云计算平台下基于改进型 DMTP 的反垃圾邮件系统设计*

刘海韬, 阳 洁†
(中南大学 信息科学与工程学院, 长沙 410083)

摘 要: 针对 DMTP(differentiated mail transfer protocol,差分邮件传输协议)存在的类别划分模糊化问题,提出了改进型 DMTP。将改进型 DMTP 与云计算结合,设计了云计算平台下基于改进型 DMTP 的反垃圾邮件系统总体模型,并详细描述了该模型中邮件传送的工作流程。通过引入对发送方类别进行判断的流程和综合云端反垃圾邮件集群系统的分析结果,该模型可以有效遏制垃圾邮件传播。最后,引用了基于 Eucalyptus 的反垃圾邮件系统构建模型。

关键词: 垃圾邮件; 差分邮件传输协议; 云计算; 简单邮件转输协议; 桉树云服务环境(Eucalyptus)
中图分类号: TP393 **文献标志码**: A **文章编号**: 1001-3695(2012)10-3809-03
doi:10.3969/j.issn.1001-3695.2012.10.053

Anti-spam system design based on improved DMTP under cloud computing platform

LIU Hai-tao, YANG Jie†
(*School of Information Science & Engineering, Central South University, Changsha* 410083, *China*)

Abstract: This paper proposed an improved DMTP (differentiated mail transfer protocol) according to the problem that the SMTA intangible categorized in DMTP, and combined this improved DMTP with cloud computing. Then it designed an overall architecture of anti-spam system and described the work flow of mail delivery in the system in detail. By differentiating sender's category and integrating results from the cloud, the proposed system can effectively restrain the spam spread. Finally, it referenced an Eucalyptus-based anti-spam system.

Key words: spam; DMTP; cloud computing; SMTP; Eucalyptus

0 引言

目前,以电子邮件为媒介的信息交流方式越来越普遍,但随之而来的垃圾邮件问题却严重干扰了正常的信息交流。据统计,邮件服务提供商每年用于处理垃圾邮件的花费超过 200 亿欧元。SMTP[1]中接收方被动接收邮件的缺点是导致垃圾邮件日益泛滥的主要原因之一。现有的反垃圾邮件技术,如黑白名单[2]、贝叶斯算法[3]、基于关键词和规则[4]等,由于其过滤方式的单一性,往往存在精准率低等问题。同时,由于垃圾邮件基于全网发送的特点,传统的反垃圾邮件系统无法进行集中、统一的管理,普遍存在成本昂贵、重复建设等问题。因此,面对海量垃圾邮件的威胁,修改 SMTP、整合网络资源,整体防御垃圾邮件是很有必要的。

Duan 等人[5]在 IM2000 协议[6]基础上提出的 DMTP 可从根本上解决接收方被动接收邮件的问题,但仍有 SMTA(sender mail transfer agent)划分类别模糊化导致用户耗费大量时间审查信封信息的不足。同时,云计算[7]的出现和发展给反垃圾邮件领域提供了新的思路,它可为反垃圾邮件系统提供一体化、可扩展的服务。因此,本文在改进 DMTP 的基础上,结合云计算防范技术,设计了云计算平台下基于改进型 DMTP 的反垃圾邮件系统模型。

1 DMTP

DMTP 基于 SIRP 模式设计,接收方可自主选择接收邮件。在 DMTP 中,RMTA(receiver mail transfer agent)根据 SMTA 的 IP 地址将其划分为三类并作不同处理:a)黑名单,关闭会话连接;b)白名单,按照 SMTP 的邮件发送流程执行;c)灰名单,由接收方根据信封信息判断是否接收邮件。由于接收方查看信封信息时间不确定,SMTA 须处于长期连通状态才能实时给接收方传送邮件,因此增加了垃圾邮件的发送成本。

