

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2012 年 9 月 7 日 (07.09.2012)



(10) 国际公布号
WO 2012/116587 A1

- (51) 国际专利分类号:
H04L 12/58 (2006.01)
- (21) 国际申请号: PCT/CN2012/070816
- (22) 国际申请日: 2012 年 2 月 1 日 (01.02.2012)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201110051222.2 2011 年 3 月 3 日 (03.03.2011) CN
- (71) 申请人 (对除美国外的所有指定国): **腾讯科技(深圳)有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED)** [CN/CN]; 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。
- (72) 发明人; 及
- (75) 发明人/申请人 (仅对美国): **王晖 (WANG, Hui)** [CN/CN]; 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。**林华尚 (LIN, Huashang)** [CN/CN]; 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。
- (74) 代理人: **北京三高永信知识产权代理有限公司 (BEIJING SAN GAO YONG XIN INTELLECTU-**

AL PROPERTY AGENCY CO., LTD.); 中国北京市海淀区学院路蓟门里和景园 A-1-102, Beijing 100088 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

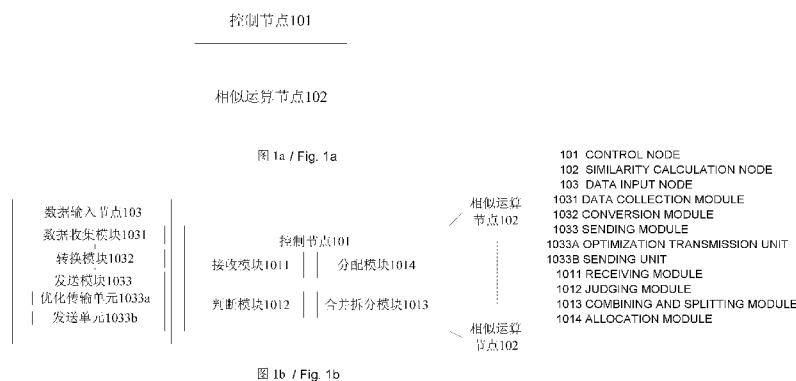
(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

(54) Title: SIMILAR EMAIL PROCESSING SYSTEM AND METHOD

(54) 发明名称: 相似邮件处理系统和方法



(57) Abstract: Disclosed are a similar email processing system and method, belonging to the technical field of networks. The system includes: a control node for receiving a sample in a preset format, judging whether or not the sample in the preset format is the final result of similarity calculation, and if it is not, then combining or splitting the sample in the preset format according to a preset standard to obtain a plurality of sub-task data packets, and distributing the plurality of sub-task data packets to a plurality of similarity calculation nodes; the plurality of similarity operation nodes for calculating the similarity relation of samples in the received sub-task data packets to obtain an intermediate result of similarity calculation, with the intermediate result of similarity calculation being a sample in a preset format, and feeding the sample in the preset format back to the control node, with the intermediate result of similarity calculation including a unique similar sample, a similarity relation and the similarity counting of the unique similar sample.

(57) 摘要:

[见续页]

WO 2012/116587 A1



本发明公开了一种相似邮件处理系统和方法，属于网络技术领域。本系统包括：控制节点，用于接收预设格式的样本，并判断所述预设格式的样本是否为相似计算最终结果，如果否，则根据预设标准对所述预设格式的样本进行合并或拆分处理，得到多个子任务数据包，将所述多个子任务数据包分配给多个相似运算节点；多个所述相似运算节点，用于对接收到的子任务数据包内的样本进行相似关系计算，获得相似计算中间结果，所述相似计算中间结果为预设格式的样本，将所述预设格式的样本反馈给所述控制节点，所述相似计算中间结果包括唯一相似样本、相似关系和所述唯一相似样本的相似计数。

相似邮件处理系统和方法

技术领域

5 本发明涉及网络技术领域，特别涉及一种相似邮件处理系统和方法。

背景技术

随着网络的发展，邮件渐渐发展成为人们日常通信的重要工具，但是，随之产生的垃圾邮件也日益增多，造成了使用者的不便，在现有技术中，采用了基于文本相似技术的反垃圾邮件体系，从统计到拦截拥有一套成熟的架构，这套系统主要基于了单机运算的模式，能够在较短时间内统计一定数量规模的邮件，从中统计获得邮件之间的相似关系和相似指数。由于这套系统能够识别出经过一定幅度变形和添加了干扰元素的垃圾邮件，因此实际应用中，无论在拦截垃圾邮件的规模，数量和准确度上都具有十分优异的指标。

在对现有技术进行分析后，发明人发现现有技术至少具有如下缺点：

15 现有技术中的相似邮件处理系统是基于单机运算模式，在能够处理的输入数据和输出数据规模上具有较大限制，对单次百万级别以上的输入数据规模存在运算速度慢，系统负载高的问题，无法实现实时，在准实时统计上由于完成时间较长也无法做到。

发明内容

20 本发明实施例提供了一种相似邮件处理系统和方法。所述技术方案如下：

一种相似邮件处理系统包括：

控制节点，用于接收预设格式的样本，并判断所述预设格式的样本是否为相似计算最终结果，如果否，则根据预设标准对所述预设格式的样本进行合并或拆分处理，得到多个子任务数据包，将所述多个子任务数据包分配给多个相似运算节点；

25 多个所述相似运算节点，用于对接收到的子任务数据包内的样本进行相似关系计算，获得相似计算中间结果，所述相似计算中间结果为预设格式，将所述相似计算中间结果反馈给所述控制节点，所述相似计算中间结果至少包括：唯一相似样本、相似关系和所述唯一相似样本的相似计数。

所述系统还包括：

30 数据输入节点，用于收集原始样本并将所述原始样本并将所述原始样本转换为预设格式，并将转换后的原始样本包作为预设格式的样本发送给所述控制节点。

所述数据输入节点包括：

数据收集模块，用于收集相似邮件处理系统服务器或服务器集群上的邮件，将所述邮件作为原始样本；

转换模块，用于将所述原始样本转换为与相似计算匹配的预设格式；

- 5 发送模块，用于为转换后的原始样本包分配任务标识，并将转换后的原始样本包作为预设格式的样本整体或分批次发送给所述控制节点。

所述发送模块包括：

优化传输单元，用于根据网络情况，将所述转换后的原始样本包分拆成多个数据包；

- 10 发送单元，用于将所述优化传输单元输出的所述多个数据包作为预设格式的样本分批次发送给所述控制节点。

所述控制节点包括：

接收模块，用于接收预设格式的样本；

- 15 判断模块，用于判断所述预设格式的样本是否满足预设条件，如果是，则所述预设格式的样本是相似计算最终结果，如果否，则所述预设格式的样本不是相似计算最终结果，并触发合并拆分模块；

