



マトリックススキャン APEX 技術概要

アンチスパム・ウイルス

マトリックススキャン APEX
アプライアンスサーバー

imatrix
アイマトリックス 株式会社

Confidential

本資料の開示を受けることを承認した者(以下「読者」)は、本資料に含まれるすべてのアイマトリックス株式会社及び関係する第三者の財務、事業計画、マーケティング、顧客、納入者、製品、技術、研究及びノウハウ等に関する情報 および、アイマトリックス株式会社から本資料の説明を受けて得たあらゆる情報(以下「情報」)を秘密にし、これをアイマトリックス株式会社から書面による承諾を得ずに、第三者に開示してはならない。「読者」は「情報」をアイマトリックス株式会社の目的とする事業への参画、関与、取引の検討もしくは製品の使用のみのために利用する。この制約は、「読者」が「情報」を開示される以前に知っていた情報、又は「読者」が法律上の手続きにより開示を要求される情報には適用されない。アイマトリックス株式会社が要請したとき、又は「読者」が事業への参画、関与もしくは取引の可能性がなくなったときは、「読者」は、本資料及び、「情報」を何らかの形で含有、構成、描写又はそれらに関連するあらゆる記録、データ、書類、媒体及びその他の品目並びにそれらの複製で、自らが所有又は管理しているものを、アイマトリックス株式会社に返却するか速やかに第三者に漏洩することなく破棄するものとする。

目次

1. はじめに	4
2. 背景	4
3. フレームワーク	4
4. 構成システム	5
5. フィルタのメリット	6
6. オンデマンドアンチスパムのスパム判定の仕組み	7
7. 検出ルールについて	10
8. スпам収集の妥当性について	11
9. 共有ホワイトリスト	13
10. マトリックススキャンの性能について	14
11. 日本検出センターの性能について	15
12. 検出率について	15
13. 日本検出センターのスパムメール収集状況	16

1. はじめに

APEX システムの技術概要を記述します。

2. 背景

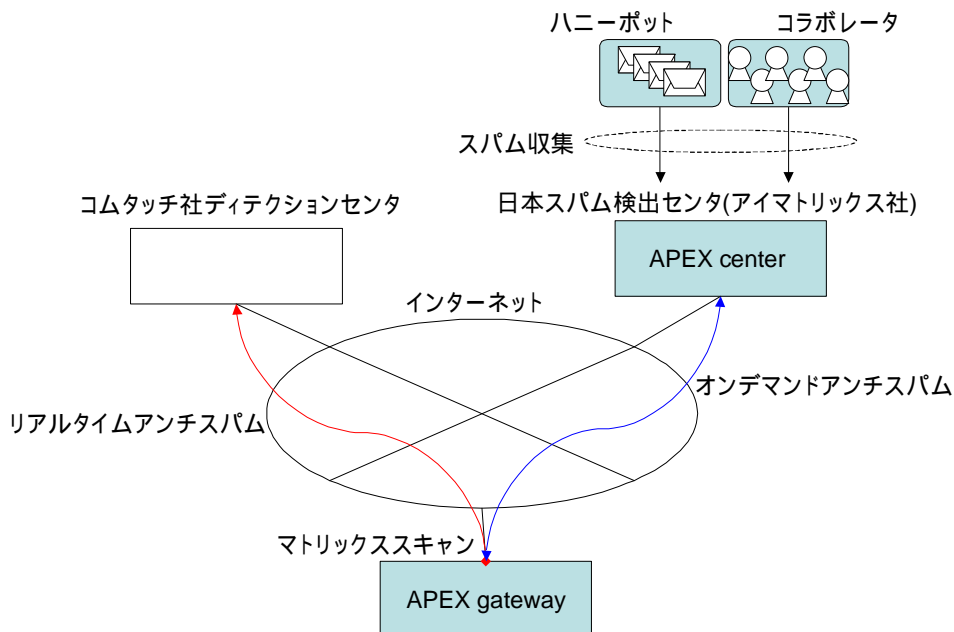
コムタッチ社 RPD テクノロジーは優秀なアンチスパムです。しかし、日本語スパムの特殊事情(アダルト/出会い系が多い。また配信量が世界レベルに達していない。など)より、日本語スパム検出率がやや劣る状態となっています。

この状況を改善するために日本スパム検出センタを開設しました。併せて、日本スパム検出センタに於いて収集した日本語スパムメールから情報を解析・抽出し、日本語スパムにフォーカスした検出方法を開発しました。

