(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 101495969 B
(45)授权公告日 2012.10.10

(21)申请号 200680019314.X
(22)申请日 2006.05.05
(30)优先权数据
60/678,391 2005.05.05 US
(85)PCT申请进入国家阶段日
2007.11.30
(86)PCT申请的申请数据
PCT/US2006/017783 2006.05.05
(87)PCT申请的公布数据
WO2006/119509 EN 2006.11.09
(73)专利权人 思科埃恩波特系统有限公司
地址 美国加利福尼亚州
(72)发明人 克雷格·斯伯罗茨 斯科特·肯尼迪
丹尼尔·昆兰 拉里·罗森斯坦
查尔斯·斯莱特
(74)专利代理机构 北京东方亿思知识产权代理有限责任公司 11258
代理人 王怡
(51)Int.Cl.
G06F 11/00(2006.01)
G06F 12/14(2006.01)
G06F 12/16(2006.01)
G06F 15/18(2006.01)
(56)对比文件
US 2005080856 A1,2005.04.14,
US 2004117648 A1,2004.06.17,
US 2004117648 A1,2004.06.17,
US 2002004908 A1,2002.01.10,
WO 2005036341 A2,2005.04.21,
审查员 陈婕

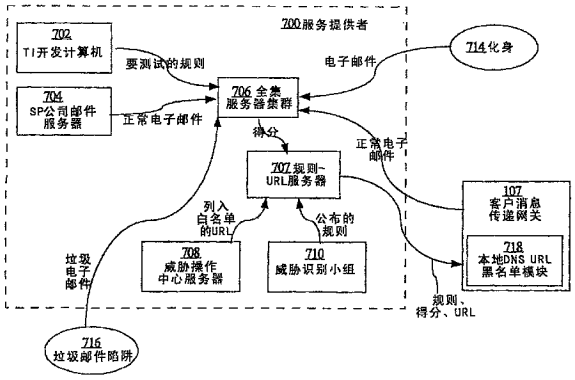
权利要求书 5 页 说明书 35 页 附图 11 页

(54)发明名称

识别电子消息中的威胁

(57)摘要

在没有病毒签名信息可用时，通过对消息内容应用试探测试并且检查发送者声望信息，来提供对计算机病毒和其他由消息承载的威胁的提早检测。结果，消息传递网关可在病毒发作早期就暂停消息递送，从而可提供充足的时间来更新能够将病毒代码从消息中剥离的防病毒检查器。为动态、灵活的威胁隔离队列提供了多种退出标准和退出动作，这允许了以非先进先出的顺序提早地释放消息。描述了一种消息扫描方法，其中通过将威胁规则只与选定的消息元素相匹配并且一旦在一个消息元素上的匹配超过了威胁阈值就停止规则匹配，可以实现从解析和扫描的提早退出。



CN 101495969 B

1. 一种用于对电子消息中新出现的威胁作出响应的装置，包括：

用于接收具有针对接收者账户的目的地地址的电子邮件消息的装置；

用于基于指定已知包含计算机病毒的消息的属性的多个规则来确定所述消息的病毒得分值的装置，其中每个规则具有与该规则中包含的标准的数目相等的权重，其中所述属性包括所述消息的文件附件的类型、所述文件附件的大小以及基于消息发送者、主题或正文以及除文件附件签名外的其他内容的一个或多个试探，并且其中所述病毒得分值是所述多个规则之中的匹配规则所返回的病毒得分值的加权平均值，该加权平均值是通过将第一加和除以第二加和来确定的，其中所述第一加和是通过对所述匹配规则所返回的病毒得分值的每一个与所述匹配规则中的相应规则的权重相乘的乘积求和而获得的，所述第二加和是通过对所述匹配规则的权重求和而获得的；

用于当所述病毒得分值大于或等于指定的阈值时，将所述消息存储在隔离队列中，而不立即将所述消息递送到所述接收者账户的装置。

2. 如权利要求 1 所述的装置，其中所述属性包括所述附件的内容的类型。

3. 如权利要求 1 所述的装置，其中所述属性包括所述消息的发送者的标识。

4. 如权利要求 1 所述的装置，其中所述试探包括将所述消息的正文的内容与携带病毒的其他消息的正文中常用的单词的字典相匹配。

5. 如权利要求 1 所述的装置，其中所述试探包括将所述消息的主题的内容与携带病毒的其他消息的主题行中常用的单词的字典相匹配。

6. 如权利要求 1 所述的装置，其中所述试探包括：

从所述消息中提取发送者标识符；

检索与所述发送者标识符相关联的声望得分值；

至少部分基于所述声望得分值来确定所述病毒得分值。

7. 如权利要求 1 所述的装置，其中所述试探包括将所述消息的文件附件的字节与唯一地标识 Microsoft 可执行文件的初始字节的规则相匹配。

8. 如权利要求 1 所述的装置，其中所述试探包括：

从所述消息中提取发送者标识符；

确定所述发送者标识符是否在本地存储的发送者黑名单中；

至少部分基于所述发送者标识符是否在所述黑名单中来确定所述病毒得分值。

9. 如权利要求 1 所述的装置，其中所述试探包括：

从所述消息中提取发送者标识符；

经由网络请求外部服务确定所述发送者标识符是否在存储的发送者黑名单中，并且接收来自所述外部服务的响应；

至少部分基于所述响应来确定所述病毒得分值。

10. 一种用于扫描消息以找出威胁的装置，包括：

用于接收并存储指定电子消息的特性的多个规则的装置，所述特性指示出与所述消息相关联的威胁，其中每个规则具有优先级值，其中每个规则与一个消息元素类型相关联；

用于接收具有针对接收者账户的目的地地址的电子邮件消息的装置，其中所述消息包括多个消息元素；

用于提取第一消息元素的装置；

用于通过仅将所述第一消息元素仅与具有对应于所述第一消息元素的消息元素类型的选定规则相匹配并且根据所述选定规则的优先级顺序，来确定所述消息的威胁得分值的装置；

用于当所述威胁得分值大于指定的阈值时，输出所述威胁得分值的装置。

11. 如权利要求 10 所述的装置，还包括用于执行以下步骤的装置：通过仅当所述威胁得分值小于指定阈值时才将其他消息元素与其他规则相匹配来确定所述消息的更新后威胁得分值。

12. 如权利要求 10 所述的装置，其中所述消息元素包括一个或多个消息头部、消息正文以及所述消息正文中的一个或多个 HTML 元素。

13. 如权利要求 10 所述的装置，其中所述第一消息元素包括一个或多个消息头部，所述选定规则仅包括头部规则，并且所述多个规则包括所述头部规则、原始正文规则和 HTML 正文规则。

14. 如权利要求 10 所述的装置，还包括用于在匹配所述选定规则中的每个规则后，测试所述威胁得分值是否大于所述指定的阈值的装置。

15. 如权利要求 10 所述的装置，还包括用于执行以下步骤的装置：通过仅当所述威胁得分值小于指定阈值时才对所述消息的正文进行解码并将所述正文与一个或多个原始正文规则相匹配，来确定所述消息的更新后威胁得分值。

