

# Postini White Paper

---

## **Spam Filtering Effectiveness**

- **Problem Definition**
- **Methodology to Address the Problem**
- **Service Components**
- **Common Questions**

## Introduction

An email provider that adopts Postini's spam filtering services for its subscribers will not encounter any noticeable performance impact to existing infrastructure. Moreover, the provider will lower total cost of ownership by reducing undesirable network traffic resulting from spam while improving end user satisfaction toward quality of service. For the end user, Postini provides the granular filtering capabilities, defined on a per-user basis, to raise the accuracy and effectiveness of mail filtering. For the long term, Postini will continually improve and evolve its spam filtering techniques to maintain the effectiveness of the service provided as the subjective nature of junk email changes. Postini will also continue evolving the range of services available through the platform.

## Objectives

This white paper serves to illustrate Postini's junk email filtering performance as found in the Junk Email Assistant. Performance is described as it applies toward two audiences:

- **Email Service Providers (ESP)** – Performance figures illustrate that the Postini Junk Email Assistant delivers a superior solution while reducing the burden on server resources and preserving an email provider's ability to maintain consistent performance. Moreover, the service also lowers costs by reducing undesirable traffic and, consequently, minimizes customer support inquiries for the long term. Postini promotes customer retention and satisfaction as email providers improve upon the overall user experience.
- **End Users** – Undesirable commercial email is primarily a nuisance to many users and produces a poor quality-of-service perception. It may have a marginal impact on productivity as users are forced to review irrelevant email messages. Junk email devalues the relevancy of electronic mail as an effective communications medium by creating an unwelcome “noise” to normal communications.

## Terminology Definitions

- **Junk Email** – Unwelcome, unsolicited email; the definition varies depending on the recipient. Also referred to as spam or unsolicited commercial email (UCE).
- **False-Positive** – During junk email filtering, a legitimate message erroneously quarantined as “junk email” due to any number of characteristics within the message.
- **Postini Junk Email Assistant** – The Postini application service that provides users with junk email filtering services, independent of the email provider's existing infrastructure.
- **Postini Message Platform** – Postini's services platform that communicates in common SMTP to the existing mail servers used by service providers with minimal configuration setup. The Postini Message Platform is the service foundation for the Junk Email Assistant and other hosted Postini applications.
- **RBL, ORBS, DUL** – Three popular consumer advocacy groups that track either known domains, points-of-entry, or specific addresses believed to promote spamming.

## The Junk Email Problem Defined

Junk email has evolved from many previous forms of messaging since the dawn of advertising. In the broadest definition, spam can be considered as an unwelcome ad. In the recent 1980s, unsolicited messages migrated from physical junk mail to faxed advertisements—and now in the digital age, junk email is delivered via electronic means. Regardless, since the rise of the Internet

and the commercialization of email, service providers have been battling this new epidemic which, under some estimates, may constitute approximately one-third (1/3) of a network's email traffic.

As the rise of junk email escalated dramatically during the 1990's, service providers (AOL, Mindspring, et. al) and email client developers (Microsoft, Qualcomm, et. al) implemented content identification rules systems for flagging and quarantining junk email both as a server-side solution and a client-side solution, respectively. At the same time, various groups were formed with the mission to identify and classify junk email (such as RBL, ORBS, DUL)—thus providing a reference collection of junk email for the fight against spam. These approaches provided only moderate success as they operated under the premise of a static rules system applied to the entire volume of diverse email traffic or, alternatively, reliance on known spammers with the inherent limitation of continuously changing spamming tactics and origins to avoid recognition.

Despite the monolithic efforts to curb the junk email flow, it is ultimately the subjectivity of junk email that prevents effective user-specific control of the problem. From an individual user's perspective, junk email is somewhat a subjective experience and definitions vary. Some people feel that any sort of advertisement from any source, legitimate or not, is considered junk email—they don't want to hear from anyone or any organization unless they know about it first. On the other hand, some are grateful to hear about new money making offers if they are currently seeking a new job. Given the broad definition of spam that may include everything from known advertisers to criminal abusers, the challenge becomes providing user-definable filtering that suits each person's criteria.

## Addressing the Problem

During the early years of commercialized Internet services, most email providers used rather draconian methods to handle spam. They would decide for a user if an email was spam, and simply bounce or delete the offending email at their mail servers before it ever reached the recipient. Given the fact they did not want to delete any valid, legitimate, email, this system was not very effective since they had to be absolutely sure an email was malicious and thus should not be forwarded—how could they really know? At the same time, popular email clients, such as Outlook and Eudora, tried including junk email filters within the email client; but then users were required to define awkward rules that were applied *after* the messages were already downloaded.

