負荷分散装置を用いたFirewallの二重化の実現とトランスペアレントプロキシの導入及びバックボーン回線負荷分散の実現

杉浦 徳宏† 山守 一徳†

本学では，Firewall 周りの負荷分散とバックボーンの負荷分散のために，負荷分散装置を導入した．まず，学外と接続する Firewall 装置の二重化(冗長化と負荷分散)を行った．実現にあたっては，Firewall を複数台の負荷分散装置で挟み込み，2 台の Firewall にトラフィックを分散させて稼働させる方式を採った．同時に，負荷分散装置の L4-7 スイッチ機能を利用して，http トラフィックを Firewall と並列に置かれたトランスペレントプロキシへ強制的に流すことによって，Firewall 装置の負荷の低減と，インターネット向け http トラフィックの低減を行った．さらに，別の負荷分散装置によってトランスペアレントプロキシからの http トラフィックを ADSL を含む複数の回線に分散し，メインバックボーンの負荷の分散を行った．導入した製品は，F5 社の BigIP と CacheFlow 社の CacheFlow である．また，各装置について二重化（冗長化と負荷分散）について検討し，装置の障害時に運用が停止することがないような構成とした．本文では，負荷分散装置の導入にあたって，冗長化と負荷分散の観点で行ったネットワーク設計の詳細を述べ，その評価を行う．また，発生した問題と各種の対策，及び今後の計画について報告する．

Realization of Dualized Firewall, Installing Transparent Proxy and Balancing Backbone Traffic by Load Balancers

SUGIURA TOKUHIRO† and YAMAMORI KAZUNORI†

The authors have installed load balancers for our firewall system. For load balancing of firewall machines and achieving the redundant system. Two firewall machines are set down between load balancers and balance traffic each other. Next, a transparent proxy that is also set down between load balancers. They drive http traffic to the transparent proxy by L4-7 switch function for reducing load of firewalls and compressing http traffic to the Internet. The load balancers are also utilized for dividing http traffic into several circuits including four ADSL ones to reduce traffic of the main backbone circuit. Also the authors take account of redundancy of load balancers themselves, so we use two in pairs i.e. four balancers totally. As a result, when any of them including firewalls and the proxy went wrong, whole traffic can run through some backup lines. This paper describes a method of network design considering load balance and redundancy, evaluation of the new system, problems in practice, possible countermeasures and the future plans.

1 はじめに

本学では，従来，情報処理センターに設置された1台のfirewallによって全学のインターネット間トラフィックの制御を行っていた．しかし，昨今のトラフィックの増大と，firewall のルールの増加によって，firewall の負荷が著しく増大し，遅延を招く結果となっていた．また，firewall が一台のため，ダウン時にはすべての学外アクセスが閉ざされるという問題も抱えていた．

また，昨今のトラフィックの増大に対して，インターネットバックボーンの増強が追いつかず，1日のうち数時間においてバックボーン回線が飽和し，極めてレスポンスの悪い状態になっていた．このためトラフィックの大半を占める http トラフィックを削減するためにプロキシを設置し利用促進に努めて来た．しかし，その利用にあたってはエンドユーザ自身が設定しなければならなかっ

† 三重大学情報処理センター
Information Processing Center, Mie Univ.

ため，利用率が上がらず削減効果は小さかった．

これらの問題を解決するために，

(1)firewall を 2 台に増やし負荷を分散させる

(2)インターネット向けトラフィックの約 8 割を占める http トラフィックを firewall に流さないようにする

(3)トランスペアレントプロキシを導入し，トラフィックの圧縮を行う

(4)ADSL 回線をバックボーン回線として導入し，http トラフィックを分散する

(5)二重化によって装置を停止せざるを得ないメンテナンスを容易にする

という方針を立て，これを実現するために負荷分散装置を導入した．また，負荷分散装置の導入にあたって，負荷分散と冗長化の観点から設計を行い，機器障害に対して堅牢なシステムとなるよう構成した．以下に，その詳細ついて述べる．

