ISBN 2354-0575

# NGHIÊN CỨU, XÂY DỰNG HỆ THỐNG GIẢM THIỂU THƯ RÁC CHO HỆ THỐNG THƯ ĐIỆN TỬ TRONG QUY MÔ TRƯỜNG ĐẠI HỌC SỬ DỤNG NGUỒN MỞ

**Nguyễn Thị Thanh Tú[1], Vũ Đình Minh[1], Nguyễn Thanh Tùng[1], Nguyễn Minh Quý[2]**

*1 Trường Đại học Bách Khoa Hà Nội*

*2 Trường Đại học Sư phạm Kỹ thuật Hưng Yên*

Ngày nhận: 01/10/2016
Ngày sửa chữa: 31/10/2016
Ngày xét duyệt: 15/11/2016

**Tóm tắt:**

*Hiện nay, trường ĐHBK Hà Nội triển khai hệ thống thư điện tử phục vụ cho khoảng 2000 cán bộ. Tuy nhiên, qua công tác theo dõi, giám sát hoạt động của hệ thống thư điện tử, chúng tôi phát hiện hệ thống thư điện tử nhận được số lượng thư ráclớn mỗi ngày. Những thư rác này ngoài những thư được sử dụng với mục đích quảng cáo, thương mại thì thư rác còn được sử dụng để tuyên truyền chống phá; là những thư rác đính kèm đường link nhiễm mã độc, virus độc hại…. Để có biện pháp phát hiện và ngăn chặn kịp thời các thư rác, đặc biệt là các thư rác có nội dung xấu độc, vi phạm pháp luật và đảm bảo hệ thống luôn ổn định giúp các thầy/cô làm việc hiệu quả trong đào tạo, giảng dạy, nghiên cứu, trao đổi hợp tác. Việc đề xuất triển khai hệ thống giảm thiểu thư rác cho hệ thống thư điện tử của trường Đại học Bách Khoa Hà Nội là rất cần thiết và cấp bách.*

***Từ khóa:*** *Spam, ASSP, Linux, BKNET, SMTP.*

## Đặt vấn đề

Trong vòng vài năm gần đây, việc sử dụng Internet nói chung và thư điện tử nói riêng ngày càng phổ biến tại Việt Nam. Một trong những hệ quả của sự phát triển này là ngày càng có nhiều thư rác gửi tới các tài khoản thư điện tử tại Việt Nam (tài khoản có đuôi .vn). Tình hình thư rác đang rất phức tạp. Công ty Điện toán và Truyền số liệu (VDC) - ISP lớn nhất Việt Nam - cho biết, thư rác hiện nay chiếm phần lớn lưu lượng email qua hệ thống máy chủ thư của ISP này.

Các thư phàn nàn gửi đến ISP nếu không giải quyết, các khách hàng của ISP đó có thể bị liệt vào danh sách đen, không gửi được email ra địa chỉ nước ngoài. Một số ISP cho biết theo đánh giá hàng năm, khách hàng của nhiều ISP ở Việt Nam thường xuyên bị tê liệt do bị liệt vào danh sách đen. Mỗi lần thoát ra khỏi danh sách này ISP phải mất một khoản phí không đáng mất.

Tại trang web Spamhaus.org (Tổ chức theo dõi các nguồn gửi thư rác), có lần vnn.vn đã có trong danh sách top 10 ISP cung cấp nhiều rác nhất. Không chỉ gây thiệt hại về tiền bạc, thư rác còn làm giảm hiệu quả làm việc, gây stress, tiêu tốn thời gian của người lao động... Những điều này cũng đồng nghĩa với việc, năng suất lao động giảm, ảnh hưởng tới tình hình hoạt động của các cơ quan, tổ chức, công ty...