Since most everyone agrees that one person's definition of spam may not hold true for another, Postini proposes an innovative way to fight the problem, without requiring users to perform an extraordinary amount of mail management. By filtering email at Postini servers using a combination of intelligent rules and public databases of known junk email and spammers, Postini can process large volumes of email automatically, each message in milliseconds, while achieving economies-of-scale. Further, neither the email provider nor its users are required to install and maintain products that tax their existing work environments.

Unlike the "one size fits all" solutions of the past that an email provider may have used, Postini allows email users to configure the Junk Email Assistant to reflect personal tastes. Postini allows each user to determine personal sensitivity towards categories of junk email types in order to temper diligent spam control against overzealous filtering, as well as more objective, cut-and-dry, methods for maintaining a personal list of always "approved" or always "blocked" senders.

### Performance Objectives: Service Provider

Email providers that deploy the Postini Junk Email Assistant will provide users with a more valuable communications medium that doesn't degrade performance. Providers will experience lower total cost of ownership through bandwidth savings while reducing customer support inquiries.

### Maintaining Current Performance

A critical aspect of deploying any new solution is minimizing the integration complexity and time-to-market, while ensuring the email provider’s existing infrastructure is not overburdened with new hardware and software. To this end, the Postini Junk Email Assistant works in conjunction with the service provider’s existing mail server to process email *before* delivery as a prior hop during mail flow. The service solution functions as an application hosted on the pre-processing Postini Message Platform. Although email delivery is not a real time communication system such as online chat, users are accustomed to receiving emails within seconds from the time sent.

Postini in-line junk email processing before delivery to the email provider’s servers can be measured within *milliseconds* for each message to prove there is no significant impact on email delivery expectations. Further, Postini junk email processing is platform-independent—email providers can expect quick time-to-market for deploying the service to an entire customer base with no downtime or integration expense, and no software or hardware to install and maintain.

### Lower Cost of Ownership: Bandwidth Savings and Thwarting Attacks

Cost savings can be measured for each service provider by measuring the percentage of junk email traffic quarantined on Postini servers and subtracting that cost percentage from the total bandwidth cost devoted to email trafficking. Since the amount of spam as a percentage of any service provider’s email bandwidth varies, savings from eliminating junk email delivery varies accordingly. Postini end-user surveys have shown up to 30% of normal email delivery categorized as junk. If the cost of email trafficking is known, it is easy to determine the cost savings over time.

Top 50 domains by messages received:			
domain	received	delivered	quarantined
*****.com	240450	181153	59297 (24.66%)
*****.com	37089	17158	19931 (53.74%)
*****.com	15273	3232	12041 (78.84%)
***.net	14205	11903	2302 (16.21%)
*****.net	12035	10052	1983 (16.48%)
*****.net	10235	8580	1655 (16.17%)
*****.com	7062	6414	648 (9.18%)
*****.com	6821	4526	2295 (33.65%)

Filter hits	
87198	bulk
35227	DNS
9403	commerce
3421	user_bad
2357	mmf
160	naughty
31	racist

• Table: Performance Report Example. (NOTE: “\*\*\*\*\*” = actual Postini customers, domains withheld.)

In addition, average bandwidth cost savings does not take into account periodic malicious attacks that can severely impact the performance of mail servers to the point of non-performance. Through Postini pre-processing, mail servers are not only prevented from trafficking junk email, they are alleviated from the consequences of denial of service attacks (“mail bombs”) and similar threats.

### Lower Cost of Ownership: Customer Service Reduction

Postini surveys of current customers have revealed that up to 15% of support inquiries are junk email-related and impact customer satisfaction of the service. By reducing the problem of unwanted email, email providers can expect customer support inquiries to be reduced accordingly while encouraging retention and decreasing possible customer churn.

