clearswift

SECURE Web Gateway v4 サイジングガイド

テクニカルガイド

第 06 版

2016 年 4 月 19 日

著作権

修正番号 1.0　2016 年 4 月

Clearswift Ltd.発行

目次

1 はじめに

Web トラフィックは、たとえユーザー数が同じであっても企業によって異なります。そこで本ガイドでは、Web セキュリティプラットフォームの適切なサイジングを行う上で、どのような要件を考慮すべきかについてご説明します。

なお、本ガイドは SECURE Web Gateway v4.3 以上のリリースを対象としています。

2 概要

Web セキュリティプラットフォームのサイジングを行う際には、数多くの要素を考慮する必要があります。たとえば、トラフィックプロファイルに基づくにしてもデータのサイズや種類は企業によって大きく異なる可能性があります。

しかし、Web セキュリティプラットフォームのサイジングを行う際には、主として考慮すべき点は次の 3 つです。

- 同時接続数
- 持続帯域幅
- 検査ポリシー

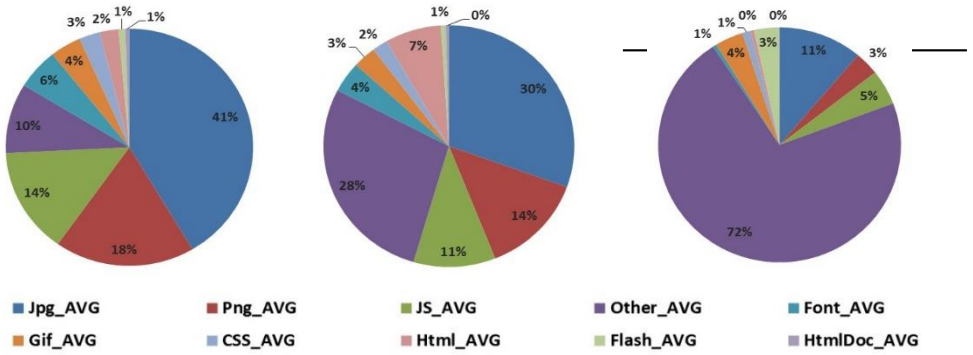
以上の 3 つはいずれも全体のパフォーマンスに大きな影響を及ぼすため、プラットフォームのサイジングの重要検討項目となります。

2.1 同時接続数

同時接続数は把握するのが難しいもののひとつであり、特に Web セキュリティプラットフォームが存在していない場合は困難です。これは、並行度の問題ですが、同時に接続ユーザーごとの接続数にも左右されます。いずれの要素も、企業のトラフィックプロファイルとユーザーのデスクトップ構成により異なります。

また、トラフィックプロファイルが異なれば、平均ページサイズやトラフィック上のデータタイプの分布状況が違ってきます。この点については、テキスト分析実行による影響という観点から考慮すべき問題です。たとえば、プロファイルがマルチメディア型の企業は大きな帯域幅を必要としますが、情報漏洩防止ポリシー実行のための処理能力はそれほど必要としません。

クリアスウィフトの顧客ベースを基に推測すると、各検討項目の最大値は次の表にまとめたようになります。

トラフィックプロファイル	標準	SaaS	マルチメディア
並行度	10%	25%	15%
ユーザーごとの接続数	5	7	9
平均ページサイズ	3.6MB	4.8MB	9MB
ページごとのデータタイプ分布			
凡例	Jpg_AVG, Png_AVG, JS_AVG, Other_AVG, Font_AVG, Gif_AVG, CSS_AVG, Html_AVG, Flash_AVG, HtmlDoc_AVG		

このように、マルチメディアコンテンツの利用が非常に多い 1500 人のユーザーを持つ組織の場合、同時接続数は 1500 x 15% x 9 = 2025 となります。一方、同じ規模の組織でも、インターネットの利用が標準型であれば同時接続数は 750 となります。

なお、この数値はあくまでも参考値であり、常に同じ数値になるとは限りません。また、組織のサイズが大きくなるほど、並行度が小さくなる傾向にあることも、検討の際の参考にする必要があります。

2.2 持続帯域幅

持続帯域幅に関しては、ピークタイムについて次の点を考慮に入れる必要があります。

- 帯域幅の使用量は、利用可能な全帯域幅のおよそ 80%～90%となるのが一般的
- HTTP/S 通信の帯域幅は大きく変動する可能性があるものの、通常は全体の使用帯域幅のおよそ 60%
- アウトバウンドの HTTP/S トラフィックは全 Web 閲覧トラフィックの 10%以下となる

したがって、実行すべき検査の種類により、分析すべき帯域幅が大きく異なることになります。たとえば、100 Mbps のインターネット接続を持つ標準型の企業は、ピーク時には通常 51 Mbps の HTTP/S 帯域幅をサポートする必要があります。しかし、DLP ポリシーを適用するのにアウトバウンドのトラフィックを検査するだけなら、ゲートウェイを通過するトラフィック量は同じでも、そのうちのわずか 5 Mbps だけを検査すれば良いことになります。