DMTP 根据 SMTA 的 IP 地址确定发送方类别,存在分类规则粒度过大的缺陷。公共邮件服务器对应的用户群类属不同类别,导致 SMTA 的类别划分模糊化。由于 SMTA 类别划分模糊化产生灰名单划分范围急剧扩张效应,从而导致用户审查信封信息的不便。可知,通过缩小分类规则粒度的方式,可在一定程度上降低信封信息量,从而降低用户审查的不便。

收稿日期: 2012-03-15; **修回日期**: 2012-04-24 **基金项目**: 国家自然科学基金资助项目(60673164)

作者简介: 刘海韬(1970-),男,湖南邵阳人,副教授,博士,主要研究方向为下一代互联网、网络安全管理;阳洁(1989-),女(通信作者),湖南衡阳人,硕士研究生,主要研究方向为下一代互联网、网络安全管理(yang0826@mail.csu.edu.cn).

2 云计算平台下基于改进型 DMTP 的反垃圾邮件系统总体设计

2.1 改进型 DMTP 的设计

2.1.1 改进型 DMTP 的实现

改进型 DMTP 通过增加 SAdd(发送方邮件地址)分类条件方式以细化分类规则粒度。该协议在 DMTP 基础上添加 251 和 252 两个返回码。新定义的返回码含义如表 1 所示。

表 1 改进型 DMTP 新定义的命令以及返回码

返回码	说明
251	RMTA 通知 SMTA 发送信封信息
252	RMTA 通知 SMTA 信封信息需进一步判定

改进型 DMTP 中 RMTA 处理信息传输请求算法如下所示:

```
require:dMTA:well-known spammer class;
require:bAdd:balck-mail-list;
require:rAdd:regular contact class;
1:receiving TCP session open request on port 25;
2:ip = get IP address of sender mail server;
3:if(ip ∈ dMTA) then
4:    reply with 554(to decline TCP session opening require);
5:else
6:    reply with 251(see table 1)
7:    accept envelope data;
8:    SAdd = get sender-ADD in envelope data;
9:    if(SAdd ∈ bAdd) then
10:       reply with 554 (to decline TCP session opening require);
11:   else if(SAdd ∈ rAdd) then
12:       reply with 220(to accept TCP session opening request);
13:       proceed as if SMTP used;
14:       else
15:       reply with 252(see table 1);
16:       reject DATA command;
17:   end if;
18:end if.
```

2.1.2 改进型 DMTP 增量式发布的实现

目前,绝大多数邮件系统支持的协议为 SMTP,短期内升级为改进型 DMTP 非常困难。面对两种协议互存的情况,可采取“发送方激励模式”在网络中增量式发布。RMTA 通过 SMTA 在 MAIL 命令中是否添加关键字“DMTP”的方式识别 SMTA 支持协议的类别。在邮件发送过程中,给类属灰白类别且不支持改进型 DMTP 的 SMTA 返回错误,只有当该 SMTA 升级后,RMTA 才向用户显示信息。改进型 DMTP 增量式发布中 RMTA 处理信息传递请求的算法如下所示:

```
require:dMTA:well-known spammer class;
1:receiving TCP session open request on port 25;
2:ip = get IP address of sender mail server;
3:if(ip ∈ dMTA) then
4:    reply with 554(to decline TCP session opening require);
5:else
6:    reply with 220(to accept TCP session opening require);
7:    proceed to the MALL command;
8:    if(find keyword "IDMTP" in the MAIL command) then
9:        proceed to DMTP;
10:   else
11:   respond to DATA command with 354;
12:   receive message;
13:   respond weth 550(permanent error);
14:   store message,send puzzle;
15:   message invisible to user;
16:   /* message becomes visible to use only after puzzle solved */
17:   end if;
18:end if.
```

2.2 系统结构模型

根据上文中对 DMTP 分析可知,由于分类规则粒度过大,导致用户审查信封信息不便。面对以上情况,可充分利用云计算的强计算力[8],将信封信息上传至云端分析。本文将改进型 DMTP 与云计算结合,设计了云计算平台下基于改进型 DMTP 的反垃圾邮件系统模型。该系统模型的体系结构如图 1 所示。

