IIJ Technology

ロードバランシング技術

～高負荷に耐えるシステムの構築～

IIJ Technology Inc.

株式会社アイアイジェイテクノロジー
プロフェッショナルサービス部
川本 信博
(kawamoto@iij-tech.co.jp)

Agenda

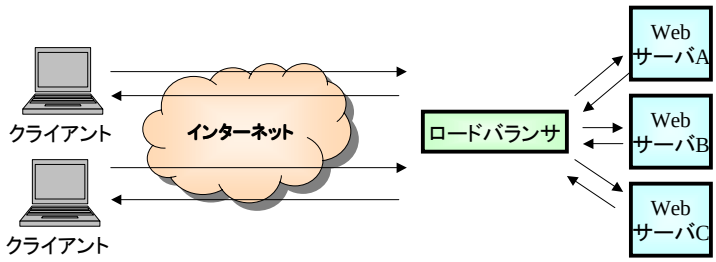
IIJ Technology Inc.

- ロードバランサの必要性
- ロードバランサの基本機能
- システム構築

ロードバランサとは

IIJ Technology Inc.

- 一般的に負荷分散装置、ロードバランサともL4/L7スイッチとも呼ばれている。
- ネットワークとサービスを提供しているサーバとの間に接続され、WWWなどのアクセスを動的にサーバに負荷分散を行う装置のこと。



ロードバランサの必要性

IIJ Technology Inc.

インターネットシステムに求められる性能

IIJ Technology Inc.

- 速いサイト ➔スケーラビリティの向上
  - アクセスの集中によるサイトのレスポンスの低下を防ぐ。
  - 8秒ルールを守る。(ブロードバンド化によりユーザの要求はさらにシビアになっている。)
- 落ちないサイト ➔アベイラビリティの向上
  - サイトの長時間にわたるシステムダウンは、ビジネス損失と直結している。
  - サイトによっては、損害賠償問題に発展する場合もある。

スケーラビリティの向上

IIJ Technology Inc.

- スケーラビリティ
  - システムの拡張能力
- 垂直拡張
  - ホスト単体を強化する。
    (CPU数を増やす、メモリ増大など)
- 水平分散
  - 同一機能のホストを複数並べて
    システム全体の能力を強化する。

アベイラビリティの向上

IIJ Technology Inc.

- アベイラビリティ（稼働率=システムの可用性）

  A=MTBF/(MTBF+MTTR)　MTBF:平均故障間隔　MTTR:平均修理時間

アベイラビリティ	ダウンタイム/年	ダウンタイム/週
99.9%	8時間45分	10分
99.99%	約52分	1分
99.999%	約5分	6秒

  システム全体の耐障害性を向上させる必要がある。
  - 機器1つ1つの耐障害性をあげることも必要だが、システム全体として、複数の機器がダウンしてもサービスが提供できるように設計するべきである。

➔複数の機器で
同一機能を提供する。水平分散

ロードバランサ導入のメリット

IIJ Technology Inc.

- スケーラビリティの向上
  - 複数サーバを利用し1つのサービスとして利用できる。
  - ロードバランサ配下のサーバを自由に追加、削除が可能。
- アベイラビリティの向上
  - サービスダウンしたサーバを検出し、サービスを提供しているホストからはずすことができる。
  - サーバのメンテナンスを行うときに、メンテナンスするサーバをサービスからはずすことにより、サービスの継続性が失われることがない。

ロードバランサ導入のデメリット

IIJ Technology Inc.

- メンテナンス負荷の増大
  - サーバが増えるため増えた分だけメンテナンス負荷が増大する。（セキュリティパッチ当てなど）
  - 複数のサーバにアクセスログが分散するため、アクセスログ解析に工夫が必要。

いずれも運用方法で工夫できるため、
ロードバランサを入れるメリットは大きい。

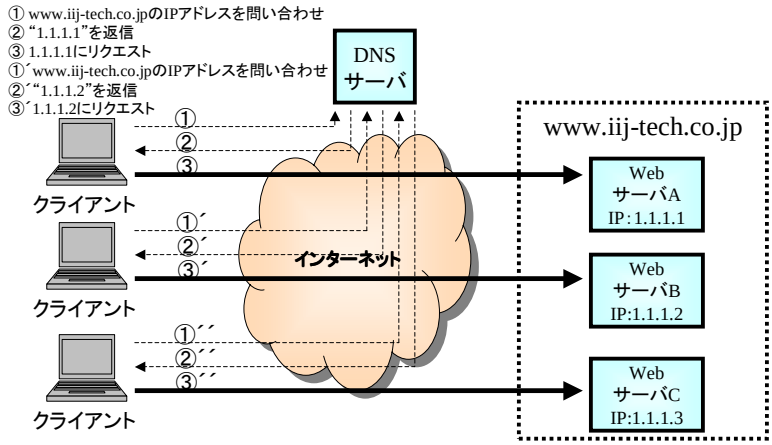
IIJ Technology Inc.

ロードバランサの基本機能

古典的な負荷分散

IIJ Technology Inc.

- DNSラウンドロビンによる負荷分散機能



DNSラウンドロビンの欠点

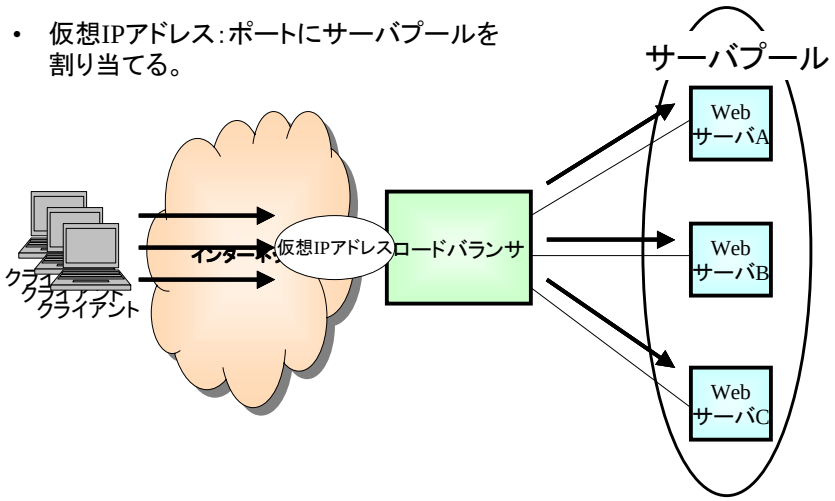
IIJ Technology Inc.

- 均等に負荷が分散されない。
- サーバダウン時でも、そのサーバにリクエストが振られてしまう。
- DNSの変更を行っても、キャッシュなどによりタイムラグが生じてしまう。

ロードバランサの基本動作(1)

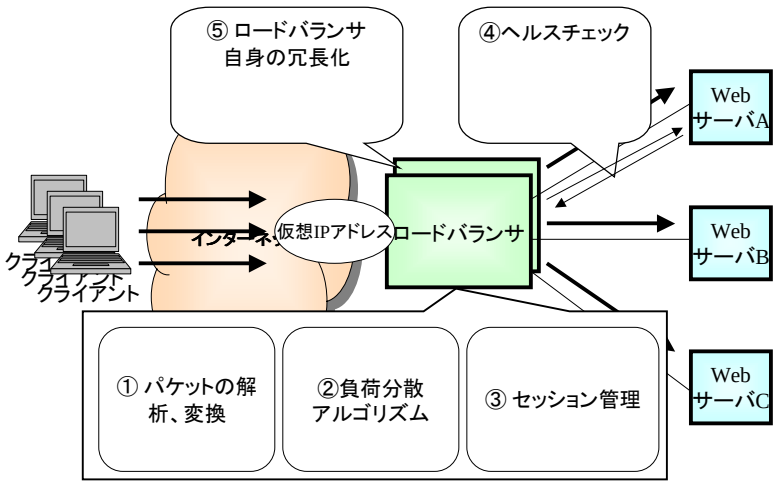
IIJ Technology Inc.

