SATBAYEV UNIVERSITY

Computing & Engineering

*Volume 3 (2025), Issue 2, 1-6*

https://doi.org/10.51301/ce.2025.i2.01

# Research of methods and algorithms for protecting network resources from spam

O. Kadyrov*, A. Batyrgaliyev

*Satbayev University, Almaty, Kazakhstan*

**Corresponding author: omarkadirov70@gmail.com*

**Abstract.** This review investigates state-of-the-art methods and algorithms for protecting network resources from spam, with a focus on e-mail and network protocol spam in corporate networks. Approaches to filtering unwanted messages are reviewed, including rule-based filtering (e.g., keywords, heuristic rules), machine learning techniques (naive Bayes, decision trees, neural networks), blocking by IP addresses and accounts, and the use of blacklists (DNSBL, URI blocklists, etc.). The advantages and limitations of each approach are described, as well as scenarios of their application in a corporate environment. A comparative table of methods is given, indicating their applicability in corporate mail systems. The article is structured as an academic review: it contains an introduction, thematic sections on filtering methods, machine learning algorithms, blocking technologies and blacklists, a comparative analysis, a conclusion and a list of references. The latest 2020-2024 research publications, reports and standards (IETF, NIST) are used to support the conclusions.

***Keywords:*** *spam, spam filtering, unwanted messages, corporate networks, mail servers, filtering rules, machine learning, naive Bayes, neural networks, IP blocking, DNSBL, blacklists, URI-blocklist.*

## 1. Introduction

Spam is the mass distribution of unsolicited messages, primarily via email. Globally, spam messages account for a significant proportion of all email traffic - as of 2023, about 160 billion spam emails are sent daily, equivalent to about 46% of all emails sent. For corporate networks, this volume of junk mail comes at the cost of lost employee productivity (filling inboxes with garbage, distractions) and increased storage and transmission costs. Moreover, spam has evolved from a simple advertising mailing to a vector of serious cyber threats: phishing messages and malicious attachments make up a significant portion of spam, which can lead to data breaches, hacks and financial losses for companies. Thus, spam traffic carries risks ranging from mere annoying noise to serious information security threats.

Over time, the methods of spam attackers have become much more sophisticated. While earlier spam could be relatively easily filtered by a list of known stop words or simple rules, today spammers actively use techniques to bypass filters: automatic generation of unique texts, obfuscation of content, use of images instead of text, distribution of mailings via botnets, etc. The frequency and sophistication of spam attacks are constantly increasing, while traditional filters often fail to adapt. This emphasizes the need to implement more advanced defenses for corporate email systems.

Spam attacks affect not only e-mail, but also other communication channels. In today's environment, unwanted messages can come via SMS, instant messaging (IM) and even Voice over IP (VoIP) - so-called SPIM (spam over instant messaging) and SPIT (spam over Internet telephony). Nevertheless, e-mail is still the main route for spam to enter corporate networks and it is this route that receives the most attention in the development of anti-spam systems. Fortunately, a substantial arsenal of anti-spam techniques has been accumulated over the past decades. In a corporate environment, a multi-layered defense against junk mail is usually implemented, combining different approaches. This paper provides an overview of these approaches, focusing on the following categories: (1) rule-based filtering and heuristics, (2) machine learning techniques for classifying spam emails, (3) blocking spam at the IP address and account level, and (4) using external blacklists (blocklists) to cut off spam. For each approach, the principles of operation, areas of effective application, and limitations are discussed. It concludes with a comparative characterization of the methods and discusses recommendations for combining them to maximize the effectiveness of spam protection of corporate network resources.

## 2. Materials and methods

### 2.1. Rule-based filtering and heuristics

One of the first and most obvious ways of automatically filtering spam are rule-based methods. Rules are pre-defined patterns that are used to detect unwanted messages. The simplest example is keyword filtering: if the subject or text of an email contains characteristic words/phrases («you win», «buy Viagra», mention of dubious financial schemes, etc.), the email is labeled spam. More advanced rule systems use dozens and hundreds of features: checking email headers for correctness, analyzing formatting (e.g., a large number of capital letters or missing return address), presence of suspicious attachments, encoding mismatches, etc. In practice, many solutions (for example, the famous open-source SpamAssassin filter) use a heuristic approach - a set of rules, each of which adds a certain number of «spaminess» points;

if the total weight exceeds a threshold, the email is recognized as spam.