Theo kết quả điều tra ban đầu của VNCERT - Bộ Bưu chính Viễn thông hơn 1/3 số người được hỏi khẳng định mỗi ngày nhận được từ 20 - 50% số thư rác mang nội dung tiếng Việt trên tổng số thư rác phải nhận mỗi ngày, khoảng 40% khẳng định trong số thư rác từng nhận có chứa các nội dung xuyên tạc, vi phạm thuần phong mỹ tục, gây hại, lừa đảo, 48,36% người dùng không có ấn tượng gì đặc biệt và 33,45% cho rằng các công ty được quảng cáo thông qua thư rác là không có uy tín thương hiệu và tiềm lực hạn chế. Chính vì sự phát triển mạnh mẽ của thư rác tại Việt Nam nên

nghị định chống thư rác chính phủ đã ra đời và bắt đầu có hiệu lực từ ngày 09-02-2009. 90% người dùng hiện nay cảm thấy khó chịu, bức xúc với thư rác, nhưng họ không biết làm gì hơn là cần mẫn xóa đi từng cái. Các cuộc khảo sát với người dùng cũng cho thấy, 3 yêu cầu đầu tiên của người dùng để đối phó với nạn này là: cần có quy định quản lý thư rác, cần có đơn vị chuyên trách và cần nâng cao nhận thức cộng đồng.

**Khó khăn gặp phải**: Cuộc chiến chống spam đầy khó khăn, điều khó khăn nhất với cuộc chiến chống thư rác qua email là tính không biên giới của Internet.Và spam muôn hình vạn trạng, địa chỉ phát tán spam cũng thiên biến vạn hóa. Việc sử dụng bộ lọc từ khóa hay chặn địa chỉ email đơn giản thường ít hiệu quả vì không theo kịp tốc độ "biến hoá" của spam. Trong cuộc chiến này cần đến sức mạnh của cả cộng đồng.Trong quá trình lọc thư rác tiếng Việt, vấn đề khó khăn nhất là phải xử lý được việc tách từ. Mặc dù, tiếng Việt gồm các ký tự La tinh nhưng tiếng Việt có những đặc trưng riêng [1, 2, 3, 4, 5].

Mặt khác, dịch vụ email (thư điện tử) là một trong những dịch vụ internet phổ dụng cho nhu cầu trao đổi thông tin hàng ngày. Tuy nhiên, thư rác (Spam mail) và Virus là một trong những phiền toái và mối lo ngại của người sử dụng dịch vụ này. Theo thống kê của một số hãng nổi tiếng trong lĩnh vực Antivirus thì mỗi ngày trên toàn thế giới có khoảng 14,5 tỷ thư rác được phát tán (chiếm đến 45% số lượng email trên toàn cầu mỗi ngày).

Trong quý I của năm 2016, theo báo cáo của Kaspersky Lab [6] số lượng thư rác đính kèm mã độc đã tăng đáng kể. Trong hai năm qua, số lượng virus trên thư rác được phát hiện trên các máy tính sử dụng sản phẩm của Kaspersky Lab dao động từ 3 đến 6 triệu. Vào cuối năm 2015 con số này bắt đầu phát triển và đầu năm 2016 đã có xu hướng tăng mạnh. Trong tháng ba, số lượng thư rác chứa virus bị phát hiện đạt 22.890.956, nhiều hơn bốn lần so với mức trung bình của cùng kỳ năm ngoái.

Hiện nay trên thế giới, nhiều mô hình đã được sử dụng để giám sát, quản lý thư điện tử bao gồm các sản phẩm mã nguồn mở như Scrollout F1, Anti-Spam SMTP Proxy (ASSP), SpamAssassin, MailScanner, SpamBayes… Nhóm nghiên cứu đã tập trung vào tìm hiểu, nghiên cứu về một số phương pháp nổi trội trên thế giới, trong đó có phương pháp Anti-Spam SMTP Proxy (ASSP) để tham khảo cho việc xây dựng, triển khai hệ thống phát hiện và xử lý thư rác cho MAIL HUST. Vì ASSP bao gồm việc triển khai các biện pháp phòng chống thư rác phổ biến như danh sách trắng, graylisting, SPF, danh sách đen DNS, và tích hợp với ClamAV và FileScan. ASSP cũng thêm các bộ lọc chính yếu thường xuyên, damping, các từ có nguồn gốc trong bộ lọc phân tích của Bayesian và hỗ trợ cho SenderBase, transparent proxying, các plug-in để khai thác OCR của các file đính kèm cho việc lọc. Việc tích hợp tất cả các tính năng này vào một máy chủ proxy SMTP làm cho công việc của máy chủ mail dễ dàng quản trị hơn nhiều so với việc cố gắng để kết hợp một loạt các công cụ cá nhân với nhau trên một máy chủ SMTP tiêu chuẩn [7, 8, 9. 10].