## Performance Objectives: Email Users

Ultimately it is the end user messaging experience that is improved upon through the adoption of the Postini Junk Email Assistant. By eliminating unwanted email, individuals are provided a more enjoyable and productive experience using email. Performance objectives include the following:

### Accuracy

Perhaps one of the most difficult aspects of measuring spam-filtering effectiveness is the assessment of accuracy. This is in large part due to the subjective nature of what may be considered undesirable email for each user and the trade-off of miscategorization. With this in mind, Postini can achieve the following effectiveness range:

- ▲ Postini filters can achieve greater than **99%** effectiveness toward quarantining undesirable junk email, defined by the user as containing suspicious content.
- ▲ Further, Postini can achieve less than **0.01%** of legitimate messages erroneously categorized and quarantined as “false-positive” due to the subjective nature of the content. \*

*\* Internal Postini testing*

Because each person can subjectively define junk email (“one man’s junk is another man’s treasure”), Postini builds in a safeguard to ensure no legitimate email is lost. Suspicious email is quarantined to a user’s personal message center hosted by Postini and the user receives a periodic email notification requesting for the messages to be reviewed. This safeguard ensures any legitimate email is not erroneously discarded for any user while also offering the opportunity to make incremental filter adjustments that can continually improve accuracy longer term.

### Transparency

Postini strives to ensure the end user experiences no undue burden through the use of Postini services. To this end, the Postini Junk Email Assistant works behind-the-scenes and in conjunction with the email provider’s existing mail servers. End users continue to use their preferred mail clients without any perceivable change in message delivery; however, the user is empowered to periodically fine-tune junk email filtering services through the Web-based interface, if desired.

### Increased Productivity and User Confidence

A reduction in undesirable email at the server level saves user download times, eliminates daily management duties of evaluating undesirable messages, and reduces the headaches spent battling the spam problem on a continual basis. Improved productivity also results in increased user confidence of the provider’s ability to deliver a valuable service while differentiating the service perception as a better value.

## Service Components

While benefiting both the service provider and end user, Postini junk email processing is designed to both achieve large-scale economies-of-scale by processing carrier levels of mail flow, while also providing catered, per-user, control to increase the effectiveness.

## Availability: Infrastructure Platform Designed for Robust Capacity and Stability

The Postini Junk Email Assistant runs as a service hosted on the Postini Message Platform, the Company’s highly scalable infrastructure for hosting value-added email applications. Applications

are developed to run on “carrier-class” hardware systems to process millions of messages per day while in tandem providing unparalleled user-specific visibility into the filtering applied to each message passing through the system. Moreover, only minor integration is required for the service providers and messages are transferred using standard SMTP. To ensure against any single points of failure, Postini services are made redundant to provide fail-over.

## Quality: Accuracy of Junk Email Identification

To ensure the Postini Junk Email Assistant accurately diverts all forms of junk email, Postini employs a number of techniques to optimize the filtering process for greater accuracy. However, it is most significantly the inclusion of user-specific definition that takes an otherwise broad-based approach and personalizes it for increased per-user effectiveness. The components that make up Postini’s Junk Email Assistant include:

- **Public Databases** – Known databases of junk email offenders help up-front to eliminate the guesswork of subjectively evaluating the message based on content alone.
- **Heuristic Rules Engine** – Postini evaluates the components of each message by applying over 400 rules to determine how closely junk email characteristics are met.
- **End-User Configuration** – By providing each user with control over application parameters, the filtering process can suit each user’s personal definition of junk email.

### Public Databases of Known Junk Email

Databases of known junk email are leveraged to eliminate the subjective guesswork up-front. These databases, such as RBL, ORBS and DUL, function much like virus pattern files to the extent that a message is identified as originating from a known offender—a blacklisted domain, point-of-entry or individual address. These public databases help eliminate the speculation before Postini’s heuristic filters are applied for subjective evaluation.

### Heuristic Engine: Rule-based Filtering

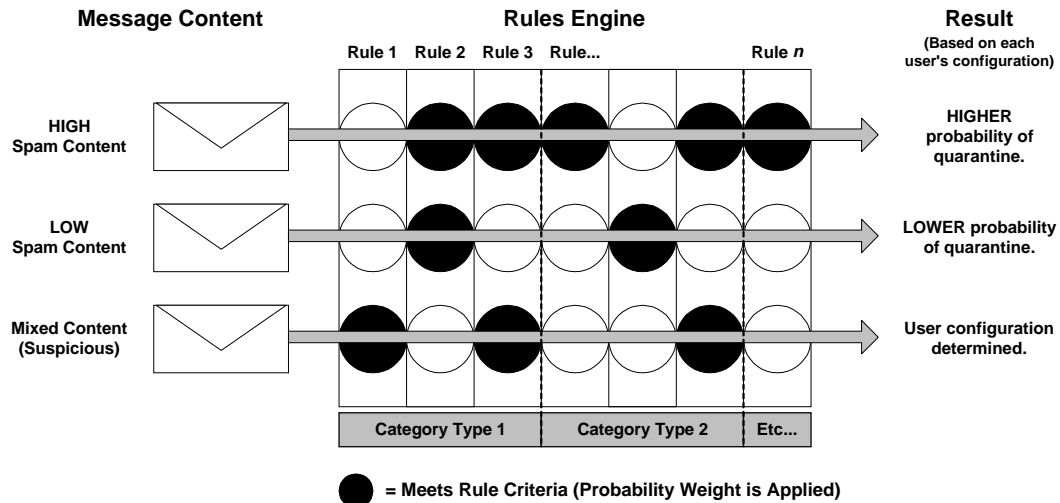
The core technology employed by Postini includes a large number of heuristic rules used for assessing email based on content type and message context. Rules are optimized and continually tuned as environmental changes dictate. Each rule indicates a unique probability a message can be defined as junk email. For example, in simple terms, the rules can appear as the following:

```
Rule 1: Excessive use of exclamation points
Rule 2: More than 3 statements of "special offer!" or "act now!"
Rule 3: "Unsubscribe" option follows unconventional formats
```