この日本スパム検出センタと日本語スパムにフォーカスしたアンチスパムソリューションが APEX です。

3. フレームワーク

APEX フレームワークの概念図を記述します。



日本スパム検出センタ(APEX center)で収集したメールからスパム判定データベースを構築し、マトリックススキャン(APEX gateway)からの要求に対してオンデマンドでスパム判定データを配信します。判定処理はマトリックススキャン(APEX gateway)で行います。

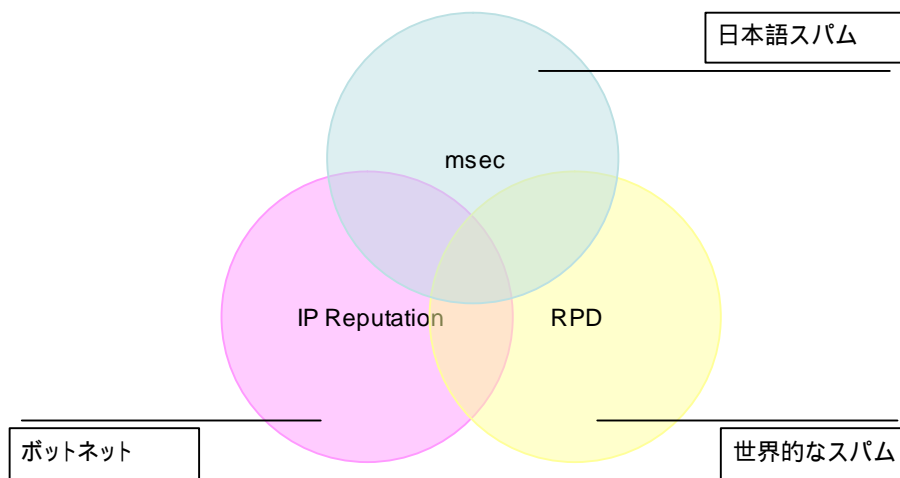
4. 構成システム

マトリックススキャン APEX は以下のシステムから構成されます。

- オンデマンドアンチスパム
 - ◇ msec(matrixscan_security_enhancement_central)
- リアルタイムアンチスパム
 - ◇ Commtouch IP Reputation Service
 - ◇ Commtouch RPD Technology

msec にはリアルタイムアンチウイルス (Zero Hour Virus Protection) も含んでいますが、本書ではアンチスパムを中心に記述するため省略します。

各アンチスパム技術は以下のような棲み分けになっています。



各フィルタは得意分野が異なっており、お互いに補完しあうようになっています。これにより、更に検出率の向上・誤検出率の低下ができます。

5. フィルタのメリット

以下に各フィルタのメリットを記述します。

- msec
 - ◇ オンデマンド
 - ◇ 日本語スパム検出
 - ◇ コラボレーション
 - ◇ 検出ルール(アルゴリズム)のダイナミックな変更
 - ◇ メンテナンスフリー

- Commtouch IP Reputation Service
 - ◇ リアルタイム
 - ◇ Commtouch RPD Technology を応用
 - ◇ ゾンビ/ボットネット検出可能
 - ◇ メンテナンスフリー

- Commtouch RPD Technology
 - ◇ リアルタイム
 - ◇ コンテンツ不問(言語、フォーマット非依存)
 - ◇ 高検出率/低誤検出
 - ◇ メンテナンスフリー

6. オンデマンドアンチスパムのスパム判定の仕組み

オンデマンドアンチスパムは日本スパム検出センタに蓄積されているスパムデータ(*1)をマトリックススキャンがメール判定時にオンデマンドで入手して判定を行います。オンデマンドを実現するためにマトリックススキャンはアウトライン(*2)を用います。

オンデマンドでスパムデータの交換を行うことのメリットは以下のとおりです。

- リアルタイム性の確保

従来型のアンチスパム/アンチウィルスのようにシグネチャをダウンロードする方法ではセンタ側の更新およびゲートウェイ側の更新間隔によるタイムラグが発生します。

オンデマンドにすることで上述の問題を回避し、最新のスパムメールを検出できるようにしています。

- 転送量の低減

従来型のアンチスパム/アンチウィルスのようにシグネチャをダウンロードする方法では一定の間隔で重複したデータや無関係なデータを冗長にダウンロードしていると思われます。