Rule-based filters are good at detecting primitive or long-known spam, but they have serious drawbacks. First, such filters are poorly adaptable to new tactics of spammers - the rule set is fixed and requires manual updating by experts as new types of spam appear. Spammers can easily bypass rigidly defined signatures by slightly changing the wording or use of symbols. Second, overly strict rules can flag legitimate emails as spam, creating False Positives. For example, a legitimate email with a legitimate promotion may be spam-filtered because of a few «trigger» words. On the other hand, too lenient rules can lead to the omission of unwanted messages (False Negatives). The hardness of rules needs to be carefully balanced, but even when tuned, rule-based filters remain less flexible than trained algorithms.

Nevertheless, the advantages of rules are transparency and relative ease of implementation. An administrator can clearly understand why an email is flagged (e.g., the «word X is in the subject line» rule is triggered) and adjust the policy if necessary. Rules work effectively to cut off spam emails at the SMTP session stage - for example, an email that does not have a correct reverse DNS record of the sender's domain or an incorrect header format can be rejected. Heuristic engines are capable of evaluating an email in real time with minimal computational effort. Such solutions are still widely used and often integrated as the first layer of filtering. However, due to their limited effectiveness against adaptive adversaries, rule-based filtering is now commonly supplemented with more intelligent methods, described below.

### 2.2. Machine learning techniques for spam detection

With the increasing complexity of spam attacks, machine learning (ML) techniques have come to the forefront to automatically train filters based on examples of spam and legitimate emails. One of the classic algorithms is a naive Bayesian classifier, first applied to spam back in the early 2000s. A Bayesian filter calculates the probability that an email is spam based on the statistical frequency of words and features observed in a large corpus of previously tagged emails. This model is able to «learn» from outgoing data: for example, if the word «discount» occurs frequently in spam and the word «report» in normal mail, the filter infers the category of the mail based on the set of words encountered. Bayesian filters have become the basis of many mail clients and servers, as they are quite simple, but more effective than hard rules due to considering probabilistic combinations of attributes.

Besides Bayesian filters, other machine learning algorithms have been successfully applied: support vector method (SVM), decision trees, logistic regression, random forest and others. In the 2010s, with the development of computing power, solutions based on deep learning - neural networks that automatically learn complex patterns of spam text - appeared. For example, a neural network can analyze the sequence of words in an entire email line, identifying non-obvious linguistic signatures of spam, or even recognize the content of images inserted into the body of an email. Modern deep learning models (convolutional networks for texts, recurrent LSTMs, transformers) have demonstrated significantly higher filtering accuracy than traditional methods. They are able to automatically extract hundreds of characteristic features from emails by training on large datasets, and thus detect spam that has not been explicitly described by rules.

The main advantage of ML algorithms is adaptability. The model can be retrained on new examples of spam, allowing it to adjust to the changing tactics of attackers. Moreover, machine learning considers a variety of factors. For example, an email can simultaneously evaluate thousands of different «signals» (words, formats, metadata) and draw a conclusion based on their combination. This is especially important because a single sign (e.g., the presence of a single suspicious word) does not mean that the email is spam. The ML model learns to weight all attributes in context. In addition, algorithms allow personalizing filtering for a particular user or organization - considering which emails a particular recipient marks as spam or not, the model can adjust the trigger threshold. Thus, the number of errors is reduced when a letter is valuable for one user and undesirable for another.

The practical results of using machine learning are impressive. The largest email services claim to block more than 99% of spam with the help of complex ML-systems. For example, Google reported that through the use of its TensorFlow platform (deep learning framework), Gmail blocks an additional ~100 million spam emails per day, reaching a filtering rate of >99.9% of junk mail. This means that out of every thousand spam emails, less than one leaks into the inbox - even though the email service handles a huge amount of traffic. Machine learning has significantly improved the detection of hard-to-detect spam: images, emails with hidden content, emails from new one-day domains, and other «slippery» categories of spam.