- 仮想IPアドレス:ポートにサーバプールを割り当てる。



ロードバランサの基本動作(2)

IIJ Technology Inc.

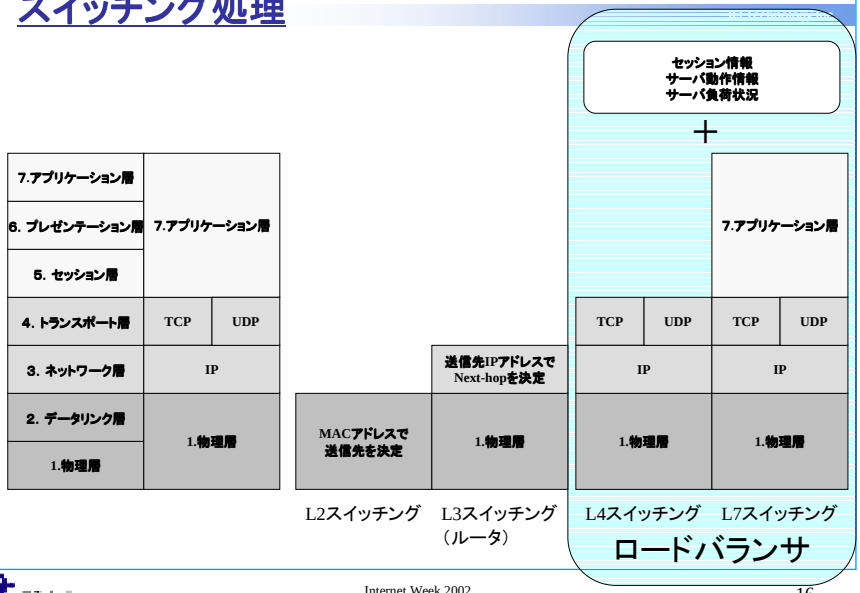


基本機能解説

IIJ Technology Inc.

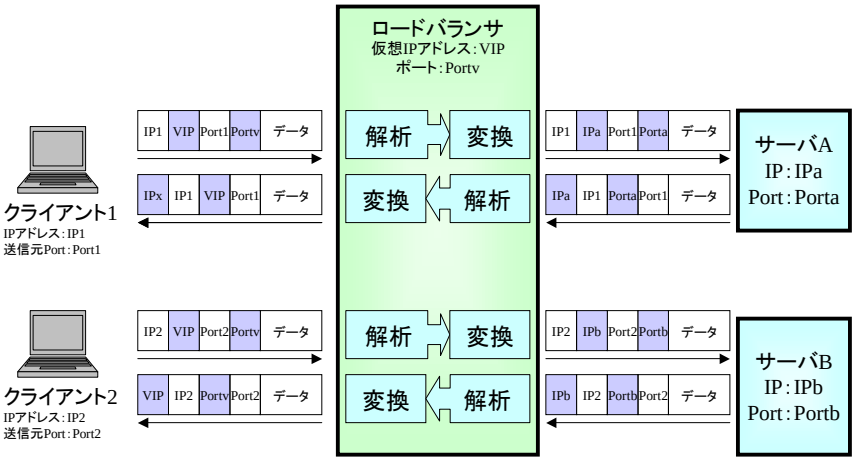
- パケット解析と変換機能
  - 負荷分散アルゴリズム
  - セッション管理
  - ヘルスチェック
  - ロードバランサ自体の冗長化

スイッチング処理



パケット解析と変換機能(1)

IIJ Technology Inc.



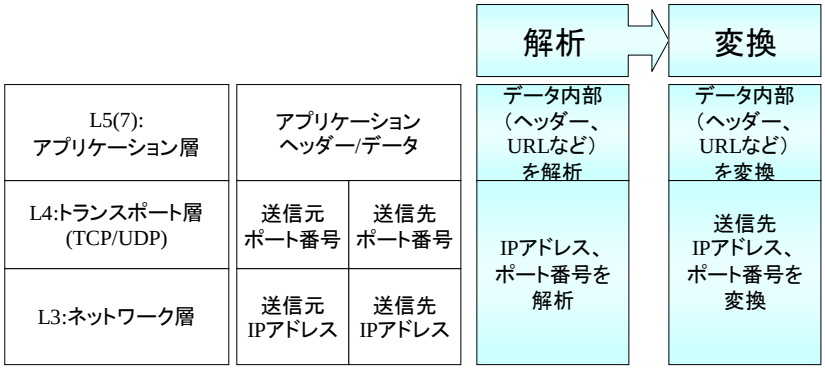
パケット解析と変換機能(2) L4スイッチング

IIJ Technology Inc.

			解析	変換
L5(7): アプリケーション層	アプリケーション ヘッダー/データ			
L4:トランスポート層 (TCP/UDP)	送信元 ポート番号	送信先 ポート番号	IPアドレス、 ポート番号を 解析	送信先 IPアドレス、 ポート番号を 変換
L3:ネットワーク層	送信元 IPアドレス	送信先 IPアドレス		

パケット解析と変換機能(3) L7スイッチング

IIJ Technology Inc.



基本機能解説

IIJ Technology Inc.

- パケット解析と変換機能
- ➢ 負荷分散アルゴリズム
- セッション管理
- ヘルスチェック
- ロードバランサ自体の冗長化

負荷分散アルゴリズム

IIJ Technology Inc.

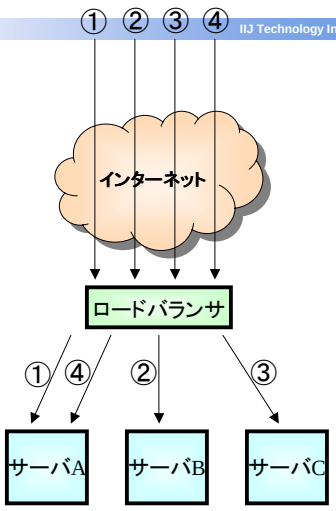
- 動的にどのサーバに振り分けるか決定するアルゴリズム
- 一般的なアルゴリズム
  - ラウンドロビン
  - 重み付け
  - 優先順位
  - 接続数
  - 応答時間
  - 複合型
  - HTTPヘッダー

アルゴリズム-ラウンドロビン

IIJ Technology Inc.

- 動作
  - クライアントからのアクセスを順番に、サーバに処理を振り分ける。

各サーバに性能差がない場合は、ラウンドロビンで負荷分散を行うことが多い。

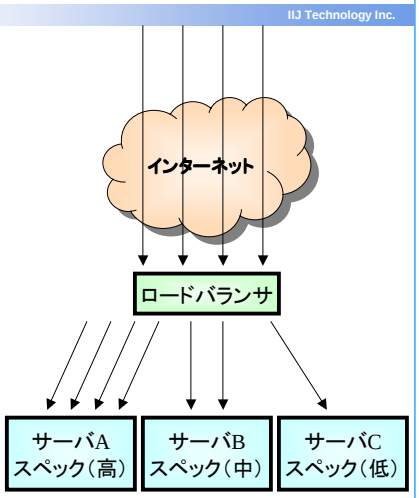


アルゴリズム-重み付け

IIJ Technology Inc.