16. 如权利要求 10 所述的装置，还包括用于执行以下步骤的装置：通过仅当所述威胁得分值小于指定阈值时才呈现标记语言消息元素并将所呈现的标记语言消息元素与一个或多个正文规则相匹配，来确定所述消息的更新后威胁得分值。

17. 如权利要求 10 所述的装置，其中所述威胁包括病毒、垃圾邮件或钓鱼式攻击中的任何一种。

18. 如权利要求 10 所述的装置，还包括：

用于当所述威胁得分值大于或等于指定的威胁阈值时，将所述消息存储在隔离队列中，而不立即将所述消息递送到所述接收者账户的装置；

用于在满足多个隔离退出标准中的任何一个时以非先进先出顺序将所述消息从所述隔离队列中释放出来的装置，其中每个隔离退出标准与一个或多个退出动作相关联；以及

用于在满足特定的退出标准时，选择并执行相关联的所述一个或多个退出动作的装置。

19. 如权利要求 18 所述的装置，其中所述隔离退出标准包括消息隔离时间限度的期满、所述隔离队列的溢出、从所述隔离队列中手工释放以及接收对用于确定所述威胁得分值的一个或多个规则的更新。

20. 如权利要求 18 所述的装置，其中所述退出动作包括从所述消息中去除文件附件并将没有所述文件附件的所述消息递送到所述接收者账户；删除所述消息；修改所述消息的主题行；以及向所述消息添加 X 头部。

21. 如权利要求 18 所述的装置，还包括用于响应于 (a) 对从所述消息隔离中手工释放所述消息的用户请求、(b) 与所述消息相关联的定时器的期满以及 (c) 所述消息隔离变满，而执行不同动作的装置。

22. 如权利要求 18 所述的装置，还包括用于执行以下步骤的装置：(a) 响应于对从所述

消息隔离中手工释放所述消息的用户请求而不加修改地将所述消息递送到所述接收者账户；以及 (b) 响应于所述消息隔离变满而从所述消息中去除文件附件并将没有所述文件附件的所述消息递送到所述接收者账户。

23. 如权利要求 18 所述的装置，还包括用于向所述消息分配期满时间值的装置，其中所述期满时间值基于向所述消息内容应用试探测试的结果而不同。

24. 如权利要求 18 所述的装置，其中所述隔离退出标准包括接收对用于确定所述威胁得分值的一个或多个规则的更新，并且其中所述退出动作包括基于更新后的规则再次确定所述消息的威胁得分值。

25. 如权利要求 18 所述的装置，其中所述退出动作包括向除所述装置外的主机上的第二隔离队列发送所述消息的拷贝。

26. 如权利要求 18 所述的装置，其中不同的一组或多组退出动作与不同的隔离退出标准相关联。

27. 如权利要求 18 所述的装置，还包括用于执行以下步骤的装置：确定所述威胁得分值何时大于或等于指定的报告阈值，并且响应于此而创建并发送警告消息到另一主机。

28. 如权利要求 18 所述的装置，其中所述威胁包括病毒、垃圾邮件或钓鱼式攻击中的任何一种。

29. 一种用于对电子消息中新出现的威胁作出响应的方法，包括：

接收具有针对接收者账户的目的地地址的电子邮件消息；

基于指定已知包含计算机病毒的消息的属性的多个规则来确定所述消息的病毒得分值，其中每个规则具有与该规则中包含的标准的数目相等的权重，其中所述属性包括所述消息的文件附件的类型、所述文件附件的大小以及基于消息发送者、主题或正文以及除文件附件签名外的其他内容的一个或多个试探，并且其中所述病毒得分值是所述多个规则之中的匹配规则所返回的病毒得分值的加权平均值，该加权平均值是通过将第一加和除以第二加和来确定的，其中所述第一加和是通过对所述匹配规则所返回的病毒得分值的每一个与所述匹配规则中的相应规则的权重相乘的乘积求和而获得的，所述第二加和是通过对所述匹配规则的权重求和而获得的；

当所述病毒得分值大于或等于指定的阈值时，将所述消息存储在隔离队列中，而不立即将所述消息递送到所述接收者账户。

30. 如权利要求 29 所述的方法，其中所述属性包括所述附件的内容的类型。

31. 如权利要求 29 所述的方法，其中所述属性包括所述消息的发送者的标识。

32. 如权利要求 29 所述的方法，其中所述试探包括将所述消息的正文的内容与携带病毒的其他消息的正文中常用的单词的字典相匹配。

33. 如权利要求 29 所述的方法，其中所述试探包括将所述消息的主题的内容与携带病毒的其他消息的主题行中常用的单词的字典相匹配。

34. 如权利要求 29 所述的方法，其中所述试探包括：

从所述消息中提取发送者标识符；

检索与所述发送者标识符相关联的声望得分值；

至少部分基于所述声望得分值来确定所述病毒得分值。

35. 如权利要求 29 所述的方法，其中所述试探包括将所述消息的文件附件的字节与唯

一地标识 Microsoft 可执行文件的初始字节的规则相匹配。

36. 如权利要求 29 所述的方法，其中所述试探包括：

从所述消息中提取发送者标识符；

确定所述发送者标识符是否在本地存储的发送者黑名单中；

至少部分基于所述发送者标识符是否在所述黑名单中来确定所述病毒得分值。

37. 如权利要求 29 所述的方法，其中所述试探包括：

从所述消息中提取发送者标识符；

经由网络请求外部服务确定所述发送者标识符是否在存储的发送者黑名单中，并且接收来自所述外部服务的响应；

至少部分基于所述响应来确定所述病毒得分值。

38. 一种用于扫描消息以找出威胁的方法，包括：

接收并存储指定电子消息的特性的多个规则，所述特性指示出与所述消息相关联的威胁，其中每个规则具有优先级值，其中每个规则与一个消息元素类型相关联；

接收具有针对接收者账户的目的地地址的电子邮件消息，其中所述消息包括多个消息元素；

提取第一消息元素；

通过仅将所述第一消息元素仅与具有对应于所述第一消息元素的消息元素类型的选定规则相匹配并且根据所述选定规则的优先级顺序，来确定所述消息的威胁得分值；

当所述威胁得分值大于指定的阈值时，输出所述威胁得分值。

39. 如权利要求 38 所述的方法，还包括：通过仅当所述威胁得分值小于指定阈值时才将其他消息元素与其他规则相匹配来确定所述消息的更新后威胁得分值。

40. 如权利要求 38 所述的方法，其中所述消息元素包括一个或多个消息头部、消息正文以及所述消息正文中的一个或多个 HTML 元素。

41. 如权利要求 38 所述的方法，其中所述第一消息元素包括一个或多个消息头部，所述选定规则仅包括头部规则，并且所述多个规则包括所述头部规则、原始正文规则和 HTML 正文规则。