However, ML methods also have limitations. Good performance requires a large corpus of labeled data (spam and «ham» - normal mail) to train the model. In an enterprise environment, this may require integration with cloud-based anti-spam services or the use of open data sets, otherwise it is difficult to train a robust model on a limited sample of company emails. In addition, complex neural network models work like a black box, and it can be difficult for an administrator to explain why an email was blocked - reducing transparency compared to simple rules. Finally, spammers also adapt to ML filters by using techniques that confound the algorithms, such as inserting random words from legitimate correspondence to statistically confuse the classifier. Despite this, the combination of machine learning techniques with traditional approaches is currently considered the most effective strategy to protect against spam. ML filters often work in the second line - after coarse rule-based and blocklist-based filtering - and provide fine-tuning to filter out the remaining complex spam cases.

### 2.3. Spam blocking by IP addresses and accounts

In addition to analyzing the content of messages, an important area of anti-spam protection is blocking at the level of network identifiers - IP addresses, sender domains and even specific accounts. The logic is simple: if a particular network node is known to be sending spam, you can prevent any connections from that node to the organization's mail server. Many corporate mail gateways support IP address blacklists - an administrator can manually or automatically add addresses that have been repeatedly used for spam attacks. For example, if a company does business only in one region, it is possible to block all incoming SMTP sessions from network ranges belonging to other geographical zones from which legitimate mail is not expected. It is also practiced to temporarily block

IPs when mass mailings are detected (rate limiting): an address that tries to send too many mails in a short period of time can be disconnected for some time.

An even simpler option is to block by the «From:» field (sender address) or by the name/address of a specific sender. Mail users often add unwanted senders to a local blacklist on their own. However, the effectiveness of such a measure is very low, as spammers usually forge the sender's address or change it constantly. Blocking on the basis of an exact address match is a «very weak heuristic» that has virtually no effect on the overall spam flow. One-day addresses bypass this defense: an attacker can generate a new fake address for each mailing. The situation is similar with blocking by username (account): spammers rarely use the same account for longer than a short time, preferring to search through many compromised or fake accounts.

Despite these limitations, IP/domain blocking remains an important element of layered defense, especially at the initial stage of incoming traffic. Its main advantage is minimal resource consumption and instantaneous action. If the source is known to be malicious, it is better to reject the connection completely before the email is received: this saves bandwidth and computational resources for content filters. Enterprise systems often use automated solutions: for example, a firewall or mail gateway can integrate with global IP reputation databases (more on them later) or have a self-blocking algorithm. The latter works as an internal spam detector for outgoing emails: if an infected computer or employee account suddenly starts sending spam from the organization, the system blocks further sending from that account/IP to prevent the company's email domain from being blacklisted on the Internet.

Thus, blocking by IP/account is effective against mass spam from fixed sources and as an emergency measure (disable the source when an attack is detected). But individually, this method does not provide sufficient protection because of the described circumvention capability. In modern conditions, spam mailings are distributed across thousands of botnet nodes that constantly change IP addresses. Temporary cloud servers or compromised email accounts on legitimate services are often used. In this regard, in corporate practice, manual maintenance of their blocking lists has given way to the use of global DNS blocklists, which will be discussed in the next section.

### 2.4. Use of blacklists (DNSBL, URI blocklists)

Blacklists (aka blocklists) are databases of known spam sources maintained by specialized organizations or communities. The most common are so-called DNSBLs (DNS-based Blackhole Lists) - lists of IP addresses and domains known to send spam, accessible through special DNS queries. The idea was proposed back in the late 1990s: a mail server receiving an e-mail can make a DNS query to a public blocklist server using the sender's IP. If the IP is found in the database as spam, the server immediately rejects or marks the email as spam. DNSBLs have proven to be an effective measure against mass spam and are used by most modern mail filters. It is a largely proactive approach: instead of analyzing the content of the email, we trust the sender's reputation tag, derived from global observations.

There are various authoritative DNSBLs. For example, the Spamhaus project maintains several lists: SBL (Spamhaus Block List) - known spammers, XBL - addresses of botnet-infected computers sending spam, PBL - ranges of dynamic addresses not intended for mail servers (home/DSL pool addresses). Other examples are blacklists from SpamCop, Barracuda, Proofpoint, etc. Most of these are open for free (with limits on the number of requests), making DNSBL a low-cost and effective way to significantly reduce the flow of incoming spam. As Spamhaus points out, using blocklists early in the email acceptance process allows the lion's share of spam to be weeded out before the resource-intensive content is analyzed, reducing the load on the infrastructure and improving filtering performance.

