(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 102394833 B
(45)授权公告日 2015.06.17

(21)申请号 201110266356.6
(22)申请日 2011.08.30
(30)优先权数据
12/872,691 2010.08.31 US
(73)专利权人 微软公司
地址 美国华盛顿州
(72)发明人 H·C·安德森 A·帕纳修克
V·S·S·雷曼尼 B·库斯
(74)专利代理机构 上海专利商标事务所有限公司 31100
代理人 顾嘉运
(51)Int.Cl.
H04L 12/58(2006.01)
(56)对比文件
US 6421709 B1,2002.07.16,
CN 101098306 A,2008.01.02,
CN 1809821 A,2006.07.26,
CN 1774706 A,2006.05.17,
审查员 张攀

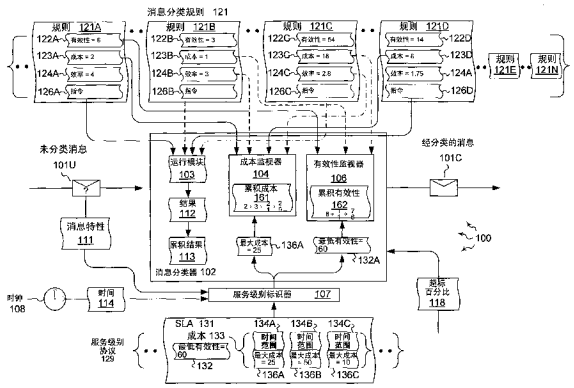
权利要求书3页 说明书12页 附图7页

(54)发明名称

自适应选择电子消息扫描规则

(57)摘要

本发明涉及用于自适应地选择电子消息扫描规则的方法、系统和计算机程序产品。本发明的实施例涉及动态（可能是不可预测地）改变分类电子消息的深度/完全性以保护抵御不期望的消息内容（例如SPAM、病毒、数字泄漏等）。维护最低有效性，并且当可用资源允许时可超过最低有效性来提供增加的保护。可以在每个消息的基础上选择可用消息分类规则的最优子集。所述规则的选择是基于可用的系统资源、最低期望有效性（例如在服务级别协议（“SLA”）中定义的）以及规则特性的。反馈循环可被用于优化所选择的分类规则子集。



CN 102394833 B

1. 在包括一个或更多处理器和系统存储器的计算机系统处，所述计算机系统包括多个电子消息分类规则 (221)，一种用于自适应地选择用于分类电子消息的规则的方法，所述方法包括：

接收一个或更多电子消息 (201U) 的动作；

对于所述一个或更多电子消息 (201U) 中的每一个，通过应用先前选择的电子消息分类规则的子集中的每个消息分类规则来计算指示该电子消息具有指定的消息特性的可能性的结果 (212) 的动作；

测量指示将每个电子消息分类规则应用于所述一个或更多电子消息中的每一个所消费的资源量的资源成本 (231) 的动作；

对于在先前选择的电子消息分类规则的子集中的每个消息分类规则，从经计算的结果 (212) 和经测量的资源成本 (231) 中为该消息分类规则合成效率度量 (232) 的动作；

将合成的效率度量与所述多个电子消息分类规则中包括的电子消息分类规则 (221) 的现有的效率度量 (224A) 进行比较的动作；

至少部分地基于合成的效率度量 (232) 与现有的效率度量 (224A) 的比较的结果来从所述多个电子消息分类规则中选择一个新的电子消息分类规则 (221) 的子集以便在对随后接收的电子消息进行分类时使用的动作。

2. 如权利要求 1 所述的方法，其特征在于，选择一个新的电子消息分类规则的子集的动作包括根据服务级别协议（“SLA”）来选择一个新的电子消息分类规则的子集的动作。

3. 如权利要求 1 所述的方法，其特征在于，选择一个新的电子消息分类规则的子集的动作包括基于效率分数重排序所述多个电子消息分类规则的动作。

4. 如权利要求 1 所述的方法，其特征在于，接收一个或更多电子消息包括接收一个或更多的电子邮件消息。

5. 如权利要求 1 所述的方法，其特征在于，接收一个或更多电子消息包括接收一个或更多的短消息服务（“SMS”）消息。

6. 如权利要求 1 所述的方法，其特征在于，接收一个或更多电子消息包括接收一个或更多的文件。

7. 如权利要求 1 所述的方法，其特征在于，所述多个电子消息分类规则被用于将电子消息分类成 SPAM 或合法的。

8. 如权利要求 1 所述的方法，其特征在于，所述多个电子消息分类规则被用于将电子消息分类成包含恶意软件或不包含恶意软件。

9. 如权利要求 1 所述的方法，其特征在于，所述多个电子消息分类规则被用于将电子消息分类成包含敏感数字信息或不包含敏感数字信息。

10. 在包括一个或更多处理器和系统存储器的计算机系统处，所述计算机系统包括多个电子消息分类规则 (121)，一种用于自适应地选择用于分类电子消息的规则的方法，所述方法包括：

接收一个或更多电子消息 (101U) 的动作；

对于所述一个或更多电子消息 (101U) 中的每一个：

将先前选择的电子消息分类规则的子集中的每个电子消息分类规则 (121A-121N) 应用于所述电子消息的动作，所述先前选择的电子消息分类规则的子集是所述多个电子消息

分类规则的子集；

对于所述先前选择的电子消息分类规则 (121A-121N) 的子集中的每个电子消息分类规则：

所述电子消息分类规则计算指示该电子消息具有指定消息特性的可能性的结果 (112) 的动作；

测量指示将所述电子消息分类规则应用于所述电子消息所消费的资源量的资源成本 (161) 的动作；

保留与将每个电子消息分类规则 (121A-121N) 应用于每个电子消息 (101U) 相关联的经计算的结果 (112) 和经测量的资源成本 (161) 的动作；

对于在先前选择的电子消息分类规则的子集中的每个消息分类规则，从所保留的经计算的结果 (112) 和经测量的资源成本 (161) 中为所述消息分类规则合成效率度量 (232) 的动作；

将合成的效率度量 (232) 与所述多个电子消息分类规则中包括的电子消息分类规则的现有的效率度量进行比较的动作；以及

至少部分地基于所述合成的效率度量与现有的效率度量的比较的结果从所述多个电子消息分类规则中选择一个新的电子消息分类规则 (121) 的子集以在对随后接收的电子消息进行分类时使用的动作。

11. 如权利要求 10 所述的方法，其特征在于，还包括在应用先前选择的电子消息分类规则的子集中的每个电子消息分类规则之前，基于经计算的效率分数来选择电子消息分类规则的所选子集的动作。

12. 如权利要求 10 所述的方法，其特征在于，还包括：

接收与将至少一个消息分类规则应用到电子消息有关的外部反馈的动作；以及

将所述外部反馈合并入来自将所述至少一个消息分类规则应用到所述电子消息的经计算的结果的动作。

13. 如权利要求 12 所述的方法，其特征在于，所接收的外部反馈指示来自将所述至少一个消息分类规则应用到所述电子消息的经计算的结果是下述项之一：假否定或假肯定。

14. 如权利要求 12 所述的方法，其特征在于，进一步包括：基于合并了所述外部反馈的经计算的结果来更新用于至少一个消息分类规则的有效性分数的动作。

15. 一种用于自适应地选择 SPAM 检测规则的系统，所述系统包括：

用于接收一个或更多的电子邮件消息 (201U) 的装置；

用于对于所述一个或更多的电子邮件消息 (201U) 中的每一个，将先前选择的 SPAM 检测规则 (221A-221N) 的子集中的每个 SPAM 检测规则应用于电子邮件消息 (201U) 的装置，所述先前选择的 SPAM 检测规则的子集是多个 SPAM 检测规则 (221) 的子集；以及

用于对于所述先前选择的 SPAM 检测规则的子集中的每个 SPAM 检测规则 (221A-221N)：

计算指示所述电子邮件消息是 SPAM 的可能性的结果 (212)；

测量指示将 SPAM 检测规则应用于所述一个或更多的电子邮件消息中的每一个所消费的资源量的资源成本 (231)；以及

从经计算的结果 (212) 和经测量的资源成本 (231) 中为所述 SPAM 检测规则合成效率

度量 (232) 的装置 ;

用于将合成的效率度量 (232) 与多个 SPAM 检测规则 (221A-221N) 中包括的 SPAM 检测规则的现有的效率度量 (224A) 进行比较的装置 ;以及

用于至少部分基于所述合成的效率度量 (232) 与现有的效率度量 (224A) 的比较的结果来选择在对随后接收的电子邮件消息进行分类时使用的新的 SPAM 检测规则 (221A-221N) 的子集的装置。

自适应选择电子消息扫描规则

技术领域

[0001] 本申请涉及自适应选择电子消息扫描规则的方法和系统。

[0002] 相关申请的交叉引用

[0003] 不适用。

背景技术

[0004] 1.背景和相关技术

[0005] 计算机系统和相关技术影响社会的许多方面。的确，计算机系统处理信息的能力已转变了人们生活和工作的方式。计算机系统现在通常执行在计算机系统出现以前手动执行的许多任务（例如，文字处理、日程安排和会计等）。最近，计算机系统彼此耦合并耦合到其他电子设备以形成计算机系统和其他电子设备可以在其上传输电子数据的有线和无线计算机网络。因此，许多计算任务的执行跨多个不同的计算机系统和/或多个不同的计算环境分布。

[0006] 在许多计算环境中，电子消息，诸如例如电子邮件消息，被用于在计算机系统用户之间合法地交换信息。然而，这些计算环境也使得用户遭受未经请求的和/或不想要的电子消息，通常称为SPAM。已经开发了许多不同的技术用于扫描并阻塞SPAM。