オンデマンドにすることで必要なデータのみ転送します。また、キャッシュや差分転送により転送量の低減を図っています。

- 情報漏えいへの配慮

コムタッチ社の方式は技術的に情報漏えいすることはありません。これは不可逆ハッシュの使用および、断片化した情報のため元メールの復元が不可能なためです。しかし、心情的には懸念が残るケースもあります。

アウトラインは外形情報のみで受信者や本文のいかなるデータも送信しません。このため、情報漏えいが技術的/心理的に発生しません。

*1: スパムデータ

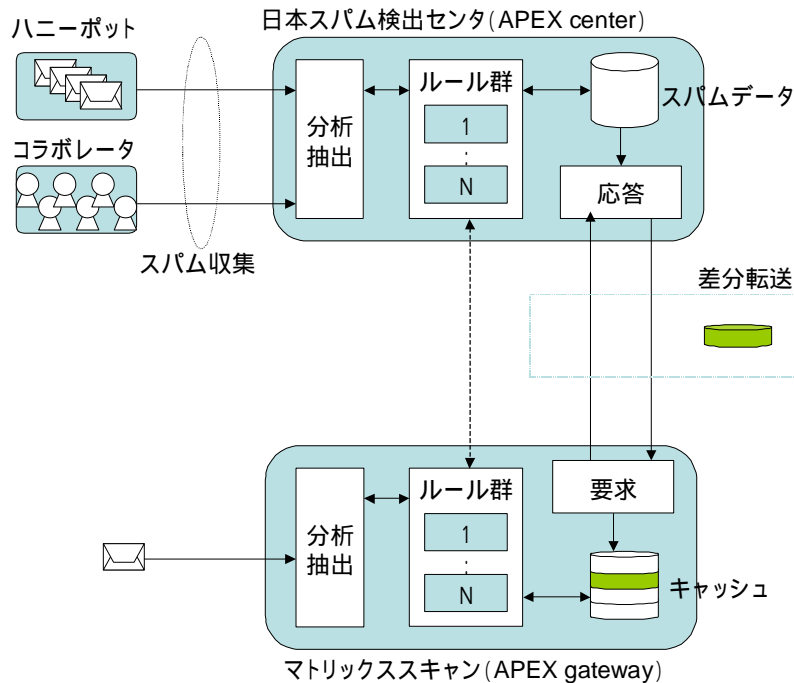
スパムメールを特定するためのデータ: 検出ルールと判定データから構成されません。

*2: アウトライン

メールから外形情報を取り出したものです。

例えば、メールボディ行数、MIME パートの有無、メールボディの形式などです。

以下にオンデマンドアンチスパムのスパム判定フローを記述します。



- 日本スパム検出センタ側スパムデータ更新のステップ
 - スパムメールからアウトラインを生成する
 - 検出ルール群から最適なルールを選択し、判定データを抽出する
 - アウトラインをインデックスとして判定データをスパムデータとして登録
- スпам判定のステップ(マトリックススキャン 日本スパム検出センタ)
 - 受信したメールからアウトラインを生成する
 - キャッシュにあるスパムデータに一致するものがあるか検索
 - 一致したとき) スпамメール
 - 一致しないとき) ステップ へ
 - 日本スパム検出センタにアウトラインに該当するスパムデータを要求
 - アウトラインに該当するスパムデータを検索/応答
 - 受信したスパムデータから一致するものがあるか検索
 - 一致したとき) スпамメール
 - 一致しないとき) 正当なメール
- 検出ルール同期のステップ

先の - のステップに於いてマトリックススキャンが知らない検出ルールが受信スパムデータにあった場合、新しい検出ルールを日本スパム検出センタからマトリックススキャンに同期する

差分転送(-):