- 動作
  - 各サーバに重み付け（ratio）をつけ、アクセスを振り分ける。

特に、性能差があるサーバを使うときに有効。



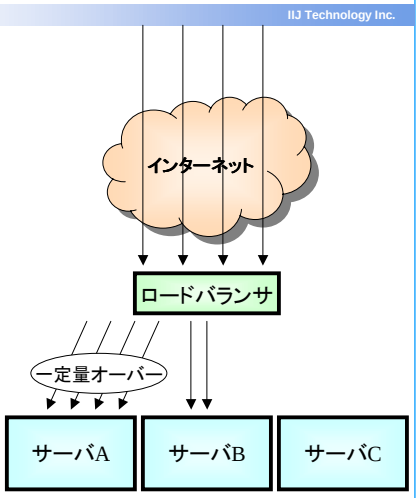
重み付け	4	2	1

アルゴリズム-優先順位

IIJ Technology Inc.

- 動作
  - サーバに優先順位（Priority）をつけ、優先順位の高いサーバのアクセスが一定以上超えた場合、次の優先順位が高いサーバに振り分ける。

プライオリティが低いサーバが負荷が低い時には、違う用途に使用可能。

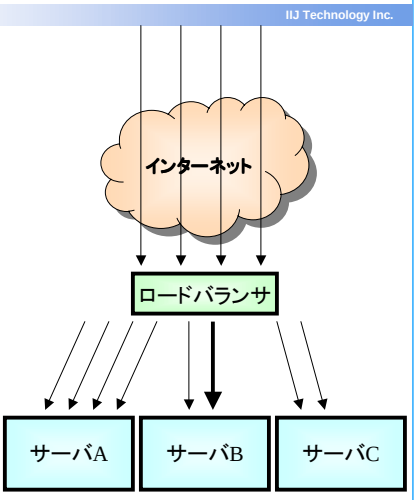


優先順位	1	2	3

アルゴリズム-接続数

IIJ Technology Inc.

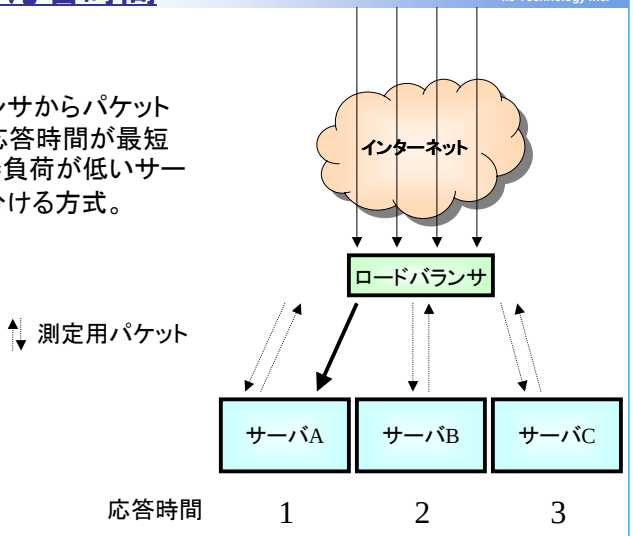
- 動作
  - セッション維持数が少ないサーバに振り分ける



アルゴリズム-応答時間

IIJ Technology Inc.

- 動作
  - ロードバランサからパケットを送信し、応答時間が最短のサーバ(=負荷が低いサーバ)に振り分ける方式。



アルゴリズム-複合型

IIJ Technology Inc.

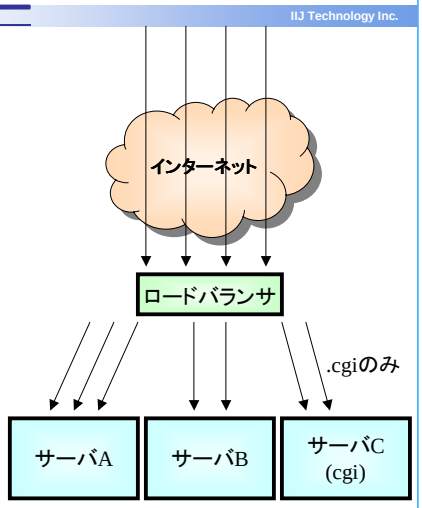
- 動作
  - 接続数＋応答時間
    - 接続数と応答時間から振り分けるサーバを決定する方式。
    - サーバの負荷状況に合わせてダイナミックに振り分けが可能。
  - ラウンドロビン＋優先順位
    - 複数のサーバをラウンドロビンで負荷分散する。トータルのセッション数が一定量を超えた場合、別のサーバに振られる。
    - “ごめんなさいページ”を表示可能。

アルゴリズム-HTTPヘッダー

IIJ Technology Inc.

- 動作
  - L7の負荷分散アルゴリズム。
  - HTTPヘッダー、URLを参照し、振り分けるサーバを決定する。
  - 拡張子が、“.cgi”のURLが含まれる場合のみcgiサーバにふる。
  - User-Agentによる振り分けも可能。

サーバによって動作を変えたいときに使用する。



基本機能解説

IIJ Technology Inc.

- パケット解析と変換機能
- 負荷分散アルゴリズム
- ➢ セッション管理
- ヘルスチェック
- ロードバランサ自体の冗長化

セッション管理（Persistence）

IIJ Technology Inc.

- 最近のWebサイトでは、ユーザのセッション管理を行うことが多い。
- ロードバランサにて、同一ユーザのアクセスを複数のサーバに振られてしまうと、セッション情報の管理をDBなどにもち、アクセスのたびにDB情報を参照しに行かなければならなくなる。=>DBに負荷が集中し、ボトルネックになる。
- 上記の問題を防ぐために、ロードバランサ側にて、同一ユーザのアクセスは同一サーバに振られるようにする必要がある。

セッション管理-送信元IPアドレス

IIJ Technology Inc.

- クライアントの送信元IPアドレスが同一の場合、同一サーバに接続する方法。
- Proxy、FW、NATにより同一ユーザの特定がしにくいため、割り振りが偏る可能性あり。
- クライアント側のネットワークが、Proxyのロードバランスを行っていて毎回送信元IPアドレスが変わる場合、セッション管理ができなくなる。

セッション管理-COOKIE

IIJ Technology Inc.

- HTTPヘッダーのCOOKIEが同一の場合、同一サーバに接続する。
- COOKIEの埋め込み方法
  - ロードバランサ側で、COOKIEにサーバIDを埋め込む。
  - サーバ側にてCOOKIEに特定の文字列を埋め込む。
- HTTPSのセッション管理は、データが暗号化されているため、そのままでは、Layer5以上の情報を利用したセッション管理は難しい。

セッション管理-SSL Session-ID

IIJ Technology Inc.

- 接続がSSLの場合、アプリケーションデータは暗号化されているため、送信元IPアドレスか、SSL Session-IDを利用したセッション管理しかできない。
- SSL Session-IDによるセッション管理は、Session-IDが同一な場合、同一サーバに接続する。
- ただし、Internet Explorerを使用するとこの設定は使えない。
  - IEの機能で、デフォルト2分に1回、SSLネゴシエイションを行うため、SSL Session-IDが変わってしまう。

基本機能解説

IIJ Technology Inc.

- パケット解析と変換機能
- 負荷分散アルゴリズム
- セッション管理
- ➢ ヘルスチェック
- ロードバランサ自体の冗長化

ヘルスチェック概要

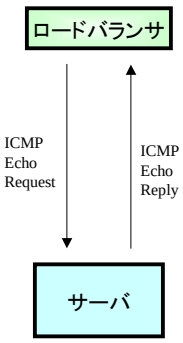
IIJ Technology Inc.