### Rule Weighting

Each rule is combined with a junk email likelihood score to indicate the relevancy of the rule. In the above example, “Rule 1” can be assigned a weighting that indicates it has low relevance, since often people simply abuse the use of punctuation. However, “Rule 3” may be given a higher weighting since it can more clearly differentiate junk email from legitimate mailing list messages.

When an email passes through Postini’s system, the entire rules set is applied within milliseconds. If the email meets the criteria of a specific rule, that weighting is added to the overall probability score and the evaluation increases the likelihood of the message being considered as junk email.



- Body of rules applied to each incoming message determines junk email probability.

For example, during the filtering process with each rule being applied...

$$\text{Rule 2}(0.05) + \text{Rule 3}(0.12) + \text{Rule 5}(0.11) + \text{Rule 248}(0.11) \dots = \text{Score: } 0.72$$

Given the possibility of a message 75% likely to be junk email, it is then evaluated against each user's personal configuration to determine scoring adjustments and the eventual consequence of the message—either quarantining the email or passing the message through for delivery.

### Probability Trade-Off Scale

Given a large body of rules, Postini has tested its rules engine for accuracy by evaluating a large body of junk *and* legitimate email. This scale balances the precision of junk email filtering with the consequences of legitimate email mistakenly quarantined as spam (“false-positive”).

There is a direct correlation between the number of junk email quarantined and the increasing likelihood of legitimate email mistakenly quarantined. For example, it may be possible to capture 99% of all junk email when a user sets all filtering options aggressively; however, a relatively *higher* number of legitimate messages may also be captured. Similarly, *less* junk email can be identified for a user when the *lowest* probability of mistaken quarantine is desired (1 per 1000 legitimate).

Given the broad range of possibilities, Postini segments a few of these “trade-offs” into user-definable configuration settings so that a sensitivity between “lenient” and “aggressive” can be chosen by the user which best reflects personal tastes for accepting filtering errors. For reference:

User Setting	“Junk” Email Caught	“Legitimate” Email Caught
Lenient	Least	Least
Moderate	Moderate	Moderate
Aggressive	Greatest	Greatest

Using a rules engine to identify an email as spam is not an unusual concept; hosting the filtering on Postini's highly scalable platform however, provides the economies-of-scale to process millions of messages with no noticeable latency for users and no impact on an email provider's infrastructure. The most significant differentiator, though, is the ability to organize the filters based on a specific user's tolerances and content preferences. This level of customization vis-à-vis a “one size fits all” solution is what provides the superior effectiveness of Postini's implementation.

## Control: User Management of Filtering Application

The most significant impact on filtering effectiveness is the user’s control over how junk email is defined. The phrase “one man’s trash is another man’s treasure” illustrates simply how junk email filtering is largely ineffective when addressed solely at the server level, invisible to each user. Although email client applications may provide more control for each user, performance and manageability issues have stalled adoption. Instead, Postini’s innovative approach provides behind-the-scenes high performance coupled with user control before email arrives at the desktop—eliminating needless junk email trafficking for users and email service providers alike.

### Web-based Message Center: Filter Configuration

Postini provides end user visibility into its system via a web-based interface for managing the personal settings related to the quality of junk email filtering. From the web, users can configure filtering parameters based on personal tastes, with an added opportunity to review any possible quarantined junk email when the content is withheld as questionably suspicious.

To tailor the filters for each user, Postini employs three methods to give users greater control over their particular junk email identification process: Sensitivity, Categorization and Explicit Approvals.