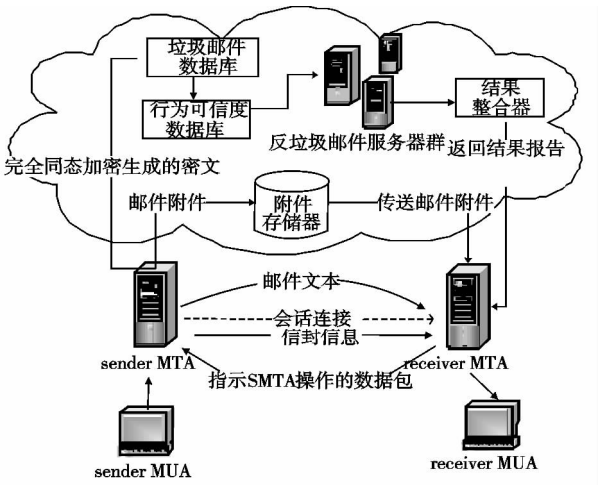

图1 云计算平台下基于改进型DMTP的反垃圾邮件系统结构模型

从图 1 可以看出,模型将本需用户审查的信封信息加密上传至云端,由云端分析,RMTA 根据云端分析结果再决定是否给用户传送信封信息。这样,通过 DMTP 与云计算的紧密结合,在减轻传统反垃圾邮件系统负荷的基础上,同时提升了系统判断精准率。

2.3 基于改进型 DMTP 邮件系统的工作流程

图 1 中,基于改进型 DMTP 邮件系统的工作流程如下:

a)SMUA(sender mail user agent)帮助用户读写文件,并将邮件传送至 SMTA。

b)SMTA 确定接收 SMUA 发送邮件后,与 RMTA 建立会话连接。

c)RMTA 根据 SMTA 的 IP 地址,将其分为两类并作不同处理:(a)黑名单,关闭会话连接;(b)其他类别,回复返回码 251。

d)SMTA 接收返回码 251 后,发送信封信息至 RMTA,信封信息中包含的参数如表 2 所示。

表 2 信封信息中包含的参数

参数	说明	参数	说明
MSID	邮件编号	RAdd	接收方邮件地址
SAdd	发送方邮件地址	BRIEF-CON	邮件简要内容

e)查询接收方的黑白名单确定发送方类别。如未出现在黑白名单中,则该发送方为灰名单。若发送方为黑名单,RMTA 断开连接;否则,RMTA 发送邮件操作数据至 SMTP。数据中包含的参数如表 3 所示。

数据中的部分参数可为空值,如发送方属于白名单,RMTA 无须为邮件在云端分配邮件文本检测地址,CPAD 可为空值。

f)SMTA 确定接收邮件处理数据后,提取 CSAD 与 CPAD 参数值进行相应处理。

(a)如果发送方属于白名单,将附件上传至 CSAD 存

储[8],文本信息直接发送至 RMTA。

(b)如果发送者属于灰名单,将附件上传至 CSAD 存储。出于对安全和隐私考虑,信封信息不能采用明文形式在网络中传输,可将其进行完全同态加密[9],将密文上传至 CPAD 分析处理。

g)云端分析结束后,RMTA 根据 CPAD 值读取云端分析结果。信封信息与云端分析结果由 RMTA 发送至 RMUA,用户根据以上数据自行决定是否提取该邮件。如果用户确定提取该邮件,RMTA 使用命令 GTML 从 SMTA 中获取邮件文本信息而附件需根据 CSAD 值从云端下载。

表3　邮件操作数据包含的参数

参数	说明
MSID	邮件编号
CSAD	附件云端存储地址
CPAD	邮件文本进行反垃圾邮件检测地址
AUTH-USER	SMTA 访问云端地址用户名
AUTH-PASS	SMTA 访问云端地址密码