所述合并拆分模块，用于根据所述相似运算节点的心跳信息，对所述预设格式的样本进行合并或拆分处理，得到多个子任务数据包；所述心跳信息用于监控和描述所述相似运算节点的空闲计算能力；

- 20 分配模块，用于将所述合并拆分模块得到的所述多个子任务数据包分别分配各个相似运算节点。

所述合并拆分模块具体用于统计所述转换后的原始样本包和所述预设格式的样本的数据关键指标，并根据配置文件登记信息和所述数据关键指标对所述转换后的原始样本包和所述预设格式的样本进行排序，并根据排序顺序将所述所述转换后的原始样本包或所述预设格式的样本进行合并或拆分处理，得到多个子任务数据包。

- 25 所述控制节点还包括：

心跳信息监控模块，用于每隔预设时长或当接收到预设格式的样本时，获取所述相似运算节点的心跳信息。

所述控制节点还用于保存并记录所述预设格式的样本，记录所述多个子任务数据包及所述子任务数据包分配的相似运算节点的映射关系，并记录所述相似运算节点的心跳信息。

- 30 所述心跳信息监控模块还用于当所述相似运算节点在预设时长内未返回心跳信息且连续未返回所述心跳信息超过预设次数，则标记所述相似运算节点崩溃，并标记所述相似运算节

点上运行的子任务数据包失败，并触发所述分配模块根据所述相似运算节点的心跳信息将标记失败的子任务数据包分配给未崩溃且空闲的相似运算节点。

一种相似邮件处理方法，包括：

5 接收原始样本和预设格式的样本，并将接收到的原始样本转换为预设格式；

 判断所述转换后的原始样本包和所述预设格式的样本是否为相似计算最终结果；

 如果否，则根据预设标准对所述转换后的原始样本包和所述预设格式的样本进行合并或拆分处理，得到多个子任务数据包；

 对每个所述子任务数据包内的样本进行相似关系计算，获得相似计算中间结果，所述相似
10 计算中间结果为预设格式的样本，反馈所述预设格式的样本，所述相似计算中间结果至少包括：唯一相似样本、相似关系和所述唯一相似样本的相似计数。

 接收原始样本和预设格式的样本，具体包括：

 收集相似邮件处理系统服务器或服务器集群上的邮件，将所述邮件作为原始样本，为所述原始样本分配任务标识；

15 根据所述预设格式的样本的任务标识判断所述预设格式的样本所属任务是否完成，如果否，则将所述预设格式的样本与所述所属任务的其他样本汇总。

 判断转换后的原始样本包和所述预设格式的样本是否为相似计算最终结果，具体包括：

 判断所述转换后的原始样本包是否满足预设条件，如果所述转换后的原始样本包满足预设条件，则所述转换后的原始样本包是相似计算最终结果，如果所述转换后的原始样本不满足
20 预设条件，则所述转换后的的原始样本包不是相似计算最终结果；

 判断所述预设格式的样本是否满足预设条件，如果所述预设格式的样本满足预设条件，则所述预设格式的样本是相似计算最终结果，如果所述预设格式的样本不满足预设条件，则所述预设格式的样本不是相似计算最终结果。

 根据预设标准对所述转换后的原始样本包和所述预设格式的样本进行合并或拆分处理，
25 得到多个子任务数据包，具体包括：

 统计所述转换后的原始样本包和所述预设格式的样本的数据关键指标，并根据配置文件登记信息和所述数据关键指标对所述转换后的原始样本包和所述预设格式的样本进行排序，并根据排序顺序将所述所述转换后的原始样本包或所述预设格式的样本进行合并或拆分处理，得到多个子任务数据包。

30 当所述预设格式的样本为至少经过一次相似计算的样本且本地服务器上存在至少两个所述预设格式的样本所属任务返回的预设格式的样本时，对所述至少两个所述预设格式的样本

所属任务返回的预设格式的样本进行合并处理。

所述预设标准至少包括下列任一项：

当所述转换后的原始样本包中的记录条目数或打成数据包后的总尺寸字节数超过预设阈值，对所述转换后的原始样本包进行拆分处理；

- 5 所述预设格式的样本中的记录条目数或打成数据包后的总尺寸字节数超过预设阈值，对所述预设格式的样本进行拆分处理。

本发明实施例提供的技术方案的有益效果是：

- 通过由控制节点对输入的样本进行合并或拆分的处理，并将得到的多个子任务数据包分配给多个相似运算节点的分布式系统来实现对千万以上级别邮件的相似处理和计算，从而提
10 高了运算速度和运算能力，降低了系统负载，可以支持实时和准实时统计与拦截的反垃圾邮件需求。

附图说明

- 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术
15 描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图 1a 是本发明实施例提供的一种相似邮件处理系统的示意图；

图 1b 是本发明实施例提供的一种相似邮件处理系统的示意图；

- 20 图 2 是本发明实施例提供的一种相似邮件处理方法的流程图；

图 3 是本发明实施例提供的一种相似邮件处理方法的流程图。

具体实施方式

- 为使本发明的目的、技术方案和优点更加清楚，下面将结合附图对本发明实施方式作进
25 一步地详细描述。

在介绍本发明提供的相似邮件处理系统之前，首先对本发明的基础知识进行简要的介绍：

- 本发明基于如下的简单常识：垃圾邮件一定在数量和规模上具有显著的规模，一定在形式上存在雷同现象，不难发现，只要我们处理和运算的速度足够快，就可以在第一时间识别出垃圾邮件（具有较大的数量规模），从而实施拦截。可见，越早发现大规模的相似的垃圾邮件，就能越早进行干预，从而越早的将垃圾邮件挡在邮箱系统外（根据统计，邮箱系统超过
30 60%的邮件为垃圾邮件）。这对用户在使用上带来的好处不言而喻，同时也可大幅降低运营成

本（带宽、存储）的压力。

实施例 1

为了提高运算速度和运算能力，降低了系统负载，本发明实施例提供了一种相似邮件处理系统，参见图 1a，该系统包括：控制节点 101 和多个相似运算节点 102。

5 其中，控制节点 101，用于接收预设格式的样本，并判断所述预设格式的样本是否为相似计算最终结果，如果否，则根据预设标准对所述预设格式的样本进行合并或拆分处理，得到多个子任务数据包，将所述多个子任务数据包分配给多个相似运算节点；

多个所述相似运算节点 102，用于对接收到的子任务数据包内的样本进行相似关系计算，获得相似计算中间结果，所述相似计算中间结果为预设格式的样本，将所述预设格式的样本
10 反馈给所述控制节点，所述相似计算中间结果至少包括：唯一相似样本、相似关系和所述唯一相似样本的相似计数。