[0007] SPAM扫描技术必须典型地协商一组度量，包括：有效性(effectiveness)、准确性、效率(efficiency)和等待时间。有效性涉及哪些范围中的SPAM可以被标识并堵塞。准确性涉及哪些范围的合法消息被不正确地标识为SPAM(例如假肯定率)。效率涉及与将消息标识为SPAM或合法的相关联的资源消耗。等待时间涉及由于扫描导致的每一个单独消息在传输上被延迟了多少时间。

[0008] 这些度量之间的平衡可以是相对复杂的任务，因为，一个领域中的改进通常意味着一个或更多其他领域中的降级。例如，更积极的反SPAM检测（增加的有效性）会导致更高的假肯定（降低的准确性），和/或由于更加复杂的处理算法而引起的更高的CPU负载（增加的资源消耗）。

[0009] 另外，这些度量的某组合经常被映射为服务提供者支持的服务级协议（“SLAs”）。例如，反SPAM服务提供者可以同意支持有效性不低于X、准确性不低于Y等。损害SLA的条款，例如在某个时间量具有低于X的有效性，会使反SPAM服务提供者遭受向消费者的某种退款。

[0010] 然而，同时，反SPAM服务典型地受到系统负载的高可变性的影响。例如，在任意给定日子、在周末或每个季度地全部时间内，SPAM的量和/或合法电子消息的量总会波动。不幸地是，这会导致服务提供者过度预配。例如，常见的设计模式是要构建具有足够能力来保证高峰负载时间的SLA的扫描服务，所述高峰负载可以比平均负载高出三到五倍。

[0011] 实际上，针对高峰负载的设计导致资源在大量时间中（可能严重地）没有充分被利用。扫描典型地包括固定数目的阶段和/或使用固定数目的扫描规则，而（如果有）也是有限考虑到可用资源。这样，在非高峰时间，固定数目的阶段和/或规则被用于扫描消息，

即使是有资源可用于进一步的扫描。这样，针对高峰负载的设计从产品销售成本的观点来看是不期望的，尽管如此，为了 SLA 还是需要这样的设计。

[0012] 当支持通常各自具有在他们的 SLA 中定义的不同度量的各种不同的服务级别时，例如，常规的消费者、高级消费者、低成本消费者等，会出现进一步复杂化。通常，高级服务供应伴随着保证更高服务级别（例如增加的准确度、更少的等待时间等）的 SLA，这要求服务提供者部分更多的计算 / 处理器资源。

[0013] 对我们来说，一种用于处理不同服务级别的设计模式是对所有服务级别通用的一种反 SPAM 服务。基于所要求的 SLA，每个服务级别被限于多个规则和 / 或消息经过的处理阶段。例如，高级消费者的电子邮件可以经历 10 个处理阶段，而基本消费者的电子邮件可仅经历 5 个处理阶段。以较低的扫描质量（例如降低的有效性）为代价，降低了服务基本消费者的成本，即使是在有资源可用于进一步扫描的情况下。除了较低的扫描质量之外，通过利用在为基本消费者提供的保护级别中的弱点，基本消费者也更易于受到针对性的攻击（系统的可预言性）。

[0014] 另一常见的模式是建立两个单独的系统，一个用于高级消费者，而另一个用于基本消费者。基于消费者类型，将每个系统设计为在服务质量和服务成本之间进行平衡。不幸地是，这种类型的系统需要加倍的基础结构，导致更高的整体成本以及为满足高峰负载时的 SLA 的常见的过度预配问题。

发明内容

[0015] 本发明涉及用于自适应地选择电子消息扫描规则的方法、系统和计算机程序产品。在一些实施例中，自适应地选择用于分类电子消息的规则。接收一个或更多电子消息。对于所述一个或更多电子消息中的每一个，将在先前选择的电子消息分类规则的子集中的每个消息分类规则应用于电子消息。先前选择的电子消息分类规则的子集是多个可用的电子消息分类规则的一个子集。

[0016] 对于先前选择的电子消息分类规则的子集中的每个电子消息分类规则，计算指示电子消息具有指定的消息特性的可能性的结果。计算资源成本，所述资源成本指示被消耗用于将所述电子消息分类规则应用到所述电子消息的资源量。保留与将每个电子邮件分类规则应用到每个电子消息相关联的经计算的结果和测量的资源成本。

[0017] 对于在先前选择的电子消息分类规则的子集中的每个消息分类规则，从所保留的经计算的结果和测量的资源成本中为该消息分类规则合成效率度量。将合成的效率度量与被包括在所述多个可用的电子消息分类规则中的电子消息分类规则的现有的效率度量进行比较。至少部分地基于合成的效率度量与现有的效率度量的比较结果从所述多个可用的电子消息分类规则中选择一个新的电子消息分类规则的子集。新的电子消息分类规则的子集用于对随后接收的电子消息进行分类。因此，可以轮换地使用消息分类规则以适应于不断变化的消息内容模式。

[0018] 提供本发明内容是为了以简化的形式介绍将在以下具体实施方式中进一步描述的一些概念。本概述不旨在标识出所要求保护的主题的关键特征或必要特征，也不旨在用于帮助确定所要求保护的主题的范围。

[0019] 本发明的附加特征和优点将在以下描述中叙述，且其一部分根据本说明书将是显

而易见的，或可通过对本发明的实践来获知。本发明的特征和优点可通过在所附权利要求书中特别指出的工具和组合来实现和获得。本发明的这些和其他特征将通过以下描述和所附权利要求书变得更加显而易见，或可通过对下文中所述的本发明的实践来领会。

附图说明

[0020] 为了描述可获得本发明的上述和其它优点和特征的方式，将通过参考附图中示出的本发明的具体实施例来呈现以上简要描述的本发明的更具体描述。可以理解，这些附图只描绘了本发明的各典型实施例，并且因此不被认为是对其范围的限制，将通过使用附图并利用附加特征和细节来描述和解释本发明，在附图中：

[0021] 图 1 示出了便于自适应地分类电子消息的示例计算机体系结构。

[0022] 图 2 示出了便于自适应地选择用于分类电子消息的规则的示例计算机体系结构。

[0023] 图 3 示出了用于自适应地分类电子消息的示例方法的流程图。

[0024] 图 4 示出了用于自适应地选择用于分类电子消息的规则的示例方法的流程图。

[0025] 图 5 示出了便于自适应的电子消息扫描和自适应地选择用于分类电子消息的规则的另一示例计算机体系结构。

具体实施方式

[0026] 本发明涉及用于自适应地选择电子消息扫描规则的方法、系统和计算机程序产品。在一些实施例中，自适应地选择用于分类电子消息的规则。接收一个或更多电子消息。对于所述一个或更多电子消息中的每一个，将在先前选择的电子邮件分类规则的子集中的每个消息分类规则应用于电子消息。先前选择的电子消息分类规则的子集是多个可用的电子消息分类规则的一个子集。

[0027] 对于先前选择的电子消息分类规则的子集中的每个电子消息分类规则，计算指示电子消息具有指定的消息特性的可能性的结果。计算资源成本，所述资源成本指示被消耗用于将所述电子消息分类规则应用到所述电子消息的资源量。保留与将每个电子邮件分类规则应用到每个电子消息相关联的经计算的结果和测量的资源成本。

[0028] 对于在先前选择的电子消息分类规则的子集中的每个消息分类规则，从所保留的经计算的结果和测量的资源成本中为该消息分类规则合成效率度量。将合成的效率度量与被包括在所述多个可用的电子消息分类规则中的电子消息分类规则的现有的效率度量进行比较。至少部分地基于合成的效率度量与现有的效率度量的比较结果来从所述多个可用的电子消息分类规则中选择一个新的电子消息分类规则的子集。新的电子消息分类规则的子集用于对随后接收的电子消息进行分类。因此，可以将消息分类规则引入所述服务或从所述服务取出以适应于不断变化的消息内容模式。

[0029] 在其它实施例中，自适应地分类电子消息。在指定时间接收从发送者发送到接收者的电子消息。基于发送者和接收者中的一个或更多来标识应用到所接收的电子消息的服务级别。

[0030] 服务级别定义了用于扫描电子消息的至少一个最低有效性值和一组最大成本值。最低有效性值表示消息分类规则的组合要具有的最低累积总有效性以满足所述服务级别。该组最大成本值中的每个最大成本值对应于不同的所分配的时间周期并且表示可用于将

消息分类规则应用于电子消息的资源总量。从该组最大成本值中选择一个最大成本值以供在基于用于所选择的最大成本值的分配的时间周期内的指定时间扫描所接收的电子消息时使用。

[0031] 将一个或更多消息分类规则应用于所接收的电子消息。每个消息分类规则具有经测量的有效性、经测量的资源成本以及考虑到经测量的资源成本基于所测量的有效性的经计算的效率。经测量的有效性表示适当地将电子消息标识为具有指定的消息特性的概率。按照效率的顺序应用一个或更多的消息分类规则直到达到服务级别中定义的最低累积总有效性。

[0032] 将每个消息分类规则应用到电子消息以生成指示该电子消息具有指定的消息特性的可能性的结果。将所应用的消息分类规则的经测量的资源成本加入所消费的资源的累积量中。所消费的资源的累积量是通过对来自所述一个或更多消息分类规则中先前应用的消息分类规则的经测量的资源成本进行求和来计算的。

[0033] 确定所消费的资源的累积量是否小于所选择的最大成本值。基于所述确定，将另外的消息分类规则应用到电子消息。当所消费的资源量小于所选择的最大成本值时，将更多的电子消息应用到所接收的电子消息中以得到高于该服务级别中定义的有效性。当所消费的资源量至少等于所选择的最大成本值时，将电子消息规则应用到另一不同的电子消息。

[0034] 本发明的各实施例可包括或利用包括诸如例如，一个或多个处理器和系统存储器之类的计算机硬件的专用或通用计算机，如以下更详细讨论的。本发明范围内的各实施例还包括用于承载或存储计算机可执行指令和 / 或数据结构的物理和其他计算机可读介质。这些计算机可读介质可以是可由通用或专用计算机系统访问的任何可用介质。存储计算机可执行指令的计算机可读介质是物理存储介质。承载计算机可执行指令的计算机可读介质是传输介质。由此，作为示例而非限制，本发明的各实施例可包括至少两种完全不同的计算机可读介质 ：计算机存储介质（设备）和传输介质。