2 負荷分散と冗長構成

図 1 に，設計した firewall 周辺の全体構成図を示す．LBL1，LBL2，LBU1，LBU2 が負荷分散装置である．LBL1 と LBL2，LBU1 と LBU2 はペアで冗長構成が組まれているが，機能的には，一台の場合となんら変わらない．これらの負荷分散装置に挟まれた領域に，firewall 機 FW1, FW2 及び，トランスペアレントプロキシ TP を並列に配した．LBL1，LBL2 の下流に SM(Summit24) があり，その下側が学内 LAN である．学内 LAN のバックボーンは多重化されているため，複数のルータ(R1～R3)が配置されている．LBU1,LBU2 から上部がバリアセグメントである．バリアセグメントには，各回線のゲートウェイ用 PC1～PC6 とルータが配置されている．RS は，SINET 上流へのルータである．バリアセグメントの一部を除いて，各機器のインターフェースはすべて SM に VLAN を切って接続している．

実際に導入した負荷分散装置は，F5 Newtworks[1]の BigIP Cache Controller（以下，BigIP と呼ぶ）である．これは，IP レベルの負荷分散機能と L4-7 スイッチとしての機能を合わせもった負荷分散措置である．内部は PC であり，スペックとしては，CPU が PentiumIII 500Mhz，メモリ 512MB である．OS として，BSD/OS をベースとしたものが使われている．

トランスペアレントプロキシ TP は，CacheFlow[2]の CacheFlow 600 である．負荷分散装置と同じく PC ベースであるが詳細は不明である．キャッシュ機能とともにコンテンツアクセラレータとしても機能する．

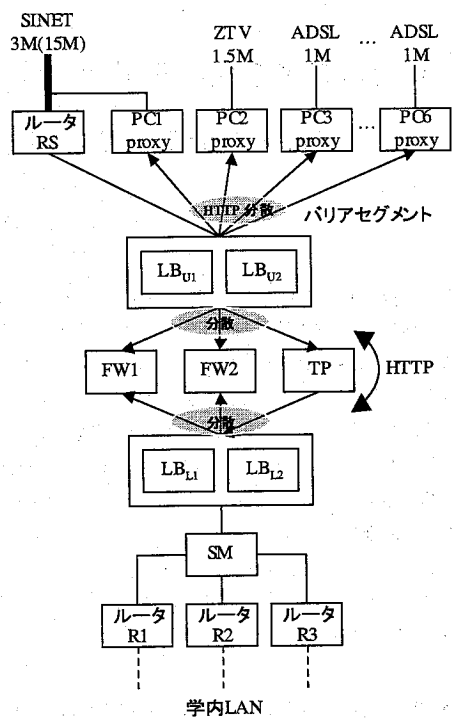

図1 全体構成図

各バックボーン回線に接続されたプロキシ用の PC は Plathome 製の PentiumIII 800Mhz，メモリ 1GB を搭載した一般的な PC である．

以下，トラフィック別に負荷分散と冗長化の観点からその経路について述べる．

2.1 http 以外のトラフィック

http 以外のトラフィックについては，上下の負荷分散装置 LBL1,2 及び LBU1,2 によって 2 台の firewall FW1, FW2 に均等に負荷分散される．上下両方向のトラフィックを分散させるために，負荷分散装置は FW の上下両方に必要である．

FW1, FW2 は負荷分散構成と同時に冗長構成にもなっているが，これらの機能は，負荷分散装置によるものである．すなわち，平時には，負荷分散装置は，両方に均等にトラフィックを分散させているが，片側の FW がダウンした場合，もう一方にすべてのトラフィックを送る．FW1, FW2 のダウン判定は，負荷分散装置からのポーリングによって行われており，ダウン判定された場合に