参见图 1b，所述系统还包括：

数据输入节点 103，用于收集原始样本并将所述原始样本并将所述原始样本转换为预设格式，并将转换后的原始样本包作为预设格式的样本发送给所述控制节点。

15 所述数据输入节点 103 包括：

数据收集模块 1031，用于收集相似邮件处理系统服务器或服务器集群上的邮件，将所述邮件作为原始样本；

转换模块 1032，用于将所述原始样本转换为与相似计算匹配的预设格式；

发送模块 1033，用于为转换后的原始样本包分配任务标识，并将转换后的原始样本包作
20 为预设格式的样本整体或分批次发送给所述控制节点。

所述发送模块 1033 包括：

优化传输单元 1033a，用于根据网络情况，将所述转换后的原始样本包分拆成多个数据包；

发送单元 1033b，用于将所述优化传输单元输出的所述多个数据包作为预设格式的样本
25 分批次发送给所述控制节点。

所述控制节点 101 包括：

接收模块 1011，用于接收预设格式的样本；

判断模块 1012，用于判断所述预设格式的样本是否满足预设条件，如果是，则所述预设格式的样本是相似计算最终结果，如果否，则所述预设格式的样本不是相似计算最终结果，
30 并触发合并拆分模块；

所述合并拆分模块 1013，用于根据所述相似运算节点的心跳信息，对所述预设格式的样

本进行合并或拆分处理，得到多个子任务数据包；所述心跳信息用于描述所述相似运算节点的空闲计算能力；

进一步地，所述合并拆分模块 1013，具体用于统计所述转换后的原始样本包和所述预设格式的样本的数据关键指标，并根据配置文件登记信息和所述数据关键指标对所述转换后的原始样本包和所述预设格式的样本进行排序，并根据排序顺序将所述所述转换后的原始样本包或所述预设格式的样本进行合并或拆分处理，得到多个子任务数据包。

分配模块 1014，用于将所述合并拆分模块得到的所述多个子任务数据包分别分配各个相似运算节点 102。

所述控制节点 101 还包括：

10 心跳信息监控模块，用于每隔预设时长或当接收到预设格式的样本时，获取所述相似运算节点的心跳信息。

所述控制节点 101 还用于保存并记录所述预设格式的样本，记录所述多个子任务数据包及所述子任务数据包分配的相似运算节点的映射关系，并记录所述相似运算节点的心跳信息。

15 所述心跳信息监控模块还用于当所述相似运算节点在预设时长内未返回心跳信息且连续未返回所述心跳信息超过预设次数，则标记所述相似运算节点崩溃，并标记所述相似运算节点上运行的子任务数据包失败，并触发所述分配模块根据所述相似运算节点的心跳信息将标记失败的子任务数据包分配给未崩溃且空闲的相似运算节点。

20 通过由控制节点对输入的样本进行合并或拆分的处理，并将得到的多个子任务数据包分配给多个相似运算节点的分布式系统来实现对千万以上级别邮件的相似处理和计算，从而提高了运算速度和运算能力，降低了系统负载，可以支持实时和准实时统计与拦截的反垃圾邮件需求。

实施例 2

25 为了提高运算速度和运算能力，降低了系统负载，本发明实施例提供了一种相似邮件处理方法，该方法的执行主体为上述实施例 1 提供的相似邮件处理系统，参见图 2，该方法包括：

201：相似邮件处理系统接收原始样本和预设格式的样本，并将接收到的原始样本转换为预设格式；

202：相似邮件处理系统判断该转换后的原始样本包和该预设格式的样本是否为相似计算最终结果；

30 203：如果否，则根据预设标准对该转换后的原始样本包和该预设格式的样本进行合并或拆分处理，得到多个子任务数据包；

如果是，则该预设格式的样本是相似计算最终结果，将该预设格式的样本作为该相似计算最终结果输出；

204: 相似邮件处理系统对每个该子任务数据包内的样本进行相似关系计算，获得相似计算中间结果，该相似计算中间结果为预设格式的样本，反馈该预设格式的样本，该相似计算中间结果包括唯一相似样本、相似关系和该唯一相似样本的相似计数。

其中，接收原始样本和预设格式的样本，具体包括：

收集相似邮件处理系统服务器或服务器集群上的邮件，将该邮件作为原始样本，为该原始样本分配任务标识；

根据该预设格式的样本的任务标识判断该预设格式的样本所属任务是否完成，如果否，
10 则将该预设格式的样本与该所属任务的其他样本汇总。

其中，判断转换后的原始样本包和该预设格式的样本是否为相似计算最终结果，具体包括：

判断所述转换后的原始样本包是否满足预设条件，如果所述转换后的原始样本包满足预设条件，则所述转换后的原始样本包是相似计算最终结果，如果所述转换后的原始样本不满足预设条件，则所述转换后的的原始样本包不是相似计算最终结果；

判断所述预设格式的样本是否满足预设条件，如果所述预设格式的样本满足预设条件，则所述预设格式的样本是相似计算最终结果，如果所述预设格式的样本不满足预设条件，则所述预设格式的样本不是相似计算最终结果。

其中，根据预设标准对该转换后的原始样本包和该预设格式的样本进行合并或拆分处理，
20 得到多个子任务数据包，具体包括：

统计该转换后的原始样本包和该预设格式的样本的数据关键指标，并根据配置文件登记信息和该数据关键指标对该转换后的原始样本包和该预设格式的样本进行排序，并根据排序顺序将该该转换后的原始样本包或该预设格式的样本进行合并或拆分处理，得到多个子任务数据包。

25 其中，当该预设格式的样本为至少经过一次相似计算的样本且本地服务器上存在至少两个该预设格式的样本所属任务返回的预设格式的样本时，对该至少两个该预设格式的样本所属任务返回的预设格式的样本进行合并处理。

所述预设标准至少包括下列任一项：

当该转换后的原始样本包中的记录条目数超过预设阈值，对该转换后的原始样本包进行
30 拆分处理；

当所述转换后的原始样本包中的记录条目数或打成数据包后的总尺寸字节数超过预设阈

值，对所述转换后的原始样本包进行拆分处理；

所述预设格式的样本中的记录条目数或打成数据包后的总尺寸字节数超过预设阈值，对所述预设格式的样本进行拆分处理。

本实施例提供的方法，与系统实施例属于同一构思，其具体实现过程详见系统实施例，这里不再赘述。

通过由控制节点对输入的样本进行合并或拆分的处理，并将得到的多个子任务数据包分配给多个相似运算节点的分布式系统来实现对千万以上级别邮件的相似处理和计算，从而提高了运算速度和运算能力，降低了系统负载，可以支持实时和准实时统计与拦截的反垃圾邮件需求。