以上の数値はあくまでも参考としてのものであり、正確なプラットフォームサイズを得るためには実際のトラフィックの数値を使用する必要があります。

2.3 検査ポリシー

情報セキュリティ上の要件は、企業によって異なります。ある企業はトラフィックの双方向についてディープ分析検査を行う必要があるかも知れませんし、別の組織はアウトバウンドのトラフィックのみを検査すれば良いのかもしれません。

検査ポリシーについては、テキスト検索式による分析の実行は、特定のデータタイプの検索よりも大きな影響を及ぼします。さらに、テキスト分析に正規表現を用いるほうが、通常の文字検索よりも影響が大きくなります。

同様に、HTTPS 検査の実行は、それを行わない場合に比べて、CPU の使用率に大きな影響を及ぼします。ユーザーごとのポリシー適用にはユーザーの認証処理が必要となり、この処理は使用されるディレクトリサービスの応答性に大きく左右されます。リソースの観点からは、この処理はメモリやCPU、ネットワークの使用率に影響を及ぼします。

2.4 その他の検討事項

SECURE Web Gateway は、クリアスウィフトが提供するサーバーをはじめ、標準的な Intel のサーバー[1] や VMware 上でも動作させることができます。使用するプラットフォームに関わらず、予想されるパフォーマンスと利用可能なリソースはほぼ同じです。

仮想化プラットフォームに対してより少ないリソースや低い性能のものを割り当てる例が見受けられますが、これはよくある勘違いの一つです。アーキテクチャーを変更してインスタンス数を増やすというのも一つの方法ですが、いずれにしろ、プラットフォームは要求される性能を提供できるものでなければなりません。

---

[1] Gateway v4 用に認証された Red Hat プラットフォームのリストについては、Red Hat Enterprise Linux 6 カタログをご覧ください：https://access.redhat.com/ecosystem

3 サイジングのガイドライン

3.1 ハードウェアのサイジング

クリアスウィフトでは、Clearswift SECURE Web Gateway をあらかじめインストールした物理的なアプライアンスを提供しています。ハードウェアプラットフォームのサイジングを検討する際には、その性能と仕様を参考にしてください。

サーバーの仕様	持続帯域幅	ピーク時の帯域幅	ピーク時の接続数
(A) 1 x Intel G3430 (デュアルコア 3.30GHz)、4GB RAM、500GB SATA@7200	25 Mbps	30 Mbps	280
(B) 1 x Intel E3-1240 v3 (クアッドコア 3.40GHz)、4GB RAM、500GB SATA@7200	60 Mbps	75 Mbps	700
(C) 2 x Intel E5-2609 v3 (ヘキサコア 1.9GHz)、8GB RAM、3x300GB SAS@15000	80 Mbps	110 Mbps	1000

重要：

- 上記の数値は HTTP トラフィックをベースにしたもので、オフボックスレポーティング機能を有効にし、プロキシーキャッシュを無効に設定した 200 Mbps のインターネットパイプを使用すると仮定しています。
- プロキシーキャッシュを有効にする場合は、SSD ドライブの使用が必須です。この場合の帯域幅は、上記のものより低くなります。

適切なサーバーを選択するには、帯域幅とピーク時接続数の両方を考慮し、それらを満足するかまたはそれ以上のハードウェア構成を考える必要があります。

なお、SECURE Web Gateway が上記の数値を実現するためには、上記に示されたストレージ性能を確保することが必要であることにご注意ください。

帯域幅や接続数に関する要件が上記で示したサーバーの上限を超える場合は、複数のサーバーを使用する方法があります。

ピークタイムに関する数値は、短時間の間だけ実現可能な最大値です。

3.2 仮想環境

Clearswift SECURE Web Gateway は、VMware 上の仮想化ゲートウェイとしてデプロイすることが可能です。性能に関する要件は他の場合と同じであり、プラットフォームが十分なリソースを提供できるようにする必要があります。

また、SECURE Web Gateway はリソースの待機時間にも敏感に影響を受けます。他の仮想化マシンとのリソース共有のためのオーバーヘッドのためにリソースへのアクセスに遅れが生じるようなことがあれば、ユーザーの Web 閲覧の使い心地が悪くなります。

仮想化マシンで良く見受けられる問題を予防するため、次の設定が推奨されます。

- VMXNET3 ネットワークアダプターを使用する
- Red Hat Enterprise Linux 6 用の VMware ツールをインストールする
- 必要なリソースを SECURE Web Gateway がいつでも利用できるように、リソース予約機能を使用する
- Gateway を実行するために必要なリソースの割り当てに柔軟性を持たせるため、小さな VM を数多く持つようにする
- VM が十分な高速応答を得られるようにする。これは、仮想化環境で動作する Gateway の性能を低下させる主要な原因の一つとなっています