Cho đến nay, trường ĐHBK Hà Nội triển khai hệ thống thư điện tử phục vụkhoảng 2000 cán bộ. Tuy nhiên, qua công tác theo dõi, giám sát hoạt động của hệ thống thư điện tử, chúng tôi phát hiện hệ thống thư điện tử nhận được số lượng thư rác rất lớn mỗi ngày (có thể lên tới hàng trăm nghìn thư rác mỗi ngày). Những thư rác này ngoài những thư được sử dụng với mục đích quảng cáo, thương mại thì thư rác còn được sử dụng để tuyên truyền phản động, chống phá Đảng, Nhà nước, làm ảnh hưởng đến uy tín chính trị của nhiều Lãnh đạo Đảng, Nhà nước.

Để có biện pháp phát hiện và ngăn chặn kịp thời các thư rác, đặc biệt là các thư rác có nội dung xấu độc, vi phạm pháp luật và đảm bảo hệ thống luôn ổn định giúp các thầy/cô làm việc hiệu quả trong đào tạo, giảng dạy, nghiên cứu, trao đổi hợp tác, việc nghiên cứu mô hình, phương pháp và xây dựng hệ thống xử lý thư rác cho hệ thống thư điện

tử của trường Đại học Bách Khoa Hà Nội là rất cần thiết và cấp bách.

**Đề xuất và xây dựng giải pháp cho hệ thống thư điện tử BKNET**

Anti-Spam SMTP Proxy (ASSP) là một Open Source phần mềm trong danh mục truyền thông được phát triển bởi John Hanna. Là dự án máy chủ nhằm mục đích tạo ra một máy chủ mã nguồn mở nền tảng độc lập SMTP Proxy thực hiện tự động một danh sách cho phép, tự học Bayesian, Greylisting, DNSBL, DNSWL, URIBL, SPF, SRS, tán xạ, quét Virus, chặn file đính kèm chứa mã độc, SenderBase và nhiều phương pháp lọc [11, 12].

Anti-Spam SMTP Proxy "ASSP" là một bộ lọc thư rác sử dụng cổng 25 (SMTP), có tùy chọn sử dụng cổng 465 (smtps) và 587 (submission), thường ở phía trước của máy chủ SMTP. ASSP chuyển tiếp hộp thoại SMTP giữa một người gửi đến và máy chủ SMTP, ngăn chặn các hộp thoại khi cần thiết. ASSP thực hiện một số kiểm tra thư rác và cấu hình cung cấp một mã lỗi 5xx trên tin nhắn thư rác SMTP ngay lập tức lại cho người gửi. Thư rác có thể bị chặn khi gửi đến hay đối tượng được gắn thẻ. ASSP cung cấp một danh sách trắng của tên người gửi cho Bayesian kiểm tra trên tiêu đề và nội dung tin nhắn xác nhận người dùng sử dụng RFC822 danh sách hay tra cứu LDAP tiếp sức từ HELO kiểm tra (khung Sender Policy) SPF kiểm tra DNSBL (DNS Block List) kiểm tra sử dụng nhiều dịch vụ danh sách block trì hoãn của các tin nhắn để phát hiện người gửi Virus.

ASSP được quản trị quản lý và gần như hoàn toàn trong suốt đối với người sử dụng. Đặc biệt, người dùng không cần phải quản lý bộ lọc thư rác hoặc các hệ thống thử thách phản ứng của riêng họ. ASSP được cấu hình sử dụng một giao diện web. Đối số admin_port cung cấp cho các cổng mạng để truy cập vào menu cấu hình, mặc định là 55555.