[0035] 计算机存储介质（设备）包括 RAM、ROM、EEPROM、CD-ROM 或其他光盘存储、磁盘存储或其他磁存储设备、或可用于存储计算机可执行指令或数据结构形式的所需程序代码装置的且可由通用或专用计算机访问的任何其他介质。

[0036] “网络”被定义为允许在计算机系统和 / 或模块和 / 或其他电子设备之间传输电子数据的一个或多个数据链路。当信息通过网络或另一个通信连接（硬连线、无线、或者硬连线或无线的组合）传输或提供给计算机时，该计算机将该连接适当地视为传输介质。传输介质可包括可用于承载计算机可执行指令或数据结构形式的所需程序代码装置且可由通用或专用计算机访问的网络和 / 或数据链路。上述的组合也应被包括在计算机可读介质的范围内。

[0037] 此外，在到达各种计算机系统组件之后，计算机可执行指令或数据结构形式的程序代码装置可从传输介质自动传输到计算机存储介质（设备）（或反之亦然）。例如，通过网络或数据链路接收到的计算机可执行指令或数据结构可被缓存在网络接口模块（例如，“NIC”）内的 RAM 中，然后最终被传输到计算机系统 RAM 和 / 或计算机系统处的较不易失性的计算机存储介质。因而，应当理解，计算机存储介质（设备）可被包括在还利用（甚至主要利用）传输介质的计算机系统组件中。

[0038] 计算机可执行指令例如包括，当在处理器处执行时使通用计算机、专用计算机、或专用处理设备执行某一功能或某组功能的指令和数据。计算机可执行指令可以是例如二进制代码、诸如汇编语言之类的中间格式指令、或甚至源代码。尽管用结构特征和/或方法动作专用的语言描述了本主题，但可以理解，所附权利要求书中定义的主题不必限于上述特征或动作。相反，上述特征和动作是作为实现权利要求的示例形式而公开的。

[0039] 本领域的技术人员将理解，本发明可以在具有许多类型的计算机系统配置的网络计算环境中实践，这些计算机系统配置包括个人计算机、台式计算机、膝上型计算机、消息处理器、手持式设备、多处理器系统、基于微处理器的或可编程消费电子设备、网络 PC、小型计算机、大型计算机、移动电话、PDA、寻呼机、路由器、交换机等等。本发明也可在其中通过网络链接（或者通过硬连线数据链路、无线数据链路，或者通过硬连线和无线数据链路的组合）的本地和远程计算机系统两者都执行任务的分布式系统环境中实施。在分布式系统环境中，程序模块可位于本地和远程存储器存储设备中。

[0040] 通常，本发明的实施例涉及动态（可能是不可预测地）改变分类电子消息的深度/完全性以保护抵御不期望的消息内容（例如 SPAM、病毒、数字泄漏等）。维护最低有效性，并且当可用资源允许时可超过最低有效性来提供增加的保护。可以在每个消息的基础上选择可用消息分类规则的最优子集。所述规则的选择是基于可用的系统资源、最低期望有效性（例如在服务级别协议（“SLA”）中定义的）以及规则特性。反馈循环可被用于优化分类规则子集。

[0041] 这样，在说明书和下述权利要求中，“消息分类”包括基于消息（或文件）特性将电子消息（例如电子邮件消息、短消息服务（“SMS”）消息、文件等）分类成不同的“类”，所述消息（或文件）特性诸如例如内容、消息大小、附件、商业对消费者领域、起源的区域、发送者、接收者、时间、日期等。

[0042] 在一些实施例中，对电子消息分类以确定（例如根据 SLA）对应于该电子消息的服务级别。服务级别定义了消息分类规则到所述电子消息的进一步应用。服务级别可以定义什么类型的服务分类规则以及多少其它服务分类规则要被应用于该电子消息。例如，在一个国家非常有效的用于分类消息的分类规则可能在另一国家对分类消息是低效的而反之亦然。

[0043] 在一些实施例中，进一步的分类涉及确定电子消息是否是不需要的和/或未经请求的电子消息（例如 SPAM），电子消息是否包含恶意软件或以其他方式被感染和/或是危险的（例如病毒、间谍软件、特洛伊木马等），在电子消息中是否泄漏了敏感信息等。例如，数字泄漏阻止（“DLP”）系统可以使用规则来确定电子消息是否包括敏感信息。

[0044] 图 1 示出了便于自适应地分类电子消息的示例计算机体系结构 100。参照图 1，计算机体系结构 100 包括消息分类器 102、服务级别标识器 107、时钟 108、超控百分比 118、消息分类规则 121 以及服务级别协议 131。所描绘的组件中的每一个都通过诸如例如局域网（“LAN”）、广域网（“WAN”）或甚至因特网之类的网络彼此连接（或者是网络的一部分）。因此，所描绘的组件中的每一个以及任何其他连接的计算机系统及其组件都可以创建消息相关数据并通过网络交换消息相关数据（例如，网际协议（“IP”）数据报和利用 IP 数据报的其他更高层协议，诸如传输控制协议（“TCP”）、超文本传输协议（“HTTP”）、简单邮件传输协议（“SMTP”）等）。

[0045] 规则121包含多个消息分类规则，诸如例如规则121到规则121N，它们可被用于分类电子消息。每个规则可以指示有效性、成本、效率，并且可以包括指令。有效性指示规则有多少可能基于使用的扫描类型准确地将消息标识为某种方式不期望的。例如，用于检测SPAM的规则的有效性可以指示该规则有多少可能检测到SPAM而没有假肯定。成本指示当运行模块运行该规则的指令时消费的（例如估计的）系统资源量。效率指示在基于有效性并参考资源消耗的情况下该规则有多高效。在一些实施例中，效率是有效性除以成本的商。执行指令以生成与分类电子消息有关的结果（例如以确定电子消息是否是SPAM，是否包含恶意软件，是否包含敏感信息等）。

[0046] 通常，配置消息分类器102以基于电子消息特性来分类电子消息。如所述，消息分类器102包括运行模块103、成本监视器104以及有效性监视器106。配置运行模块103以运行包括在所接收规则中的指令（例如脚本或其它可执行代码）。指令产生可以用作分类电子消息的数据点的单独的结果。例如，单独的结果可以指示电子消息是否是不想要的和/或未经请求的电子消息（例如SPAM），是否是受感染的或危险的，是否包含敏感信息等。运行模块103可以从运行许多不同的规则中累积单独的结果。消息分类器102随后可以使用所累积的单独的结果来分类消息。

[0047] 配置成本监视器104以跟踪与扫描电子消息相关联的正在进行的资源成本。当规则运行时，成本监视器104维护针对电子消息运行的任意规则的总资源成本。在一些实施例中，当每个规则运行时，将该规则的成本加入任意先前运行规则的资源成本。

[0048] 配置有效性监视器106以跟踪正在进行的扫描电子消息的有效性。当规则运行时，有效性监视器106维护针对电子消息运行的任意规则的总有效性。在一些实施例中，当每个规则运行时，将该规则的有效性加入任意先前运行规则的有效性。

[0049] 服务级别协议129包含多个SLA，包括SLA 131。每个SLA包括最低有效性和一个或更多成本。每个成本可应用于指定的日期/时间范围。最低有效性表示当扫描电子消息时（即使在资源消费超标时）将要达到的累积有效性（即多个分类规则的有效性总和）。表1的表是基于消费者类型的每个SLA的有效性的示例

[0050]

消费者类型	最低有效性
基本消费者	75
高级消费者	100

[0051] 表1

[0052] 表1指示对于基本消费者的最低有效性（即来自多个分类规则的应用的累积有效性）是75，而对于高级消费者的最低有效性是100。在SLA中分配最低有效性时，其它因素也可被考虑。

[0053] 所述一个或更多成本各自包括时间范围和最大成本。每个时间范围/最大成本对表示当在该时间范围内接收消息时将被考虑用于将规则应用到该消息的最大资源成本。对于不同的服务级别，时间范围/最大成本对可以不同或相同。在一些实施例中，可以在公共的可访问表中分配时间范围/最大成本对，以便所述时间范围/最大成本对对于许多SLA

都是相同的。在另外的实施例中，可以诸如例如通过将时间范围 / 最大成本对包括在 SLA 中以在每个 SLA 的基础上分配时间范围 / 最大成本对。表 2 是时间范围 / 最大成本对的示例

[0054]

一天中的时间	最大成本
高峰小时	50
普通小时	75
非高峰小时	100

[0055] 表 2

[0056] 表 2 指示在高峰小时、普通小时和非高峰小时期间应用分类规则的最大资源成本为 50、75 和 100。也可考虑其他因素。

[0057] 最大成本可以随时间改变。如果消息分类服务器将另外的硬件加入并因此具有更好的计算能力，则最大成本数字可上升。另一方面，如果服务加入了另外的消费者或者如果负载突然增加，则最大成本数字会下降。

[0058] 在一些实施例中，最低有效性相对于最大成本更加重要地被考虑。在这些实施例中，超过最大成本的资源可以被消费以确保达到最低有效性。如果使用比最大成本少的资源达到最低有效性，则可以应用另外的分类规则以增加有效性直到达到或超过了最大成本。

[0059] 配置服务级别标识器 107 以标识对应于所接收的电子消息的服务级别。基于消息特性和时间 / 日期，服务级别标识器 107 可以从服务级别协议 131 中标识合适的 SLA。时钟 108 可以维护日期和一天中的时间，并在接收到电子消息时将该信息发送给服务级别标识器 107。服务级别标识器可以将该消息的最低有效性和最大成本发送给消息分类器 102。对于每个消息分类规则，消息分类器 102 可以将累积有效性与最高有效性进行比较并将累积的成本与最大成本进行比较以确定要将哪些分类规则和多少分类规则应用于所接收的消息。