マトリックススキャンはキャッシュに保存されているスパムデータを除いて日本スパム検出センタにスパムデータを要求します。

日本スパム検出センタは新着または更新されたスパムデータが存在したとき、差分スパムデータとしてマトリックススキャンに応答します。

差分転送により冗長なスパムデータによる通信を低減します。

7. 検出ルールについて

オンデマンドアンチスパムでは検出ルール(アルゴリズム)を用いてスパムメールを検出するデータを抽出/判定します。

通常のアンチスパム/アンチウィルスゲートウェイでは検出するアルゴリズムを変更する場合は、センタ側/ゲートウェイ側ともにバージョンアップ等が必要となります。

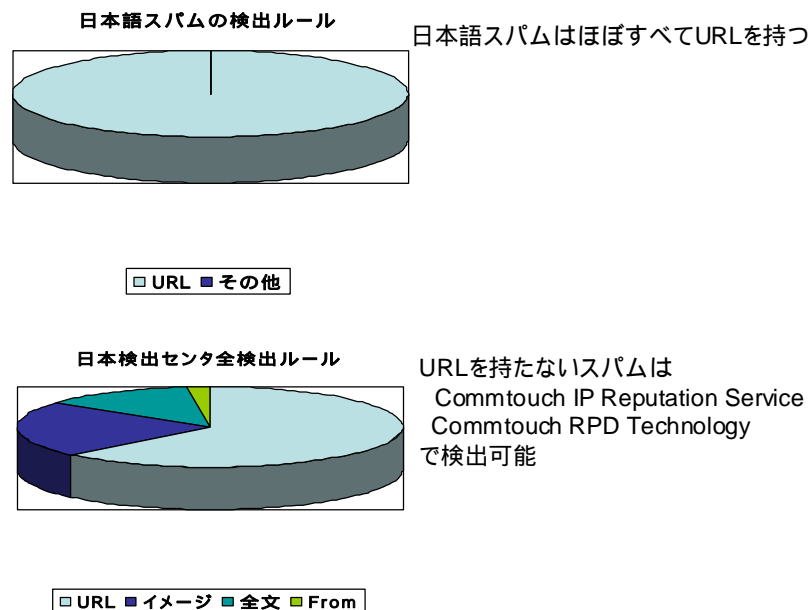
オンデマンドアンチスパムでは検出ルールをダイナミックに適用する仕組みを実装しています。これは変遷するスパマーに柔軟に対応するためです。

既存の検出ルールで検出できない新たなスパムメールが発生した際にAPEXモニタリングチームにより新しい検出ルールが日本スパム検出センタに適用されます。新しい検出ルールは、日本語スパム検出センタに適用されたマトリックススキャン側にシームレスに適用されます。

オンデマンドアンチスパムは日本語スパムへの柔軟な対応を目標に設計されました。検出ルールはセマンテック/非セマンテックな解析を行い、検出データを抽出します。

現時点(2007/7)で日本スパム検出センタに蓄積されている日本語スパムメール約99%が何らかのURLを持ちます。

オンデマンドアンチスパムでは上述のような日本語スパムを検出可能です。また、その他の検出ルールも実装しています。



8. スпам収集の妥当性について

日本スパム検出センタはハニーポットおよびコラボレータからの報告によってスパムデータの拡充を行っています。

妥当性の検証は、まず系統的に APEX 監視システムで行います。APEX 監視システムで判断できないメールは APEX モニタリングチームにより解決します。

APEX 監視システム：

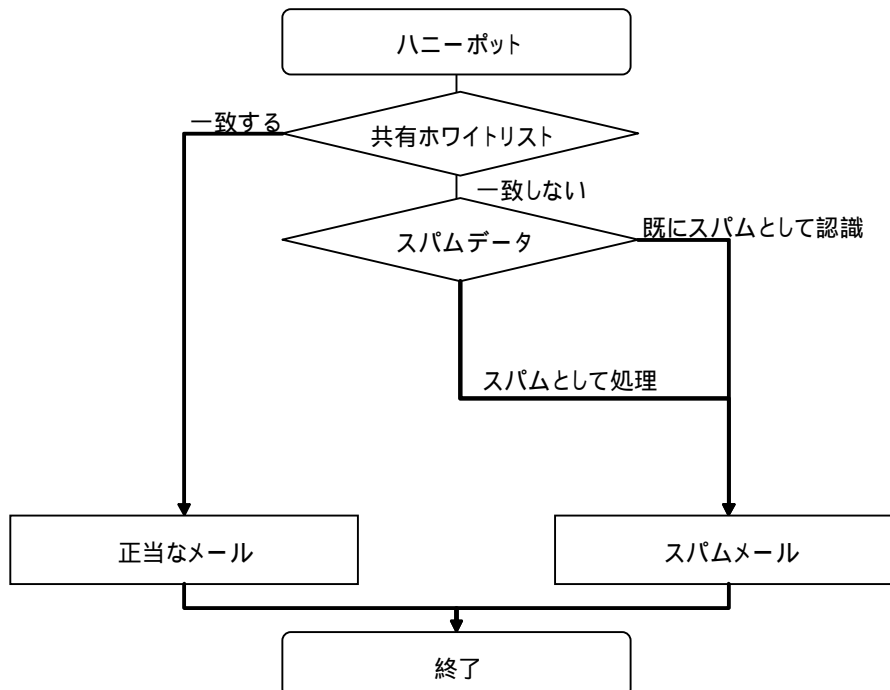
報告元の信頼度(報告回数と報告の成否率)と報告されたメールの再発度によって自動的に判断します。例えば APEX モニタリングチームからの報告は報告の意図に副う確率が高いです。また、同じスパムデータを持つ報告が単位時間あたりに再発すると信頼度が高いと判断します。