### Sensitivity Levels

With the accuracy trade-offs described previously, a compromise exists between the possibility of a message being junk email and incorrect categorization of “legitimate” messages. Rather than make a blind decision for all users, Postini empowers each user to determine a preferred tolerance. Although sensitivity can be segmented into infinite combinations, Postini simplifies the process based on generalizations most users have toward erroneously quarantined, legitimate, messages.

**Filter Sensitivity**

Select a level of protection that meets your needs.

If too many legitimate emails are being quarantined, adjust the sensitivity of the setting to a more lenient setting or add specific senders to your approved senders list.

Lenient  FEWER suspicious email messages are directed to the Message Center.

Moderate  MORE suspicious email messages are detected, but errs on the side of delivering messages rather than directing to the Message Center.

Aggressive  MOST suspicious emails are detected, but trying to keep your inbox as clean as possible, this setting could direct valid messages to the Message Center.

• User Filter Sensitivity Control.

The tradeoff is simple: If the user wishes Postini filtering to be more diligent, there is greater likelihood that some legitimate mail will also be misclassified and, thus, quarantined from delivery.

### Categorization






To help improve filtering further, Postini provides visibility for each user into the types of filters that can be applied to each message. Again, with simplicity in mind, Postini provides four generalized categories that meet the needs of most users.

In this case, the entire body of rules is classified into distinct groups around sexual content, moneymaking schemes, racial insensitivity and special offers—with a general category for the

majority of known bulk email. Again, the level of granularity can be taken to any point; however, Postini has found that most users' tastes generally follow these primary areas of concern.

### Category Filters

Select the specific categories of junk email that you want to block from your inbox. The Postini "Bulk Mail" filter will block most general-purpose spam.

Sexually Explicit 	<input checked="" type="checkbox"/>	Don't be subject to someone else's tastes - Block inappropriate sexually-oriented messages.
Get Rich Quick 	<input type="checkbox"/>	Bothersome schemers and scammers - Block the money-making offers and high-risk investments messages.
Racially Insensitive 	<input checked="" type="checkbox"/>	Insulting jokes and perspectives - Block offensive messages about sex, creed, religion, and race.
Special Offers 	<input checked="" type="checkbox"/>	"Too good to be true" email messages often are just that - Block annoying advertisements.
Bulk Mail 	<input type="checkbox"/>	Commercially unsolicited email is blocked automatically by activating Postini junk email filtering.

• User Category Definition Control.

By choosing to activate any particular category of filters, users employ only the specific rules used to determine that type of content. From a previous example, "Rule 2: More than 3 statements of 'special offer!' or 'act now!'"—along with other similar rules in this category—would not be taken into consideration for identifying "Special Offers" if the user does not have that set of rules ("Category Filters") selected within his or her personal configuration.

### Explicit Approval or Blocking

To complement each user's category tastes and sensitivity tolerance, Postini increases filter effectiveness further by providing a means to explicitly *block* or *approve* known sender addresses or domains. This allows users to specifically quarantine or pass through email when the sender is legitimate, but the content is questionable ("approve")—or alternatively, the sender is unwelcome, although the content may be considered reasonably legitimate ("block").

For example, a user may subscribe to a weekly electronic newsletter featuring discounts at a popular sporting goods store. Although the emails explicitly include advertising, the recipient specifically wants to know about the new deals. In this case, if the newsletter is being quarantined as junk email, the user can simply add the newsletter email address to his or her always "approved" list and it will not be blocked further. Similar messages originating from other addresses will continue to be filtered as usual.

### Web-based Message Center: Junk Email Review

To compensate for the possibility that some suspicious email will fall into this gray area of legitimate addresses that may send suspicious content, Postini provides a web-based Message Center for reviewing all email. This ensures that any questionable email is not immediately discarded, rather, pending review for each user before discarding.

**Message Center**

Postini Quarantined the Following Junk Email Messages

[View Trash](#)
Message 1 - 4 of 4

<input type="checkbox"/>	Sender	Subject	Filter	Date
<input type="checkbox"/>	kimlu@UBBG.ETF.BG.AC.yu	Take Off My BLACK TEDDIE Tonight		Wed 01-31
<input type="checkbox"/>	R-6-139574-329818-2-12070-UK1-...	Calling all Juventus fans!		Wed 01-31
<input type="checkbox"/>	news@mx0.j2.com	Free DSL from j2 and Winfire		Wed 01-31
<input type="checkbox"/>	millions1cd@aol.com	Introducing.....		Tue 01-30

[Select All](#)
[Unselect All](#)

[View Trash](#)
Message 1 - 4 of 4

- User review of suspicious, quarantined, email ensures no legitimate email is ever lost.