实施例 3

为了提高运算速度和运算能力，降低了系统负载，本发明实施例提供了一种相似邮件处理方法，该方法的执行主体为上述实施例 1 提供的相似邮件处理系统的各个不同节点，该相似邮件处理系统包括数据输入节点、控制节点和相似计算节点，在本实施例中以该相似邮件处理系统中包含数据输入节点、控制节点、4 个相似计算节点为例进行说明，需要说明的是，控制节点既可以接收原始样本进行转换，也可以接收来自数据输入节点的样本，并由数据输入节点进行转换，在本发明实施例中，以数据输入节点进行转换为例进行说明，参见图 3，该方法的一个实施例具体包括：

301：数据输入节点中的数据收集模块收集相似邮件处理系统服务器或服务器集群上的邮件，将该邮件作为原始样本；

其中，该数据输入节点用于收集原始样本并将该原始样本并将该原始样本转换为预设格式，并将转换后的原始样本包作为预设格式的样本发送给该控制节点。

本领域技术人员可以获知，该数据输入节点可以为能够与控制节点通信的一台服务器，还可以为多台服务器组成的服务器集群。

302：数据输入节点中的转换模块将该原始样本转换为与相似计算匹配的预设格式；

需要说明的是，在后续进行相似计算时，为提升处理速度与方便记录处理结果，需要对原始样本进行转换，该转换是根据后续的相似计算节点上配置的相似计算算法进行的，需转换为该相似计算算法对应的数据格式。其中，该相似计算算法可以为多种，本发明对此不做限定。

303：数据输入节点中的发送模块为转换后的原始样本包分配任务标识，并将转换后的原始样本包作为预设格式的样本整体或分批次发送给该控制节点；

其中，分配任务标识是为了使系统正在运行的任务透明化，技术人员可以通过任务标识

获知当前系统正在运行的是哪些任务，并可以当需要终止某项任务时，控制节点可以根据任务标识向正在运行该任务的子任务的相似运算节点发送终止指令。

可选地，根据该预设格式的样本的任务标识判断该预设格式的样本所属任务是否完成，如果否，则将该预设格式的样本与该所属任务的其他样本汇总。

- 5 具体地，当原始样本的规模超过一定值，例如 1G 时，发送模块中的优化传输单元根据网络情况，将该转换后的原始样本包分拆成多个数据包；并由发送单元将该优化传输单元输出的该多个数据包作为预设格式的样本分批次发送给该控制节点，占用较少的内存和带宽资源。

需要说明的是，数据输入节点可以为控制节点的一部分，其转换格式的功能也可以由控制节点进行，当控制节点包含该功能时，数据输入节点负责收集邮件，并将邮件打包作为原始样本发送给控制节点，控制节点接收到原始样本后，扫描原始样本，将原始样本转换为预设格式的样本，进行步骤 305 的判断后，当该预设格式的样本不是相似计算最终结果时，统计预设格式的关键数据指标（包括数据包尺寸或记录条目等指标），根据样本的配置信息（包括每个包包括的记录条数或每个包的尺寸），根据关键数据指标进行排序，将排序后的排列拆分或合并成多个子任务数据包。上述的步骤是对原始样本的处理。

- 15 304：控制节点的接收模块接收预设格式的样本，该预设格式的样本包括转换后的原始样本包和由相似计算节点反馈的相似计算中间结果；

其中，控制节点用于接收预设格式的样本，并判断该预设格式的样本是否为相似计算最终结果，如果否，则根据预设标准对该预设格式的样本进行合并或拆分处理，得到多个子任务数据包，将该多个子任务数据包分配给多个相似运算节点；

- 20 后续步骤中出现的预设格式的样本根据其来源和经过的处理步骤的不同，可以分为经过数据输入节点转换的转换后的原始样本包和未经数据输入节点转换的预设格式的样本，而对于控制节点来说，控制节点接收到的数据都为预设格式，因此，后续不对转换后的原始样本包和预设格式的样本进行区别，而统称为预设格式的样本。

需要说明的是，在接收样本时，分 2 种情况：

- 25 1、所有样本一次性输入，任务的生命周期在本次输入数据的相似运算完成后达到结束点，相似关系只覆盖本次输入的样本；

2、样本分开多次传输，任务生命周期较长或无终止时间，需要输出的相似关系数据要覆盖所有输入数据，并且能够即输出已经传输完毕的样本部分之间的相似结果，无需等待所有样本全部传输完再启动相似计算过程；

- 30 需要说明的是，该控制节点是整套系统中的控制部分，该控制节点还用于处理来自数据输入节点的请求，在本实例中，该请求用于请求对预设格式的样本进行相似计算处理，为了

保障安全性，控制节点可以对该请求的合法性进行验证，当请求验证合法时，再对接收到的预设格式的样本进行处理。该控制节点一般为一台服务器，在热备情况下，可由两台或更多。

进一步地，该控制节点还用于保存并记录该预设格式的样本，记录该多个子任务数据包及该子任务数据包分配的相似运算节点的映射关系，并记录该相似运算节点的心跳信息。

5 305：控制节点的判断模块判断该预设格式的样本是否满足预设条件；

如果是，则该预设格式的样本是相似计算最终结果，将该预设格式的样本作为该相似计算最终结果输出；

如果否，则该预设格式的样本不是相似计算最终结果，并执行步骤 306；

其中，预设条件是指样本的相似计数达到预设阈值且此样本包已经过滤并剔除掉独立样本，独立样本是指未与其他任何样本有相似关系的；或经过相似计算后并未发现新的相似关系，例如，输入 1000 个样本，经过计算后没有可合并的样本，仍然为 1000 个样本。

其中该预设条件为技术人员根据系统的承载能力或其他要素设定的，本发明实施例不做具体限定。

15 在一个实施例中，当预设格式的样本为转换后的原始样本包时，该转换后的原始样本包内的记录条目之间的差异很大，无需进行相似计算，此时，该转换后的原始样本包即可以作为相似计算最终结果。

306：控制节点的合并拆分模块根据该相似运算节点的心跳信息，对该预设格式的样本进行合并或拆分处理，得到多个子任务数据包；

20 其中，该心跳信息用于监控和描述该相似运算节点的空闲计算能力，包括：其 CPU 或内存的配置情况和计算能力与当前正在运行的任务列表。心跳信息监控模块用于每隔预设时长或当接收到预设格式的样本时，获取该相似运算节点的心跳信息。具体地，心跳信息监控模块每隔预设时长（例如 1 分钟）向相似运算节点发送心跳信息请求或当控制节点接收到预设格式的样本时触发心跳信息监控模块向相似运算节点发送心跳信息请求，相似计算节点接收到心跳信息请求时，向控制节点反馈当前正在运行的子任务列表等信息。心跳信息监控模块
25 保存反馈的心跳信息，定期监控所有相似计算节点的状况，并监控正在运行的子任务的完成情况，包括正在运行、结束或异常失败等，用于在分派子任务数据包和相似计算节点崩溃时的查询处理。