[0060] 超标百分比 118 定义了即使已经达到最低有效性并且已经达到或超过最大成本时要应用于电子消息的另外的分类规则的某个百分比。超标百分比 118 允许原本要被跳过（例如由于它们的效率）的分类规则有时被执行。在一些实施例中，超标百分比 118 指示在消息分类规则 121 中将要被应用于电子消息的每个规则的百分比。

[0061] 图 3 示出了用于自适应地分类电子消息的示例方法 300 的流程图。方法 300 将参考计算机架构 100 的组件和数据来描述。

[0062] 方法 300 包括在指定时间接收电子消息的动作，该电子消息从发送者发送到接收者（动作 301）。例如，消息分类器 102 可以在（由时钟 108 所指示的）时间 114 接收消息 101U。消息 101U 可以包括包含了发送者地址和接收者地址的消息特性 111。

[0063] 方法 300 包括基于下述一个或更多项来标识可应用于所接收的电子消息的服务级别的动作：发送者和接收者、定义至少一个最低有效性值和一组最大成本值的服务级别，最低有效性值表示消息分类规则的组合要具有的以满足所述服务级别的最低累积总有效

性，该组最大成本值中的每个最大成本值对应于不同的分配的时间周期，每个最大成本值表示可用于将消息分类规则应用于电子消息的资源总量（动作 302）。例如，服务级别标识器 107 可以接收消息特性 111 和时间 114。基于消息特性 111，（例如发送者和 / 或接收者地址）服务级别标识器 107 可以将 SLA 131 标识为可用于分类消息 101U。

[0064] 如所述，SLA 131 定义了最低有效性 132 和成本 133。成本 133 包括时间范围 / 最大成本对，包括时间范围 134A/ 最大成本 136A、时间范围 134B/ 最大成本 136B、时间范围 134C/ 最大成本 136C 等。

[0065] 方法 300 包括从该组最大成本值中选择一个最大成本值以供在基于在所选择的最大成本值的所分配的时间周期内的指定时间扫描所接收的电子消息时使用的动作（动作 303）。例如，服务级别标识器 107 可以确定时间 114 在时间范围 134A 内。作为响应，服务级别标识器 107 可以选择最大成本 136A 以供在扫描未分类的消息 101U 时使用。

[0066] 服务级别标识器 107 可以将最低有效性 132 和最大成本 136A 发送给消息分类器 102。消息分类器 102 可以使用最低有效性 132 和最大成本 136A 来确定何时要停止将消息分类规则应用于未分类的消息 101U。

[0067] 方法 300 包括将一个或更多消息分类规则应用于所接收的电子消息的动作，每个消息分类规则具有经测量的有效性、经测量的资源成本以及考虑到经测量的资源成本基于经测量的效果的经计算的效率，经测量的有效性表示适当地将电子消息标识为具有指定的消息特性的概率，按照效率的顺序应用一个或更多的消息分类规则直到达到服务级别中定义的最低累积总有效性。例如，消息分类器可以按照效率的顺序应用规则 121 中的规则直到达到最低有效性 132（即 60）。

[0068] 对于所描述的规则，情况可能是效率 124A（即 4）对于规则 121 中规则来说是最高的。这样，规则 121A 是第一个应用于未分类消息 101U 的规则。在应用规则 121A 之后，累积的有效性 162 是 8，等于有效性 122A。消息分类器 102 确定 8 小于 60，因此，另外的分类规则被应用以达到最低有效性 132A。

[0069] 情况可能是效率 124B（即 3）对于规则 121 中规则来说是次最高的。这样，规则 121B 是下一个应用于未分类消息 101U 的规则。在应用规则 121B 之后，累积的有效性 162 是 11，等于有效性 122A 加上有效性 122B 的总和。消息分类器 102 确定 11 小于 60，因此，另外的分类规则将被应用以达到最低有效性 132A。

[0070] 情况可能是效率 124C（即 2.8）对于规则 121 中规则来说是最高的。这样，规则 121C 是下一个应用于未分类消息 101U 的规则。在应用规则 121C 之后，累积的有效性 162 是 66，等于有效性 122A 加上有效性 122B 加上有效性 122C 的总和。消息分类器 102 确定 66 大于 60，因此不需要应用另外的分类规则以满足 SLA131。

[0071] 对于被应用的一个或更多消息分类规则中的每个，方法 300 包括将消息分类规则应用于电子消息以生成指示该电子消息具有指定的消息特性的可能性的动作（动作 305）。例如，运行模块 103 可以对未分类消息 101U 执行指令 126A 以生成结果 112。结果 112 指示未分类消息 101U 是不想要的和 / 或未经请求的电子消息、是受感染的或危险的消息、包含敏感信息等的可能性。运行模块 103 可以将结果 112 存储在累积结果 113 中。指令 126B 和 126C 可对未分类消息 101U 相似地执行以生成结果。这些结果也可以被存储在累积结果 113 中。

[0072] 对于所应用的一个或更多消息分类规则中的每个，方法 300 包括将所应用的消息分类规则的经测量的资源成本加入到所消费的资源的累积量中，所消费的资源的累积量是通过对从一个或更多消息分类规则中先前应用的消息分类规则的经测量的资源成本进行求和来计算的。例如，在应用规则 121A、121B 和 121C 之后，累积成本 161 是 21，等于成本 123A 加上成本 123B 和成本 123C。

[0073] 方法 300 包括确定所消费的资源的累积量是否少于所选的最大成本值的动作（动作 307）。例如，成本监视器 104 可以确定累积成本 104 是否少于最大成本 136A。方法 300 包括基于所述确定，将另外的消息分类规则应用于电子消息的动作（动作 308）。例如，消息分类器 102 可以基于累积的成本 161 是否少于最大成本 136A 来将另外的消息分类规则应用于电子消息。

[0074] 如在计算机体系结构 100 中所述，在达到最低有效性 132A 之后，累积成本 161（即 21）少于最大成本 136A（即 25）。这样，可以将另外的分类规则应用于未分类的消息 101U 以增加对未分类的消息 101U 进行分类的有效性。

[0075] 例如，情况可能是效率 124D（即 1.75）对于规则 121 中规则来说是次最高的。这样，规则 121D 是下一个应用于未分类消息 101U 的规则。这样，运行模型 103 可以针对未分类消息 101U 执行指令 126D 以生成结果并将结果存储在累积结果 113 中。在应用规则 121D 之后，累积成本 161 转变为 29，等于成本 123A 加上成本 123B 加上成本 123C 加上成本 123D。（有效性 122D 实质上被忽略，因为已经达到了最低有效性 132A）。因为累积成本 161（即 29）超过了最大成本 136A（即 25），所以不将另外的规则应用于未分类的消息 101U。

[0076] 或者，如果在应用规则 121C 之后达到了最低有效性 132A，累积成本 161 已经等于或大于最大成本 136A（即 25），则不将另外的规则应用于未分类的消息 101U 然而，最低有效性还是达到了。

[0077] 当达到最低有效性 132A 并且达到或超过了最大成本 136A 时，消息分类器 102 可以参考超标百分比 118。消息分类器 102 可以使用超标百分比 118 来确定是否要将另外的分类规则应用于未分类的消息 101U。如果是，消息分类器 102 将来自消息分类规则 121 中的一个或更多（或所有剩余的）规则，诸如例如规则 121E，应用于未分类的消息 101U。超标百分比 118 的使用允许原本未使用或限制使用的消息分类规则的性能（例如有效性和成本）被评估并适当地被改变。基于这些改变，可以增加这些分类规则的使用频率。例如，评估较旧规则的性能将揭示该较旧规则由于不断变化的 SPAM 模式现在更加有效。

[0078] 当不将另外的规则应用于未分类的消息 101U 时，消息分类器 102 可以使用累积结果 113 来分类未分类的消息 101U。例如，从累积结果 113 中，消息分类器 102 可以将未分类的消息 101U 分类为合法消息或不想要和 / 或未经请求的消息（例如 SPAM），分类为包括或不包括恶意软件，分类为包括或不包括敏感信息等。信息分类器 102 可以将经分类的消息 101C 输出以指示所述分类。

[0079] 一旦输出经分类的消息 101C，消息分类器 102 可以转变为分类下一电子消息。

[0080] 图 2 示出了便于自适应地选择用于分类电子消息的规则的示例计算机体系结构 200。参照图 2，计算机体系结构 200 包括消息分类器 202、消息分类规则 221 和规则选择和重排序模块 216。所描绘的组件中的每一个都通过诸如例如局域网（“LAN”）、广域网（“WAN”）或甚至因特网之类的网络彼此连接（或者是网络的一部分）。因此，所描绘的

组件中的每一个以及任何其他连接的计算机系统及其组件都可以创建消息相关数据并通过网络交换与消息相关数据（例如，网际协议（“IP”）数据报和利用 IP 数据报的其他更高层协议，诸如传输控制协议（“TCP”）、超文本传输协议（“HTTP”）、简单邮件传输协议（“SMTP”）等）。

[0081] 规则 221 包含多个消息分类规则，诸如例如规则 221 到规则 221N，它们可被用于分类电子消息。类似于规则 121，规则 221 中的每个规则指示了有效性、成本、效率，并且可包括指令。

[0082] 通常，配置消息分类器 202 以基于电子消息特性来分类电子消息。例如，消息分类器可以接收未分类消息 201U 作为输入并生成经分类的消息 201C 作为输出。经分类的消息 201C 中的每个消息可以被分类为，例如指示该消息是否是 SPAM、是否包含恶意软件、是否包含敏感信息等。