If an email is legitimate, users can deliver the message from their Message Center with the added opportunity to add the sender to an “approved” list for the future. In this way, a user’s personal filtering becomes more accurate over time—again, without impacting email servers or the desktop.

**Approve Senders**

Select sender(s) to add to your Approved Senders list:

messaging.today@messagingonline.com

- Legitimate email quarantined at Postini can be forwarded and the sender approved for subsequent email.

## Common Questions

**If a user has a configuration set for “Aggressive” filtering, could he or she still receive a message that is considered junk email or otherwise offensive?**

In general, a few suspicious words will not trigger a filter—rather, the combination of phrases, sender/recipient addresses, contextual reference, and legitimate content determines the probability an email is considered spam. For example, if you receive an email with, “My new sports car is quite sexy,” it is unlikely this would trigger the “sexual content” filter; but if other vulgar terms or an explicit URL are present, the likelihood increases. The challenge is toward providing a rules system accurate enough to be effective, but flexible enough to be catered to every user’s personal tastes. However, with more indicators, there is greater possibility someone’s minimum thresholds are met.

**If a user belongs to a legitimate mailing list, how will the filtering work?**

Postini does not generally consider mailing lists to be junk email. Mailing lists tend to be explicitly subscribed to by a user. If the user belongs to a list that repeatedly and periodically sends a message from a known address, the user should unsubscribe if the email is undesired—if this is impossible to do, Postini can be configured to always “block” that address.

There is the possibility, however, that mailing list messages could be mistakenly quarantined as spam, especially when they contain advertising. To compensate, Postini employs a number of behind-the-scenes techniques to remedy the “gray areas” that arise. For example, Postini maintains a list of addresses for industries such as financial, online ticketing, auction sites, and so forth to ensure a higher likelihood these types of message will pass through.

**What ultimately decides whether a message is quarantined or passed through?**

Although a message may be assigned a probability of being junk email regardless of a user’s configuration, the action of quarantining or passing through the message is determined based on the user’s specific settings. One user may have the action of quarantining lowered if certain category filters are turned off (some filters won’t be applied toward the probability). For other users, a low sensitivity setting may determine that the message is passed through. Each user has a personal criteria for junk email tolerance and Postini provides the tools to cater to each individual.

**What happens when Postini makes a mistake and quarantines legitimate email?**

Realizing that many messages fall into a gray area of being considered suspicious, Postini does not blindly discard any questionable messages. Through a personal web-based interface, each user can visit her Postini-hosted Message Center to review suspicious email. If a user never receives any suspicious email, he can simply go about his business and never have to worry about spam. However, if an email is quarantined at Postini, the user will receive a notification at least once a week asking to review the suspicious email—in this way, the user can wait until a significant number of messages has been gathered rather than dealing with daily, if not hourly, disruption.

**Is the processing of a user’s personal email private and safe?**

Yes. All of Postini’s filtering components are automated and process each message within milliseconds before handing it off to its next destination. Messages that are held as suspicious are kept in each user’s personal, secure, quarantine until review—regardless, they are deleted after a certain time period if the user does not wish to review the suspicious messages. Postini systems are hosted at the same data centers that run many of the world’s premier web properties, and thus have equal, if not better, security safeguards than most email service providers maintain.

For all legitimate email that is not identified as spam, Postini passes the email to the destination server directly after the few milliseconds required process the email.

## Summary

Postini provides an innovative and highly effective approach to junk email filtering that delivers the power of “carrier-class” hardware with granular, per user, control of filtering processes. By hosting the services off-site, Postini saves the service provider from trafficking junk email that can constitute approximately 30% of total email traffic and even protect servers from denial of service or junk email storm attacks. Moreover, each user does not have to spend time managing junk email on the desktop; rather, more time can be spent towards productivity.

Due to the subjective nature of email, Postini gives users the option to set personal sensitivity tolerances as well as determine the categories of junk email to quarantine. Junk email is diverted to a user’s personal Message Center for review to ensure that suspicious email is not mistakenly categorized as spam. By employing a heuristic rules system, Postini can accurately determine the probability of a message being considered junk email; however, it is the user’s specific configuration that ultimately determines whether a message is passed through.

The Postini Junk Email Assistant, through both a comprehensive rules system, combined with user-specific settings, provides the highest effectiveness of any similar solution on the market today and helps service providers deliver a higher level of customer satisfaction to subscribers.