングサイトのコンテンツ検査も行っている。本稿では、当センターのウェブホスティングサービスの概要と運用状況、及びホスティングサイトのコンテンツ検査について紹介する。

2. ホスティングサービス概要

2.1 システム構成

ウェブホスティングサーバは 3 台構成で運用している。サーバの 1 台あたりの仕様は、表 2 に示す通りである。

機種	富士通社製 PRIMERGY RX300 S5
CPU	Intel Xeon X5570 (クアッドコア/2.93GHz) × 2
RAM	12GB
HDD	1TB (RAID 5)
OS	RedHat Enterprise Linux 5.8
Httpd	Apache 2.2 系
PHP	PHP5.3 系
PHP 以外	RedHat 搭載版
データベース	MySQL (RedHat 搭載版)

表 2 ホスティングサーバの 1 台あたりの仕様

ホスティングサイトの 3 台のサーバへの振り分けは、目的・アクセス頻度・サーバへの負荷・管理面等を総合的に考慮して、下記のように振り分けている。

[1 台目 （以後、www1 と呼ぶ）]
全学ウェブ（学内からのアクセス用）と学内組織（表 1 の 2 番）の公式ウェブサイト 71 サイトを収容。これらのサイトは、SSL 証明書を利用した https での接続を提供する場合があるため、各サイトに固有の IP アドレスを割り当て、DNS 名 ←→ IP アドレスが 1 対 1 で対応する運用としている。

[2 台目 （以後、www2 と呼ぶ）]
表 1 の 3～6 番 のサイト（合計 98 サイト）を収容。これらのサイトは、SSL 証明書を利用しないため、1 つの IP アドレスに複数の DNS 名を対応させる仮想ホスト方式で運用している。

[3 台目 （以後、www3 と呼ぶ）]
全学ウェブ（学外からのアクセス用）のみを収容。学外向け本学公式ウェブサイトであるので、IPv6 Native でもアクセスできるようにしている。

1 サイトあたりの割り当て Disk 容量は、概ね 5GB 以内としている。また、コンテンツのアップロードに関しては、不正アップデートを防ぐため、ウェブサイト管理者が申請時に届け出た IP アドレス (学内端末 2 台まで指定可能。ただし、学外から指定端末へのリモートアクセスは禁止）からの SFTP（SCP, SSH）アクセスのみを許可するという対策を取っている。

2.2 CPU ロードアベレージとネットワークトラフィック

図 1, 2, 3 に示しているのが、それぞれ www1, www2, www3 の 1 分間及び 5 分間平均の CPU ロードアベレージである。縦軸の単位はパーセント（%）であり、緑色が 1 分間平均、青色が 5 分間平均を表している。図 4, 5, 6 に示しているのは、それぞれ www1, www2, www3 の 5 分間平均のネットワークトラフィックである。縦軸の単位は kB/s であり、緑色がサーバ ➔ クライアントへのダウンロードトラフィック、青色がその逆のアップロードトラフィックである。図の横軸は時系列を示しており、期間は、2012 年 7 月 19 日（木）01 時～7 月 20 日（金）09 時である。図 1, 2, 3 の CPU ロードアベレージについては、図の左側が 7 月 19 日 01 時に対応しているが、図 4, 5, 6 のネットワークトラフィック については、図の右側が 7 月 19 日 01 時に対応している。CPU ロードアベレージ、ネットワークトラフィックの双方とも、平日の振る舞いは、曜日によって大きな違いはない。