- ロードバランサは定期的に、「ヘルスチェック」を行い、サーバが稼動しているか確認をしている。サーバがダウンしたと判断された場合、新しいアクセスはサーバに振られないようになる。
- ヘルスチェックの種類には、以下のものがある。
  - PING監視　（L3の監視）
  - TCP監視　（L4の監視）
  - アプリケーション監視　（L7の監視）
  - 作りこみ

ヘルスチェック-PING監視

IIJ Technology Inc.

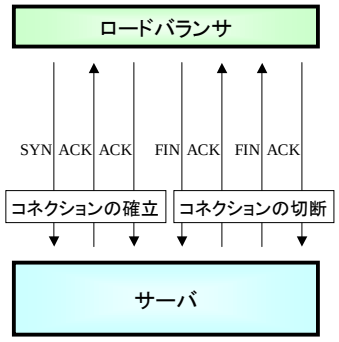
- サーバにPINGを行い応答があればサーバが稼動していると判断する。
- ネットワークの到達性しか監視できない。



ヘルスチェック-TCP監視

IIJ Technology Inc.

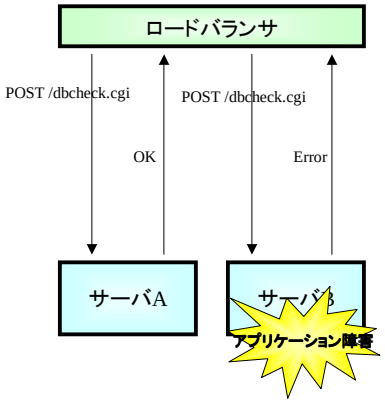
- サーバのサービスを提供しているTCPポートに対して、接続の確認を行う。
- TCP接続確認しか監視できないため、アプリケーションが正常な値を返してきているか分からない。



ヘルスチェック-アプリケーション監視

IIJ Technology Inc.

- アプリケーションの動作にあわせた監視を行う。
  - HTTP
  - FTP
  - SMTP
  - POP3



ヘルスチェック-UDP

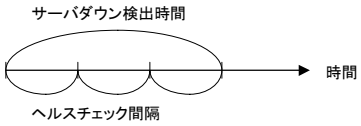
IIJ Technology Inc.

- UDPポートの監視は、難しい。
  - アプリケーション側で別TCPポートを用意し、そのTCPポートに対して監視を行う。
  - アプリケーション監視、作りこみによる監視を行うしかない。

ヘルスチェックの注意点

IIJ Technology Inc.

- ヘルスチェック間隔



サーバダウン検出時間=ヘルスチェック間隔×回数

  - ヘルスチェック間隔を短くするとサーバダウン検出時間が長くなる。
  - ヘルスチェックの間隔が短いとサーバ負荷が高くなったとき、ダウンと誤認してしまう。
  - Webサーバの場合、10secx3回、15secx3回くらい。
- アプリケーションログに影響
  - アプリケーションログに、ヘルスチェックのログが残ってしまう。

基本機能解説

IIJ Technology Inc.

- パケット解析と変換機能
- 負荷分散アルゴリズム
- セッション管理
- ヘルスチェック
- ➢ ロードバランサ自体の冗長化

ロードバランサ自体の冗長化

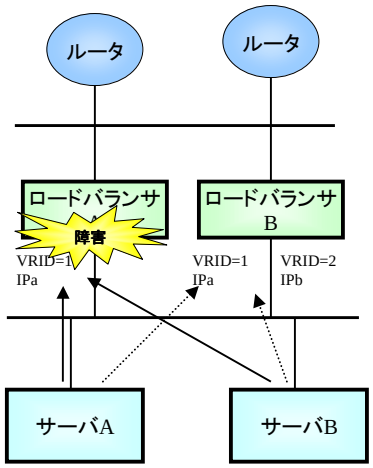
IIJ Technology Inc.

- ロードバランサ自体の冗長化の手法として、以下の2つが挙げられる。
  - VRRP(Virtual Router Redundancy Protocol:RFC2338)を使用した冗長化構成
  - 共有IPアドレス＋シリアル接続によるHeart Beat
    サーバアプライアンス型のロードバランサに多い。

冗長化(1) VRRP

IIJ Technology Inc.

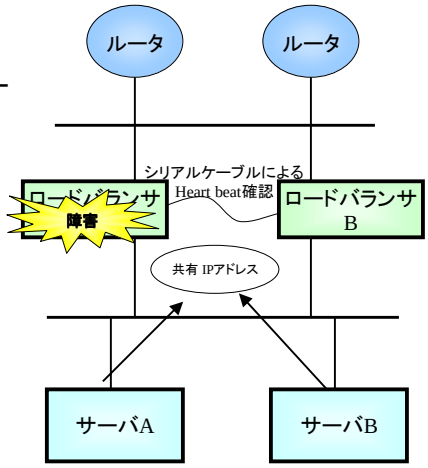
- VRRPを利用した冗長化
  - ロードバランサAから送信されるVRRP Advertisement（広告）パケットが障害により、送信不可になる。
  - ロードバランサBが、ロードバランサAのIP、MACを引き継ぐ。
  - 復旧後、プライオリティが高いVRRP Advertisement（広告）パケットがロードバランサAから送信されるようになるため、もとに戻る。



冗長化(2) 共有IPアドレス＋シリアル

IIJ Technology Inc.

- 共有IPアドレス＋シリアルケーブルによるHeart Beat
  - 通常、共有IPアドレスは、ロードバランサAにて立ち上げている。
  - ロードバランサA障害により、シリアルケーブル経由のHeart Beat確認が失敗する。
  - ロードバランサBは、共有IPアドレスをアップする。
  - ロードバランサ復旧後、手動にてもとにもどす。



IIJ Technology Inc.

システム構築編

構築編

IIJ Technology Inc.

- ネットワークトポロジー設計
  - フラットベース vs NATベース
  - VLANの分け方
  - 冗長化設計
- 事例
  - Webサイト
  - FWロードバランス

ネットワークアーキテクチャの違い

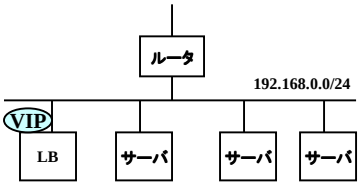
IIJ Technology Inc.

- フラットベース構成（L2接続）
  - VIPと各ノード（サーバ）が同一サブネット上にある構成
- NATベース（L3接続）
  - VIPと各ノード（サーバ）が、異なるサブネット上にある構成

フラットベースアーキテクチャ

IIJ Technology Inc.

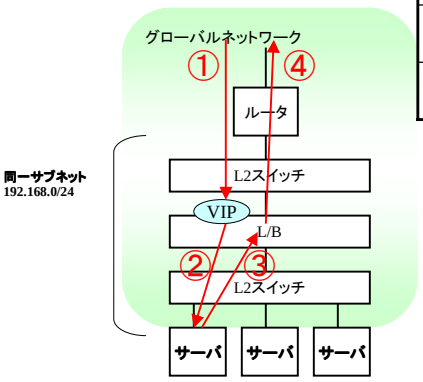
- VIPとサーバが同一サブネット上に存在するアーキテクチャ
- 2つの設計方法がある。
  - ブリッジ構成
  - ルート構成
- メリット
  - 同一サブネット上で構成できるため、構成がシンプル
  - サーバ起点の通信の場合、NATせずに直接通信が可能
- デメリット
  - グローバルアドレスを使用する場合、アドレスを消費してしまう。
  - セキュリティ対策は別途必要。



フラットベースアーキテクチャ：ブリッジ構成

IIJ Technology Inc.