2.4　云端反垃圾邮件集群系统的过滤机制

云计算是近年来互联网领域兴起的一个热点,实现了计算在客户间的共享,显著提高了处理器和存储设备的利用率,也避免了用户对信息系统的重复建设。云端反垃圾邮件集群系统基于免疫机制原理[10]设计,采取多步骤相结合方式综合分析确定邮件类别。

a)垃圾邮件数据库。该数据库中保存已识别的垃圾邮件密文,云端首先将待检测的邮件密文与垃圾邮件数据库相匹配,如果匹配一致,则可确定该邮件为垃圾邮件,直接返回分析结果;否则,邮件需继续检测。云端通过建立垃圾邮件数据库可共享信息资源,极大节省了系统资源。

b)行为可信度数据库。该数据库中保存发送方的行为可信度指数,如果发送方的行为可信度低于接收方设定值,就把该邮件判定为垃圾邮件;否则交给集群系统继续检测。实时降低被集群系统检测出的垃圾邮件发送方可信度。

c)反垃圾邮件服务器群。综合采用多种不同技术(如蜜罐技术[11]、贝叶斯算法、基于关键词和规则过滤[12]、URL 分析[13]等)的过滤引擎组成庞大服务器动态分析平台,可显著增加垃圾邮件的召回率和精确率。

d)结果整合器。整合过滤引擎群的分析结果对邮件进行判断,设定一个阈值,如果大于这个阈值就可判断该邮件为垃圾邮件。结果整合器还包含了对过滤引擎群的管理功能,如果反垃圾邮件服务器群中某过滤引擎由于执行异常或者失败导致不能反馈结果,则可根据总结果的子集整合结果报告。

3　基于 Eucalyptus 的反垃圾邮件系统构建模型

本文设计了云计算平台下基于改进型 DMTP 的反垃圾邮件系统模型。在系统基于云计算的构建模型这一块,可选择 Eucalyptus[14,15]。Eucalyptus 是目前云计算领域发展快速且成熟的开源软件基础设施之一,主要是用来通过计算集群或工作站群来实现弹性的、实用的云计算。本文在分析 Eucalyptus 体系结构的基础上,引用了基于 Eucalyptus 的反垃圾邮件系统构建模型。该系统的构建模型如图 2 所示。

由图 2 可知,

a) Anti-Spam CLC(cloud controller)层,系统主要的控制器组件,负责管理整个系统。SMTA 和 RMTA 可与 CLC 通信,由 CLC 负责将请求传递给相应的组件,并收集来自这些组件的响应发送至 SMTA 或 RMTA。

b) CC(cluster controller)层,负责管理整个虚拟实例网络。CC 根据收集 NC(node controller)的资源状态信息,将虚拟机中实例的执行请求调度到 NC。