42. 如权利要求 38 所述的方法，还包括：在匹配所述选定规则中的每个规则后，测试所述威胁得分值是否大于所述指定的阈值。

43. 如权利要求 38 所述的方法，还包括：通过仅当所述威胁得分值小于指定阈值时才对所述消息的正文进行解码并将所述正文与一个或多个原始正文规则相匹配，来确定所述消息的更新后威胁得分值。

44. 如权利要求 38 所述的方法，还包括：通过仅当所述威胁得分值小于指定阈值时才呈现标记语言消息元素并将所呈现的标记语言消息元素与一个或多个正文规则相匹配，来确定所述消息的更新后威胁得分值。

45. 如权利要求 38 所述的方法，其中所述威胁包括病毒、垃圾邮件或钓鱼式攻击中的任何一种。

46. 如权利要求 38 所述的方法，还包括：

当所述威胁得分值大于或等于指定的威胁阈值时，将所述消息存储在隔离队列中，而不立即将所述消息递送到所述接收者账户；

在满足多个隔离退出标准中的任何一个时以非先进先出顺序将所述消息从所述隔离队列中释放出来，其中每个隔离退出标准与一个或多个退出动作相关联；并且

在满足特定的退出标准时，选择并执行相关联的所述一个或多个退出动作。

47. 如权利要求 46 所述的方法，其中所述隔离退出标准包括消息隔离时间限度的期满、所述隔离队列的溢出、从所述隔离队列中手工释放以及接收对用于确定所述威胁得分值的一个或多个规则的更新。

48. 如权利要求 46 所述的方法，其中所述退出动作包括从所述消息中去除文件附件并将没有所述文件附件的所述消息递送到所述接收者账户；删除所述消息；修改所述消息的主题行；以及向所述消息添加 X 头部。

49. 如权利要求 46 所述的方法，还包括：响应于 (a) 对从所述消息隔离中手工释放所述消息的用户请求、(b) 与所述消息相关联的定时器的期满以及 (c) 所述消息隔离变满，而执行不同动作。

50. 如权利要求 46 所述的方法，还包括：(a) 响应于对从所述消息隔离中手工释放所述消息的用户请求而不加修改地将所述消息递送到所述接收者账户；以及 (b) 响应于所述消息隔离变满而从所述消息中去除文件附件并将没有所述文件附件的所述消息递送到所述接收者账户。

51. 如权利要求 46 所述的方法，还包括向所述消息分配期满时间值，其中所述期满时间值基于向所述消息内容应用试探测试的结果而不同。

52. 如权利要求 46 所述的方法，其中所述隔离退出标准包括接收对用于确定所述威胁得分值的一个或多个规则的更新，并且其中所述退出动作包括基于更新后的规则再次确定所述消息的威胁得分值。

53. 如权利要求 46 所述的方法，其中所述退出动作包括向除所述方法外的主机上的第二隔离队列发送所述消息的拷贝。

54. 如权利要求 46 所述的方法，其中不同的一组或多组退出动作与不同的隔离退出标准相关联。

55. 如权利要求 46 所述的方法，还包括：确定所述威胁得分值何时大于或等于指定的报告阈值，并且响应于此而创建并发送警告消息到另一主机。

56. 如权利要求 46 所述的方法，其中所述威胁包括病毒、垃圾邮件或钓鱼式攻击中的任何一种。

识别电子消息中的威胁

技术领域

[0001] 本发明一般地涉及检测电子消息中的威胁 (threat)，例如计算机病毒 (virus)、垃圾邮件 (spam) 和钓鱼式攻击 (phishing attack)。本发明更具体而言涉及用于对电子消息中新出现的威胁作出响应、管理承载威胁的消息的隔离队列以及扫描消息以找出威胁的技术。

背景技术

[0002] 本部分中描述的方法可以实行，但不一定是以前已经构想出或实行过的方法。因此，除非这里另有指明，否则本部分中描述的方法并不是本申请的权利要求的现有技术，并且并不因为被包括在本部分中就被承认是现有技术。

[0003] 在连接到公共网络的计算机中由消息承载的病毒的经常发作已经成为一个严重的问题，尤其对于具有较大的专用网络的商业企业来说更是如此。由于雇员生产力的浪费、购买附加硬件和软件的资本投入、因许多病毒破坏共享目录上的文件而造成的信息丢失以及因许多病毒从用户计算机附加并发送随机文件而造成的对隐私和机密的侵害，可能导致数千美元的直接和间接成本。

[0004] 此外，来自病毒的损害会在非常短的一段时间中发生。从病毒发作之时到在可检测和阻止受病毒感染的消息的企业邮件网关处公布和部署病毒定义之时之间的时间中，企业网络中的很大一部分机器可能已受到感染。“发作”和“规则部署”之间的时间窗口通常是五 (5) 个小时或更多。缩短反应时间将会是非常有价值的。

[0005] 在大多数的病毒发作中，可执行的附件现在充当着病毒代码的载体。例如，在过去三年的 17 次主要病毒发作中，13 个病毒是通过电子邮件附件发送的。通过电子邮件附件发送的 13 个病毒中的 12 个是通过危险的附件类型发送的。因此，一些企业网络邮件网关现在阻止所有类型的可执行文件附件。

[0006] 显然，作为响应，病毒编写者现在正在隐藏可执行文件。病毒编写者越来越多地将已知的危险文件类型隐藏在看起来无害的文件中。例如，病毒编写者可将可执行文件嵌入在由 WinZIP 和其他归档工具生成的那类 .zip 文件内。这种 .zip 文件经常被企业用来压缩和共享较大的文件，因此大多数企业不愿意或者不能够阻止 .zip 文件。还可能将可执行文件嵌入在 Microsoft Word 和一些版本的 Adobe Acrobat 中。

[0007] 基于前述原因，很明显需要一种用于管理病毒发作的改进方法。用于防止递送大量不请自来的商业电子邮件（“垃圾邮件”）和包含其他形式的威胁（例如钓鱼式攻击）的消息的现有技术也被认为是不充分的。用于扫描消息以找出威胁的现有技术也被认为是效率低下并且需要改进的。