SECURE Web Gateway が VMware で動作するか物理的なハードウェア上で動作するかに関わらず、性能に関する要件は同じであることにご注意ください。

3.3 Gateway Reporter

Gateway Reporter に必要なサーバーの仕様は、必要とされるストレージ容量によって決まります。ストレージ容量は、監査データの保存日数と全 Gateway における監査トランザクション数の積として求められます。

保存日数、現在のデータベースのサイズ、および過去 7 日間における一日のトランザクション処理数は、システム > システムの設定 > レポートデータ設定で確認できます。

保存されるトランザクションデータは、1 件あたりおよそ 600 バイトです。以上の数値を基にして、必要なディスク容量を計算することができます。たとえば、1 日あたり 270,500 トランザクションを 60 日間保存する場合は、

> 270,500 トランザクション x 60 日間 x 600 バイト = 9,738MB (9.7GB) のディスク容量

3.4 検査ポリシーに関する検討事項

デプロイメント後の課題として、SECURE Web Gateway による処理の負担を増大させ、結果としてパフォーマンスに影響を及ぼすポリシーやシステム構成について考慮する必要があります。以下の項では、この点についてのガイダンスと解決法について解説します。

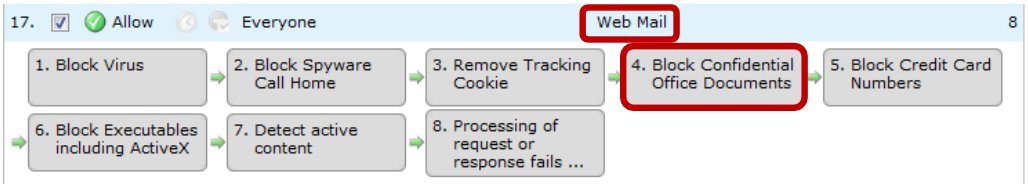
3.4.1 テキスト分析

テキスト分析のコンテンツ規則は非常に強力で、Web コンテンツや添付ファイル内のキーワードやフレーズを検索するのに活用できます。さらにこの規則は、テキスト中の一定のパターン（顧客参照番号など）を検索するための複雑な正規表現を定義することも可能です。しかし、正規表現による処理は、「トップシークレット」などの通常のキーワード検索よりも CPU に負担をかけます。

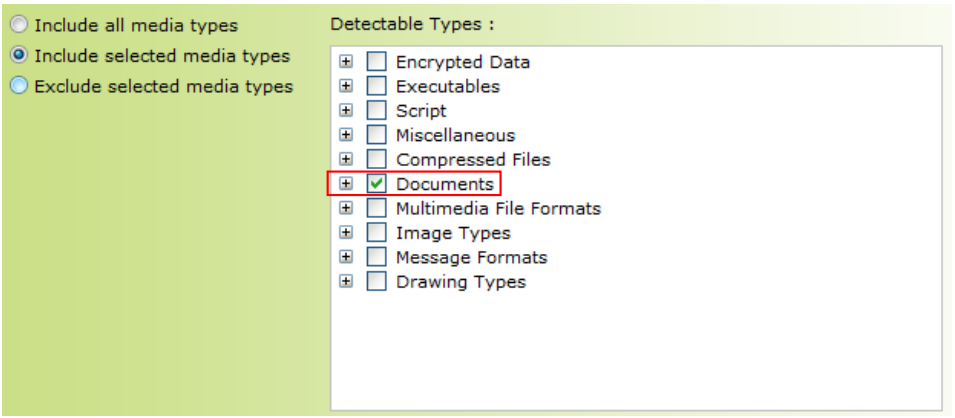
SECURE Web Gateway では、通常のテキスト検索を行う際に、Web コンテンツをすべて検索するのではなく、Web 転送の特定の部分だけを対象として検索できます。サイトの種別やファイル形式、文書中の部分、さらには検索方法を限定することで、処理に要する負荷や誤検出の発生リスクが削減できます。たとえば、社外秘の文書中の機密に該当するフレーズを検索するのであれば、アウトバウンドの Web トラフィックだけを対象とすれば良いわけです。