The DNSBL mechanism is standardized and documented in industry guidelines (for example, IETF RFC 5782 describes the structure and use of DNS blocklists and whitelists). In essence, a blocklist provides a DNS-implemented reputation database. Blacklisted addresses are typically discovered by aggregating complaints, analyzing spam trap logs, and other methods. For example, many blocklist providers maintain their own networks of «honeypots» - unused email addresses published in the public domain as traps for spammers. Any email that arrives at such an address is obviously spam, and information about its sender is automatically added to the database. Spam messages from users (feedback loop), results of automatic scanning of malicious campaigns, etc. are also considered. For example, every day the Spamhaus system receives data on analyzing ~9 billion SMTP connections and ~3 million new domains being evaluated for undesirability. Based on this Big Data, a mix of automated (including ML) and expert methods are used to generate list updates sent out in near real-time.

In addition to IP-based DNSBLs, there are URL/domain blocklists, also known as URI Realtime Blacklists (URIBL) or SURBLs. These are aimed at detecting spam via malicious links within emails. The idea is that even if a spammer sends emails from constantly changing addresses, they often advertise the same fraudulent site or link. URIBL aggregates domains that have been seen inside spam emails and allows the filter to mark the email as spam if a blacklisted URL is found in the body of the email. Unlike DNSBL, it analyzes the content of the message (URI) rather than the connection parameters. Such lists (e.g., Project SURBL, URIBL.com) complement traditional DNSBLs and improve the detection of phishing and advertising emails in terms of email content.

There are significant advantages to using external blacklists. First, a corporate email system does not need to collect global data on spammers on its own - it is enough to turn to an authoritative list that accumulates information from all over the world. Secondly, speed and simplicity: IP verification via DNS takes milliseconds and can be performed at the SMTP protocol stage (before receiving the message body), which saves traffic and processing time. Third, most DNSBLs are free or conditionally free, which is beneficial even for small organizations.

However, there are limitations. No blocklist provides 100% coverage: new spammers or freshly infected botnet nodes will not be in the database for some time, so some spam will pass the filter. It is especially difficult for blocklists to resist hailstorm or snowshoe tactics, when mailings come from thousands of scattered IPs with low intensity - such sources are more difficult to list quickly. On the other hand, there are false positives: a legitimate mail server may have been temporarily blocklisted (e.g., due to hacking and spamming in the past) and its emails will be rejected. Administrators are forced to

monitor the reputation of their servers and take delisting measures if they get on the DNSBL. Potential attack vectors against the blocklists themselves are also described: attackers may try to intentionally blacklist the addresses of bona fide senders by flooding the traps with emails from the victim's fake address. In 2023, researchers demonstrated the so-called HADES attack scenario, when by exploiting a vulnerability in the trap infrastructure, they managed to add not malicious but ordinary mail servers to popular DNSBLs, causing them to be blocked en masse. This demonstrates the need for careful and responsible use of blocklists: many ISPs have data verification policies and whitelisting mechanisms for critical services to mitigate this risk.

For enterprise mail systems, connecting to multiple robust DNSBLs and URIBLs is the de facto standard. In combination with local rules and trainable filters, blacklists provide the first line of defense, cutting off a significant portion of outright spam on the fly. Practice shows that competent use of DNSBL can block 70% to 90% of unwanted emails before they are analyzed by the content, dramatically reducing the overall load on the anti-spam system and improving the quality of filtering.

## 3. Results and discussion

### 3.1. Comparative analysis of methods and their applicability

Each of the approaches considered has its own area of effectiveness. Below are summarized the main characteristics of spam protection methods in the corporate environment (Table 1), indicating their advantages, disadvantages and typical application:

***Table 1. Comparison of Spam Defense Methods in Enterprise Networks***

| Approach | Advantages | Limitations | Application in corporate environment |
|---|---|---|---|
| Rule-based and heuristic filtering | Transparent decisions; low computational overhead; easy to configure for known spam patterns. | Labor-intensive rule maintenance; poor adaptation to new spam types; risk of false positives. | Basic filtering level: initial filtering of primitive spam. Effective combined with other methods. |
| Machine learning methods | High detection accuracy; adaptive to evolving spam; identifies complex patterns; personalization possible. | Requires large training datasets; computationally intensive; "black box" nature of decisions; vulnerability to adversarial spam. | Main filtering layer in corporate email servers or cloud email services (e.g., Gmail, Outlook 365). |
| IP/User-based blocking | Instant blocking of known sources; reduces system load; simple implementation; emergency measure against compromised accounts. | Spammers dynamically change IP/user; constant updates required; risk of blocking legitimate senders. | Initial firewall-level protection to block known spam IPs/domains, temporary measure during attacks. |
| Blacklists (DNSBL, URIBL) | Global coverage; low query cost; regular updates; efficiently filters mass spam before content analysis. | Incomplete coverage; risk of false positives; dependency on external services. | Standard corporate email gateway component; used during SMTP connection and content analysis. |

As the table shows, there is no perfect one-size-fits-all method. Rules provide a foundation of transparency and speed, but fail to keep up with volatile spam. Machine learning gives high coverage and flexibility, but requires resources and data. IP-based blocking is effective against known attackers, but powerless against distributed botnets. External blocklists are great at filtering mass spam but can miss targeted attacks or create operational hassles unblocking erroneously entered addresses.

For corporate networks, the most productive is the combined use of all these approaches. For example, a practical anti-spam system can function as follows: when an incoming connection is attempted, the sender's IP address is immediately checked by DNSBL (if blacklisted - reject). If the IP is clean, the e-mail is accepted and the content is checked by the rules (elementary signs, links to prohibited domains) - if gross criteria are met, the e-mail is quarantined. The remaining messages are analyzed by the ML-module, which considers more subtle features and makes a final decision on spam classification. Additionally, if outbound spamming is detected within an organization (e.g., an infected computer sends many emails), the account-based blocking mechanism suspends the sending and notifies administrators. This layered approach achieves the best results: each method overrides the shortcomings of the other. According to current reports, combining multiple technologies can filter out 99-99.9% of unwanted mail while minimizing false alarms.

## 4. Conclusions

The growth in the volume and sophistication of spam in recent years has presented corporate users with serious cybersecurity challenges. This paper provides an overview of the main methods for protecting network resources from spam - from traditional rules and heuristics to the latest machine learning algorithms and global reputation services. The comparative analysis shows that only a combination of these approaches can provide sustainable and effective protection. A multi-layered strategy, including filtering at the network stage (IP and DNSBL blocking), content analysis by rules and statistical models, as well as intelligent classification by ML algorithms, seems to be the most effective for corporate email systems.

Current trends point to further integration of methods: for example, there are solutions where deep learning models take over some of the tasks previously solved by rules, or systems that automatically generate new rules based on spam detected by the ML model. In addition, collaboration between organizations to share data on spam threats is developing, which improves the relevance of blacklists and training samples. Standards and guidelines (IETF, NIST, etc.) recommend a comprehensive approach that includes both technical filtering measures and administrative measures (phishing recognition training for employees, acceptable mail usage policies).

We can conclude that the spam problem is evolving, but at the same time the means of counteraction are improving. Advanced machine learning algorithms can now filter out the vast majority of unwanted messages. Nevertheless, spam continues to be an «arms race» that requires constant attention. Enterprise networks should implement converged solutions that combine the best available techniques and

regularly update them to keep pace with emerging threats. Only with multiple layers of protection and flexible adaptation to new types of spam can reliable filtering be ensured and communications within an organization remain secure and productive.

### References


[1] EmailToolTester. (2024). Spam Statistics 2025: Survey on Junk Email, AI Scams & Phishing

[2] Roumeliotis, K.I., et al. (2024). Next-Generation Spam Filtering: Comparative Fine-Tuning of LLMs, NLPs, and CNN Models for Email Spam Classification. *Electronics, 13(11), 2034*

[3] Kumaran, N. (2019). Spam does not bring us joy – ridding Gmail of 100 million more spam messages with TensorFlow. *Google Cloud Blog*