附图说明

[0008] 在附图中以示例方式而非限制方式示出了本发明，图中类似的标号指代相似的元件，其中：

[0009] 图 1 是根据一个实施例用于管理计算机病毒发作的系统的框图。

[0010] 图 2 是根据一个实施例由病毒信息源执行的生成可疑消息计数的过程的流程图。

[0011] 图 3 是根据一个实施例图示出基于病毒发作信息的消息处理的数据流程图。

[0012] 图 4 是根据一个实施例确定病毒得分值的方法的流程图。

[0013] 图 5 是根据一个实施例图示出用于管理病毒发作的一组规则的应用的流程图。

[0014] 图 6 是图示出其上可实现一个实施例的计算机系统的框图。

[0015] 图 7 是可用在用于阻止“垃圾邮件”消息的方法中或用于其他种类的电子邮件扫描过程的系统的框图。

[0016] 图 8 是在假设的示例性病毒发作中时间与受感染的机器数目之间的关系图。

[0017] 图 9 是用于重新扫描可能包含病毒的消息的方法的流程图。

[0018] 图 10 是实现上述逻辑的消息传递网关中的消息流模型的框图。

[0019] 图 11 是利用提早退出方法执行消息威胁扫描的过程的流程图。

具体实施方式

[0020] 描述了一种用于管理计算机病毒发作的方法和装置。在下面的描述中，为了进行说明，阐述了许多具体细节以便全面地理解本发明。但是，对于本领域的技术人员来说很明显的是，没有这些具体细节也能实现本发明。在其他情况下，以框图形式示出了公知的结构和设备，以避免不必要地模糊本发明的主题。

[0021] 这里根据下面的提纲来描述实施例：

[0022] 1.0 综述

[0023] 2.0 病毒发作控制方法一第一实施例 - 结构和功能概述

[0024] 2.1 网络系统和病毒信息源

[0025] 2.2 对可疑消息进行计数

[0026] 2.3 基于病毒发作信息来处理消息

[0027] 2.4 生成病毒发作信息

[0028] 2.5 使用病毒发作信息

[0029] 2.6 附加特征

[0030] 2.7 示例性使用情形

[0031] 3.0 用于阻止垃圾邮件消息的方法

[0032] 3.1 从垃圾邮件扫描中提早退出

[0033] 3.2 垃圾邮件扫描裁决缓存

[0034] 4.0 基于消息试探 (Heuristics)、发送者信息、动态隔离操作和细颗粒规则的病毒检测方法

[0035] 4.1 利用消息试探进行检测

[0036] 4.2 基于发送者的病毒检测

[0037] 4.3 包括重新扫描的动态隔离操作

[0038] 4.4 细颗粒规则

[0039] 4.5 消息传递网关与服务提供者的通信

[0040] 4.6 传出白名单模块

[0041] 5.0 实现机构 - 硬件概述

[0042] 6.0 扩展和替换

[0043] 1.0 综述

[0044] 在前述背景技术部分中所确定的需求以及在下面的描述中将显现出来的其他需求和目的在本发明中得以实现，本发明在一个方面中包括一种方法，该方法包括：接收具有针对接收者账户的目的地地址的电子邮件消息；基于指定已知包含计算机病毒的消息的属性的一个或多个规则来确定所述消息的病毒得分值，其中所述属性包括所述消息的文件附件的类型、所述文件附件的大小以及基于消息发送者、主题或正文以及除文件附件签名外的其他内容的一个或多个试探；当所述病毒得分值大于或等于指定的阈值时，将所述消息存储在隔离队列中，而不立即将所述消息递送到所述接收者账户。

[0045] 在另一方面中，本发明提供了一种方法，该方法包括：接收具有针对接收者账户的目的地地址的电子邮件消息；确定所述消息的威胁得分值；当所述威胁得分值大于或等于指定的阈值时，将所述消息存储在隔离队列中，而不立即将所述消息递送到所述接收者账户；在满足多个隔离退出标准中的任何一个时以非先进先出顺序将所述消息从所述隔离队列中释放出来，其中每个隔离退出标准与一个或多个退出动作相关联；并且在满足某个特定的退出标准时，选择并执行相关联的一个或多个退出动作。

[0046] 在另一方面中，本发明提供了一种方法，该方法包括：接收并存储指定电子消息的特性的多个规则，所述特性指示出与所述消息相关联的威胁，其中每个规则具有优先级值，其中每个规则与一个消息元素类型相关联；接收具有针对接收者账户的目的地地址的电子邮件消息，其中所述消息包括多个消息元素；提取第一消息元素；通过仅将所述第一消息元素仅与具有对应于所述第一消息元素的消息元素类型的选定规则相匹配并且根据所述选定规则的优先级顺序，来确定所述消息的威胁得分值；当所述威胁得分值大于指定的阈值时，输出所述威胁得分值。

[0047] 在这些方面中，在没有病毒签名信息可用时，通过对消息内容应用试探测试并且检查接收者声望信息，来提供对计算机病毒和其他由消息承载的威胁的提早检测。结果，消息传递网关可在病毒发作早期就暂停消息递送，从而可提供充足的时间来更新能够将病毒代码从消息中剥离的防病毒检查器。为动态、灵活的威胁隔离队列提供了多种退出标准和退出动作，这允许了以非先进先出的顺序提早地释放消息。描述了一种消息扫描方法，其中通过将威胁规则只与选定的消息元素相匹配并且一旦在一个消息元素上的匹配超过了威胁阈值就停止规则匹配，可以实现从解析和扫描的提早退出。

[0048] 2.0 病毒发作控制方法 - 第一实施例 - 结构和功能概述

[0049] 2.1 网络系统和病毒信息源

[0050] 图 1 是根据一个实施例用于管理计算机病毒发作的系统的框图。其身份和位置通常未知的病毒发送者 100 通过在具有承载病毒的可执行文件附件的电子邮件或电子邮件中向公共网络 102（例如因特网）发送被病毒感染的消息。该消息或者被寻址到或者被病毒的动作传播到多个目的地，例如病毒信息源 104 和垃圾邮件陷阱 (spamtrap)106。垃圾邮件陷阱是用于收集关于不请自来的电子邮件消息的信息的电子邮件地址或电子邮件邮箱。病毒信息源 104 和垃圾邮件陷阱 106 的操作和实现在下文中更详细论述。为了图示一个简单的示例，图 1 只示出了采取病毒信息源 104 和垃圾邮件陷阱 106 形式的两个目的地，但是

在实际实施例中可能存在任意数目的这种病毒信息源。

[0051] 病毒发送者 100 可从公共来源或者通过向少量已知地址发送病毒并使得病毒传播来获得病毒信息源 104 和垃圾邮件陷阱 106 的网络地址。

[0052] 病毒信息处理器 108 可通信地耦合到公共网络 102，并且可从病毒信息源 104 和垃圾邮件陷阱 106 接收信息。病毒信息处理器 108 实现下文进一步描述的某些功能，包括从病毒信息源 104 和垃圾邮件陷阱 106 收集病毒信息、生成病毒发作信息以及将病毒发作信息存储在数据库 112 中。

[0053] 消息传递网关 107 通过防火墙 111 或其他网络元件直接或间接地从公共网络 102 耦合到专用网络 110，该专用网络 110 包括多个末端站 120A、120B、120C。消息传递网关 107 可与为专用网络 110 处理电子邮件的邮件传送代理 109 相集成，或者邮件传送代理可以单独部署。例如，可从 IronPort Systems，Inc.，San Bruno，California 购得的 IronPort MessagingGateway Appliance(MGA)（例如 C60、C30 或 C10 型）可实现邮件传送代理 109、防火墙 111 和这里针对消息传递网关 107 描述的功能。

