基于双变量多项式以及空间部署信息的对称密钥建立方案 *

许 亮[1], 马行坡[2], 张 玲[1]
(1. 湖南第一师范学院 信息科学与工程系, 长沙 410205; 2. 中南大学 信息科学与工程学院, 长沙 410083)

摘 要: 为了提高建立密钥的安全性,将基于双变量多项式的密钥建立方案和传感器节点的部署信息结合起来,提出了一种新的密钥建立方案——BPIKS。经分析和证明显示,该方案既具备双变量多项式密钥建立方案的优点又能够避免其不足,能达到既不增加网络开销又提高密钥安全性的目的。

关键词: 无线传感器网络; 对称密钥; 双变量多项式; 部署信息
中图分类号: TP393.08 **文献标志码**: A **文章编号**: 1001-3695(2010)07-2614-04
doi:10.3969/j.issn.1001-3695.2010.07.060

Pair-wise key establishing scheme for sensor network using deployment knowledge and bivariate polynomials

XU Liang[1], MA Xing-po[2], ZHANG Ling[1]
(1. *Dept. of Information Science & Engineering, Hunan First Normal University, Changsha* 410205, *China*; 2. *School of Information Science & Engineering, Central South University, Changsha* 410083, *China*)

Abstract: This paper made use of the deployment knowledge of the sensors and the bivariate polynomials, proposed a new pair-wise key establishing scheme, which was named BPIKS. The analysis and proof at the last part of this paper, declare that BPIKS possesses of all the advantages of the bivariate polynomial scheme and can evade its disadvantages.
Key words: wireless sensor network; symmetric key; bivariate polynomial; deployment knowledge

0 引言

以提供安全、可靠的保密通信为目标的密钥管理方案和协议的设计是无线传感器网络安全最为重要、最为基本的研究领域。目前,国内外学者提出的许多密钥建立方案都是基于密钥预分配的,实验表明,密钥预分配方案是适合传感器网络的、可行性好的对称密钥建立方案。已经提出的密钥预分配方案有基于密钥池的密钥预分配方案、基于对称矩阵的密钥预分配方案、基于多项式的密钥预分配方案和基于初始密钥的密钥预分配方案等。有些密钥预分配方案还与节点的空间部署信息相结合,以提高随机密钥预分配方案的安全性并减少网络开销。文献[1]中提出的 E-G 方案以及文献[2]中提出的 q-composite 方案属于基于密钥池的随机密钥预分配方案;文献[3,4]提到的方案属于基于对称矩阵的密钥预分配方案;文献[5,6]提出的方案是基于对称双变量多项式的随机密钥预分配方案;文献[7]中提出的密钥预分配方案是基于初始密钥的。本文提出的密钥预分配方案 BPIKS 是基于双变量多项式以及节点的空间部署信息的。双变量多项式随机密钥预分配方案具有计算量小、存储开销少、无通信开销等优点,然而其安全性不够好,故本文将多项式随机密钥预分配方案与节点的部署信息结合,提出了基于多项式以及空间部署信息的对称密钥建立方案,以达到既要保留多项式随机密钥预分配方案的优点又要提高其安全性的目的。

1 双变量密钥预分配方案回顾

多项式密钥预分配方案首先由 Blundo 等人[6]提出,其主要思想是在传感器节点布置之前由密钥发布服务器为每一个节点分配一个 t 度多项式片段,每个节点都能利用该多项式片段以及其他 $t-1$ 个节点的 ID 独立地计算出与其他 $t-1$ 个节点的通信密钥。原始的密钥预分配方案主要是用来为一组节点建立通信密钥的,本文只考虑两个节点之间对称密钥的建立,因此只介绍双变量多项式密钥预分配方案。

一个 t 度双变量多项式 $f(x,y)$ 可以定义为

$$f(x,y)=\sum_{i,j=0}^{t}a_{ij}x^iy^j$$

系数 $a_{ij}(0\leqslant i,j\leqslant t)$ 随机取自于有限集 $GF(Q)$,Q 是能足够容纳一个加密密钥的大素数。网络中每一个节点都有一个惟一的标志。在节点布置以前,密钥发布服务器会对每一个节点初始化,为每一个节点分配一个多项式片段。节点 p 被分配的多项式片段为 $g_p(y)$,该多项式片段可以通过以下方式获得:

$$g_p(y)=f(p,y)$$

收稿日期: 2009-11-15; **修回日期**: 2010-03-24 **基金项目**: 湖南省教育厅科研资助项目(09C231)

作者简介:许亮(1982-),女,湖南岳阳人,讲师,硕士,主要研究方向为计算机软件技术及应用、算法研究(dysfxul@163.com);马行坡(1980-),男,河南郑州人,博士研究生,主要研究方向为无线传感器网络密钥管理、分布式数据存储;张玲(1979-),女,湖南常德人,讲师,硕士,主要研究方向为智能信息采集.