需要说明的是，控制节点和所有的相似计算模块之间维持 TCP 长链接。

30 进一步地，本发明实施例中，所述预设格式的样本中的记录条目数或打成数据包后的总尺寸字节数超过预设阈值，对所述预设格式的样本进行拆分处理，具体地，当预设格式的样本必须满足如下几个方面中的任一条时，需对样本进行拆分处理：

- 1、样本已经按照数据关键指标排序；
- 2、记录条目数超过预设阈值，如 10 万；
- 3、打成数据包后的数据包尺寸超过预设阈值，如 1G；

进一步地，本发明实施例中，当样本必须满足如下几个方面中的任一条时，需对样本进

5 行合并处理：

1、样本在排序后，相似的记录条目只出现在此数据关键指标的某个连续范围内，或以较高概率出现；

2、根据数据关键指标在完成相似计算，经过唯一化样本步骤（即只保留一个样本，但记录合并掉的所有样本与此唯一样本之间的相似指数），保持不变；

10 3、一个任务标识在其生命周期内，存在多次和较慢的原始数据提交过程时，必定发生一部分已经先行计算相似的情况，或在数据量较大，一次需分发多个子任务数据包并接收对应的相似运算结果时，当所述预设格式的样本为至少经过一次相似计算的样本且本地服务器上存在至少两个所述预设格式的样本所属任务返回的预设格式的样本时，对所述至少两个所述预设格式的样本所属任务返回的预设格式的样本进行合并处理。

15 需要说明的是，合并运算处理到后期，会出现全部的唯一相似样本数量仍然庞大的情况，此时若仍然按照上面方法处理，会陷入一个分拆合并的死循环过程，当唯一相似样本数量超过预设阈值，为避免陷入死循环，根据不同的情况进行处理，具体如下：

1、丢弃相似计数较小的样本，例如，丢弃全部相似计数小于 5 的样本；

20 2、若经过一轮相似计算后，若某个子任务数据包中的样本之间均不存在相似关系，则标记此部分子任务数据已经达到了最终计算状态，不在参与后续的合并和分拆过程，直至这个任务标识有新的输入数据传入并排序在这个子任务数据包的数据范围内；

3、经过的计算次数越多，则丢弃的阈值应该逐步增大；

4、当全部子任务均达到最终状态或经历的运算次数达到一个阈值，则不再进行下一轮运算，标记此部分原始输入数据已经全部计算完成，本次相似计算任务完成。

25 307：控制节点的分配模块将该合并拆分模块得到的该多个子任务数据包分别分配各个相似运算节点；

本领域技术人员可以获知，在步骤 305 的分配时已经考虑到了各个相似计算节点的计算能力，所以各个相似计算节点接收到的数据包大小和包含条目可以不一致。

30 需要说明的是，如果当前相似运算节点无法处理所有的子任务数据包，可以先分配一部分，等待相似运算节点的心跳信息显示该相似运算节点空闲，再将后续的子任务数据包分配出去，一个相似计算节点上可以分配有一个或多个子任务数据包。

308: 相似计算节点接收一个或多个子任务数据包, 并对接收到的子任务数据包内的样本进行相似关系计算, 获得相似计算中间结果, 该相似计算中间结果为预设格式的样本, 将该预设格式的样本反馈给该控制节点, 执行步骤 304, 直到该样本所属任务完成。

进一步的, 当控制节点接收到预设格式的样本时, 根据其任务标识判断该样本所属任务中的子任务数据包是否都已经反馈, 如果是, 则该次任务结束, 如果否, 将该反馈的预设格式的样本和后续输入的样本再进行合并或拆分, 并再次分配给相似计算节点进行相似计算。

该相似计算中间结果至少包括唯一相似样本、相似关系和该唯一相似样本的相似计数, 还可以包括其他信息。相似关系是指样本之间的相似指数, 例如, 样本 A 与 B 之间不相似, 则其相似关系为 $\text{Sim}(A, B) = 0$ 。

10 在本实施例中, 相似计算节点只负责每个数据包内部条目的相似计算, 并将每个数据包的相似计算中间结果反馈给控制节点, 而不对数据包之间进行处理。且运算节点单元负责进行具体的相似计算任务, 除了数据的输入和输出外, 不对原始数据进行任何改变。

其中, 相似计算节点可以为不同 CPU 计算能力的服务器, 并可以使用一个或几个相似计算的核心算法;

15 优选地, 为了避免系统信息过于繁杂, 相似计算节点不会主动上报自己的心跳信息, 只在收到心跳信息请求后才返回必要的信息给控制节点。

优选地, 每个任务具有最长运行时间限制, 即如果运算时间超过指定秒数, 则该任务作废, 此时只有部分相似样本完成了相似运算, 根据子任务的配置信息来决定是否需要返回未完成的结果给控制节点。在子任务运行期间, 当接收到控制节点发出了终止指令, 则该运算立即停止并立即丢弃; 当子任务运算完毕, 由相似计算节点发请求给控制节点, 返回结果数据, 具备超时重试机制; 即当相似计算节点发送的请求在预设时长内未接收到控制节点的反馈时, 则重新发送, 当重新发送次数超过预设次数, 则认为控制节点崩溃。若发生相似计算节点崩溃, 相似计算节点内的数据和未完成的子任务不做恢复处理, 在相似计算节点恢复响应后, 等待新的运算请求;

25 下面给出一个简化后的实例来示意如何获得海量输入原始样本之间的完整相似关系:

原始样本中含有 ABCDEFGHI9 个样本, 根据数据关键指标排序后, 拆分成 3 个包, 分别为:

1 号包	A	B	C
2 号包	D	E	F
3 号包	G	H	I

经过第一轮派发和样本反馈后，得到如下结果：

数据包	相似关系	相似计数
1 号包	$S(B, A)=0.9$ $S(C, A)=0.7$	count (A)=3
2 号包	$S(E, D)=0.8$ $S(F, D)=1$	count (D)=3
3 号包	$S(H, G)=0.66$ $S(I, G)=1$	count (G)=3

3 个子任务均已经完成并返回结果，准备进行第二轮派发，由于数据量少，经过合并后不需要再次拆分：

4 号包	A	D	G
------	---	---	---

将这个数据包作为新的子任务派发后，得到下面结果：

4 号包	$S(D, A)=0.9$ G	count (A)=6 count (G)=3
------	--------------------	----------------------------

- 5 单独的 G 代表无任何样本与他相似。由于只有一个包，且运算完毕，本次请求已经全部处理好。此时，整理后的唯一相似样本和全部的相似关系如下：

样本列表	样本计数	相似关系
A	count (A)=6	$S(B, A)=0.9$
G	Count (G)=3	$S(C, A)=0.7$ $S(E, D)=0.8$ $S(F, D)=1$ $S(H, G)=0.66$ $S(I, G)=1$ $S(D, A)=0.9$