は速やかに前述の切り替えが行われる．また，ダウンした装置が復帰した場合には速やかに振り分け対象に戻され，トラフィックを等分する通常の運用形態に戻る．

2.2 http トラフィック

学外からの http リクエストは，負荷分散装置 LBL1,2 の L4-7 スイッチ機能によって，トランスペアレントプロキシ TP に向けられる．トランスペアレントプロキシ TP は，リクエストを上部の負荷分散装置 LBU1,2 によるバーチャルサーバにフォワードする．LBU1,2 は，このリクエストを受け PC1～6 に振り分ける．PC1～6 は，squid によるアプリケーションプロキシで http リクエストを受け，各回線より取得する．回線として，SINET(メインバックボーン)，ZTV 回線，ADSL 回線 4 本の計 6 回線である．SINET 回線は，ATM メガリンクサービスによる専用線で SINET 名古屋ノードまで 3Mbps(2001 年 11 月より 15Mbps に増速されている)で接続されている．ZTV 回線は商用プロバイダ ZTV[3]の光回線で容量は 1.5Mbps である．ADSL 回線は，アクセス回線としてフレッツ ADSL を利用し 2 種類の商用プロバイダ経由で接続している．最終的に http トラフィックはこのいずれかの回線より送受信されることになる．回線配分比は，等分で運用している．

http トラフィックについても，冗長構成が採られており，トランスペアレントプロキシ TP がダウンした場合には，FW1,2 にすべての http トラフィックが振られ，運用を続けることができる．回線用プロキシがダウンした場合には，その回線に対する振り分けは行われなくなるが，多重化されているため問題はない．

2.3 負荷分散装置の冗長化

BigIP は，BSD/OS を載せた PC であり，ハードディスクなどの装置を持つ．このため，故障は避けられず対策を講じておく必要がある．そこで，BigIP を 2 台構成にすることによって冗長化することにした．BigIP の冗長化では，一台がアクティブとなって動作し，もう一方の装置は，ホットスタンバイとなる．従って，2 台構成であることを利用して，BigIP そのものの負荷を分散させるというようなことはできない．

スタンバイ側は常にアクティブ側を監視し，障害時にはスタンバイ側が速やかにアクティブに切り替わる．アクティブ側とスタンバイ側で常にコネクションテーブルを共有しているため，切り替えにおいてコネクションが切れることはなく，ユーザは切り替わったことを全く意識することがない．

3 評価

3.1 Firewall の負荷分散効果

firewall には富士通製 Workstation GP-400S に OS として Solaris2.6 を載せたものに firewall ソフトウェアとして Firewall-1 をインストールして利用している．負荷分散の効果としては，データを取っていないため定量的な評価が難しい，定性的に次のようなことが挙げられる．

(1)firewall の負荷が減り遅延が減少した

従来，ping などでパケットロスなどが起こっていたものが全くなくなった．

(2)firewall のログが減少し，不正利用等の発見が容易になった

従来，http のコネクション数が莫大であったためログの大半が http コネクションであり，処理しきれない程に膨らんでいた．検索等をかけるのにも大変な処理能力を必要としていたが，現在では容易に検索等をかけることができるようになった．

(3)firewall を片側ずつ安易に停止でき，メンテが容易になった

従来，停止を伴うメンテナンスについては，全学外通信が途絶えてしまうため，事前に学内へその旨広報し，正確に日時を決めて実行する必要があった．しかし，現在，片側ずつ止められるようになったため，緊急のメンテナンスの際にも告知の必要がなく，きわめて容易に運用できるようになった．

表 1 キャッシュ効果

	objects	bytes
from cache	61.4%	38.4%
from source	19.3%	36.4%
Non-cacheable	19.1%	25.1%

3.2 トランスペアレントプロキシの効果

表 1 に，TP のキャッシュヒット率を示す．オブジェクトベースでは，かなりのヒット率を示しているが，転送量の面ではキャッシュヒットは低い．これは，キャッシュされ易い通常のホームページを構成するようなオブジェクトは比較的小さく，キャッシュされない数 10MB 以上の大きなファイルのダウンロードによる転送量が支配的にな