テキスト検索による性能低下を抑えるための方法としては、次のように検索分野を限定する方法があります。

- 特定の種類のサイトやドキュメントに限定する

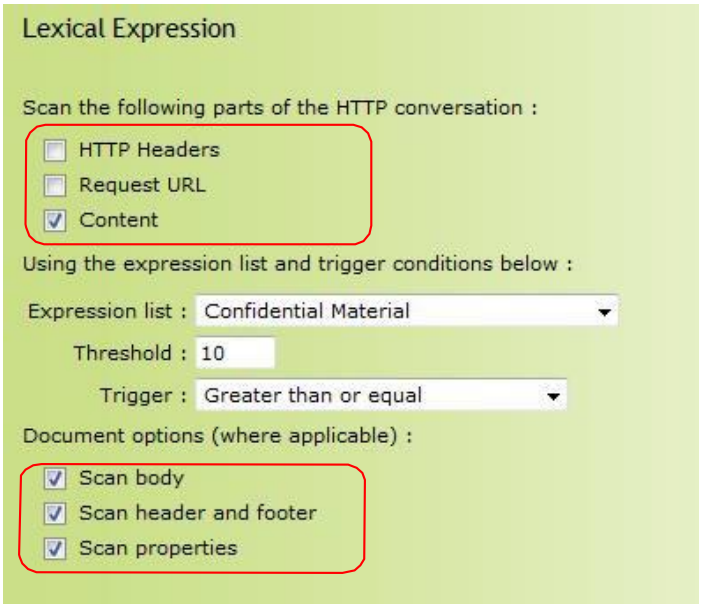


- 特定のファイル形式に限定する

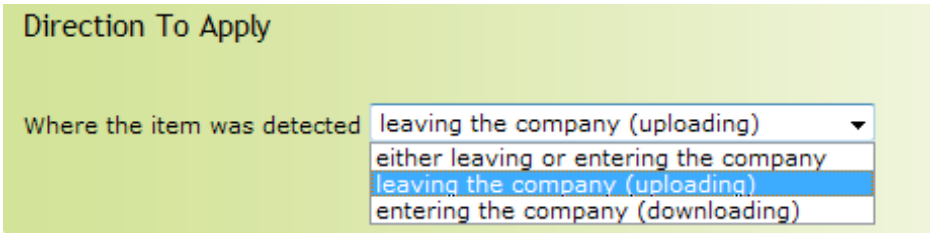


- Web ページ、文書の内容、URL、HTTP ヘッダー、または文書のヘッダー／フッターやプロパティのみに限定する

注：「HTTP ヘッダー」や「要求 URL」の選択は、ほとんど必要ありません。すべての HTTP ヘッダーやすべての URL で特定のフレーズ検索を行うと、パフォーマンスに影響を与えます。これらを選択する際には十分な検討が必要です。



- ディレクション ― インバウンドの検査も必要ではありますが、情報漏洩は外に向かってのみ起こるものです。



3.4.2 データベースの最適化

データベースの最適化には、2 つの側面があります。

1. データベースインデックスの再構築

   デフォルトでは、インデックスの再構築は毎週土曜日の午後 9 時に実行されます。この日時に設定されている理由は、Web プロキシーのパフォーマンスに与える影響が最も少ない業務時間外であるためです。

2. データベースの縮小

   データベースの縮小とは、データベース中の削除された行が占有している不要なディスク領域を解放することです。このオプションは、クリアスウィフトのカスタマーサポートによる明確な指示がない限り有効にしないでください。

4 サイジング例

以下に示すサイジング例では、前項で使用した参考値を使い、組織に必要とされる適切なハードウェアプラットフォームとインスタンスを導き出していますので、参考にしてください。

4.1 マーケティング企業：2,000 ユーザー、100 Mbps インターネット接続

マーケティング企業はマルチメディアコンテンツの使用量が多いため、マルチメディア型のプロファイルに分類されます。インターネット接続は 100 Mbps であり、その使用率を 80%とします。また、全トラフィック中の Web 閲覧トラフィックを 65%とします。

以上の要件から、この企業には次の仕様が必要となります。

- **接続数**：2000 (ユーザー数) x 15% (並行度) x 9 (ユーザーごとの接続数) =**2700**
- **帯域幅**：100 Mbps x 80% x 65% =**52 Mbps**

必要な持続帯域幅は(C)タイプのサーバーで対応可能ですが、接続数は上限を超えています。したがって、(C)タイプのサーバーが 3 台必要であるということが分かります。

4.2 標準的企業：2000 ユーザー、100 Mbps インターネット接続

この企業では継続的にインターネットが必要となる活動は特になく、標準型のプロファイルに分類されます。インターネット接続は 100 Mbps であり、その使用率を 95%とします。また、トラフィックの大部分は地方支社との間の WAN 接続であり、Web 閲覧に使用されるトラフィックは 55%とします。

この企業に必要な仕様は以下の通りです。

- **接続数**：2000 (ユーザー数) x 10% (並行度) x 5 (ユーザーごとの接続数) =**1000**
- **帯域幅**：100 Mbps x 95% x 55% =**52.25 Mbps**

したがって、この企業のトラフィック要件を満たすためには、接続数と必要な帯域幅の両方を満たす(C)タイプのサーバー1 台で十分であるということが分かります。