Hình 1. *Mô hình đề xuất*

**Bước 1**: Cài đặt hệ thống ASSP trên Windows server 2012 R2. Cài đặt môi trường gồm những phần mềm sau:

- Visual C++ 2008 Redistributables: là một môi trường phát triển tichsh ợp (IDE) được sử dụng để tạo ra các ứng dụng Windows trong C, C++, và các ngôn ngữ lập trình C++/CLI. Nó cung cấp cho các nhà phát triển một ứng dụng duy nhất mà họ có thể viết, chỉnh sửa, kiểm tra và gỡ lỗi mã cho sản phẩm của mình. Môi trường lập trình bao gồm quyền truy cập vào rất nhiều thư viện mã được chia sẻ, cho phép các nhà phát triển sử dụng mã đã được phát triển cho các thủ tục cụ thể thay vì phải viết lại từ đầu.

- Active Perl:công cụ đọc các file Perl đuôi *.pl vì các exploit thường được viết bằng Perl. Nó còn được sử dụng để thi hành các lệnh thông qua các file *.pl.

- Win32 OpenSSL: OpenSSL không chỉ là SSL. Nó có khả năng về các thuật toán băm (message digest), mã hóa và giải mã các tệp, các chứng nhận số, các chữ ký số và các số ngẫu nhiên OpenSSL không chỉ là API, nó cũng có một công cụ sử dụng dòng lệnh. Công cụ dòng lệnh này có

**ISBN 2354-0575**

thể làm những thứ tương tự như API, nó có khả năng kiểm tra các máy chủ và các máy khách SSL.

**Bước 2**: Kết nối hệ thống ASSP và Máy chủ Email:

- hình trên giao diện website theo địa chỉ IP của máy chủ ASSP và máy chủ Email và các port SMTP, SSL theo SMPT Listten Port, Destination, Relay Host, Relay Port.

- Tạo 2 file domain.txt, relayips.txt; Domains.txt chứa tất cả các tên miền được chấp nhận trên máy chủ email và Relayips.txt chứa tất cả các IP cho phép chuyển tiếp qua hộp thư.

- Thay đổi NAT trên tường lửa để gửi port TCP/25 (SMTP) cho ASSP thay cho máy chủ mail.

- Nhấp vào SMTP Connections trên thành menu trái, sẽ có một tab riêng biệt hiển thị thông tin các kết nối SMTP trực tiếp trong thời gian thực.

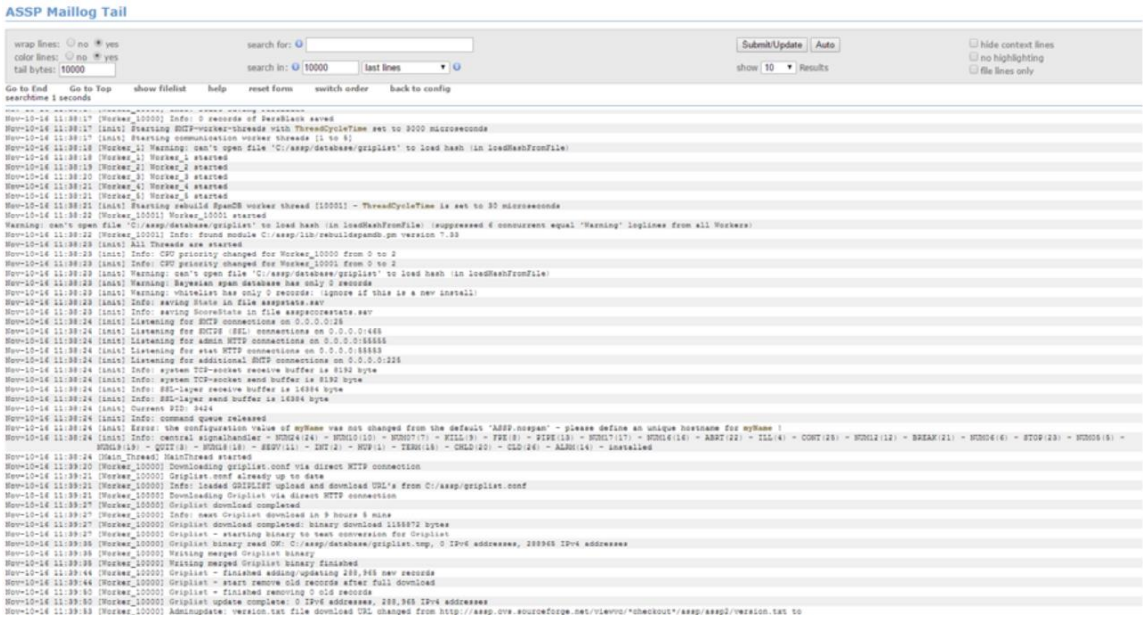
**Bước 3**: Tạo các luồng xử lý thư rác:

Cú pháp: rule number <=>rule_type <=>recipient <=>replace_with <=>when_sender <=>next_rule_if <=>jump_target'.

Kết quả đạt được

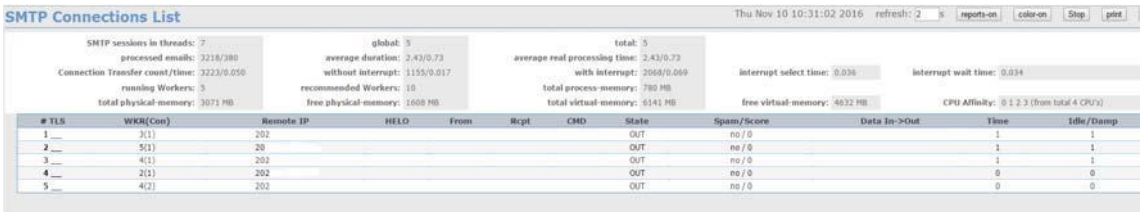
Những kết quả đạt được trong quá trình thực nghiệm từ hệ thống máy chủ của trường đại học Bách khoa là rất khả quan. Hệ thống log đã có thể ghi lại những sự kiện xảy ra trên server một cách hiệu quả như ngày giờ đăng nhập vào server, các tiến trình xử lý thư điện tử xảy ra trên server, và đưa ra các thông báo cảnh báo về hệ thống rất chi tiết, rõ ràng và cụ thể.

Qua quá trình thực hiện thu thập log của máy chủ ASSP đã thu thập được các kết quả tốt. Dựa vào hệ thống phân tích đã cho thấy được những loại thư rác, thư chứa mã độc gửi vào thư điện tử của BKNET như là các thư quảng cáo, mua bán bất động sản cho đến các thư gửi virus mã độc CTBLocker là một dạng mã độc tống tiền. Thư đượcgửi dưới dạng một tập tin ở dạng đính kèm, khi mở tập tin này, máy tính của người dùng sẽ bị kiểm soát, các tệp dữ liệu dạng Word, Excel… sẽ bị mã hóa và không thể mở ra được. Hacker sẽ đòi một khoản tiền lớn từ người dùng bị nhiễm để gửi khóa g iải mã




Hình 2. *Log cho thấy các hoạt động xử lý thư, thay đổi cấu hình, quản trị người dùng của ASSP*

**ISBN 2354-0575**




Hình 3. *Thông tin các kết nối SMTP trực tiếp trong thời gian thực*

Ngoài ra hệ thống có chức năng hiển thị các kết nối SMTP trực tiếp trong thời gian thực cùng các thông tin địa chỉ IP người gửi,địa chỉ thư người gửi, địa chỉ người nhận; có phải spam không, số điểm chấm spam… như trong hình 3 ở trên.