APEX モニタリングチーム：

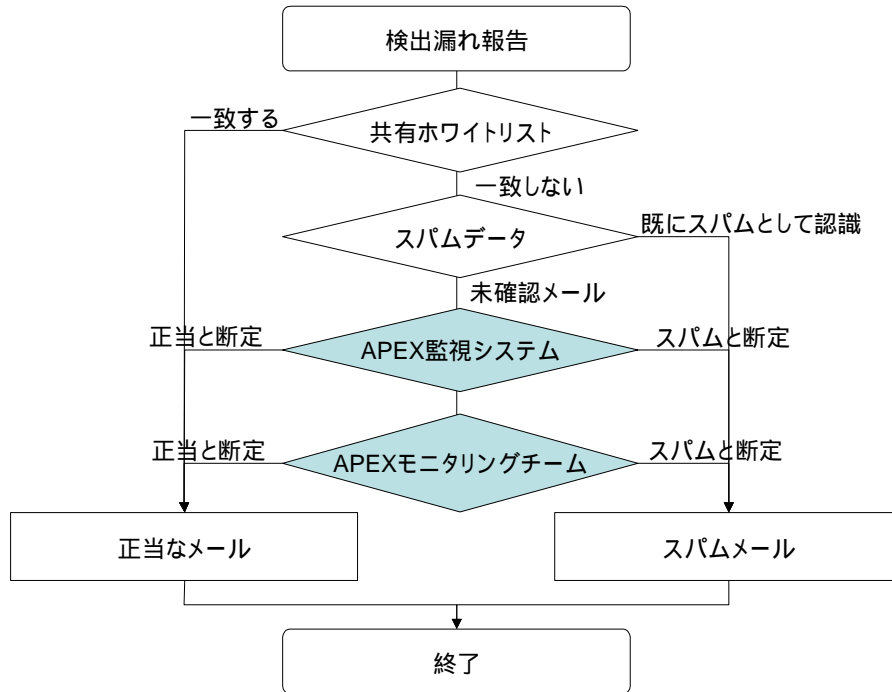
日本スパム検出センタを運用しているメンバーによる知識・経験から判断します。

以下にスパムデータ拡充時の妥当性検証について記述します。

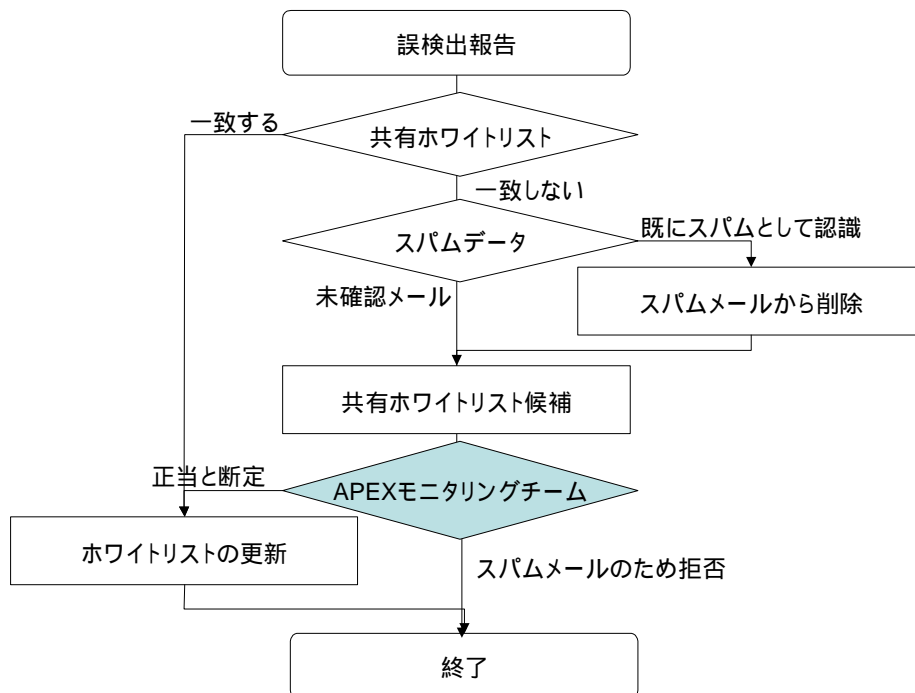
- ハニーポットにスパムメールを受信したときの妥当性チェック・システム



● 検出漏れ報告を受け付けたときの妥当性チェック・システム



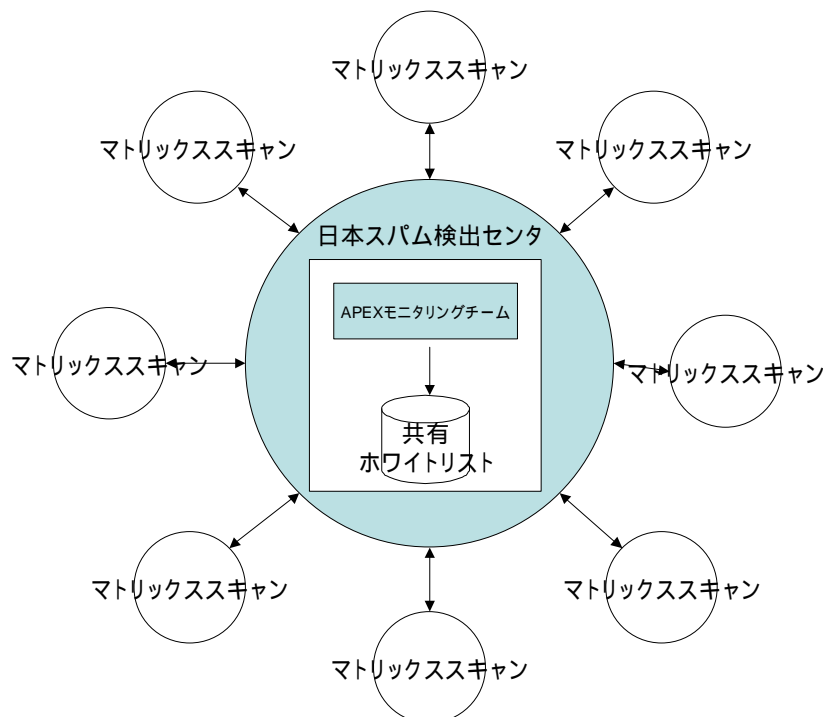
● 誤検出報告を受け付けたときの妥当性チェック・システム



9. 共有ホワイトリスト

日本スパム検出センタでは、共有のホワイトリストを持つことにより、リアルタイムメールプロテクション全体の誤検出率の低減を行います。共有ホワイトリストはコラボレーションモデルで収集されます。

収集された共有ホワイトリストへの登録依頼は APEX モニタリングチームによって精査(*1)され共有ホワイトリストへの登録を行います。



*1: APEX モニタリングチームによる精査

以下の項目について調査/対処します。

- 既存の共有ホワイトリストとの適合、最適化
- IP アドレスから得られる情報(DNS, whois, Reputation, ドメイン世代など)
- ドメインや差出人(From ヘッダ) から推測される Web サイトの調査

マトリックススキャンで共有ホワイトリストの使用する/しないは設定により変更可能です。また、共有ホワイトリストの登録依頼は Web ユーザインタフェースから行います。共有ホワイトリストに関しての設定/操作方法は『マトリックススキャン APEX オペレーション・マニュアル』を参照願います。

10. マトリックススキャンの性能について

リアルタイムメールプロテクションは1通のメールを最大3つのフィルタでスキャンします。アプライアンスモデルによりハードウェアスペックが異なります。また、メールサイズやトラフィック量など環境に依存することも多くあります。また設定などは標準値のまま測定を行いました。

このため、チューニングによりさらに処理能力を向上させることは可能です。あくまで従来製品との性能比として参照願います。

以下に性能を記述します。

【注意】 弊社が何ら性能を保証するものではありません。

● 測定方法

ハードウェア(使用モデル):	マトリックススキャン Model-1000
メールデータ:	スパムメール 50 種類 正当なメール 50 種類
多重度:	100 コネクション