节点 p 在其存储空间内存储了 t 个系数 $g_j(0\leqslant j\leqslant t)$:

$$g_j=\sum_{i=0}^{t}a_{ij}p^i(0\leqslant j\leqslant t)$$

其中:p 表示节点 ID,g_j 是多项式 $f(p,y)$ 中 y^j 的系数。双变量多项式有一个对称性质:$f(x,y)=f(y,x)$。

节点 p 和 q 为了建立彼此之间的对称密钥,它们首先互相交换节点 ID,然后令 $y=p$,节点 q 计算 $f(q,y)$ 的值;令 $y=q$,节点 p 计算 $f(p,y)$ 的值。因为 $f(p,q)=f(q,p)$,p 和 q 计算出来的值是相同的,并且可以此值作为它们之间的通信密钥。

多项式密钥预分配的优点是密钥的建立没有通信开销,缺陷是它具有 k 安全性质。根据文献[7],k 度双变量密钥预分配方案只有当被捕获的节点数少于 k 时是安全的。也就是说,当被捕获的节点个数少于 k 时,网络是安全的;而当被捕获的节点个数大于 k 时,将所有被捕获节点的信息集中起来就可以将双变量多项式的系数推导出来,这样,再通过监听获得的传感器节点 ID 信息,网络中所有节点之间的通信密钥都可以被计算出来。

2　BPIKS 方案

本文将双变量多项式密钥预分配方案与无线传感器网络空间部署信息相结合,提出了一种新的密钥预分配方案。依据文献[8],可行的节点部署模型分为三种:混合网格模型(blend grid model)、边界网格模型(border grid model)和分散网格模型(scatter grid model)。本文采用的节点部署模型在边界网格模型的基础上作了修改,更改后的边界网格模型如图 1 所示。为了方便描述,本文首先定义初始网格的概念。

定义　把一个大的矩形网络区域划分成 $m\times n$ 个大小相同的矩形区域,其中的每一个矩形区域称之为初始网格。

把每个初始网格中所包含的区域分为三类:第一类区域为包含在初始网格之中同时又不与其他网格相邻的矩形区域;第二类区域为第一类区域边线到初始网格边线投影之间所夹矩形区域;剩下的为第三类区域。部署在第一类区域内的节点简称为第一类节点,其邻居节点主要分布在其所在的初始网格内;部署在第二类区域内的节点简称为第二类节点,其邻居节点主要分布在相邻的两个初始网格内;部署在第三类区域内的节点简称为第三类节点,其邻居节点主要分布在相邻的四个初始网格内。

2.1　密钥预分配阶段

假设网络部署区域为矩形区域,包含有 $m\times n$ 个初始网格,每个初始网格区域内包含一个第一类区域。密钥设置服务器首先在有限集 F_g(g 为能足够容纳一个密钥的大素数)上产生 $m\times n$ 个双变量多项式,其中每一个初始网格与其中的一个双变量多项式相对应。第 k 个初始网格对应的双变量多项式为 $f_k(x,y)=\sum_{i,j=0}^{t}k_{ij}x^iy^j$。为了保证这种对应关系,本文将这 $m\times n$ 个初始网格按照一定顺序进行编号,并将每个初始网格的编号作为其对应双变量多项式的 ID 号。对于每一个待部署的节点来说,一方面,需要依据其对应的初始网格所对应的双变量多项式计算多项式片段并下载到该传感器节点上;另一方面,如果节点对应的小矩形区域(初始网格内节点所在的小矩形区域)有相邻的初始网格的话,密钥设置服务器还需要根据相邻的初始网格所对应的双变量多项式分别计算多项式片段,并下载到该节点上。对于要部署在第一类区域内的节点,其对应的小矩形区域有零个相邻的初始网格;对于要部署在第二类区域内的节点,其所对应的小矩形区域可能与零个或者一个初始网格相邻;对于要部署在第三类区域内的节点,该节点所对应的小矩形区域可能与零个、一个或者三个初始网格相邻。