Từ những dữ liệuthu thập được đã giảm thiểu tối đa được các vấn đề về an toàn thông tin, tài liệu của người dùng tránh mất mát dữ liệu, lộ tài khoản khi bị dính mã độc từ một thư điện tử chứa mã độc hay hacker dùng chính những thư đã chiếm được quyền để gửi mã độc đi các thư điện tử khác trong và ngoài trường. Giúp hệ thống được an toàn và vận hành ổn định hơn.

**Kết luận**

Nhóm đã nghiên cứu, triển khai thành công hệ thống xử lý ngăn chặn nhằm giảm thiểu thư rác ASSP trên cơ sở xây dựng và tích hợp một số hệ thống mã nguồn mở, tại Trường Đại học Bách Khoa Hà Nội. Hệ thống đã và đang được sử dụnggiúp giảm thiểu số lượng thư rác, thư chứa mã độc giúp cho các cán bộ yên tâm sử dụng hệ thống thư điện tử, cũng như tránh được những bức xúc khi phải nhận nhiều thư rác, cũng như việc có một hệ thống xử lý thư phía trước giúp chia tải việc phải xử lý và gửi thư cho hệ thống email server giúp hệ thống chạy ổn định hơn.Từ đó, kết hợp với các phương pháp lọc thống kê và các kỹ thuật phát hiện ra các máy bị nhiễm mã độc gửi thư trong trường làm ảnh hưởng đến hoạt động của hệ thống thư điện tử kịp thời xử lý và ngăn chặn, đảm bảo an toàn và thông tin thông suốt trên toàn hệ thống.

**Tài liệu tham khảo**

[1]. Dinh Dien, “Tu Tieng Viet”, Proceeding of ICMLC2002 Conference, Beijing, November 2002.

[2]. Dinh Dien, Hoang Kiem, Nguyen Van Toan, “Vietnamese Word Semntation”, The sixth Natural Language Processing Pacific Rim Sympossium, Tokyo, Japan 2001

[3]. Foo S., Li H, “Word Segmentation and Its Effect on Information Retrieval”, Information Processing & Management: Anh International Journal, 2004

[4]. H. Nguyen, T. Vu, N. Tran, K. Hoang, “Internet and Genertics Algorithm-base text Categorization for Documents in Vietnamese”, Research, Innovation and Vision of the Future, the 3rd International Conference in Computer Science, (RIVF 2005), Can Tho, Viet Nam 2005

[5]. Le An Ha, “A method for word segmnetation in Vietnamese”, Proceedings of Corpus Linguistics, Lancaster, UK, 2003.

[6]. Kaspersky Lab.

[7]. Amit Sharma, Bayesian Mail Filter for detecting spam, 2008

[8]. Graham, P., A plan for Spam, 2008

[9]. OReilly.SpamAssassin.Jul.2004.eBook-DDU. Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

**ISBN 2354-0575**

[10]. Delany SJ, P Cunningham & B Smyth (2006) ECUE: A Spam Filter that Uses Machine Learning to track Concept Drift, in: Proc of the 17th Eur. Conf. on Artificial Intelligence (PAIS stream), p627-631.

[11]. http://www.openspf.org

[12]. http://www.linuxlinks.com/article/20130407093613870/Anti-SpamTools.html.

## RESEARCHING AND BUILDING AN OPEN SOURCE ANTI-SPAM (SPAM-AVOIDANCE) FRAMEWORK FOR UNIVERSITY'S EMAIL SYSTEM

**Abstract:**

*In recent years, the email system of Hanoi Univeristy of Science and Technology (HUST) has become more and more important in exchanging information and research activities. However, based on the monitoring and managing system of email service, we found that the email system received a huge of spam per day.Beside using these spam (messages) for advertising and trade purpose, they are also used for propaganda purpose unsuitable and illegal content. Furthermore, they are often attached malicious link, malicious virus ...So, detecting and preventing spam are necessary to ensure the stability of email system and assist the teachers working efficiently in training, teaching, research, exchanging information. Therefore, we proposed a research proposal and built anti-spam solutions for email system of HUST.*

***Keywords:*** *Spam, ASSP, Linux, BKNET, SMTP.*