[0054] 在一个实施例中，消息传递网关 107 包括病毒信息逻辑 114，用于从病毒信息处理器 108 获得病毒发作信息并且根据在消息传递网关处设置的方针来处理以末端站 120A、120B、120C 为目的地的消息。如这里进一步描述的，病毒发作信息可包括若干类型的信息中的任何一种，其中包括（但不限于）病毒得分值和将病毒得分值与和病毒相关联的消息特性关联起来的一个或多个规则。如参考图 3 进一步描述的，这种病毒信息逻辑可与消息传递网关 107 的内容过滤器功能相集成。

[0055] 在一个实施例中，病毒信息逻辑 114 被实现为消息传递网关 107 中的独立的逻辑模块。消息传递网关 107 利用消息数据来调用病毒信息逻辑 114，并且接收作为响应的裁决。该裁决可以基于消息试探。消息试探为消息评分并确定消息是病毒的可能性。

[0056] 病毒信息逻辑 114 部分基于消息的参数来检测病毒。在一个实施例中，病毒检测是基于以下之中的任何一个或多个来执行的：包含可执行代码的邮件的试探；不匹配的消息头部的试探；来自未知的开放转发器的邮件的试探；具有不匹配的内容类型和扩展名的邮件的试探；来自动态用户列表、列入黑名单的主机或已知声望不佳的发送者的邮件的试探；以及发送者真实性测试结果。发送者真实性测试结果可由从公共网络接收发送者 ID 值的逻辑来生成。

[0057] 消息传递网关 107 还可包括防病毒检查器 116、内容过滤器 118 和防垃圾邮件逻辑 119。防病毒检查器 116 例如可包括 Sophos 防病毒软件。内容过滤器 118 提供这样的逻辑，该逻辑用于根据与专用网络 110 相关联的方针来限制对在消息主题或正文中包含不可接受的内容的消息的递送或接受。

[0058] 防垃圾邮件逻辑 119 扫描传入的消息，以根据邮件接受方针确定它们是否是不想要的，例如传入消息是否是不请自来的商业电子邮件，并且防垃圾邮件逻辑 119 应用方针以限制对任何不想要的消息的递送、对任何不想的消息进行重定向或者拒绝接受任何不想要的消息。在一个实施例中，防垃圾邮件逻辑 119 扫描消息，并且为每个消息返回 0 到 100 之间的得分，该得分指示出该消息是垃圾邮件或另一类型的不想要的电子邮件的概率。得分范围与可由管理员定义的可能的垃圾邮件和很可能的垃圾邮件的阈值相关联，其中对于所述可能的垃圾邮件和可能的垃圾邮件，用户可应用下文进一步描述的一组指定动作。在

一个实施例中，得分为 90 或更高的消息是垃圾邮件，得分为 75-89 的消息是疑似垃圾邮件。

[0059] 在一个实施例中，防垃圾邮件逻辑 119 至少部分基于声望信息来确定垃圾邮件得分，该声望信息是从数据库 112 或者例如来自 IronPort Systems，Inc. 的 SenderBase 这样的外部声望服务获得的，它指示出消息的发送者是否与垃圾邮件、病毒或其他威胁相关联。扫描可包括将 X 头部记录在经扫描的消息中，该 X 头部证实消息已被成功地扫描，并且包括标识出为消息匹配的规则的被扰乱的字符串。扰乱可包括基于私钥和单向散列算法来创建规则标识符的散列。扰乱确定了只有指定的一方（例如图 7 的服务提供者 700）才能够对匹配的规则进行解码，从而提供系统的安全性。

[0060] 专用网络 110 可以是与商业企业相关联的企业网络，或者是任何其他形式的需要更强的安全性或保护的网络。公共网络 102 和专用网络 110 可使用例如 TCP/IP 这样的开放标准协议来进行通信。

[0061] 病毒信息源 104 可以包括介于公共网络 102 和另一专用网络（为清楚起见没有示出）之间的消息传递网关 107 的另一实例，用于保护该另一专用网络。在一个实施例中，病毒信息源 104 是 IronPort MGA。垃圾邮件陷阱 106 与一个或多个电子邮件地址或电子邮件邮箱相关联，该一个或多个电子邮件地址或电子邮件邮箱与一个或多个域相关联。垃圾邮件陷阱 106 被建立来用于接收不请自来的电子邮件消息或者说“垃圾邮件”以便分析或报告，并且通常不用于传统的电子邮件通信。例如，垃圾邮件陷阱可以是例如“dummyaccountforspam@mycompany.com”这样的电子邮件地址，或者垃圾邮件陷阱可以是这样的电子邮件地址的集合：这些电子邮件地址被聚集成一个邮件交换 (MX) 域名系统 (DNS) 记录，其中接收到的电子邮件信息被提供给该记录。邮件传送代理 109 或者另一 IronPort MGA 的邮件传送代理可以是主机垃圾邮件陷阱 106。

[0062] 在一个实施例中，病毒信息源 104 生成并向病毒信息处理器 108 提供信息以用于管理计算机病毒发作，并且病毒信息处理器 108 可从垃圾邮件陷阱 106 获得信息以用于同样的目的。例如，病毒信息源 104 生成接收到的具有可疑附件的消息的计数，并且将该计数提供给病毒信息处理器 108，或者允许外部过程检索该计数并将它们存储在专用数据库中。消息传递网关 107 还可通过下述方式来充当病毒信息源：检测具有与病毒相关联或者因其他原因而可疑的指示的消息，创建在特定的一段时间中接收到的可疑消息的计数，并且周期性地将该计数提供给病毒信息处理器 108。

[0063] 作为具体示例，这里描述的功能可实现为综合性消息数据收集和报告设备的一部分，所述设备例如是来自 IronPort Systems，Inc. 的 SenderBase 服务。在该实施例中，病毒信息处理器 108 可从病毒信息源 104 和垃圾邮件陷阱 106 检索或接收信息，生成具有可疑附件或其他病毒指示符的消息的计数，并且利用计数更新数据库 112 并生成病毒发作信息以供消息传递网关 107 的病毒信息逻辑 114 在以后检索和使用。与 SenderBase 服务相关的方法和装置在 2004 年 5 月 28 日递交的 Robert Brahms 等人的题为“TECHNIQUES FOR DETERMINING THE REPUTATION OF AMESSAGE SENDER”的共同未决的申请 No. 10/857,641 中有所描述，特此通过引用将其全部内容并入，就如同在这里完整阐述了一样。

[0064] 作为附加或替换，病毒信息源 104 可包括在万维网上的“spamcop.net”域可访问到的 SpamCop 信息服务，或者 SpamCop 服务的用户。病毒信息源 104 可包括一个或多个因特网服务提供者或其他大量邮件接收者。