给定任意一个节点 q,其对应的初始网格编号为 k。如果 q 将要被部署在第一类区域中,密钥设置服务器将把 $g_p^k(y)=f_k(p,y)$ 下载到该节点当中;如果 q 将要被部署在第二类区域中,并且该区域有一个相邻的编号为 b 的初始网格,密钥设置服务器将把 $g_p^k(y)=f_k(p,y)$ 和 $g_p^b(y)=f_b(p,y)$ 下载到该节点当中;如果 q 将要被部署在第三类区域中,并且该区域与三个编号分别为 d、e、f 的初始网格相邻,密钥设置服务器将把 $g_p^k(y)=f_k(p,y)$、$g_p^d(y)=f_d(p,y)$、$g_p^e(y)=f_e(p,y)$ 和 $g_p^f(y)=f_f(p,y)$ 下载到该节点当中。另外,与多项式片段相对应的双变量多项式 ID 号也同时要下载到传感器节点当中。

2.2　传感器节点部署阶段

当密钥设置服务器完成对每个节点下载密钥材料信息后,开始进行传感器节点的部署。对于待部署区域而言,其中的每个初始网格包含一个第一类区域、四个第二类区域和四个第三类区域。节点部署时采用人工或机器人分组进行,初始网格中每一类区域中的每一个小区域对应一组传感器节点。图 1 中,整个区域被划分成 2×2 个初始网格。在 4 号初始网格内,g 为第一类区域,x、y、z、w 为第二类区域,r、s、u、v 为第三类区域。g、z、x、v 四个区域与下载了根据 4 号网格所对应双变量多项式计算出的多项式片段的传感器节点组相对应;y、u 两个区域与同时下载了根据 3、4 号网格所对应双变量多项式计算出的多项式片段的传感器节点组相对应;w、s 两个区域与同时下载了根据 2、4 号网格所对应多项式计算出的多项式片段的传感器节点组相对应;区域 r 与同时下载了根据 1~4 号网格所对应多项式计算出的多项式片段的传感器节点组相对应。部署时对应的传感器节点组部署到对应的区域中。

2.3　密钥直接建立阶段

传感器节点被部署到指定区域以后,进入对称密钥建立阶段。传感器节点首先与其邻居节点交换自身 ID 信息及其所下载多项式片段对应的双变量多项式的 ID 信息。如果两个相邻节点之间有一个相同的多项式 ID 号,则这两个节点可依据该多项式对应的多项式片段计算对称密钥,将计算出的对称密钥作为两个节点的通信密钥;如果两个相邻节点之间有多个相同的多项式 ID 号,则这两个节点可先分别依据这些多项式对应的多项式片段计算对称密钥,然后取计算出来的多个对称密钥的异或值作为这两个节点的通信密钥。节点 A 和 B 建立对称密钥的通信过程表示如下:

$$A\rightarrow B:\quad \mathrm{ID}_A,\mathrm{SET}_A^f$$

$$B\rightarrow A:\quad \mathrm{ID}_B,\mathrm{SET}_B^f$$

其中:SET_A^f 和 SET_B^f 分别表示节点 A 和 B 所下载多项式片段对应的双变量多项式 ID 号集合,该集合中可能有一个、两个或者四个 ID 号。假设 A、B 之间有两个相同的双变量多项式 ID 号,分别为 i_1 和 i_2,则节点 A 按照以下方式计算对称密钥:

$$K_{AB}=g_A^{i1}(\mathrm{ID}_B)\oplus g_A^{i2}(\mathrm{ID}_B)$$

则节点 B 按照以下方式计算对称密钥:

$$K_{BA}=g_B^{i1}(\mathrm{ID}_A)\oplus g_B^{i2}(\mathrm{ID}_A)$$

由 2.1 节所述可知，$g_A^{i1}(\mathrm{ID}_B)=g_B^{i1}(\mathrm{ID}_A)$，$g_A^{i2}(\mathrm{ID}_B)=g_B^{i2}(\mathrm{ID}_A)$，故有 $K_{AB}=K_{BA}$。