将这个结果记录在磁盘文件或数据库当中，可随时被查阅，整个处理过程结束。

- 在实际运行中，会出现相似计算节点崩溃的情况，当所述相似运算节点在预设时长内未返回心跳信息且连续未返回所述心跳信息超过预设次数，则标记所述相似运算节点崩溃，并
- 10 标记所述相似运算节点上运行的子任务数据包失败，并触发所述分配模块根据所述相似运算节点的心跳信息将标记失败的子任务数据包分配给未崩溃且空闲的相似运算节点。下面以一个例子进行说明：

本实施例中，该相似邮件处理系统包括一控制节点和 4 个相似计算节点，该 4 个相似计算节点分别为 Node1、Node2、Node3 和 Node4，正在运行的子任务数据包为 P1、P2、P3 和 P4，其相似计算节点上正在运行的子任务数据包可以见下表 1。

表 1

节点	Node1	Node2	Node3	Node4
任务	P1、P2	P3	P4	——

5

当前在控制节点向该 4 个相似计算节点发送心跳信息请求，获取的心跳信息见下表 2，

表 2

节点	Node1	Node2	Node3	Node4
状态	正在运行 P1、P2	——	P4 运行结束	空闲

其中，Node2 在预设时长内未反馈心跳信息，且在请求超过预设次数后，Node2 仍未反馈心跳信息，则认为 Node2 崩溃，从上一次正常的心跳信息表 3 中中查询 Node2 运行的任务，

10

表 3

节点	Node1	Node2	Node3	Node4
状态	正在运行 P1、P2	正在运行 P3	正在运行 P4	空闲

由表 3 可以获知，Node2 在崩溃时正在运行 P3，且由表 2 可以知道 Node4 空闲，且 Node3 已经运行完毕，Node4 和 Node3 中，Node3 的运算能力较强，而 P3 数据量较大，则将 P3 分配给 Node3 重新进行相似计算。

在实际运行中，还会出现控制节点崩溃的情况，正常情况下，控制节点会定时通过 LOG
15 保存一份子任务信息列表，通过跟重构的子任务列表对比，可以找到需要派发和崩溃时派发失败的那分子任务，从而能够恢复崩溃前的大致状态。这种情况包括控制节点崩溃，相似计算节点正常运作。此时，相似计算节点在短时间内回报的运算结果请求将全部超时，但由于有超时重试直到成功为止的机制，已经分派出去的子任务信息和数据均保持完整，当控制节点恢复服务后，相似计算节点回报请求会被正常接收和处理。另外，控制节点恢复启动后
20 立即通过心跳服务来收集此刻正在运行的子任务情况，结合控制节点的 LOG 数据可以重新构造子任务列表。需要注意，在极端情况下，这里存在丢失部分信息的可能性，丢失的信息可能为已经接受了相似计算请求，但还没来得及拆分或已经拆分但没来得及派发的那一部分。

通过由控制节点对输入的样本进行合并或拆分的处理，并将得到的多个子任务数据包分配给多个相似运算节点的分布式系统来实现对千万以上级别邮件的相似处理和计算，从而提

高了运算速度和运算能力，降低了系统负载，可以支持实时和准实时统计与拦截的反垃圾邮件需求。

本发明实施例提供的上述技术方案的全部或部分可以通过程序指令相关的硬件来完成，所述程序可以存储在可读取的存储介质中，该存储介质包括：ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

5

以上所述仅为本发明的较佳实施例，并不用以限制本发明，凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

权 利 要 求 书

1、一种相似邮件处理系统，其特征在于，包括：

5 控制节点，用于接收预设格式的样本，并判断所述预设格式的样本是否为相似计算最终结果，如果否，则根据预设标准对所述预设格式的样本进行合并或拆分处理，得到多个子任务数据包，将所述多个子任务数据包分配给多个相似运算节点；

多个所述相似运算节点，用于对接收到的子任务数据包内的样本进行相似关系计算，获得相似计算中间结果，所述相似计算中间结果为预设格式，将所述相似计算中间结果反馈给所述控制节点，所述相似计算中间结果至少包括：唯一相似样本、相似关系和所述唯一相似
10 样本的相似计数。

2、根据权利要求1所述的系统，其特征在于，所述系统还包括：

数据输入节点，用于收集原始样本并将所述原始样本并将所述原始样本转换为预设格式，并将转换后的原始样本包作为预设格式的样本发送给所述控制节点。

3、根据权利要求2所述的系统，其特征在于，所述数据输入节点包括：

15 数据收集模块，用于收集相似邮件处理系统服务器或服务器集群上的邮件，将所述邮件作为原始样本；

转换模块，用于将所述原始样本转换为与相似计算匹配的预设格式；

发送模块，用于为转换后的原始样本包分配任务标识，并将转换后的原始样本包作为预设格式的样本整体或分批次发送给所述控制节点。

20 4、根据权利要求3所述的系统，其特征在于，所述发送模块包括：

优化传输单元，用于根据网络情况，将所述转换后的原始样本包分拆成多个数据包；

发送单元，用于将所述优化传输单元输出的所述多个数据包作为预设格式的样本分批次发送给所述控制节点。

5、根据权利要求1所述的系统，其特征在于，所述控制节点包括：

25 接收模块，用于接收预设格式的样本；

判断模块，用于判断所述预设格式的样本是否满足预设条件，如果是，则所述预设格式的样本是相似计算最终结果，如果否，则所述预设格式的样本不是相似计算最终结果，并触发合并拆分模块；

所述合并拆分模块，用于根据所述相似运算节点的心跳信息，对所述预设格式的样本进
30 行合并或拆分处理，得到多个子任务数据包；所述心跳信息用于监控和描述所述相似运算节

点的空闲计算能力；

分配模块，用于将所述合并拆分模块得到的所述多个子任务数据包分别分配各个相似运算节点。

5 6、根据权利要求5所述的系统，其特征在于，所述合并拆分模块具体用于统计所述转换后的原始样本包和所述预设格式的样本的数据关键指标，并根据配置文件登记信息和所述数据关键指标对所述转换后的原始样本包和所述预设格式的样本进行排序，并根据排序顺序将所述转换后的原始样本包或所述预设格式的样本进行合并或拆分处理，得到多个子任务数据包。

7、根据权利要求5所述的系统，其特征在于，所述控制节点还包括：

10 心跳信息监控模块，用于每隔预设时长或当接收到预设格式的样本时，获取所述相似运算节点的心跳信息。

8、根据权利要求7所述的系统，其特征在于，所述控制节点还用于保存并记录所述预设格式的样本，记录所述多个子任务数据包及所述子任务数据包分配的相似运算节点的映射关系，并记录所述相似运算节点的心跳信息。