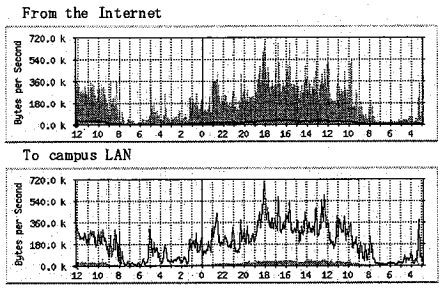

図2 TPトラフィック

っているためと思われる．

図2にTPのトラフィックを示すが，その傾向はこのグラフからも見て取れる．上図がTPのインターネット側のインターフェースで，下図が学内LAN側のインターフェースのトラフィックである．この上下のトラフィックの差が転送量からみたキャッシュの効果を表しているということになる．しかし，見てわかるとおり，上下の図でトラフィックにほとんど差がない．定量的には，図の一日の平均でインターネットからTPへは201kB/s，TPから学内へは188kB/sであった．これは，トラフィックが増加してしまっており当初期待していたものとはまったく逆の結果である．数ヶ月単位の平均でもほぼ同じ結果となっている．

この原因としては，TPが保持しているコンテンスの更新を行うために，必要以上にインターネット側にアクセスしていることと，CacheFlowの持つコンテンツアクセラレータとしての機能から先読みのためにトラフィックが増えていることが理由として考えられる．

また，キャッシュの効果を測定するために，TPで全くキャッシュを行わない設定にした場合の結果を表2に示す．キャッシュを行わない場合，一日の転送量が減少してしまっている．本来，TPから学内の転送量は変化しないはずであり（同一日に調査することはできないので必ずしも変化しないとはいえない），インターネットからTPへのトラフィックがどう変動するかを測定することによってキャッシュの効果を計るのが目的であった．しかし，本実験により得られた総転送量が減ってしまうという結果から，キャッシュしないことによる負荷増量に対してどこかにボトルネックがあると考えられる．まず，キャッシュしないことにより，小さなリクエストがすべてインターネット側に出て行くために，トータルでスループットが低下したことが考えられる．実際，TPのactive connection数や，TPからのリクエストを受ける上側のプロキシのリクエスト数は数10%程度増加している．しかし，主原因がどこであるかは現在調査中である．

また，実際のインターネットへのアクセスのレスポンスについては定量的なデータがないが，体感的にはキャッシュありの場合の方がずっとレスポンスがよく，快適である．これは，保持キャッシュの自動更新や先読み機能が有効に働き，高いキャッシュヒット率を達成している結果と言える．これらの恩恵を受けながら，転送量は微増のままであると考えることもでき，当初の目的とは異なるものの，回線に余裕がある現在では結果としてよい方向だと言える．

表2 非キャッシュ時トラフィック(一日平均)

	Cache	No-cache
From the Internet	191kB/s	184kB/s
To campus LAN	188kB/s	164kB/s

3.3 ADSL回線の効果

従来，本学では，SINET 3MBpsをメインバックボーンとし，商用回線であるZTV 1.5Mbpsをhttp専用としてプロキシを立て利用していた．しかし，SINET回線は一日の数時間に渡って飽和するような状況であり，極めて危機的な状況であった．一方，http用のZTV回線は，プロキシの利用率が低いために十分に利用されていたとはいえない状態であった．そこで，負荷分散装置によってhttpトラフィックを強制的に拾い上げ，複数の回線に分散させることにした．回線としては，従来からのZTV専用線に加えて，コストパフォーマンスに優れるADSL回線を4回線導入し増強した．

ADSL回線は，収容局からの距離や物理的な回線品質によってスループットが変化する，また収容局からプロバイダまでの地域IP網内がベストエフォートである，といった制約を持っている．このため低価格ではあるが，実際にどの程度実用になるかについては全くの未知であった．本学の場合，ADSL回線単体としては，いずれも最高1.2Mbps，平均1Mbps程度であることがわかった．また，複数回線の導入にあたっては，同時利用によって地域IP網の容量を食い合ってしまい，結果として回線数分のスループットが得られないことが懸念されたが，予備実験より，そうした影響はほとんどないことが確かめられた．また，頻繁に回線断が起こるというようなことが報告され