如此，A、B 之间便建立了对称密钥。

2.4　密钥间接建立阶段

当两个相邻节点之间没有相同的双变量多项式 ID 号时，为了建立对称密钥，它们首先会寻找一条能够到达对方的安全路径（这条路径上任一链路的两个节点之间都已经建立了对称密钥），然后其中一个节点将自身产生的对称密钥经安全路径传输给另外一个节点。如图 2 所示，节点 A 与 C 相邻，节点 C 与 B 相邻，并且它们之间都已通过直接密钥建立方式建立了对称密钥；节点 A 与 B 相邻，但 A、B 之间没有相同的多项式 ID，无法直接建立对称密钥。节点 A 与 B 建立对称密钥的过程如下：

a）节点 A 随机产生一个密钥 K_{AB}。

b）节点 A 用 A 与 C 之间的对称密钥加密并传输给节点 C：$A\rightarrow C:E_{K_{AC}}(K_{AB})$。

c）节点 C 先用 K_{AC} 解密 $E_{K_{AC}}(K_{AB})$ 获得 K_{AB}，再用 K_{CB} 对 K_{AB} 加密并传输给节点 B：$C\rightarrow B:E_{K_{CB}}(K_{AB})$。

d）节点 B 收到 $E_{K_{CB}}(K_{AB})$ 后，用 K_{BC} 解密就可获得 K_{AB}。

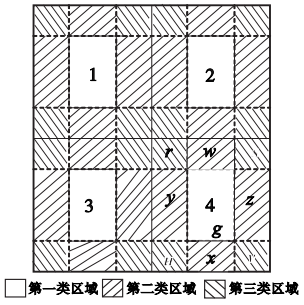

图1　节点部署区域种类划分

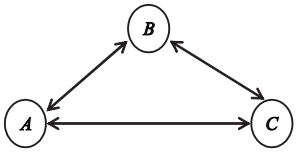

图2　对称密钥的间接建立

3　方案性能分析

3.1　安全连通性分析

当一个节点与其邻居节点建立了对称密钥时，称它们之间的链路是安全的。本文所进行的安全连通性分析主要是分析安全链路的条数占整个传感器网络链路总条数的比率。从两个方面进行分析：a）分析传感器网络在进行直接对称密钥建立过程后网络中安全链路条数占总链路条数的比率；b）分析不能直接建立对称密钥的相邻节点之间间接建立对称密钥后网络中安全链路条数占总链路条数的比率。假定传感器节点的通信半径为 R，传感器网络各类区域中小矩形的最小宽度为 w，对于第一阶段的分析，本文给出以下引理：

定理 1　当 $w>R/2$ 时，节点在进行直接密钥建立阶段后，安全链路数占网络总链路数的比率为 100%。

证明　先分析第一类节点。由于 $w>R/2$，即 $R<2w$，第一类节点的邻居节点一定分布在其所在初始网格以及与该初始网格相邻的外围第二类、第三类区域范围内，而此范围内的节点均包含该初始网格所对应的双变量多项式片段，故第一类节点与其邻居节点都能直接建立对称密钥。再分析第二类节点，由于 $R<2w$，第二类节点的邻居节点可能分布在以下三个区域：a）其所在的初始网格；b）与其所在的小矩形区域相邻的初始网格；c）与其所在初始网格相邻的外围小矩形区域。区域 a）中节点都含有该第二类节点所在初始网格对应的双变量多项式片段，区域 b）中节点都含有与该第二类节点所在小矩形区域相邻的初始网格所对应的多项式片段，而这两类多项式片段该第二类节点都拥有，因此第二类节点也可以与其所有的邻居节点直接建立对称密钥。对于第三类节点，同样，由于 $R<2w$，第三类节点的邻居节点都分布在与该节点所在小矩形区域有一个公共顶点的初始网格中，而该类节点包含所有这些初始网格所对应多项式的多项式片段，因此，第三类节点也可以与其邻居节点直接建立对称密钥。

定理 2　在节点密度达到每个小矩形区域至少存在一个节点的条件下，节点在经过间接密钥建立阶段后，安全链路数占网络总链路数的比率为 100%。

证明　首先定义安全路径。如果一条路径上的所有链路都是安全链路，则该路径就是安全路径。只要传感器网络中任意两个节点之间都有一条安全路径，那么经过间接密钥建立阶段以后，任意相邻节点之间都可以建立一条安全链路。因此要证明定理 2，只需要证明任意两个节点之间都有一条安全路径。如果一个传感器网络去掉不安全链路之后仍然是连通的，则称该传感器网络是安全连通的。一个传感器网络是安全连通的，说明任意两个传感器节点之间都存在一条安全路径。由于位于同一初始网格内的节点都有根据相同的双变量多项式计算而来的多项式片段，同一初始网格内的任意相邻节点之间都可以建立一条安全链路，即同一初始网格内的传感器节点是安全连通的；又因为任意两个相邻初始网格内都存在一对相邻的第二类小矩形区域和两对相邻的第三类小矩形区域，而相邻的同类小矩形区域内的节点之间又根据相同多项式计算得来的多项式片段，可以建立安全链路，相邻的两个初始网格可以通过这些区域作为桥梁建立安全链路。因此，整个传感器网络都是安全连通的，定理 2 得证。