[0083] 如所述，消息分类器 102 包括运行模块 203 和效率合成器 214，运行模块 203 进一步包括资源监视器 213。配置运行模块 203 以运行包括在所接收规则中的指令（例如脚本或其它可执行代码）。指令产生可以用作分类电子消息的数据点的单独的结果（可能受外部用户反馈的影响）。资源监视器 213 可以监视（例如本质上实时地）在规则执行期间各种所消费的资源（例如系统存储器、处理器、网络带宽等）的量。

[0084] 效率合成器 214 可以接收结果以及对所消费的资源的指示，并为所应用的规则合成更新的效率。所应用的规则的结果和所消费的资源也可以用于更新该规则的有效性和 / 或成本以与合成的效率相一致。

[0085] 这样，每个分类规则的成本和有效性可以被测量为值，在特定时间点时（例如当被应用时）测量，并且可以随时间改变。随 SPAM 模式和内容的发展，分类规则可以变得更加有效或更不有效。如果特定的历史 SPAM 活动经历在数量上复兴，则较旧的规则可以突然变得更加有效。而且，当软件升级或优化时，可以降低规则的成本。

[0086] 规则选择和重排序模块 216 可以（例如基于效率）从消息分类规则 221 中选择以应用于电子消息的规则。规则选择和重排序模块 216 还可以（例如基于效率）对消息分类规则 221 进行排名。

[0087] 图 4 示出了用于自适应地选择用于分类电子消息的规则的示例方法 400 的流程图。方法 400 将参考计算机架构 200 的组件和数据来描述。

[0088] 方法 400 包括接收一个或更多电子消息的动作（动作 401）。例如，消息分类器 202 可以接收未分类消息 201U。

[0089] 对于一个或更多电子消息中的每一个，方法 400 包括将在先前选择的电子消息分类规则的子集中的每个消息分类规则应用于电子消息，先前选择的电子消息分类规则的子集是多个电子消息分类规则的一个子集（动作 402）。例如，消息分类器 202 可以（例如基于 SLA 中的最低有效性和最大成本，并且还可以是超标百分比）将规则 221A-221C 应用于未分类消息 201U 中的每个消息。

[0090] 对于先前选择的电子消息分类规则的子集中的每个电子消息分类规则，方法 400 包括电子消息分类规则计算指示该电子消息具有指定的消息特性的可能性的结果的动作（动作 403）。计算资源成本，所述资源成本指示被消耗用于将所述电子消息分类规则应用到所述电子消息的资源量。例如，运行模块 203 可以对未分类消息 101U 执行指令 226A 以

生成结果 212。结果 212 可以（例如基于消息所分配的接收者）指示 201U 中的消息是不想要的电子消息、是感染的或危险的电子消息、包含敏感信息等的可能性。规则 221B 和 221C 的结果也可以被计算。

[0091] 可以将外部反馈（例如来自用户）合并入计算的结果。例如，外部反馈 261 可以被合并入结果 212。基于用户对有效性的感知，外部反馈可以提高或降低所计算的有效性。当合适时，也可以将外部反馈合并入规则 221B 和 221C 的经计算的结果。

[0092] 在一些实施例中，提交含有未捕获的 SPAM、恶意软件或敏感信息的电子消息（假否定）以及被分类为包括 SPAM、恶意软件或敏感信息的合法消息（假肯定）以进一步分析。这种类型的反馈也可以用于微调有效性分数。

[0093] 对于先前选择的电子消息分类规则的子集中的每个电子消息分类规则，方法 400 包括测量指示被消费用于将电子消息分类规则应用于电子消息的资源量的资源成本的动作（动作 404）。例如，资源监视器 213 可以测量指示通过对来自 201U 的消息执行指令 226A 消费的所消耗的资源 231 的量的资源成本。规则 221B 和 221C 的资源消费成本也可以被测量。

[0094] 方法 400 包括保留与将每个电子邮件分类规则应用到每个电子消息相关联的经计算的结果和经测量的资源成本的动作（动作 405）。例如，消息分类器 202 可以将结果 212 和所消费的资源 231 连同对未分类的消息 201U 中其它消息执行规则 221A 的资源成本一起保留。还可以保留对未分类的消息 201U 中的消息执行规则 221B 和 221C 的结果和资源成本。

[0095] 对于在先前选择的电子消息分类规则的子集中的每个消息分类规则，方法 400 包括从所保留的经计算的结果和经测量的资源成本中为该消息分类规则合成效率度量的动作（动作 406）。这样，对于规则 221A、221B 和 221C 中的每一个，效率合成器 214 可以从保留的经计算的结果和经测量的资源成本中合成效率度量。例如，对于规则 221A，效率合成器 214 可以从结果 212 和消费者资源 231 以及从来自将规则 221A 应用到未分类的消息 201U 中的其它消息的经计算的结果和经测量的资源成本中合成经合成的效率 232。也可以为规则 221B 和 221C 合成效率。

[0096] 消息分类器 202 随后可以用合成的效率 232 取代 224A。有效性 222A 和成本 223A 在适当时也可以被升级以与合成的效率 232 相一致。规则 221B 和 221C 的效率、有效性和成本也可以在适当时被更新。

[0097] 方法 400 包括将合成的效率度量与被包括在所述多个电子消息分类规则中的电子消息分类规则的现有的效率度量进行比较的动作（动作 407）。例如，合成的效率 232 可以与其它消息分类规则 221 中包含的效率进行比较。规则 221B 和 221C 的合成的效率也可以与其它消息分类规则 221 中包含的效率进行比较。

[0098] 方法 400 包括至少部分地基于合成的效率度量与现有的效率度量的比较结果来从所述多个电子消息分类规则中选择一个新的电子消息分类规则的子集以便在对随后接收的电子消息进行分类时使用的动作（动作 408）。例如，基于合成的效率，规则 221A、221B 和 221C 相对于彼此以及相对于消息分类规则 221 中的其它规则变得更高效或更低效。这样，在（例如基于 SLA）选择用于分类电子消息的一个新的规则子集时，可放弃规则 221A、221B 和 221C 中的一个或更多。

[0099] 在一些实施例中，在采用与分类消息（例如捕获 SPAM、恶意软件、敏感信息等）的每一规则的相对有效性和运行这些规则的实际观察到的成本有关的实况测量时，消息分类规则的成本和有效性被连续地重新测量（例如在一整天中）。对于较高效规则来说，当其对更多消息运行时，存在有关规则的有效性和成本的更多数据。对于较低效的规则来说，超标百分比（或随机机率）诸如例如 1％，提供了收集更新成本和有效性信息的至少一个基线量。由于重新计算了成本和有效性，用于对规则进行排序的效率分数也要被重新计算。后续的电子消息使用基于更新的分数所选择的消息分类规则来分类。

[0100] 当写新规则（例如用于捕捉新类型的 SPAM 或恶意软件）时，规则资料库的大小增长了。可以用有效性分数 0 和成本分数 1 来引入新引入的规则，这会导致效率分数为 0 并将该规则放置在列表的最尾部。随着时间流逝，当根据超标百分比将该规则应用于消息时，最后可累加足够的现实世界的数据来为该新规则计算更加现实的成本和有效性值，并且因此更加合适的效率分数。由于效率分数被重新计算，这些新规则将自动迁移到它们在列表中的最优顺序。

[0101] 随着时间逝去，规则资料库会变得太大以至于无法基于当前超标百分比（例如 1％）来可行地运行。因此，可以加入另一等级 (tier)，其中，基于减少的超标百分比，诸如例如 0.1％，来运行具有小于大概 0.1 的效率分数的规则。这样的低效的规则的结果甚至可能不用于分类消息，相反，仅用于生成经更新的成本和有效性信息。

[0102] 图 5 示出了便于自适应的电子消息扫描和自适应地选择用于分类电子消息的规则的示例计算机体系结构 500。

[0103] 接收未分类的邮件 501U。将发送者 / 接收者信息发送给消费者类 531。消费者类 531 标识在分类未分类的邮件 501U 时要使用的最低有效性。在阶段 541 中，从运行规则 511A、511B 和 511C 中实现消费者类 531 的最低有效性。在阶段 542 中，在资源可用性 504 允许时就机会性地运行包括规则 511D 的一个或更多另外的规则。在阶段 543 中，不运行最多至规则 511N 的一个或更多其它规则的决策基于随机机会 518 而超标，并且这些最多至规则 511N 的一个或更多规则被运行。基于规则 511A 到 511N 的结果，输出经分类的邮件 501C（例如作为 SPAM 或合法的）。

[0104] 由每个规则的运行时程序 503 来为规则 511A 到 511N 收集性能数据。更新的成本分数被写回到规则 511A 到 511N 中。为规则 511A 到 511N 将每个规则的结果 512 确定为肯定（例如是 SPAM）或否定（例如是合法的）。外部反馈 561 被合并以标识结果中的假肯定和假否定。更新的有效性分数被写回到规则 511A 到 511N 中。重新计算效率，并基于该重新计算的效率重新排序规则。

[0105] 本发明可具体化为其它具体形式而不背离其精神或本质特征。所描述的实施例在所有方面都应被认为仅是说明性而非限制性的。因此，本发明的范围由所附权利要求书而非前述描述指示。落入权利要求书的等效方案的含义和范围内的所有改变被权利要求书的范围所涵盖。