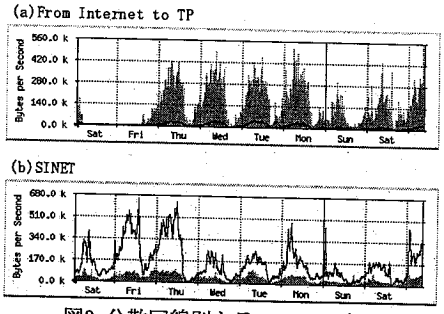

図3 分散回線別トラフィック変化

Mon: 全回線利用　　Thu: SINETのみ
Tue: キャッシュなし　Fri: FW経由
Wed: ADSL系のみ　　(朝9時に回線切り替え)

ていたが，本学の回線の場合そのようなことは全くない．

実際に，利用する回線を変化させた場合のトラフィックの様子を図3に示す．図3(a)が，TPのインターネット側インターフェースのトラフィック，すなわち，TPへ入ったhttpトラフィックを示す．図3(b)が，SINETのトラフィックである．一週間にわたって曜日毎に次のように変化させた．月曜は，通常の運用形態で全回線を利用した(SINETへのhttp配分比は他回線の2倍である)．火曜は，全回線を利用するがTPでキャッシュしない設定，水曜はADSL系(ZTV回線含む)のみを利用，木曜は，SINETのみ，金曜は，httpトラフィックをTPではなくFW1,2側に流しSINETだけを利用した．図3(a)より，インターネット側から流れ込むhttpトラフィックは各曜日ともほぼ同じであったことがわかる．図3(b)で水曜のトラフィックを基準として，そこからの減少分が，ADSL系によって補助された分である．SINETだけを利用した場合と，ADSL系だけを利用した場合に学内へのトラフィックにほとんど違いがないことから，ADSL系回線がSINETの代わりとして十分に機能したことがわかる．

4 問題点

導入からこれまで運用してきた中で発生した問題として以下のものがある．

(1)BigIP (負荷分散装置)は，動的な経路情報が受けられない
(2)負荷分散プロセスのみがダウンした場合に，スタンバイ側へ切り替わらない
(3)NATホストとの間の無駄トラフィックの発生
(4)IPSec等すべてのIPプロトコルを負荷分散できるわけではない
(5)特定URLについては，トランスペアレントプロキシを経由しないようにする必要がある
(6)Firewallやトランスペアレントプロキシそのものへのアクセスについては個別対応する必要がある

以下，それぞれについて詳細を述べる．

(1)BigIPは，動的な経路情報が受けられない

本学の経路情報は，ルータ間はOSPFによって交換され，ルータより末端はRIPによって配信されている．しかし，BigIPは，標準の状態ではこれらの動的な経路情報を受けることができない．OSは，BSD/OSであるのでgatedを動かすことは可能であるが，意図した通りに動くかどうかは全くの未知であるため，実際の利用はためらわれた(routedは標準装備せず)．本学では，BigIP直下の学内バックボーンが最大で3重化された経路をもつため，OSPFが受けられることが必須である．結局，BigIPで経路情報を受けることは諦め，基幹ルータとBigIPの間に，SM(Summit24)によってセグメントを一つ追加し，動的な経路制御はSMで行うことにした．

(2)負荷分散プロセスのみがダウンした場合に，スタンバイ側へ切り替わらない

BigIPの冗長化では，片側がホットスタンバイ状態でアクティブ側を監視しており，アクティブ側の障害を検出した場合には速やかにに切り替わる．この障害判定としては，(i)アクティブ側がping(ICMP)に応答しない，(ii)リンクダウン，の2種類によって行われる．負荷分散機能は，kernelといくつかのプロセスによって実現されているため，負荷分散プロセスのみがダウンしkernelが生きている場合には，pingに応答してしまう．この場合，障害とは判定されないため切り替わらない．ハード障害の場合には，kernelも応答しなくなるため，このような障害検出方法で問題ないと思われるが，実際にはプロセスのみが止まってしまうことも考えられるため，よりロバストな障害検出が必要だと思われる．今後のバージョンアップで改善されることを期待したい．

(3)NATホストとの間の無駄トラフィックの発生

BigIP導入後，負荷分散装置LBL1,2を挟んでトランスペアレントプロキシTPと一部のNAT(IP masquerade)ホストの間で，syn, ackが莫大に増殖するという問題が発生している．tcpdumpで見てみると，NATホストからのsynに対して，BigIP