[0065] SenderBase 和 SpamCop 服务提供了强大的用于检测病毒的数据源。这些跟踪通过垃圾邮件陷阱地址、末端用户投诉报告、DNS 日志和第三方数据源来跟踪关于每日的数百万消息的信息。该数据可被用于利用这里的方法迅速地检测病毒。具体而言，发送到合法地址或者垃圾邮件陷阱地址并且没有被防病毒扫描器识别为病毒的具有特定附件类型的消息的数目（相对于正常级别）提供了如下的提早警告指示符：基于防病毒扫描器尚不知晓并且尚不能够检测到的一种新病毒，已经发生了病毒发作。

[0066] 在另一替换实施例中，作为对这里的自动方法的补充，病毒信息源 104 可包括对由信息服务顾问或分析者或者外部来源所获得的数据的手工审查。例如，监视来自防病毒卖家、第三方卖家、安全邮件列表、垃圾邮件陷阱数据和其他来源的警告的人类管理员在大多数情况下都可早在病毒定义被公布之前就能检测到病毒。

[0067] 一旦基于病毒发作信息识别出了病毒发作，例如消息传递网关 107 这样的网络元件就可提供各种选项，用于基于消息是病毒的概率来处理消息。当消息传递网关 107 与邮件传送代理或者邮件网关相集成时，网关可立即对该数据作出反应。例如，邮件传送代理 109 可延迟将消息递送到专用网络 110 中，直到从防病毒卖家接收到病毒更新并且病毒更新被安装在消息传递网关 107 上为止，从而使得延迟的消息可在病毒更新被接收到之后被防病毒检查器 116 所扫描。

[0068] 延迟地消息可被存储在隔离队列 316 中。隔离队列 316 中的消息可根据进一步描述的各种方针被释放和递送、被删除或者在递送之前被修改。在一个实施例中，在消息传递网关 107 中建立了多个隔离 316，并且对于被管理的专用网络 110 中的计算机 120A、120B 等等的每个接收者账户，有一个隔离与之相关联。

[0069] 虽然在图 1 中没有示出，但病毒信息处理器 108 可包括或者可通信地耦合到病毒发作操作中心 (VOOC)、接收病毒得分 (RVS) 处理器或者两者。VOOC 和 RVS 处理器可与病毒信息处理器 108 相分离，但可通信地耦合到数据库 112 和公共网络 102。VOOC 可实现为配有人员的中心，其中每天 24 小时、每周 7 天都有员工监视由病毒信息处理器 108 收集并存储在数据库 112 中的信息。配备给 VOOC 的员工可采取手工动作，例如发出病毒发作警告、更新存储在数据库 112 中的信息、公布病毒发作信息以便消息传递网关 107 可访问病毒发作信息以及手工地发起将病毒发作信息发送到消息传递网关 107 和其他消息传递网关 107 的操作。

[0070] 此外，配备给 VOOC 的员工可将邮件传送代理 109 配置为执行某些动作，例如递送“软反弹 (soft bounce)”。软反弹是在邮件传送代理 109 基于邮件传送代理 109 可访问的一组规则而返回接收到的消息时执行的。更具体地说，当邮件传送代理 109 通过接受来自发送者的电子邮件消息而完成 SMTP 事务时，邮件传送代理 109 基于邮件传送代理 109 可访问的一组存储的软件规则来确定接收到的消息是不想要的或者不可递送的。响应于确定接收到的消息是不想要的或者不可递送的，邮件传送代理 109 将消息返回到由发送者指定的反弹电子邮件地址。当邮件传送代理 109 将消息返回给发送者时，邮件传送代理 109 可从消息中剥离任何附件。

[0071] 在一些实现方式中，病毒发作信息是响应于由员工（例如配备给 VOOC 的员工）采取的手工动作而提供或公布的。在其他实现方式中，病毒发作信息是根据病毒信息处理器、VOOC 或 RVS 的配置而自动提供的，然后病毒发作信息和所采取的自动动作被 VOOC 处的员工

所审查，如果该员工认为必要或者需要的话可以进行修改。

[0072] 在一个实施例中，VOOC 或系统组件处配备的员工根据一个实施例可以基于多种因素来确定消息是否包含病毒，这些因素例如是 (a) 接收带附件的消息的模式、(b) 接收到的消息的附件的危险特性、(c) 公布的卖家病毒警告、(d) 增加的邮件列表活动性、(e) 消息的基于来源的危险特性、(f) 与接收到的消息的来源相关联的动态网络地址的百分比、(g) 与接收到的消息的来源相关联的计算机化主机的百分比以及 (h) 可疑容量模式的百分比。

[0073] 上述因素中的每一个可包括多种标准。例如，接收到的消息的附件的危险特性可以基于以下考虑：附件的文件名有多可疑、该文件是否与多个文件扩展名相关联、附加到接收消息的相似文件大小的量、附加到接收消息的相似文件名的量以及已知病毒的附件的名称。接收带附件的消息的模式可以基于以下考虑：包含附件的消息的数目的当前比率、接收到的具有危险附件的消息的数目趋势以及报告带附件的消息增加的客户数据源、病毒信息源 104 和垃圾邮件陷阱 106 的数目。

[0074] 此外，关于消息是否包含病毒的确定可以基于从客户端发送来的信息，例如可利用电子邮件消息将信息从用户报告到系统，该电子邮件消息在安全的环境中在系统处被接收，从而使得如果系统的消息接收者被病毒感染的话则该消息接收者会被尽可能地配置为防止计算机病毒扩散到系统的其他部分。

[0075] RVS 处理器可被实现为一个自动化系统，该系统生成将被提供给消息传递网关 107 和其他消息传递网关 107 的病毒发作信息，该病毒发作信息例如采取针对各种附件类型的病毒得分值形式或者采取将病毒得分值与消息特性关联起来的一组规则的形式。

[0076] 在一个实施例中，消息传递网关 107 包括裁决缓存 115，该裁决缓存 115 提供对来自防病毒检查器 116 和 / 或防垃圾邮件逻辑 119 的裁决值的本地存储，以便在接收到重复的消息时再次利用。裁决缓存 115 的结构和功能在下文中进一步描述。在一个实施例中，消息传递网关 107 包括日志文件 113，该日志文件 113 可存储与消息传递网关的功能相关的统计信息或状态消息。可被记录的信息的示例包括消息裁决和作为裁决的结果而采取的动作；在消息上匹配的规则（以扰乱的格式）；关于发生了扫描引擎更新的指示；关于发生了规则更新的指示；扫描引擎版本号等等。

[0077] 2.2 对可疑消息进行计数

[0078] 图 2 是根据一个实施例生成可疑消息计数的过程的流程图。在一个实施例中，图 2 的步骤可由病毒信息源执行，例如由图 1 的病毒信息源 104 执行。