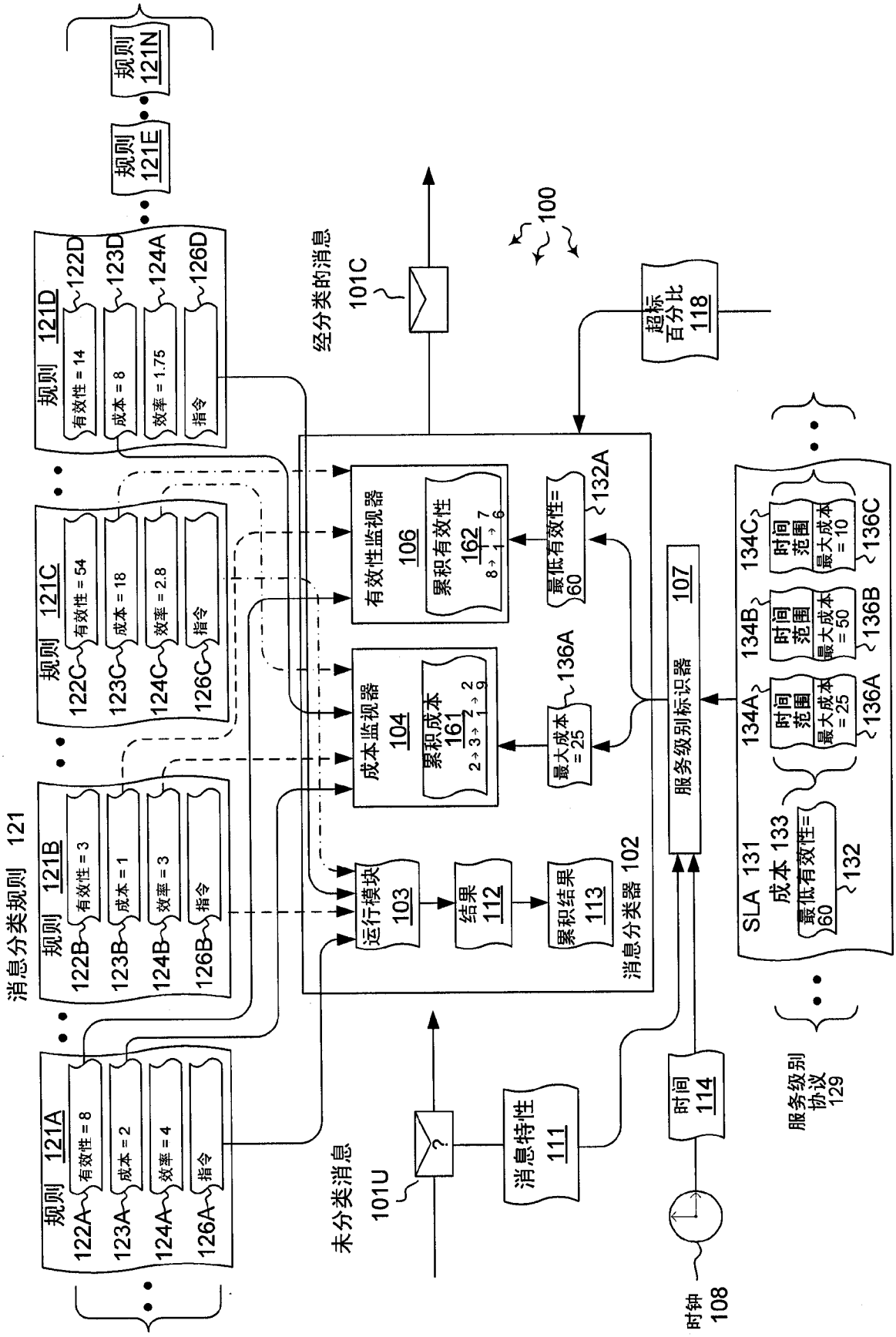

消息分类规则 121
规则 121A
有效性 = 8
成本 = 2
效率 = 4
指令
122A
123A
124A
126A
规则 121B
有效性 = 3
成本 = 1
效率 = 3
指令
122B
123B
124B
126B
规则 121C
有效性 = 54
成本 = 18
效率 = 2.8
指令
122C
123C
124C
126C
规则 121D
有效性 = 14
成本 = 8
效率 = 1.75
指令
122D
123D
124A
126D
规则 121E
规则 121N
未分类消息 101U
?
消息特性 111
时钟 108
时间 114
运行模块 103
结果 112
累积结果 113
消息分类器 102
成本监视器 104
累积成本 161
2→3→2→1→9
最大成本 = 25
136A
有效性监视器 106
累积有效性 162
8→1→7→6
最低有效性 = 60
132A
经分类的消息 101C
100
超标百分比 118
服务级别标识器 107
SLA 131
成本 133
最低有效性 = 60
132
时间范围
最大成本 = 25
136A
134A
时间范围
最大成本 = 50
136B
134B
时间范围
最大成本 = 10
136C
134C
服务级别协议 129

图 1

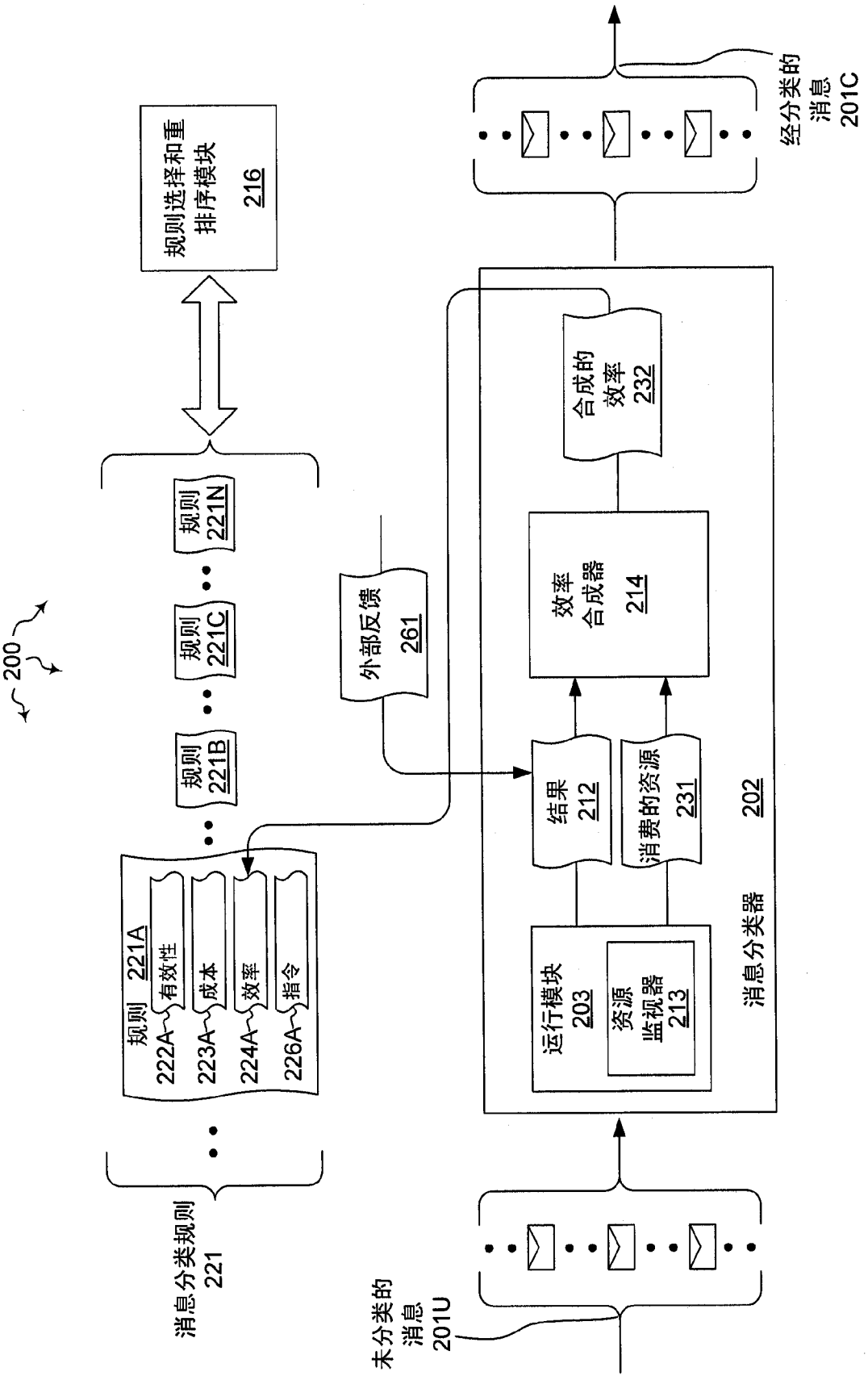

200
规则选择和重排序模块 216
规则 221N
规则 221C
规则 221B
规则 221A
有效性 222A
成本 223A
效率 224A
指令 226A
消息分类规则 221
外部反馈 261
结果 212
消费的资源 231
合成的效率 232
效率合成器 214
运行模块 203
资源监视器 213
消息分类器 202
未分类的消息 201U
经分类的消息 201C