- LBがブリッジとして動作



ルータ	192.168.0.1
LB	192.168.0.2
	VIP：192.168.0.10
サーバ	192.168.0.100-102
	デフォルトルート：192.168.0.1（ルータ）

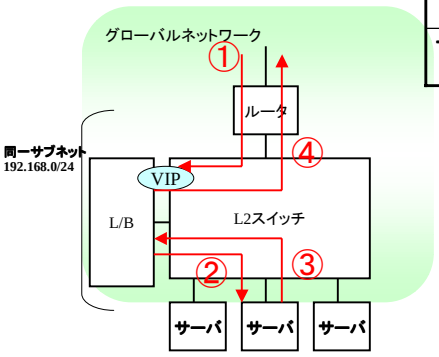
ブリッジ構成時の動作

① SrcIP：X　DstIP：VIP(192.168.0.10)
ロードバランサで受け、分散先を決定
② SrcIP：X　DstIP:サーバ(192.168.0.100)
サーバでリクエストを処理後、レスポンスを返す。
③ SrcIP：サーバ(192.168.0.100)　DstIP: X
途中で、ブリッジとなっているLBを経由し、SrcIPをVIPに変換。
④ SrcIP:VIP　DstIP：X

フラットベースアーキテクチャ：ルート構成

IIJ Technology Inc.

- 同一サブネット上に、LBが接続される構成



ルータ	192.168.0.1
LB	192.168.0.2
	VIP：192.168.0.10
サーバ	192.168.0.100-102
	デフォルトルート：192.168.0.2（LB）

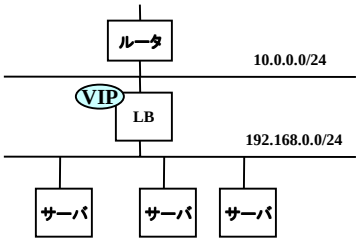
ルート構成時の動作

① SrcIP：X　DstIP：VIP(192.168.0.10)
ロードバランサで受け、分散先を決定
② SrcIP：X　DstIP:サーバ(192.168.0.100)
サーバでリクエストを処理後、レスポンスを返す。
③ SrcIP：サーバ(192.168.0.100)　DstIP: X
サーバのデフォルトルートとなっているLBを経由し、SrcIPをVIPに変換。
④ SrcIP:VIP　DstIP：X

NATベースアーキテクチャ

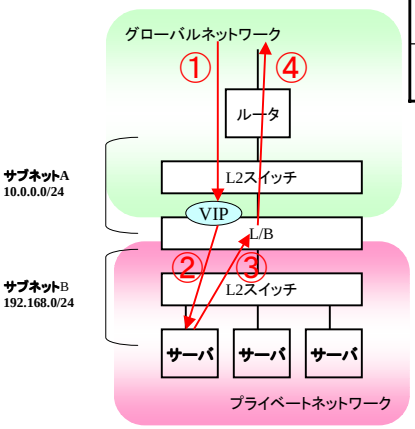
IIJ Technology Inc.

- VIPと各ノード（サーバ）が、異なるサブネット上にある構成
- LBは、ルータのような動作を行う。
- メリット
  - サーバ側のアドレスに、プライベートアドレスを使用できる。
  - VIPで指定したポート以外通さないため、セキュリティがあがる。
- デメリット
  - 管理するサブネットが増える。
  - サーバ起点の通信の場合、必ずNAT等のアドレス変換が必要。



NATベースアーキテクチャ

IIJ Technology Inc.



ルータ	10.0.0.1
LB	External：10.0.0.2
	Internal：192.168.0.1
	VIP：10.0.0.10
サーバ	192.168.0.100-102
	デフォルトルート：192.168.0.1（LB）

ルート構成時の動作

① SrcIP：X　DstIP：VIP(10.0..0.10)
ロードバランサで受け、分散先を決定
② SrcIP：X　DstIP:サーバ(192.168.0.100)
サーバでリクエストを処理後、レスポンスを返す。
③ SrcIP：サーバ(192.168.0.100)　DstIP: X
サーバのデフォルトルートとなっているLBを経由し、SrcIPをVIPに変換。
④ SrcIP:VIP　DstIP：X

フラットベース（L2) vs NATベース（L3）

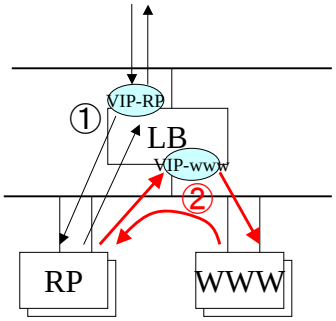
IIJ Technology Inc.

- 通信要件に応じて使い分けが必要
  - サーバ起点の通信時に、NATできない場合=＞フラットベース
  - グローバルアドレスが豊富ではない場合 =＞ NATベース
  - ファイアウォールを導入しない場合=＞NATベース（推奨）

Bounce back通信の設計①

IIJ Technology Inc.

- 折り返し（Bounce back）通信を行う場合、同一サブネット上にサーバを設置すると通信が成り立たない。
- 例として、Reverse Proxyの設計では、、、



① VIP-RP通信

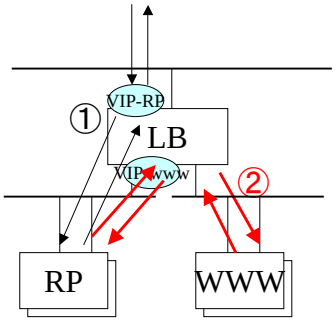
② VIP-WWW通信
1：SrcIP：RP DstIP：VIP-www
2：SrcIP: RP DstIP: www
3：SrcIP：www DstIP：RP

RPは、送信先IPとは別のIPアドレスからレスポンスが返ってくるため、通信が成り立たない。

Bounce back通信の設計②

IIJ Technology Inc.

- Bounce back通信を確立するためには、RPとwwwとのサブネットを分ける必要がある。



① VIP-RP通信

② VIP-WWW通信
1：SrcIP：RP DstIP：VIP-www
2：SrcIP: RP DstIP: www

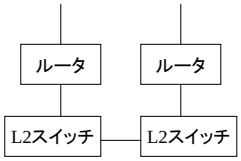
3：SrcIP：www DstIP：RP
4：SrcIP: VIP-www DstIP: RP

必ず、LBを通過するように設計機能、用途に応じてサブネットを分けておく。

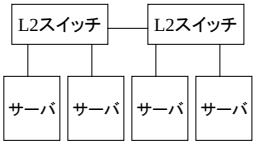
LBを用いたネットワークの冗長化設計①

IIJ Technology Inc.

- 上位接続の冗長化
  - 回線の二重化
  - ルータの二重化
  - L2スイッチの二重化



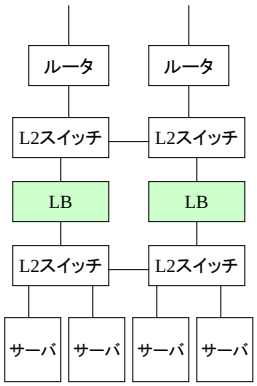
- サーバ接続の冗長化
  - L2スイッチの二重化

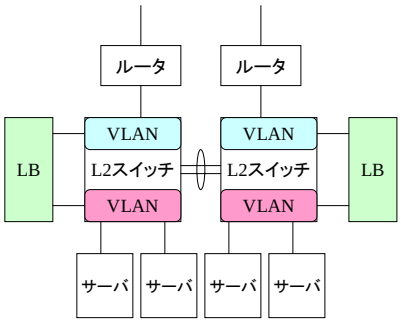


LBを用いたネットワークの冗長化設計②

IIJ Technology Inc.