CPU ロードアベレージは、一時的に 100%を超える場合もあるが、平均的には、30%以下であり、余裕のある安定した運用となっていることがわかる。

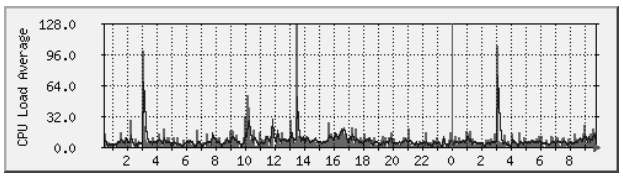

図 1　www1 の CPU ロードアベレージ

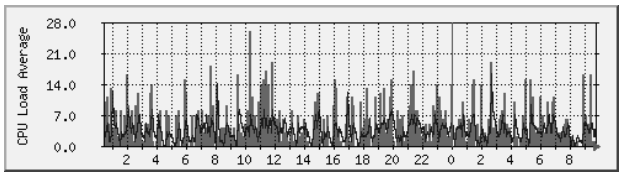

図 2　www2 の CPU ロードアベレージ

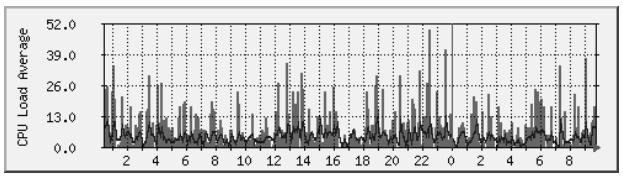

図 3　www3 の CPU ロードアベレージ

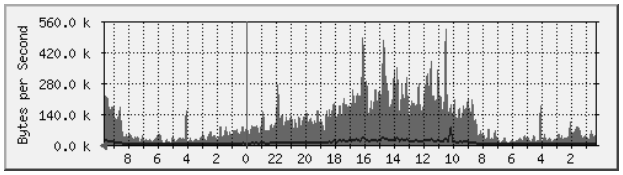

図 4　www1 のネットワークトラフィック

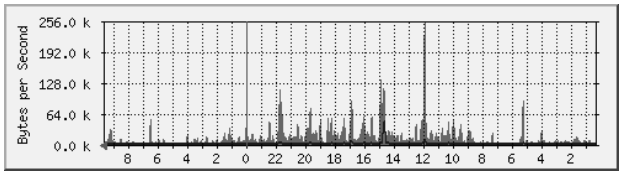

図 5　www2 のネットワークトラフィック

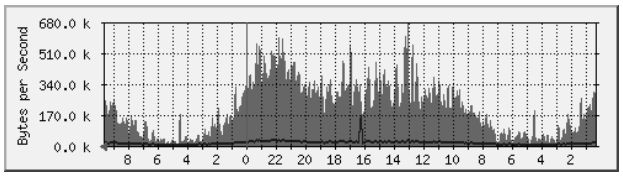

図 6　www3 のネットワークトラフィック

ネットワークトラフィックについては、平均的には、全学ウェブ（学外からのアクセス用）である www3 が最も多いが、それでも 700kB/s 以下である。紙面の都合上、図は記載していないが、休日のネットワークトラフィックは、www1, www2 が平日の半分以下に減少するのに対して、www3 は平日と大きくは変わらないという特徴がある。

2.2 アクセス状況

3 台のウェブサーバに対するアクセス状況を大まかに把握するため、各サーバにリクエストされた GET メソッド（クライアントからのファイルのダウンロード要求）の 1 日当たりの平均数を調べた結果を表 3 に示す。GET メソッドの集計期間は、2012 年 7 月 8 日（日） ～14 日（土）の 1 週間であり、平日平均は、7 月 9 日 ～13 日、休日平均は、7 月 8 日と 14 日 の平均数である。

	GET メソッド数（平日平均）	GET メソッド数（休日平均）
www1-www3 合計	1,417,723 (100%)	878,792 (100%)
www1-A 学内	155,030	21,274
www1-A 学外	229,379	174,451
www1-A 合計	384,409 (27.1%)	195,725 (22.3%)
www1-B 学内	263,833	21,176
www1-B 学外	896	934
www1-B 合計	264,729 (18.7%)	22,110 (2.5%)
www2 学内	17,090	1,095
www2 学外	36,463	22,083
www2 合計	53,553 (3.8%)	23,178 (2.6%)
www3 学内	700	213
www3 学外	714,332	637,566
www3 合計	715,032 (50.4%)	637,779 (72.6%)

表 3　各ウェブサーバに対する GET メソッドのリクエスト数の一日当たりの平均数。集計期間は、2012 年 7 月 8 日（日） ～ 7 月 14 日（土）の 1 週間。

表 3 の中で、www1-A は、www1 に対する GET メソッドの内、全学ウェブへのアクセスを除いた 71 サイトの分であり、www1-B は、www1 に対する GET メソッド

の内、全学ウェブに対する GET メソッドを示す。表中の「学内」は、学内からのアクセス、「学外」は学外からのアクセスを意味する。全学ウェブに対するアクセスは、学外からは www3 へ、学内からは www1 へ DNS ベースで振り分けているので、URL 名でアクセスする限り、www1 への学外からのアクセス、及び www3 への学内からのアクセスは発生しないはずであるが、IP アドレスベースで直接アクセスされる場合もあるためゼロにはならない。表 3 からわかることは以下の 5 つである。