15 9、根据权利要求7所述的系统，其特征在于，所述心跳信息监控模块还用于当所述相似运算节点在预设时长内未返回心跳信息且连续未返回所述心跳信息超过预设次数，则标记所述相似运算节点崩溃，并标记所述相似运算节点上运行的子任务数据包失败，并触发所述分配模块根据所述相似运算节点的心跳信息将标记失败的子任务数据包分配给未崩溃且空闲的相似运算节点。

20 10、一种相似邮件处理方法，其特征在于，包括：

接收原始样本和预设格式的样本，并将接收到的原始样本转换为预设格式；

判断所述转换后的原始样本包和所述预设格式的样本是否为相似计算最终结果；

如果否，则根据预设标准对所述转换后的原始样本包和所述预设格式的样本进行合并或拆分处理，得到多个子任务数据包；

25 对每个所述子任务数据包内的样本进行相似关系计算，获得相似计算中间结果，所述相似计算中间结果为预设格式的样本，反馈所述预设格式的样本，所述相似计算中间结果至少包括：唯一相似样本、相似关系和所述唯一相似样本的相似计数。

11、根据权利要求10所述的方法，其特征在于，接收原始样本和预设格式的样本，具体包括：

30 收集相似邮件处理系统服务器或服务器集群上的邮件，将所述邮件作为原始样本，为所

述原始样本分配任务标识；

根据所述预设格式的样本的任务标识判断所述预设格式的样本所属任务是否完成，如果否，则将所述预设格式的样本与所述所属任务的其他样本汇总。

12、根据权利要求 10 所述的方法，其特征在于，判断转换后的原始样本包和所述预设格式
5 式的样本是否为相似计算最终结果，具体包括：

判断所述转换后的原始样本包是否满足预设条件，如果所述转换后的原始样本包满足预设条件，则所述转换后的原始样本包是相似计算最终结果，如果所述转换后的原始样本不满足预设条件，则所述转换后的原始样本包不是相似计算最终结果；

判断所述预设格式的样本是否满足预设条件，如果所述预设格式的样本满足预设条件，
10 则所述预设格式的样本是相似计算最终结果，如果所述预设格式的样本不满足预设条件，则所述预设格式的样本不是相似计算最终结果。

13、根据权利要求 10 所述的方法，其特征在于，根据预设标准对所述转换后的原始样本包和所述预设格式的样本进行合并或拆分处理，得到多个子任务数据包，具体包括：

统计所述转换后的原始样本包和所述预设格式的样本的数据关键指标，并根据配置文件
15 登记信息和所述数据关键指标对所述转换后的原始样本包和所述预设格式的样本进行排序，并根据排序顺序将所述转换后的原始样本包或所述预设格式的样本进行合并或拆分处理，得到多个子任务数据包。

14、根据权利要求 10 所述的方法，其特征在于，当所述预设格式的样本为至少经过一次相似计算的样本且本地服务器上存在至少两个所述预设格式的样本所属任务返回的预设格式的
20 的样本时，对所述至少两个所述预设格式的样本所属任务返回的预设格式的样本进行合并处理。