- NATベースLBの冗長化
- 上位、下位のスイッチを同一スイッチで実現

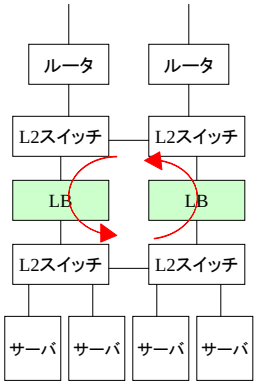




LBを用いたネットワークの冗長化設計③

IIJ Technology Inc.

- フラットベース（ブリッジ）LBの冗長化

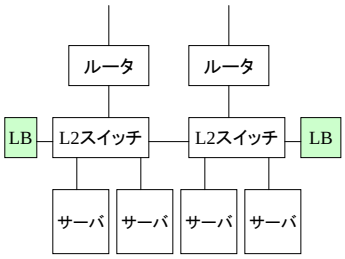


- L2接続を行うため、ブリッジングループが発生。
- L2スイッチ、LBでSTPを動作させる必要あり。
- STPは遅いため、お勧めできない。
- MST（IEEE802.1s）、RSTP（IEEE802.1w）を使用すれば約3sec以内で切り替わるが、LBではまだ実装されていないものがほとんど。

LBを用いたネットワークの冗長化設計④

IIJ Technology Inc.

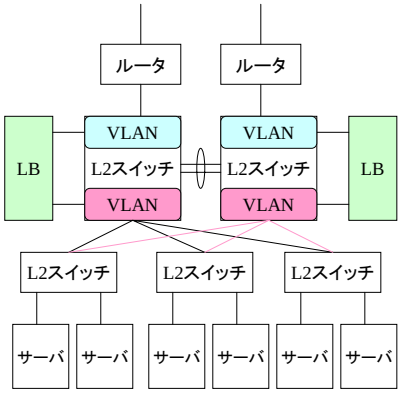
- フラットベース（ルート構成）LBの冗長化



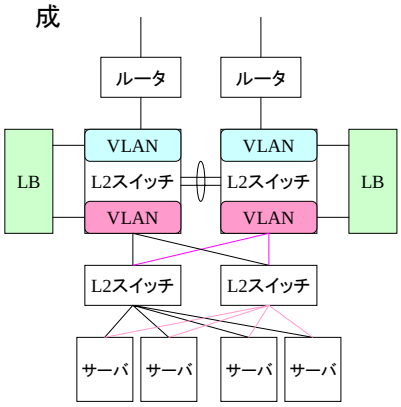
LBを用いたネットワークの冗長化設計⑤

IIJ Technology Inc.

- 大規模なネットワークの冗長化
  - サーバ台数が多いとき



- サーバ接続スイッチのダウン、サーバNIC障害を考慮した構成



ネットワーク設計まとめ

IIJ Technology Inc.

- フラットベースvsNATベース
  - 通信要件、セキュリティ要件により決定させる。
- サブネットの分け方
  - サーバ機能、種類によって分ける。
  - 分けすぎは、管理するネットワークが多くなるのでほどほどに
- 冗長化設計
  - フラットベース、NATベースでは物理接続が同じでも論理設計が異なるので、要注意。
  - STPの設計に注意する。
- その他
  - LBには、LB以外の仕事をなるべくさせないように設計させる
    - ルーティングは、ルータへ
    - STPは、L2スイッチのみで

構築編

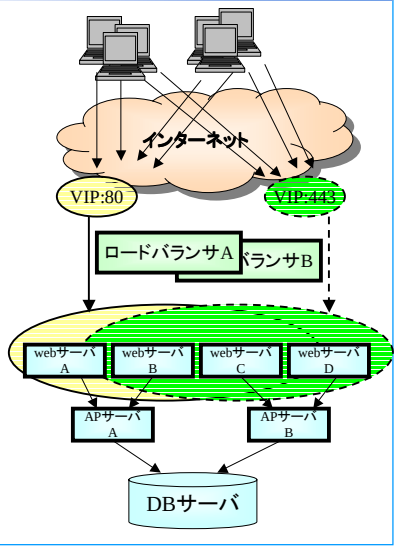
IIJ Technology Inc.

- ネットワークトポロジー設計
  - フラットベース vs NATベース
  - VLANの分け方
  - 冗長化設計
- 事例
  - Webサイト
  - FWロードバランス

Webサイト：一般的なWebシステム

IIJ Technology Inc.

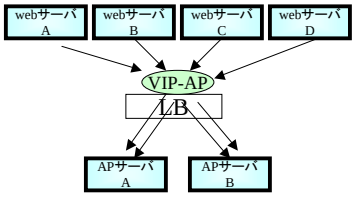
- ロードバランサの標準的な機能にて実現。
- ヘルスチェックの注意点
  - 三層構造（Web-AP-DB）となっているときは、Webサーバだけが対象となるヘルスチェックだけでは不十分。
  - APサーバ、DBサーバと連携していることを確認する必要がある。
  - APサーバ、DBサーバと通信した結果がでる動的ページを生成し、内容を確認できるヘルスチェックを行うこと。



Webサイト：APサーバロードバランス(1)

IIJ Technology Inc.

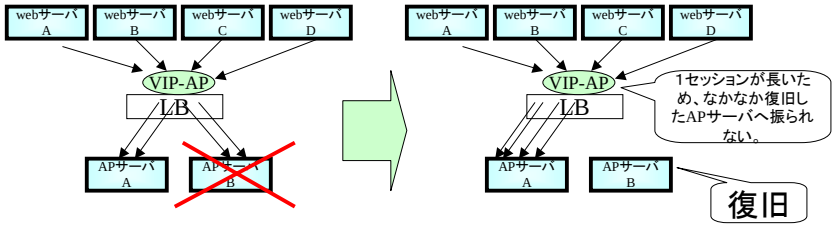
- WebサーバからAPサーバへのロードバランス設計
- 注意点
  - APサーバへ接続されるWebサーバの数が限られている
  - サーバ数は、Webサーバ≧APサーバ



Webサイト：APサーバロードバランス(2)

IIJ Technology Inc.

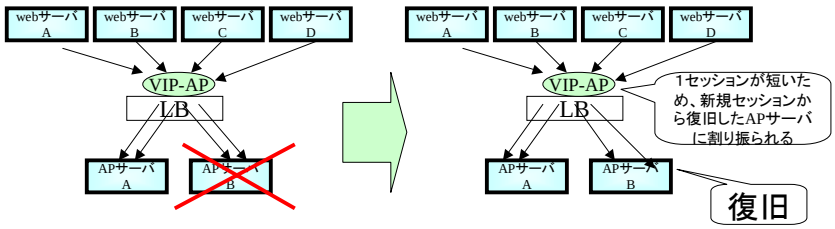
- SrcIPでセッション維持してしまうと、問題あり
  - APサーバのダウンから復旧した場合でも、セッションが維持されたままなので、復旧したサーバに振られない。



Webサイト：APサーバロードバランス(3)

IIJ Technology Inc.