图 2

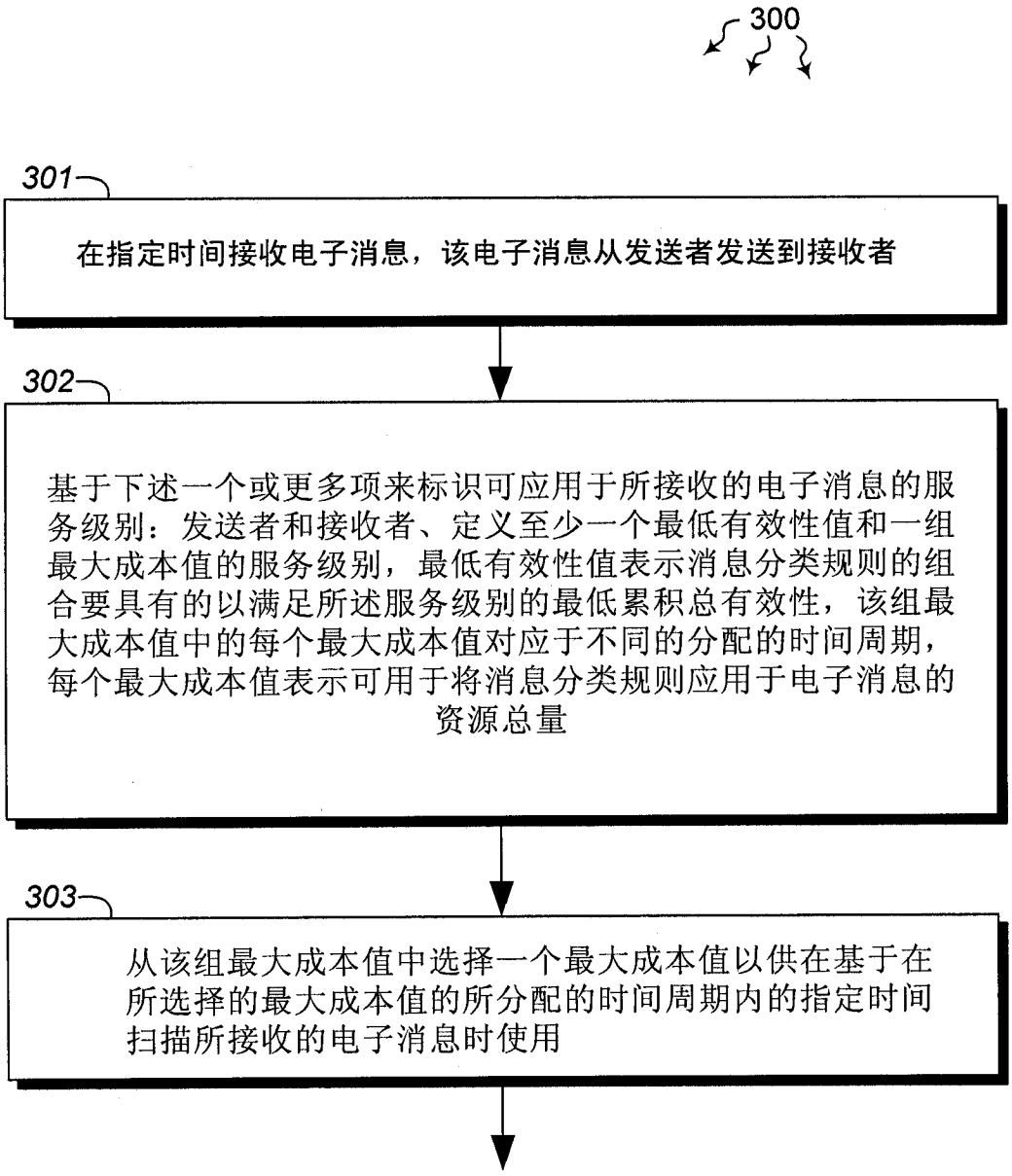

300
301
在指定时间接收电子消息，该电子消息从发送者发送到接收者
302
基于下述一个或更多项来标识可应用于所接收的电子消息的服务级别：发送者和接收者、定义至少一个最低有效性值和一组最大成本值的服务级别，最低有效性值表示消息分类规则的组合要具有的以满足所述服务级别的最低累积总有效性，该组最大成本值中的每个最大成本值对应于不同的分配的时间周期，每个最大成本值表示可用于将消息分类规则应用于电子消息的资源总量
303
从该组最大成本值中选择一个最大成本值以供在基于在所选择的最大成本值的所分配的时间周期内的指定时间扫描所接收的电子消息时使用

图 3

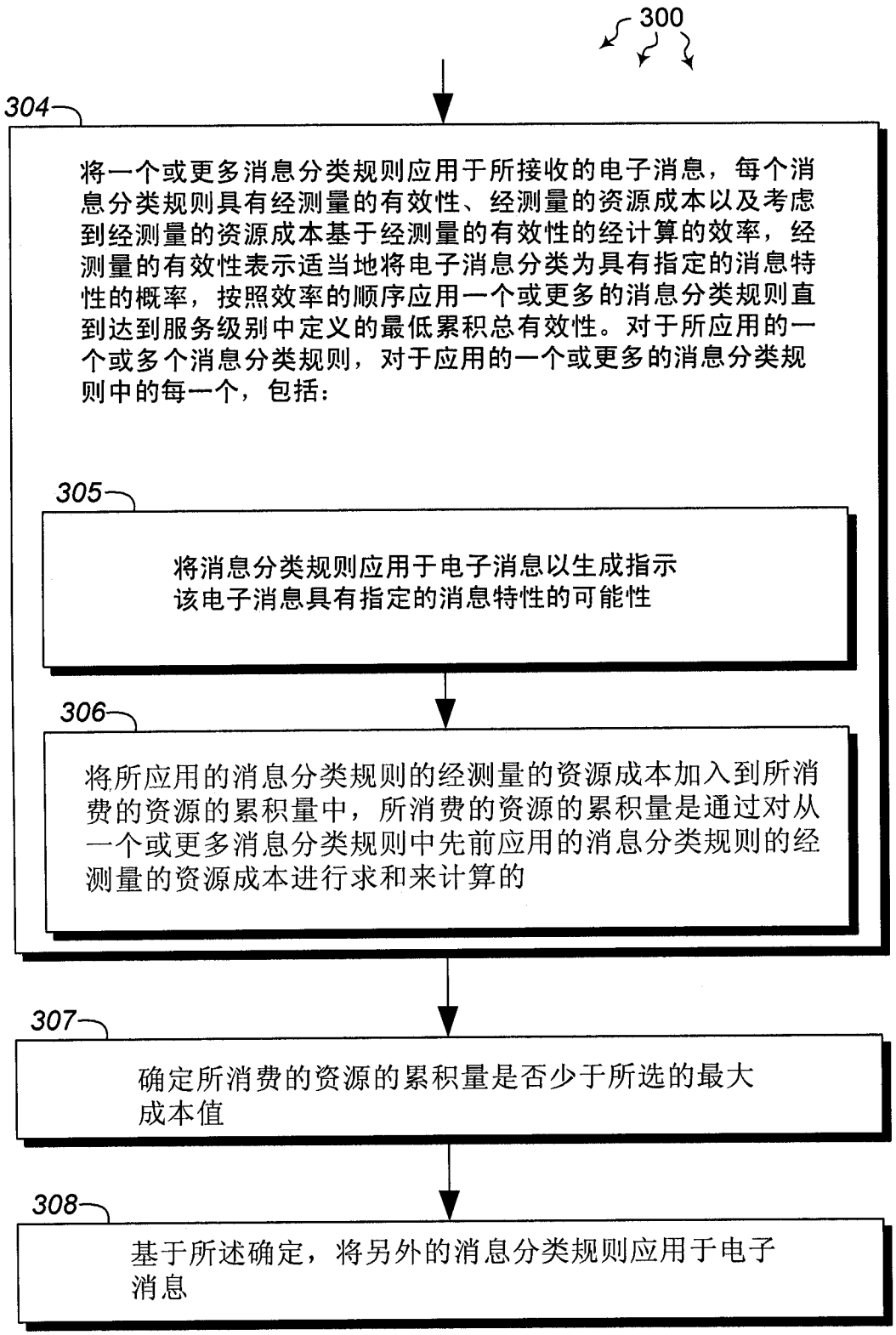

300
304
将一个或更多消息分类规则应用于所接收的电子消息，每个消息分类规则具有经测量的有效性、经测量的资源成本以及考虑到经测量的资源成本基于经测量的有效性的经计算的效率，经测量的有效性表示适当地将电子消息分类为具有指定的消息特性的概率，按照效率的顺序应用一个或更多的消息分类规则直到达到服务级别中定义的最低累积总有效性。对于所应用的一个或多个消息分类规则，对于应用的一个或更多的消息分类规则中的每一个，包括：
305
将消息分类规则应用于电子消息以生成指示该电子消息具有指定的消息特性的可能性
306
将所应用的消息分类规则的经测量的资源成本加入到所消费的资源的累积量中，所消费的资源的累积量是通过对从一个或更多消息分类规则中先前应用的消息分类规则的经测量的资源成本进行求和来计算的
307
确定所消费的资源的累积量是否少于所选的最大成本值
308
基于所述确定，将另外的消息分类规则应用于电子消息