● 測定結果

従来のマトリックススキャン: Commtouch RPD Technology のみ	約 60 通/秒
マトリックススキャン APEX: すべてのフィルタ(* 1)	約 50 通/秒

従来に比べ 15%~20%程度性能が低下します。しかし、メール流量から見れば十分な性能が確保できています。(通常のメールサーバの平均的なメール流量は 20 通/秒程度が基準となります)

* 1:すべてのフィルタ

msec

Commtouch IP Reputation Service

Commtouch RPD Technology

11. 日本検出センターの性能について

日本検出センターはサーバ1台あたりマトリックススキャンの 10 倍の処理能力を持ちます。現在、マトリックススキャンの出荷されているモデルからピーク時のメールトラフィックを計算し、そのピーク時のリクエストを処理するのに十分なサーバを確保しています。

サーバ稼動状況はリアルタイムに APEX モニタリングチームにより監視されています。あらゆる兆候をいち早く把握し、必要な場合はサーバをスケールアウトします。サーバスケールアウトに伴うサービスの停止はありません。

12. 検出率について

リアルタイムメールプロテクションの検出率を以下に示します。

【注意】 弊社が何ら性能を保証するものではありません。

● 測定方法

弊社メールアカウントのうち比較的スパムが多くなるメールアカウントをサンプリング調査しました。検出率の測定は瞬間的な大量のスパムメールを行う必要は無く、ある程度継続した期間中に多くの種類を検出できるか否かになるという観点からです。

瞬間(ある期間や時間のみフォーカス)検出率 100%はありえますが、一定期間を通して測定した場合、検出率 100%はありえません。

● 測定結果

約 20%の検出率向上(マトリックススキャン APEX : マトリックススキャン)

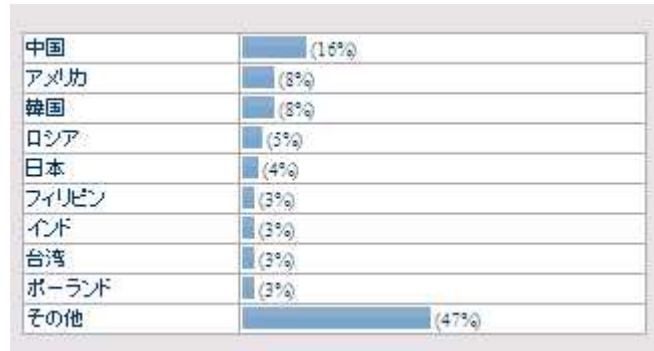
マトリックススキャン APEX の詳細

リアルタイムメールプロテクション:		98.8%
	msec:	87.5%
	Commtouch IP Reputation:	50.7%
	Commtouch RPD Technology:	83.1%

Commtouch RPD Technology は本来スパムメールとして扱うべきバルク判定を検出率に含めていません。あるフィルタの不得意なスパムメールを他のフィルタが補うため全体で検出率が向上しています。

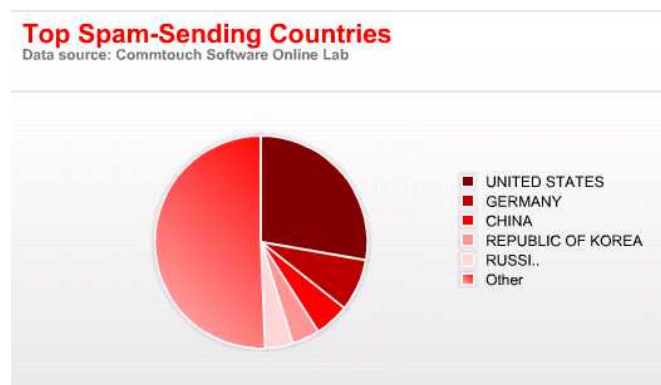
13 日本検出センターのスパムメール収集状況

日本検出センターで収集しているスパムメールの発信国別状況は以下の通りです。



2007/7 現在

コムタッチ社ディテクションセンターで収集しているスパムメールの発信国別状況は以下の通りです。



2007/7 現在 (<http://www.commtouch.com>)

集計期間などの要素に依存するため、一概には言えませんが、日本検出センターは、日本語スパムにフォーカスしているため、傾向として世界中のスパムメールと異なる傾向を示していると考察します。