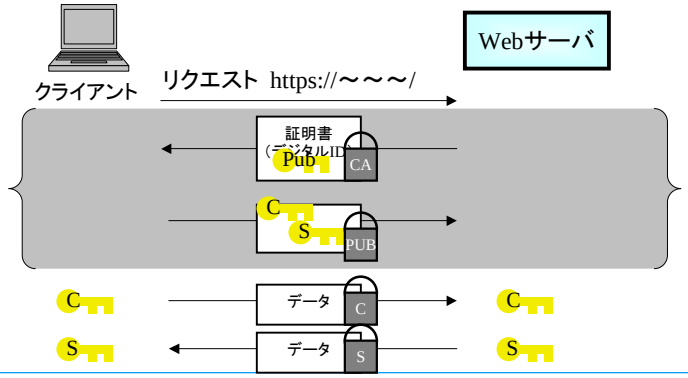
- アプリケーションで使用するセッション情報によるセッション維持をする必要あり。
  - COOKIE情報
  - URLにsession IDを埋め込むなど
- 送信元IPアドレス以外のセッション情報で、セッション維持を行うことにより、APサーバダウンからの復旧時にも復旧したサーバに通信が割り振られるようになる。



Webサイト：SSLセッション維持の問題(1)

IIJ Technology Inc.

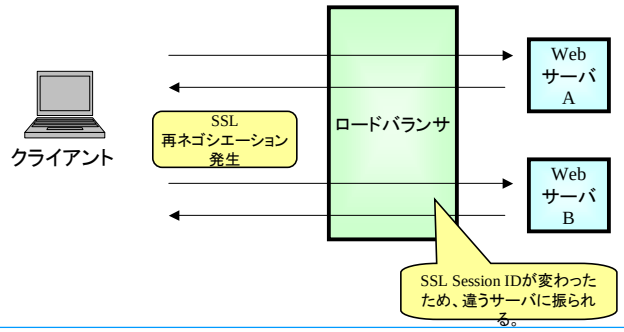
- ひとつのクライアントからのSSL（HTTPSなど）の通信が、複数のサーバに分散されてしまうと、振られるサーバが変わるたびに、SSLネゴシエーションが発生してしまう。



Webサイト：SSLセッション維持の問題(2)

IIJ Technology Inc.

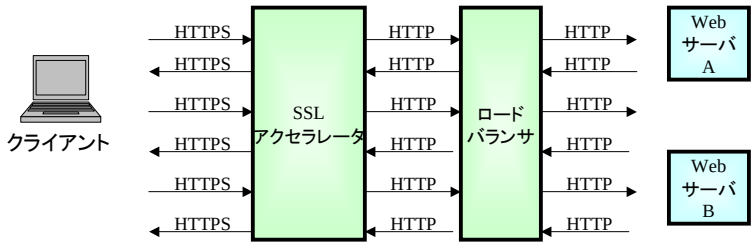
- SSLネゴシエーション時につけられるSSL Session IDを利用してセッション維持を行うと、再ネゴシエーションが発生した場合、SSL Session IDが変わってしまい、アプリケーション側のセッション情報が失われる。



Webサイト：SSLセッション維持の問題(3)

IIJ Technology Inc.

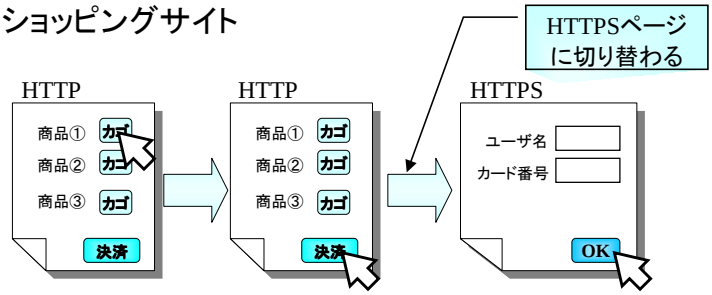
- 解決するには、
- SSLアクセラレータを入れる。
  - ブラウザ-SSLアクセラレータ間でのみSSL通信を行うため、Webサーバが変わったために起こるSSL再ネゴシエーションは、発生しない。
  - ロードバランサには、暗号化を解いた状態（HTTP）で通信されるため、アプリケーションに合ったセッション維持方法を選択できる。
- SSLアクセラレータ付のロードバランサもある。



Webサイト：HTTP、HTTPSを利用するサイト(1)

IIJ Technology Inc.

- 問題
  - HTTPページからHTTPSページへ変わったときのセッション管理
- 例）ショッピングサイト



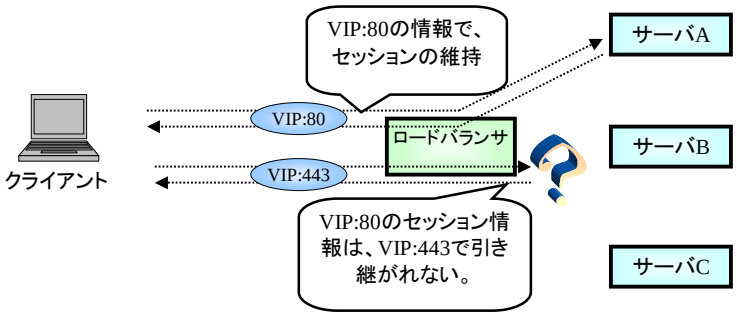
(1) 商品をカゴに入れる。

(2) 商品が確定したので、決済をする。

(3) 決済のため、個人情報を入力。

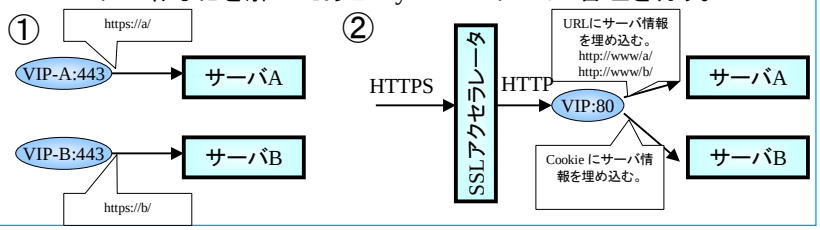
Webサイト:HTTP、HTTPSを利用するサイト(2)

- ほとんどのロードバランサは、VIP:PORTの組でセッション情報をもっている。



Webサイト:HTTP、HTTPSを利用するサイト(3)

- 解決策
  - 別のVIP:PORT同士でセッション情報の共有化が行えるロードバランサを使用する。
  - サイトの作り方を工夫する。
    HTTPSに切り替わるリンクに細工をする。
    - ① 各サーバに対応したVIPを定義する方法
    - ② Cookie、URLにサーバのIDを埋め込み、SSLアクセラレータで暗号化を解いたあとLayer7のセッション管理を行う。



Webサイト：HTTP、HTTPSを利用するサイト(4)

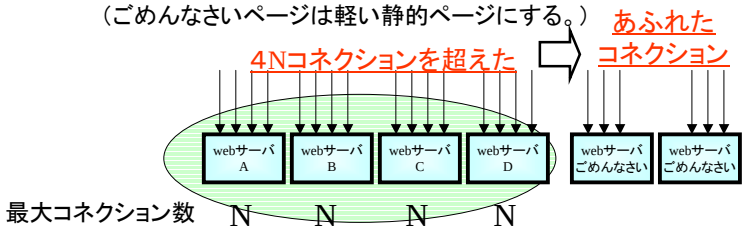
IIJ Technology Inc.

- 究極的な解決策
  - Webアプリケーションのつくりで、サーバに依存したセッション管理を止める。
  - 最近のAPサーバは、APサーバ同士でセッション情報のやり取りができるため、個々のWebサーバ、APサーバに依存しないシステムができる。

Webサイト：サーバ過負荷時の動作

IIJ Technology Inc.