图 3（继续）

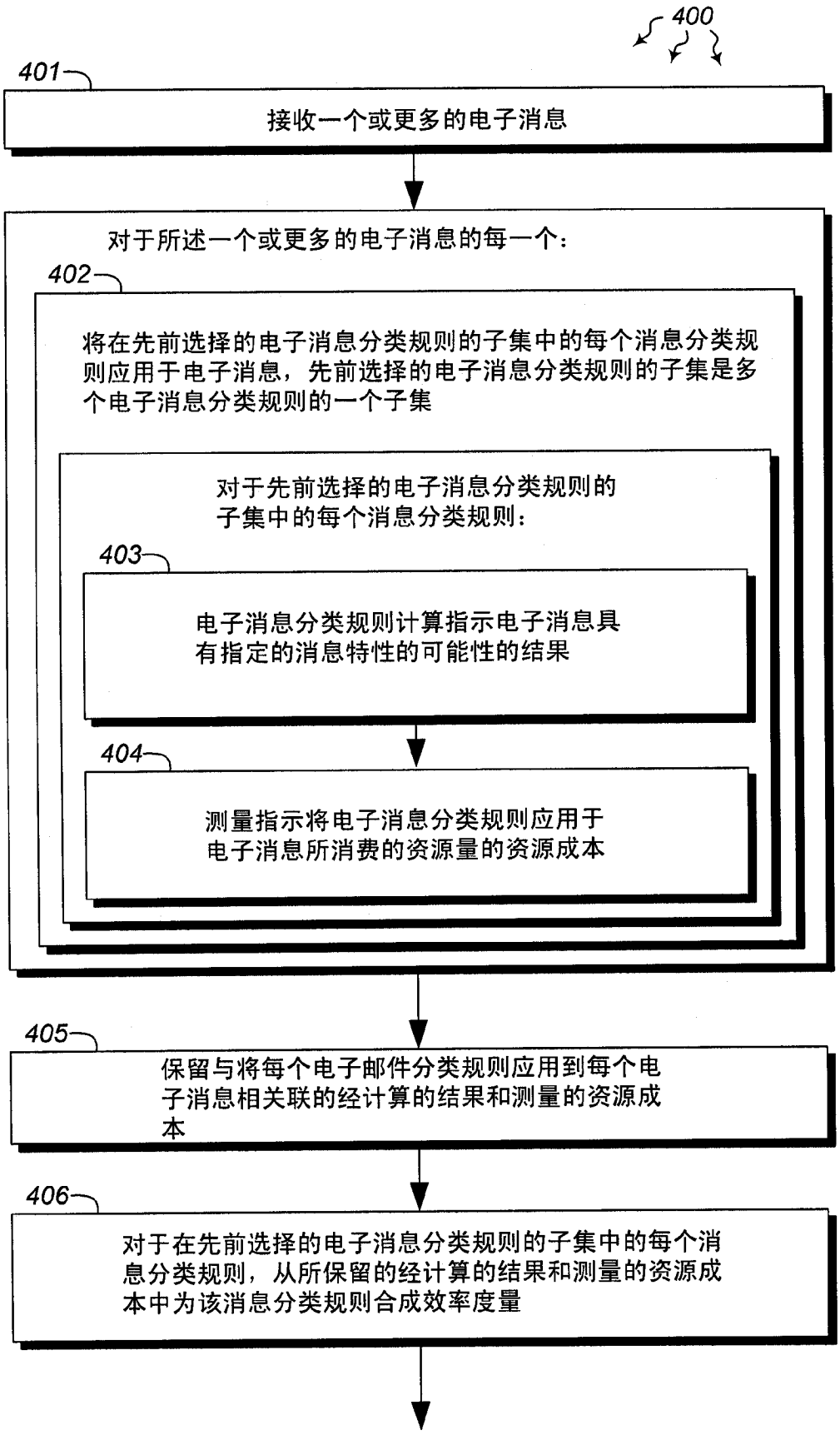

400
401
接收一个或更多的电子消息
对于所述一个或更多的电子消息的每一个：
402
将在先前选择的电子消息分类规则的子集中的每个消息分类规则应用于电子消息，先前选择的电子消息分类规则的子集是多个电子消息分类规则的一个子集
对于先前选择的电子消息分类规则的子集中的每个消息分类规则：
403
电子消息分类规则计算指示电子消息具有指定的消息特性的可能性的结果
404
测量指示将电子消息分类规则应用于电子消息所消费的资源量的资源成本
405
保留与将每个电子邮件分类规则应用到每个电子消息相关联的经计算的结果和测量的资源成本
406
对于在先前选择的电子消息分类规则的子集中的每个消息分类规则，从所保留的经计算的结果和测量的资源成本中为该消息分类规则合成效率度量

图 4

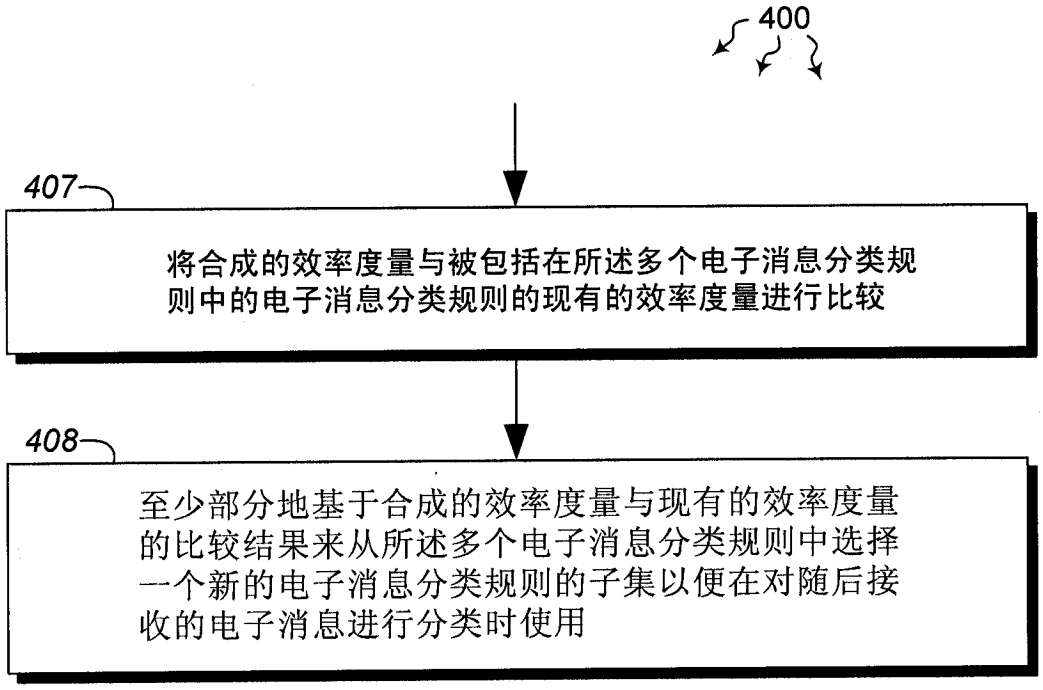

400
407
将合成的效率度量与被包括在所述多个电子消息分类规则中的电子消息分类规则的现有的效率度量进行比较
408
至少部分地基于合成的效率度量与现有的效率度量的比较结果来从所述多个电子消息分类规则中选择一个新的电子消息分类规则的子集以便在对随后接收的电子消息进行分类时使用

图 4（继续）

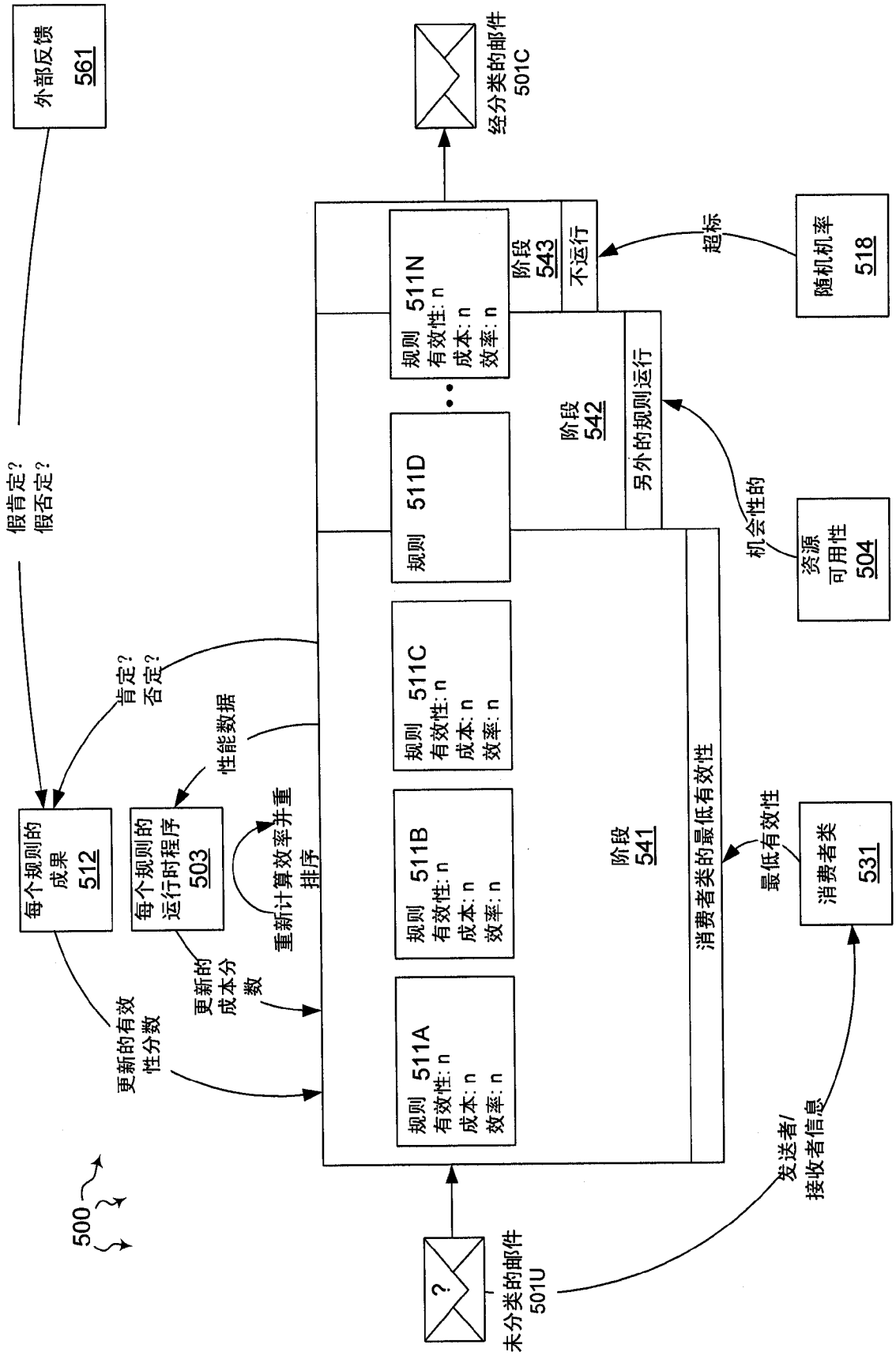

外部反馈
561
经分类的邮件
501C
规则 511N
有效性: n
成本: n
效率: n
阶段
543
不运行
超标
随机机率
518
规则 511D
阶段
542
另外的规则运行
机会性的
资源
可用性
504
假肯定?
假否定?
肯定?
否定?
性能数据
每个规则的
成果
512
每个规则的
运行时程序
503
重新计算效率并重
排序
规则 511C
有效性: n
成本: n
效率: n
规则 511B
有效性: n
成本: n
效率: n
阶段
541
消费者类的最低有效性
最低有效性
消费者类
531
更新的有效
性分数
更新的
成本分
数
规则 511A
有效性: n
成本: n
效率: n
500
?
未分类的邮件
501U
发送者/
接收者信息

图 5