15、根据权利要求 10 所述的方法，其特征在于，所述预设标准至少包括下列任一项：

当所述转换后的原始样本包中的记录条目数或打成数据包后的总尺寸字节数超过预设阈值，对所述转换后的原始样本包进行拆分处理；

25 所述预设格式的样本中的记录条目数或打成数据包后的总尺寸字节数超过预设阈值，对所述预设格式的样本进行拆分处理。



图 1a



图 1b

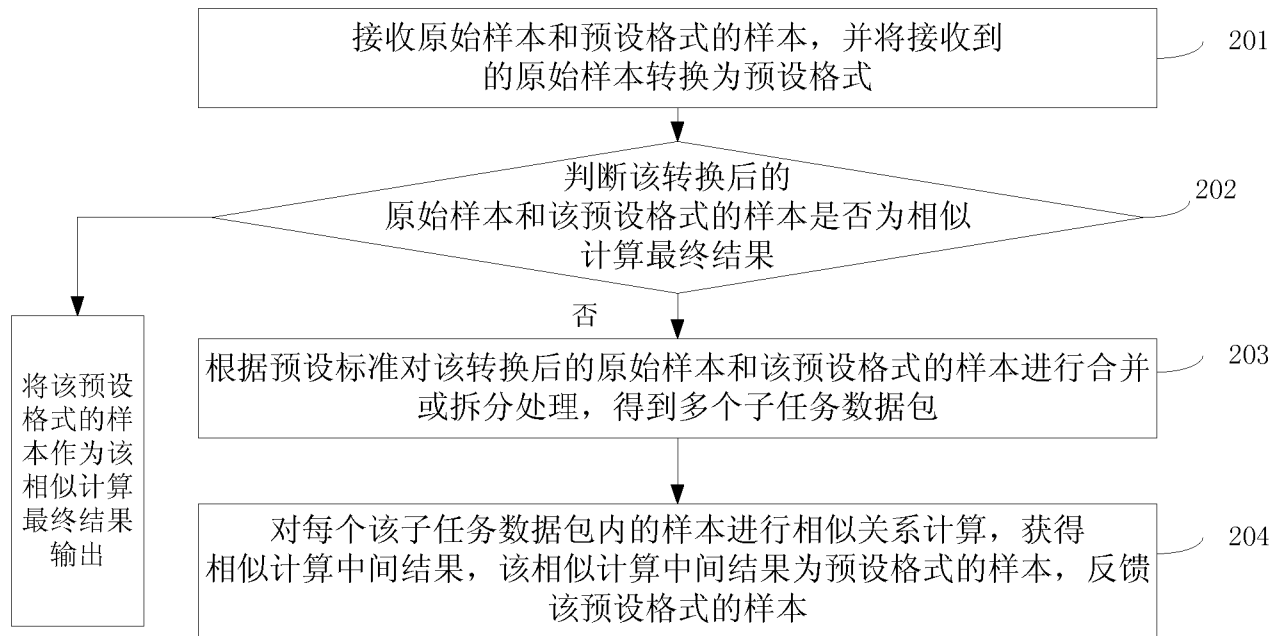


图 2

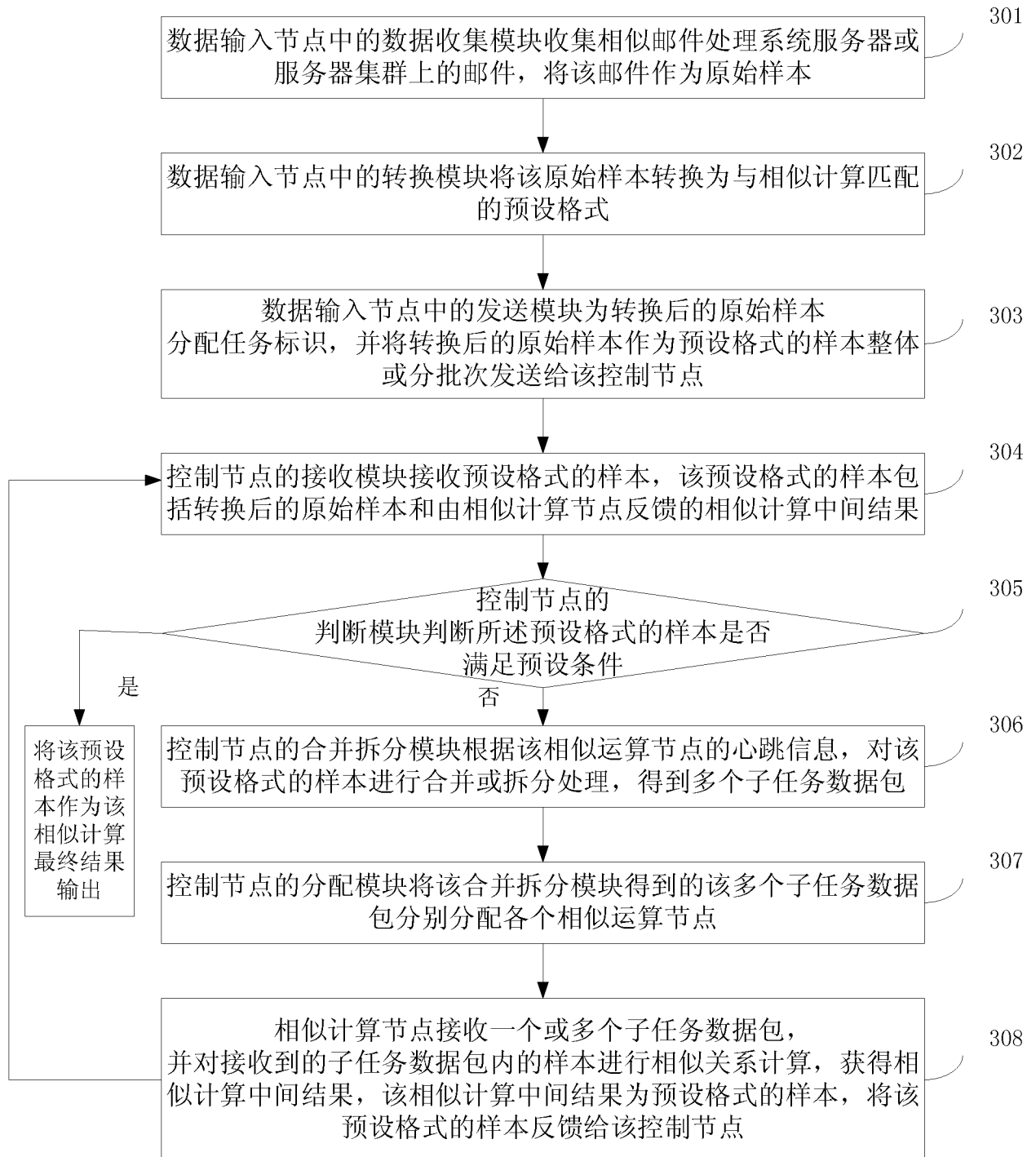


图 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2012/070816

A. CLASSIFICATION OF SUBJECT MATTER

H04L 12/58(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CPRSABS, CNTXT, CNKI, VEN: spam, junk, ?mail, similarity, combin+, split+, divid+, sampling, distributed, mass, million

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN1922837A (SYMANTEC CORP) 28 Feb. 2007 (28.02.2007) the whole document	1-15
A	CN101159704A (UNIV ZHEJIANG) 09 Apr. 2008 (09.04.2008) the whole document	1-15
A	US7590694B2 (GOZOOM COM INC) 15 Sept. 2009 (15.09.2009) the whole document	1-15

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
“A” document defining the general state of the art which is not considered to be of particular relevance	“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
“E” earlier application or patent but published on or after the international filing date	“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	“&” document member of the same patent family
“O” document referring to an oral disclosure, use, exhibition or other means	
“P” document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 17 Apr. 2012 (17.04.2012)	Date of mailing of the international search report 26 Apr. 2012(26.04.2012)
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10)62019451	Authorized officer YU, Ruifu Telephone No. (86-10) 62411248

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2012/070816

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN1922837A	28.02.2007	US7831667B2	09.11.2010
		US2011055343A1	03.03.2011
		WO2004105332A2	02.12.2004
		TW200527863A	16.08.2005
		JP4598774B2	15.12.2010
		JP2007503660A	22.02.2007
		EP1649645A2	26.04.2006
		US2005108339A1	19.05.2005
		US2005108340A1	19.05.2005
		WO2004105332A3	10.03.2005
		US2005132197A1	16.06.2005
CN101159704A	09.04.2008	NONE	
US7590694B2	15.09.2009	US2005160148A1	21.07.2005

国际检索报告

国际申请号

PCT/CN2012/070816

A. 主题的分类

H04L 12/58(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC: H04L

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CPRSABS, CNTXT, CNKI, VEN: 垃圾邮件、垃圾短信、垃圾信息、垃圾短消息、相似、合并、拆、分、样本
分布式、海量、千万; spam, junk, ?mail, similarity, combin+, split+, divid+, sampling, distributed, mass, million

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	CN1922837A (布赖特梅有限公司) 28.2 月 2007 (28.02.2007) 全文	1-15
A	CN101159704A (浙江大学) 09.4 月 2008 (09.04.2008) 全文	1-15
A	US7590694B2 (GOZOOM COM INC) 15.9 月 2009 (15.09.2009) 全文	1-15



其余文件在 C 栏的续页中列出。



见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇
引用文件的公布日而引用的或者因其他特殊理由而引
用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了
理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的
发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件
结合并且这种结合对于本领域技术人员为显而易见时,
要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

17.4 月 2012 (17.04.2012)

国际检索报告邮寄日期

26.4 月 2012 (26.04.2012)

ISA/CN 的名称和邮寄地址:

中华人民共和国国家知识产权局

中国北京市海淀区蓟门桥西土城路 6 号 100088

传真号: (86-10)62019451

受权官员

于瑞甫

电话号码: (86-10) 62411248

国际检索报告

关于同族专利的信息

国际申请号

PCT/CN2012/070816

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
CN1922837A	28.02.2007	US7831667B2	09.11.2010
		US2011055343A1	03.03.2011
		WO2004105332A2	02.12.2004
		TW200527863A	16.08.2005
		JP4598774B2	15.12.2010
		JP2007503660A	22.02.2007
		EP1649645A2	26.04.2006
		US2005108339A1	19.05.2005
		US2005108340A1	19.05.2005
		WO2004105332A3	10.03.2005
		US2005132197A1	16.06.2005
CN101159704A	09.04.2008	无	
US7590694B2	15.09.2009	US2005160148A1	21.07.2005