[0079] 在步骤 202 中，接收到消息。例如，病毒信息源 104 或消息传递网关 107 接收由病毒发送者 100 发送的消息。

[0080] 在步骤 204 中，确定消息是否危险。在一个实施例中，如果病毒信息源 104 或消息传递网关 107 处的病毒检查器扫描了消息而没有识别出病毒，但消息也包括具有已知危险的文件类型或扩展名的文件附件，则该消息被确定为危险。例如，可以认为 MS Windows(XP Pro) 文件类型或扩展名 COM、EXE、SCR、BAT、PIF 或 ZIP 是危险的，因为病毒编写者通常将这种文件用于恶意的可执行代码。前述只是可能被认为危险的文件类型或扩展名的示例；存在多于 50 种已知的不同文件类型。

[0081] 还可以通过下述方式作出关于消息可疑的确定：从消息中提取源网络地址，例如源 IP 值，并且向 SenderBase 服务发出查询以确定是否已知该源与垃圾邮件或病毒相关联。

例如，在确定消息是否可疑时可考虑由 SenderBase 服务提供的声望得分值。如果消息是从与已知受到危害的、有发送病毒历史的或者最近才开始向因特网发送电子邮件的主机相关联的 IP 地址发送来的，则该消息也可被确定为是可疑的。确定还可以基于以下因素中的一个或多个 :(a) 直接附加到消息的文件附件的类型或扩展名、(b) 压缩文件、档案、.zip 文件或直接附加到消息的另一文件内包含的文件的类型或扩展名以及 (c) 从附件获得的数据指纹。

[0082] 此外，关于可疑消息的确定可以基于可疑消息的附件的大小、可疑消息的主题的内容、可疑消息的正文的内容或者可疑消息的任何其他特性。一些文件类型可能嵌入有其他文件类型。例如，“.doc”文件和“.pdf”文件可能嵌入有其他图像文件类型，例如“.gif”或“.bmp”。在确定消息是否可疑时可以考虑宿主文件类型内的任何嵌入文件类型。可疑消息的特性可被用于制定规则，这些规则被提供给消息传递网关 107 并且包括与一个或多个这种特性相关联的病毒得分值。

[0083] 在步骤 206 中，如果消息可疑，则当前时间段的可疑消息计数被递增。例如，如果消息具有 EXE 附件，则具有 EXE 附件的消息的计数被递增 1。

[0084] 在步骤 208 中，报告可疑消息的计数。例如，步骤 208 可包括向病毒信息处理器 108 发送报告消息。

[0085] 在一个实施例中，病毒信息处理器 108 实时地连续接收许多报告，例如步骤 208 的报告。当接收到报告时，病毒信息处理器 108 利用报告数据来更新数据库 112，并确定和存储病毒发作信息。在一个实施例中，病毒发作信息包括根据下文参考图 4 进一步描述的子过程来确定的病毒得分值。

[0086] 2.3 基于病毒发作信息来处理消息

[0087] 图 3 是根据一个实施例图示出基于病毒发作信息的消息处理的数据流程图。在一个实现方式中，图 3 的步骤可由 MGA 来执行，例如由图 1 中的消息传递网关 107 来执行。有利的是，通过执行图 3 所示的步骤，可以在消息被肯定地确定包含病毒之前对其作出反应。

[0088] 在块 302，向消息应用内容过滤器。在一个实施例中，应用内容过滤器包括检查消息主题、其他消息头部值和消息正文，确定内容值是否满足了用于内容过滤的一个或多个规则并且在规则得到满足时采取一个或多个动作（例如可在内容方针中指定）。块 302 的执行是可选的。因此，一些实施例可执行块 302，而其他实施例可能不执行块 302。

[0089] 另外，在块 302，检索病毒发作信息以用于后续的处理步骤。在一个实施例中，在块 302，实现图 3 的消息传递网关 107 可周期性地向病毒信息处理器 108 请求当时的病毒发作信息。在一个实施例中，消息传递网关 107 利用防止未经授权方访问病毒发作信息的安全通信协议，大约每五 (5) 分钟从病毒信息处理器 108 检索病毒发作信息。如果消息传递网关 107 不能检索病毒发作信息，则网关可使用存储网关中的最近可用的病毒发作信息。

[0090] 在块 304 中，向消息应用防垃圾邮件过程，并且根据垃圾邮件方针标记或处理看起来是不请自来消息的消息。例如，垃圾邮件消息可被无声地丢弃、被移动到指定的邮件箱或文件夹，或者消息的主题可被修改以包括例如“可能是垃圾邮件”这样的标记。块 304 的执行是可选的。因此，一些实施例可执行块 304，而其他实施例可以不执行块 304。

[0091] 在块 306 中，向消息应用防病毒过程，并且标记出看起来在消息或文件附件中包含病毒的消息。在一个实施例中，来自 Sophos 的防病毒软件实现块 306。如果消息被确定

为病毒，则在块 308 中，消息被删除、隔离在隔离队列 316 或者根据适当的病毒处理方针以其他方式被处理。

[0092] 或者，如果块 306 确定消息不是病毒，则在块 310 中，执行测试以确定以前是否对消息进行过病毒扫描。如这里进一步说明，在先前已对消息进行病毒扫描之后，可从后面的块再次到达块 306。

[0093] 如果在块 306 中消息以前已被进行过病毒扫描，则图 3 的过程假定已经利用在识别出病毒发作时成功识别病毒所必需的所有模式、规则或其他信息对防病毒过程 306 进行了更新。因此，控制传递到块 314，其中以前被扫描过的消息被递送。如果在块 310 中确定消息以前未被扫描过，则过程继续到块 312。

[0094] 在块 312 中，执行测试以确定在块 302 获得的病毒发作信息是否满足指定的阈值。例如，如果病毒发作信息包括病毒得分值 (VSV)，则检查该病毒得分值以查明该病毒得分值是否等于或大于阈值病毒得分值。

[0095] 阈值是由管理员命令在配置文件中指定的，或者是在一个单独的过程中从另一机器、过程或来源接收的。在一个实现方式中，阈值对应于消息包含病毒或者与新的病毒发作相关联的概率。接收到高于阈值的得分的病毒受到由操作者指定的动作，例如执行将消息隔离在隔离队列 316 中的动作。在一些实现方式中，将单个指定阈值用于所有消息，而在其他实现方式中，基于不同的特性使用多个阈值，从而管理员可以基于消息传递网关接收到的消息的类型以及对于相关联的消息接收者来说什么是正常的或者不那么危险的，来相对于其他消息更小心地对待某些消息。在一个实施例中，基于 0 到 5 的病毒得分标度，使用默认的阈值 3，其中 5 是最高的危险（威胁）级别。

[0096] 例如，病毒发作信息可包括病毒得分值，并且网络管理员可确定允许的阈值病毒得分值并将该阈值病毒得分值广播到所有消息传送代理或者执行图 3 的过程的其他处理器。又例如，病毒发作信息可包括将病毒得分值与指示出病毒的一个或多个消息特性关联起来的一组规则，并且基于这里参考图 5 描述的方法，可以基于消息的匹配规则来确定病毒得分值。