① GET メソッドの一日当たりの平均数は、平日は約 142 万回、休日は約 88 万回（平日の約 60%）である。
② www1-A, www1-B, www2, www3 の平日平均数はそれぞれ、全体の 27.1 %, 18.7 %, 3.8 %, 50.4 %である。つまり、平日の GET メソッド全体の約半数は、学外からの全学ウェブに対するリクエストによるものである。
③ www1-B と www3 の平日平均の合計は、全体の 69.1%であり、全 GET メソッドの約 7 割が全学ウェブへのリクエストとなっている。
④ 学内からのアクセスにおける www1-A, www1-B, www2 の休日平均は、それぞれ、21,274、 21,176、1,095 であり、これは平日平均の 13.7%, 8.0%, 6.4% に相当し、大幅に減少する。
⑤ www1 全体、www2 の休日平均は、それぞれ、217,835、23,178 であり、平日平均と比べて半分以下の値となっているが、www3 の場合は、637,779 であり、平日平均と大きくは変わらない。これは、ネットワークトラフィックの振る舞いと一致している。

3. コンテンツ検査

3.1 クライアント型ハニーポットシステムによるコンテンツ検査

ウェブサイトが改竄されていないかどうかの確認は、ウェブサイト管理者の責務として利用規約に定めているので、本来であれば情報基盤センターが実施する必要はない。しかし、ウェブサイトのコンテンツをウェブデザイン会社に外注しているサイトが多数あり、JAVA, Perl, PHP, SQL 等のプログラミング言語で記載された SSI, CGI ファイルまでウェブサイト管理者自身で把握・検査をすることは事実上不可能である。また、コンテンツ管理システム（CMS）を利用してウェブブラウザ経由でコンテンツの更新を行う場合、CMS 用ユーザー ID/パスワード が盗まれてしまうと、正規のウェブサイト管理者としてアクセスされてしまうので、コンテンツ自体に脆弱性がなくても改竄されてしまう危険性がある。

そこで、当センターでは、本学の情報・物理セキュリティ研究拠点及びその共同研究者である（株）セキュアブレイン社の協力の下、ホスティングサイトのコンテンツがサイト閲覧者を攻撃する悪性サイトに改竄されていないかどうかの検査を定期的に実施している。コンテンツ検査は 2 種類の方法で実施しており、1 つは、クライアント型ハニーポットシステム（以後、CHP システムと呼ぶ）を利用した検査であり、もう一つは、セキュアブレイン社の「法人向け gred セキュリティサービス（3.2 章で説明）」を利用した検査である。

CHP システムとは、同システムを搭載した PC から検査対象ウェブサイトにアクセスし、「マルウェア等の意図しない不正なファイル」がクライアント PC 側に自動的に作成（ダウンロード）されないかどうかを検査する仕組みのことである。CHP システムのプログラムは、セキュアブレイン社が開発したものをベースにし、情報・物理セキュリティ研究拠点側で改良を加えたものである。CHP システムによるウェブアクセスの巡回方法は以下のとおりである。

① 検査対象ウェブサイトのトップページ（通常は、index.html ）にアクセスし、そこから再帰的にリンクを抽出し、アクセスしていく。
② リンクは、Anchor タグの HREF 属性で、相対パスで指定されているものを抽出する。絶対パスを抽出しないのは、検査対象サーバ以外の外部 URL まで 検査し、 検査が際限なく続く可能性を防ぐためである。
③ 各ページは、閲覧後 60 秒の待機時間を空けて、次のページの検査をする。60 秒間待機するには、クライアントからのアクセス後、時間をおいてマルウェアのダウンロードを開始させるコンテンツが存在するためである。

CHP システムは、ウェブブラウザやプラグインの脆弱性を利用した PC への攻撃を実行するような不正プログラムやマルウェアが検査対象ウェブサイトに仕掛けられているかどうかを検知するためのシステムなので、CHP システムを搭載した検査用クライアント PC は、脆弱性対策が適用されておらず、マルウェア対策ソフトウェアも搭載していない状態の PC を使用する必要がある。具体的には、Windows XP Professional SP2 で Microsoft Update が全く適用されていない状態の PC で行う。ウェブブラウザは IE6 を使用する 。このため、検査用 PC は、学内 LAN ではなく、商用の ISP に接続し、その上で学外からホスティングサイトの検査を実施することとしている。