c) NC 层,负责管理过滤引擎。

d) engine 层是垃圾邮件过滤引擎,负责执行邮件的具体过滤任务。

e) Anti-Spam database 包括垃圾邮件数据库和行为可信度数据库。

f) W(Walrus)层,管理 SMTA 和 RMTA 对系统内的存储服务的访问。SMTA 和 RMTA 须通过 W 层对附件进行操作。

g) SC(storage controller)层,SC 与 W 联合工作,用于存储和访问附件信息。

h) Mail attachments database, RMTA 从该数据库中下载附件。

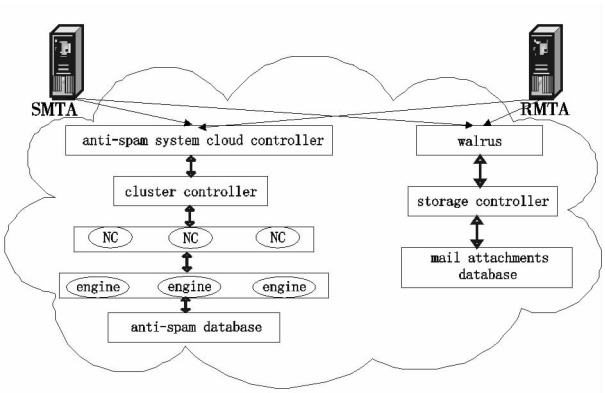

图2　基于Eucalyptus的反垃圾邮件系统构建模型

4　结束语

本文分析了在 SMTP 上扩展的 DMTP 的不足,提出了改进型 DMTP,并通过对云计算构建模型的研究,设计了云计算平台下基于改进型 DMTP 的反垃圾邮件系统的结构模型与基于 Eucalyptus 的反垃圾邮件系统构建模型。该模型可解决目前 SMTP 中接收方被动接收邮件的缺点和传统反垃圾邮件系统无法综合管理、产生的垃圾邮件发送成本低和反垃圾邮件系统重复建设等问题。目前该系统仍只处于原型阶段,还存在诸多不完善问题,如要投入实际应用还需投入大量人力、物力来实现。

参考文献:

[1] SMALLBERG D. RFC 876, SMTP: survey of SMTP implementations [S/OL]. (1983-09). http://datatracker. ietf. org/doc/rfc876/.

[2] 林丹宁. 反垃圾邮件关键技术研究与实现[D]. 杭州:浙江大学, 2007:15-19.

[3] ISAAC B, JAP W J, SUTANTO J H. Improved Bayesian anti-spam filter implementation and analysis on independent spam corpuses [C]//Proc of International Conference on Computer Engineering and Technology. Washington DC: IEEE Computer Society, 2009:326-330.

[4] ALMEIDAL T A, YAMAKAMI A. Content-based spam filtering[C]// Proc of International Joint Conference on Neural Networks. 2010: 1-7.

[5] DUAN Zhen-hai, DONG Ying-fei, KARTIK G. DMTP: controlling through message delivery differentiation [J]. Science Direct Computer Networks, 2007, 51(10): 2616-2630.　(下转第 3823 页)

(上接第 3811 页)

[6] BERNSTEIN D. Internet mail 2000 (IM2000)[EB/0L]. (2000). http://cr. yp. to/im2000. html.

[7] AUN M T B, BOK-MIN G O I, KIM V T H. Cloud enabled spam filtering services: challenges and opportunities [C]//Proc of IEEE Conference on Sustainable Utilization and Development in Engineering and Technology. 2011:63-68.

[8] WEI Wei, MA Xiao-song, YU Ting. EMFS: email-based personal cloud storage[C]//Proc of the 6th IEEE International Conference on Networking, Architecture, and Storage. 2011: 248-257.

[9] GOMATHISANKARAN M, TYAGI A, NAMUDURI K. HORNS: a homomorphic encryption scheme for cloud computing using residue number system [C]//Proc of the 45th Annual Conference on Information Sciences and Systems. 2011:1-5.

[10] 张泽明, 罗文坚, 王煦法. 一种基于人工免疫的多层垃圾邮件过滤算法[J]. 电子学报, 2007,20(3):406-414.

[11] LI Hong-xia, CHEN Jun-ming, JIN Xin. An outlook on network hone-ypot[C]//Proc of International Conference on Computer Science and Service System. 2011:48-52.

[12] GANSTERER W, ILGER M, STRAUB J, *et al*. Anti-spam met-hods-state-of-the-art [R]. Vienna: Faculty of Computer Science, University of Vienna, 2005: 1-99.

[13] GEORGIOU E, DIKAIAKOS M D, STASSOPOULOU A. On the properties of spam-advertised URL addresses [J]. Journal of Network and Computer Applications, 2008, 31(4):966-985.

[14] NURMI D, WOLSKI R, GRZEGORCZYK C, *et al*. The Eucalyptus open-source cloud-computing system [C]//Proc of the 9th IEEE/ACM International Symposium on Cluster Computing and the Grid. Washington DC: IEEE Computer Society, 2009: 124-131.

[15] KHAN I, REHMAN H, ANWAR Z. Design and deployment of a trusted eucalyptus cloud[C]//Proc of the 4th IEEE International Conference on Cloud Computing. Washington DC: IEEE Computer Society, 2011:380-387.