3.2　网络抗节点俘获能力分析

节点俘获有两种攻击类型，即节点盲俘获和有目的的节点俘获。节点盲俘获指的是攻击者随机从网络中俘获 k 个节点；有目的的节点俘获指的是攻击者了解区域划分，集中在某一区域俘获节点，利用被俘获节点上的信息计算该区域节点所对应的多项式，从而破解其他利用该多项式建立的节点之间的通信密钥。

首先分析 BPIKS 方案的抗节点盲俘获能力。假设整个传感器网络的面积为 S_{total}，总的节点数为 N_{total}，一个初始网格以及与其有公共边或者有公共顶点的外围第二类和第三类区域总面积为 S_f。在节点均匀分布条件下，同一个多项式对应的节点数为 $N_f=\dfrac{S_f}{S_{\text{total}}}\times N_{\text{total}}$，则任意俘获网络中的 t 个节点并利用其中的信息计算出一个双变量多项式的概率为

$$p=\frac{C^t N_f}{C^t N_{\text{total}}}$$

在有目的的节点捕获这种情况下，攻击者可以集中捕获某一小区域内的节点。当初始网格内的第一类区域有 t 个节点被俘获时，该初始网格对应的双变量多项式可以由这 t 个节点上的信息计算出来，这样该初始网格中第一类区域内的节点与其邻居节点建立的对称密钥将会被破解；当初始网格内的某一第二类小矩形区域有 t 个节点被俘获时，则该初始网格以及与该小矩形相邻的初始网格所对应的双变量都会被计算出来，这两个初始网格之中第一类区域和相邻的第二类区域、第三类区

域内的节点与其邻居节点之间建立的对称密钥将会被破解;当初始网格内的某一第三类小矩形区域有 t 个节点被俘获时,则该小矩形区域所在初始网格以及与该小矩形区域相邻的或者有公共顶点的初始网格内所有节点与其邻居节点建立的对称密钥都会被破解。

综合以上两种情况,当网络中有 t 个节点被俘获时,攻击者只能破解很少部分网格内节点的对称密钥,与单纯基于双变量多项式的对称密钥建立方案相比,安全性有了很大的提高。

3.3　计算、通信、存储开销分析

BPIKS 方案的计算、通信和存储开销都比较小。在计算开销方面,节点只需要计算 1 ~4 个多项式的值,这取决于两相邻节点之间相同多项式 ID 的个数;在通信开销方面,节点之间需要交换各自存储的多项式片段对应的多项式 ID 列表,而此列表的内容很少,每个列表中的 ID 数不超过四个,因此节点之间的通信开销很少;在存储开销方面,对取自于有限集 F_q 的多项式而言,其对应的每个多项式片段所占的存储空间为 $(t+1)\log q$ 位,如果多项式 ID 占用空间为 l 位,则节点被占用的存储空间不超过 $4(t+1)\log q+4l$ 位。

3.4　网络规模扩展性分析

由于传感器网络包含大量的传感器节点,密钥预分配方案必须要支持网络扩展。按照本文提出的密钥方案,无论网络规模多大,节点所保存的多项式片段不会超过四个,密钥建立时需要交换的多项式 ID 列表中不超过四个 ID 值,计算对称密钥时需要计算的多项式的值也不超过四个,因此,当网络规模增大时,节点的计算、通信、存储开销不会增加。另外,由于一个初始网格对应一个双变量多项式,当被捕获的节点数达到 t 个时,不安全链路被局限在局部区域,不会对全网造成危害。因此,本文提出的密钥预分配方案具有很好的网络扩展性。

4　结束语

无线传感器网络技术在军事、环境、医疗、家庭和其他的商用领域有很高的应用价值。本文所述的密钥预分配方案是在双变量多项式密钥预分配方案的基础上结合节点的部署知识提出的,新提出的密钥预分配方案具有安全性高,计算、存储、通信开销低,支持网络扩展等特点,很适合大规模无线传感器网络的安全应用。

参考文献:

[1] ESCHENAUER L, GLIGOR V. A key management scheme for distributed sensor networks[C]//Proc of the 9th ACM Conference on Computer and Communications Security. New York: ACM Press, 2002:41-47.