[4] Stith, M. (2020). DNSBLs for email filtering – an introduction. *APNIC Blog*

[5] Spamhaus. (2022). DNSBLs and Email Filtering – How to Get It Right

[6] Security StackExchange. (2015). Is it appropriate to block a sender's email address? – Answer by user Zaph

[7] Li, R., et al. (2024). HADES Attack: Understanding and Evaluating Manipulation Risks of Email Blocklists. *NDSS Symposium*

[8] Levine, J. (2010). DNS Blacklists and Whitelists. *IRTF RFC 5782*

[9] MDaemon Documentation. (n.d.). URI Blocklists (URIBL) – SecurityGateway 10.5


# Желілік ресурстарды спамнан қорғау әдістері мен алгоритмдерін зерттеу


О. Кадыров*, А. Батыргалиев

*Satbayev University, Алматы, Қазақстан*

**Корреспонденция үшін автор: omarkadirov70@gmail.com*



**Аңдатпа.** Шолу электрондық поштаға және корпоративтік желілердегі желілік протоколдардағы спамға баса назар аудара отырып, желілік ресурстарды спамнан қорғаудың заманауи әдістері мен алгоритмдерін зерттейді. Қажетсіз хабарламаларды сүзу тәсілдері қарастырылады, соның ішінде ережелерге негізделген сүзу (мысалы, кілт сөздер, эвристикалық ережелер), Машиналық оқыту әдістері (аңғал Байес, шешім ағаштары, нейрондық желілер), IP мекенжайлары мен есептік жазбалар арқылы бұғаттау және қара тізімдерді пайдалану (DNSBL, URI блок-тізімдері және т.б.). Әр тәсілдің артықшылықтары мен шектеулері, сондай-ақ оларды корпоративті ортада қолдану сценарийлері сипатталған. Корпоративтік пошта жүйелерінде олардың қолданылуын көрсететін әдістердің Салыстырмалы кестесі келтірілген. Мақала академиялық шолу ретінде құрылымдалған: кіріспе, сүзу әдістері, Машиналық оқыту алгоритмдері, құлыптау технологиялары және қара тізімдер, салыстырмалы талдау, қорытынды және әдебиеттер тізімі бойынша тақырыптық бөлімдер бар. Қорытындылардың негіздемесі ретінде 2020-2024 жылдардағы ең жаңа ғылыми басылымдар, есептер мен стандарттар (IETF, NIST) пайдаланылды.



***Негізгі сөздер:*** *спам, спамды сүзу, қажетсіз хабарламалар, кәсіпорын желілері, пошта серверлері, сүзу ережелері, Машиналық оқыту, аңғал Байес, нейрондық желілер, IP блоктау, DNSBL, қара тізімдер, URI блок тізімі.*


# Исследование методов и алгоритмов защиты сетевых ресурсов от спама


О. Кадыров*, А. Батыргалиев

*Satbayev University, Алматы, Казахстан*

**Автор для корреспонденции: omarkadirov70@gmail.com*



**Аннотация.** В обзоре исследуются современные методы и алгоритмы защиты сетевых ресурсов от спама, с акцентом на электронную почту и спам в сетевых протоколах в корпоративных сетях. Рассматриваются подходы к фильтрации нежелательных сообщений, включая фильтрацию на основе правил (например, ключевые слова, эвристические правила), методы машинного обучения (наивный Байес, деревья решений, нейронные сети), блокировку по IP-адресам и учетным записям, а также использование черных списков (DNSBL, URI-блоклисты и др.). Описаны преимущества и ограничения каждого подхода, а также сценарии их применения в корпоративной среде. Приведена сравнительная таблица методов с указанием их применимости в корпоративных почтовых системах. Статья структурирована как академический обзор: содержит введение, тематические разделы по методам фильтрации, алгоритмам машинного обучения, технологиям блокировки и черным спискам, сравнительный анализ, заключение и список литературы. В качестве обоснования выводов использованы новейшие научные публикации 2020–2024 гг., отчеты и стандарты (IETF, NIST).

***Ключевые слова:*** *спам, фильтрация спама, нежелательные сообщения, корпоративные сети, почтовые серверы, правила фильтрации, машинное обучение, наивный Байес, нейронные сети, блокировка IP, DNSBL, черные списки, URI-блоклист*.

Received: 09 March 2025
Accepted: 15 June 2025
Available online: 30 June 2025