このCHPシステムを用いた検査において、現時点では、ホスティングサイト上に不正なコンテンツは発見されていない。CHPシステムを使用して巡回調査中の検査PCの画面を参考のため図7に示しておく。

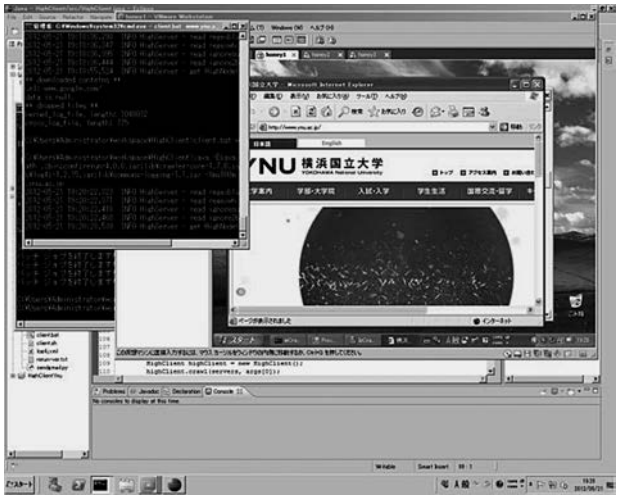

図 7　CHP システムを使用して巡回調査中の検査 PC の画面

3.2　セキュアブレイン社の法人向け gred セキュリティサービスによる検査

CHPシステムによるコンテンツ検査は、マルウェア等のクライアント攻撃型プログラムがウェブサイト上に仕掛けられているかどうかを検知する仕組みであるので、『フィッシングサイトの埋め込み』や『SQLインジェクション』といった改竄・危険性を検知することはできない。このため、当センターでは、CHPシステムに加えて、セキュアブレイン社の「法人向けgredセキュリティサービス（以後、gred サービスと呼ぶ）」を利用したコンテンツ検査を1ヶ月に一回の割合で実施している。gredサービスは、フィッシングサイトの埋め込みやSQLインジェクションといった不正コンテンツをも検知するシステムである。検査方法は、CHPシステムとは違い、シグネチャやルールベースでの検知となっている。このgredサービスは、同社が有償サービスで一般向けに提供しているものであるが、共同研究の関連もあり、セキュアブレイン社の御厚意により、同サービスを無償でご提供頂いている次第である。この検査においても、現時点ではホスティングサイト上に不正なコンテンツは発見されていない。

なお、セキュアブレイン社が提供しているgredサービスには、この法人向けgredセキュリティサービスとは別に、個人向けの「gredでチェック」という無料サービスも存在する。この無料版gredもフィッシングサイトやアクセスすると危険なサイトを判定する機能がある。従って、わざわざ有償版のgredサービスで検査する必要はないのではないかと考える方もいるかもしれないが、この無料版 gred はあくまで、有償版の gred サービスのデモという位置づけであるため、必ずしも、信頼できる結果が得られるとは限らないことに留意しておく必要がある。その理由は、無料版gredに限らず、このような「公開型動的解析サービス（任意の利用者から情報を受け付け、解析し、その結果を利用者に提供するシステム）」の多くは、検査プログラムが検査対象の URL にアクセスする際の IP アドレスが容易に露見してしまい、検査対象ウェブサイト側で検査を回避する設定を施すことが可能であるためである[2]。信頼できる検査結果を得たいのであれば、有償版のgredサービスを受けるべきであろう。

4.　おわりに

本稿では、当センターが学内利用者に提供しているウェブホスティングサービスの概要と運用状況、及びホスティングサイトのコンテンツ検査に関する紹介を行った。最後に、現在取り組んでいる課題について述べておく。上記で紹介したとおり、CHP システムは、「意図しないファイルの作成」を検知するシステムであるが、実は、その他に、クライアント PC 側の「レジストリの改変」や「プロセスの起動」も検知することができる。しかし、どのようなレジストリ改変やプロセス起動を悪性とみなすかという判定基準がまだ検討段階であるため、これらの検査は行っていない。この判定基準については、現在、情報・物理セキュリティ研究拠点が課題として取り組んでいるところである。

参考文献

[1]　http://www.securebrain.co.jp/

[2]　笠間貴弘、織井達憲、吉岡克成、松本勉、情報処理学会論文誌、「公開型マルウェア動的解析システムに対するデコイ挿入攻撃の脅威」vol. 52,　No.9, 2761 - 2774　(2011)