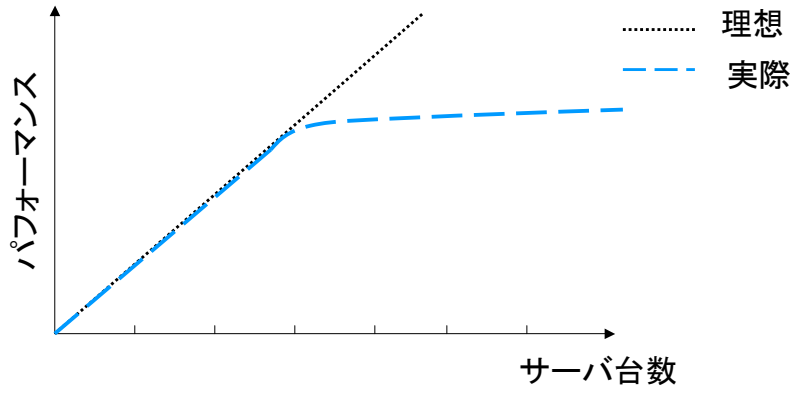
- サーバが過負荷状態（CPU使用率100%、メモリ不足、I/Oなど）におちいると、MAXのパフォーマンスが出ない。不安定な状態になる。
- 過負荷になったときの対応をあらかじめ準備しておく。
  - => サーバの最大コネクション数を制限
  - => ごめんなさいサーバを用意。
    （ごめんなさいページは軽い静的ページにする。）



Webサイト：サーバ増設による落とし穴(1)

IIJ Technology Inc.

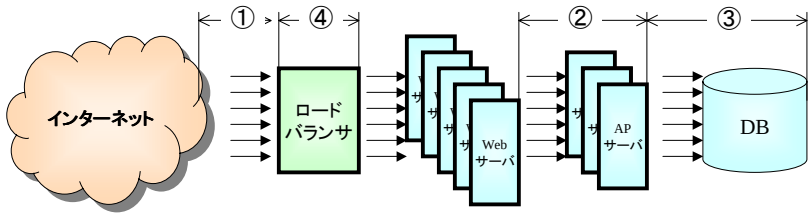
- ロードバランサに頼って、Webサーバを増設すればシステム全体のパフォーマンスがあがるというわけではない。



Webサイト：サーバ増設による落とし穴(2)

IIJ Technology Inc.

- 考えられる要因
  - ① ネットワークのボトルネック
  - ② APサーバのボトルネック
  - ③ DBサーバのボトルネック
  - ④ アーキテクチャの問題
  - ⑤ ロードバランサ自体の負荷



サイト全体のパフォーマンスを向上させる指針

IIJ Technology Inc.

- 解決策
  - ① ネットワーク
    - 増速するしかない
  - ② APサーバ
    - CPU、メモリの追加。上記機種にアップグレード。
    - APサーバ台数を増やす。
  - ③ DBサーバ
    - CPU、メモリの追加。上記機種にアップグレード。
    - DBサーバを用途によって、分ける。
  - ④ アーキテクチャの問題
    - WebサーバとDBサーバの二層構造の場合、Webサーバ、APサーバ、DBサーバの三層構造にしてみる。
    - Web、cgiサーバを兼用している場合、cgiサーバと静的コンテンツサーバを分けてみる。Cacheサーバの導入を検討する。

サイト全体のパフォーマンスを向上させる指針

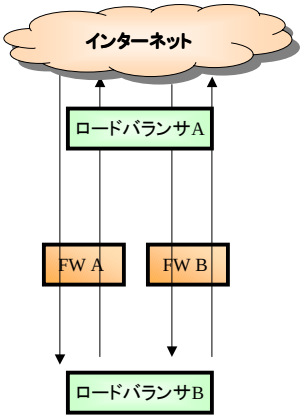
IIJ Technology Inc.

- 解決策
  - ⑤ ロードバランサ
    - ロードバランサ自体の負荷の軽減
      - Layer7スイッチングは遅いため、やめてみる。
      - なるべく処理がかからないヘルスチェックに変更する。
      - 最大セッション数のボトルネックの場合、スイッチ型の場合、ポートを分ける。サーバアプライアンス型の場合、メモリ増設もしくは上位機種にアップグレード。

ファイアウォールの負荷分散

IIJ Technology Inc.

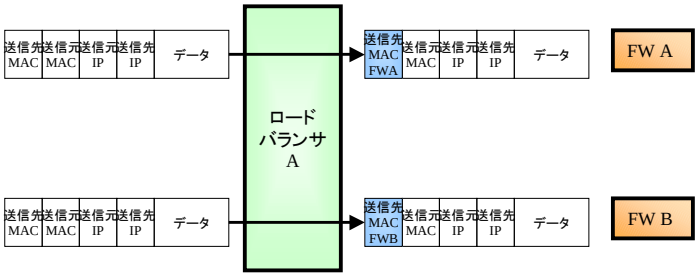
- ロードバランサでファイアウォールをはさむ。(サンドイッチ構成)
- 行きと帰りのパケットが同じFWを通らなければならない。
- この構成で、FWの負荷分散をするときは、FWでアドレス変換(NAT)は行わない。



ファイアウォール負荷分散時の動作(1)

IIJ Technology Inc.

- ロードバランサAは、透過(Transparent)モード。
  - アドレス変換、ポート変換を行わない。送信先MACアドレスのみを変更。

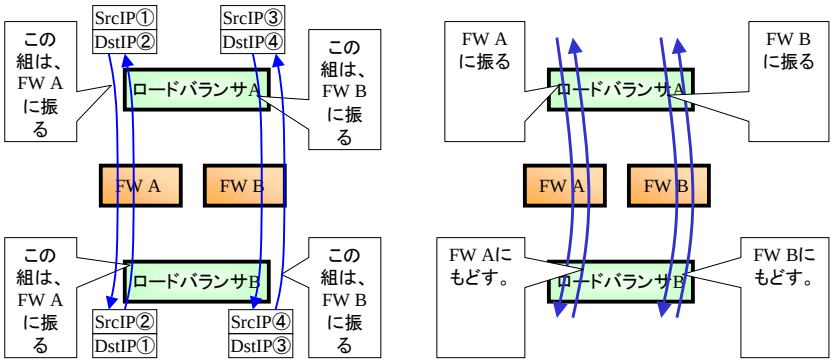


ファイアウォール負荷分散時の動作(2)

IIJ Technology Inc.

• 行きと帰りのパケットが同じFWを通るようにするには、

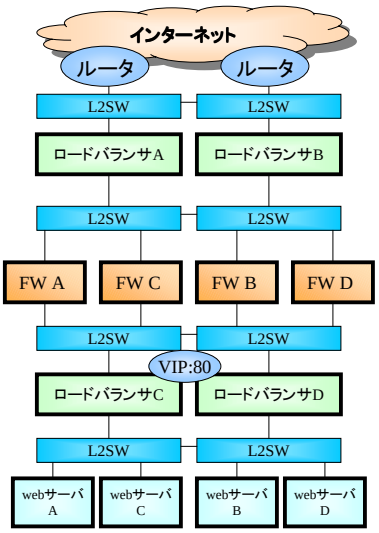
- HASH機能
  - SrcIPアドレス、DstIPアドレスの組で通るFWを決める。
- LastHop機能
  - 送られてきたFWに返す。



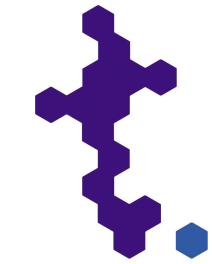
FWを入れた冗長化構成

IIJ Technology Inc.

- FW配下のロードバランサは、Webサーバの負荷分散も行う。



ありがとうございました。



IIJ Technology

Network Initiative