が複数の ack を返し，さらに，NAT ホストは syn を再送するという状態が頻発している．これにより，トラフィックがおよそ 10 倍程度にまで増大している．現在メーカにて調査中である．

(4) IPSec 等すべての IP 層プロトコルを負荷分散できるわけではない

BigIP は，IP 以外の全ての IP 層プロトコルを負荷分散できるわけでない．例えば，IPSec については負荷分散することができず，forward 動作のみ，すなわち特定の FW を通過する設定となってしまう．本学には，学外の組織と IPSec によって VPN を張っているユーザが存在するため，IPSec を負荷分散できることが望ましいが，トラフィック的にはわずかであるためあまり問題ではない．しかし，これは別に厄介な問題を孕んでいる．すなわち，forward 動作の場合，2 台の firewall のうち片一方だけが IPSec を受け持つことになるため，その一台を停止させることができないということになる．当初の目的(5)としてメンテナンスのために容易に装置を止めることができるようなシステムを目指していただけに，これは非常に問題である．

この対策として，BigIP の時期バージョンの OS(4.0)では，すべての IP 層プロトコルを負荷分散させることができるようになるということで，バージョンアップに期待するところである．

(5) 特定 URL については，トランスペアレントプロキシを経由しないようにする必要がある

本学のプロキシシステムの場合，前述のように http リクエストは，最終的に複数ある回線用プロキシのいずれかから出て行くことになる．このため，リクエストの度に使用されるソース IP アドレスが変わることになり，一部の有料サイトなどで認証が通らず問題となる．また，トランスペレントプロキシ経由では，表示が崩れたり，全くアクセスができないなどの問題が発生する場合もある．このため，URL によってはトランスペアレントプロキシを迂回して firewall 側に流す必要がある．BigIP は，L4-7 スイッチ機能を有しているので，http リクエストのヘッダを読んで振り分け条件を変更することができる．そこで，この機能を利用して振り分けを行っている．

(6) Firewall やトランスペアレントプロキシそのものへのアクセスについては個別対応する必要がある

負荷分散装置 LBL1,2, LBU1,2 に挟まれた領域に存在するホスト（FW1, FW2, TP）そのものにアクセスしたい場合もある．例えば，FW1 にリモートからルールを追加したり，トランスペレントプロキシの設定をリモートから変更する場合など．また，逆に，内部のホストからメールを送ったり，時刻合わせを行うためにアクセスする場合などもある．これらのアクセスは，そのままでは負荷分散装置を通過するときに，firewall 側に負荷分散されてしまうため問題が起こる．この対策としては，各利用プロトコルごとに個別に負荷分散装置に経路を設定する他はなく，煩雑である．

5 おわりに

負荷分散装置とトランスペアレントプロキシの導入により，当初目標としていた，firewall の負荷分散について実現することができた．また，http トラフィックを拾い上げ，トランスペアレントプロキシを経由させ，さらに複数の回線に分散させることができた．回線として ADSL 回線を複数導入し増大し続ける回線需要を十分に満たすことができた．しかし，トランスペアレントプロキシによるトラフィックの削減については，当初の期待と異なり，効果がほとんどなかった．しかし，レスポンスの向上には貢献しており，回線容量に余裕がある現在ではトラフィック削減をする必要はなく，コンテンツアクセラレータとして十分に機能しているといえる．

今後の課題としては，http 以外のトラフィックの ADSL 回線への分散が挙げられる．現在，ブロードバンド化によってストリーミングや p2p などの http 以外の大容量トラフィックが急激に増加しており，これらの需要にこたえるためには http 以外についても分散させる必要がある．また，現在，TP から上部の各回線用プロキシへと多段接続になっているため，ここがボトルネックになっている可能性がある．これらに対処するために，回線用プロキシを NAT ルータに変更し，プロキシを TP のみの 1 段にすることよって高速化することを計画している．

参考

[1]F5 Network, Inc. http://www.f5.com
[2]CacheFlow Inc. http://www.cacheflow.com
[3]株式会社 ZTV, 三重県津市あつの台 4-7-1