[2] CHAN H, PERRIG A, SONG D. Random key predistribution schemes for sensor networks[C]//Proc of IEEE Symposium on Security and Privacy. Washington DC: IEEE Computer Society, 2003:197-213.

[3] DU W, DENG J, HAN Y S, *et al*. A pairwise key pre-distribution scheme for wireless sensor networks[C]//Proc of the 10th ACM Conference on Computer and Communications Security. New York: ACM Press, 2003:42-51.

[4] BLOM R. An optimal class of symmetric key generation systems[C]//Proc of the EUROCRYPT Workshop on Advances in Cryptology: Theory and Application of Cryptographic Techniques. New York: Springer-Verlag, 1984:335-338.

[5] LIU D, NING P. Establishing pair-wise keys in distributed sensor networks[C]//Proc of the 10th ACM Conference on Computer and Communications Security. New York: ACM Press, 2003:52-61.

[6] BLUNDO C, SANTIS A D, HERZBERG A, *et al*. Perfectly secure key distribution for dynamic conferences[J]. Information and Computation, 1998, 146(1):1-23.

[7] 王国军,吕婷婷,过敏意. 无线传感器网络中基于临时初始密钥的密钥管理协议[J]. 传感技术学报, 2007, 20(7):1581-1586.

[8] YU Bo, CAO Xiao-mei, HAN Peng, *et al*. Flexible deployment models for location-aware key management in wireless sensor networks[C]//Proc of the 8th Asia Pacific Web Conference. Berlin: Springer-Verlag, 2006:343-354.

(上接第 2610 页)像边缘检测结果中提取出一个 24 维的直方图统计量用于刻画图像的内容特征,并采用灰度直方图和四个已经被广泛验证了的颜色统计特征来刻画图像的色彩特征。

与基于 OCR 的图像检测方法相比,本文提出的模型不依赖于对图像文字内容信息的提取,因此对各种图像文字识别干扰技术免疫;与其他采用图像基础特征的相关检测方法相比,本文提出的模型对图像文件所包含信息的描述更为全面也更为系统。通过在两组图像型垃圾邮件公开数据集上的交叉验证结果表明,所提出的特征模型能够准确区分图像型垃圾邮件与正常邮件图片,其分类准确率优于近期报道的多数相关方法的实验结果,误报率也稳定在较低水平。该实验结果验证了本文所提出的图像型垃圾邮件特征模型的有效性,同时表明该方法具有良好的应用推广前景。

参考文献:

[1] BLANZIERI E, BRYL A. A survey of learning-based techniques of e-mail spam filtering[J]. Artificial Intelligence Review, 2008, 29(1):63-92.

[2] FUMERA G, PILLAI I, ROLI F. Spam filtering based on the analysis of text information embedded into images[J]. Journal of Machine Learning Research, 2006, 7(12):2699-2720.

[3] 万明成,耿技,程红蓉,等. 图像型垃圾邮件过滤技术综述[J]. 计算机应用研究, 2008, 25(9):2579-2582.

[4] CHAN P K, LIPPMANN R P. Machine learning for computer security[J]. Journal of Machine Learning Research, 2006, 7(12):2669-2672.

[5] DREDZE M, GEVARYAHU R, ELIAS-BACHRACH A. Learning fast classifiers for image spam[C]//Proc of the 4th Conference on E-mail and Anti-Spam. 2007:487-493.

[6] WAN Ming-cheng, ZHANG Feng-li, CHENG Hong-rong, *et al*. Text localization in spam image using edge features[C]//Proc of International Conference on Communications, Circuits and Systems. 2008:838-842.

[7] 张引,潘云鹤. 复杂背景下文本提取的彩色边缘检测算子设计[J]. 软件学报, 2001, 12(8):1229-1235.

[8] BYUN B, LEE C H, WEBB S, *et al*. An anti-spam filter combination framework for text-and-image emails through incremental learning[C]//Proc of the 6th Conference on E-mail and Anti-Spam. 2009.

[9] WANG Zhe, JOSEPHSON W, LV Qin, *et al*. Filtering image spam with near-duplicate detection[C]//Proc of the 4th Conference on E-mail and Anti-Spam. 2007.

[10] ARADHYE H B, MYERS G K, HERSON J A. Image analysis for efficient categorization of image-based spam e-mail[C]//Proc of the 8th International Conference on Document Analysis and Recognition. Washington DC: IEEE Computer Society, 2005:914-918.