[0097] 由管理员设置的阈值病毒得分值的值指示出何时发起消息的延迟递送。例如，如果阈值病毒得分值是 1，则实现图 3 的消息传递网关将会在病毒信息处理器 108 所确定的病毒得分值较低时延迟消息的递送。如果阈值病毒得分值是 4，则实现图 3 的消息传递网关将会在病毒信息处理器 108 所确定的病毒得分值较高时延迟消息的递送。

[0098] 如果没有超过指定的阈值得分值，则在块 314 中，消息被递送。

[0099] 如果在块 312 中确定超过了阈值病毒得分值，并且在块 310 中确定消息以前未被扫描过，则消息被置于发作隔离队列 316 中。每个消息被标记以指定的保持时间值或者期满日期时间值，其代表消息被保持在发作隔离队列 316 中的时间段。发作隔离队列 316 的目的是将消息递送延迟足够长的时间，以便能够更新防病毒过程 306 使其考虑到与所检测到的病毒发作相关联的新病毒。

[0100] 保持时间可具有任何需要的持续长度。示例性的保持时间值可以在一 (1) 小时到二十四 (24) 小时之间。在一个实施例中，提供了十二 (12) 小时的默认保持时间值。管理员可通过向实现这里的过程的消息传递网关发出命令来随时将保持时间改变到任何优选的保持时间值。因此，保持时间值是用户可配置的。

[0101] 可提供一个或多个工具、特征或用户接口以允许操作者监视发作隔离队列和隔离的消息的状态。例如，操作者可获得当前被隔离的消息的列表，并且该列表可标识出队列中的每个消息被隔离的原因，例如满足指定阈值的消息的可应用病毒得分值或者为消息匹配的一组规则中的一个或多个规则。总结信息可由消息特性（例如文件附件的类型）提供，或者如果一组规则被使用的话则由可用规则提供。可提供一个工具来允许操作者 p 审查队列中的每个消息。可提供另一特征来允许操作者搜索满足一个或多个标准的隔离消息。可提供另一特征来模拟被处理的消息（这可被称为“跟踪”消息），以确保已正确地执行消息传递网关的配置并且正在根据病毒发作过滤器适当地处理传入的消息。

[0102] 此外，可提供一个工具，该工具示出来自病毒信息处理器、VOOC 或 RVS 的关于已识别出的特殊或重大病毒危险或威胁的一般警告信息。另外，在 MGA 中可包括工具，用于在警告被发出时联络与 MGA 相关联的一个或多个员工。例如，当消息被隔离时、当一定数目的消息已被隔离时或者当隔离队列的容量已被充满或者已经达到指定的级别时，自动电话或寻呼系统可联络指定的个人。

[0103] 消息可以以图 3 中标记为 316A、316B、316C 的路径所指示的三种方式退出发作隔离队列 316。如路径 316A 所示，当消息的指定保持时间期满时，该消息可以正常期满。结果，对于正常期满，在一个实现方式中，发作隔离队列 316 充当 FIFO（先进先出）队列。然后，基于下述假定，消息被传送回防病毒过程 306 以进行重新扫描：在保持时间期满之后，已经利用检测可能在消息中的病毒所必需的任何模式文件或其他信息更新了防病毒过程。

[0104] 如路径 316B 所示，可将消息从发作隔离队列 316 中手工释放出来。例如，响应于由管理员、操作者或者其他机器或者过程发出的命令，一个或多个消息可被从发作隔离队列 316 中释放出来。在手工释放时，在块 318 中，例如当操作者可能已经接收到指示出特定种类的消息肯定受到病毒感染的离线信息时，执行重新扫描或删除消息的操作者决定；在这种情况下，操作者可以选择在块 320 删除消息。或者，操作者可能在保持时间值期满之前就已经接收到指示出防病毒过程 306 已经响应于病毒发作而被用新的模式或其他信息所更新的离线信息。在这种情况下，操作者可选择通过将消息发送回防病毒过程 306 进行扫描来重新扫描消息，而不等待保持时间期满，如路径 319 所示。

[0105] 又例如，操作者可搜索当前保持在发作隔离队列 316 中的消息，以识别一个或多个消息。这样识别出的消息可被操作者选择来由防病毒过程 306 扫描，例如为了测试防病毒过程 306 是否已被用足以检测病毒发作中涉及病毒的信息所更新。如果对所选消息的重新扫描成功地识别了病毒，则操作者可手工地释放发作隔离队列中的一些或全部消息，以便释放的消息可被防病毒过程 306 重新扫描。但是，如果防病毒过程在所选的测试消息中未检测到病毒，则操作者可等待到稍后的时间并且重新测试一条测试消息或另一消息以确定防病毒过程 306 是否已被更新到能够检测病毒，或者操作者可以等待并且在消息的期满时间期满时让消息被释放。

[0106] 如路径 316C 所示，消息也可能例如因为发作隔离队列 316 已满而提早期满。溢出方针 322 被应用到提早期满的消息。例如，溢出方针 322 可要求消息被删除，如块 320 中所示。又例如，溢出方针 322 可要求消息的主题被附加上关于消息很可能包含病毒的适当危险警告，如块 324 所示。例如，诸如“可能被感染”或者“疑似病毒”之类的消息可被附加到主题，例如附加在消息主题行的结束或开头处。具有附加的主题的消息经由防病毒过程 306

被递送，并且因为消息以前已被扫描过，因此该过程从防病毒过程 306 继续通过块 310，然后消息被递送，如块 314 所示。

[0107] 虽然为了清楚起见在图 3 中未示出，但可以应用附加的溢出方针。例如，溢出方针 322 可要求去除消息的文件附件，然后递送已经剥离了文件附件的消息。可选地，溢出方针 322 可要求只剥离那些超过特定大小的文件附件。又例如，溢出方针 322 可要求当发作隔离队列 316 已满时允许 MTA 接收新的消息，但在 SMTP 事务期间接受消息之前利用 4xx 临时错误来拒绝消息。

[0108] 在一个实施例中，根据路径 316A、316B、316C 对消息的处理可以由用户针对隔离队列的全部内容来配置。或者，这种方针可以由用户针对每个消息来配置。

[0109] 在一个实施例中，块 312 还可包括在从病毒信息处理器 108 获得的病毒发作信息满足指定阈值时，例如在病毒得分值达到或超过指定的阈值病毒得分值时生成并发送警告消息到一个或多个管理员。例如，在块 312 发送的警告消息可包括一封电子邮件，该电子邮件指定其病毒得分已被改变的附件类型、当前病毒得分、先前病毒得分、当前阈值病毒得分以及从病毒信息处理器 108 接收到该类附件的病毒得分的最近更新的时间。

[0110] 在另一实施例中，图 3 的过程涉及每当隔离队列中的消息总数超过管理员设置的阈值时或者当已经超过特定量或百分比的隔离队列存储容量时生成并发送警告消息到一个或多个管理员。这种警告消息可指定隔离队列大小、已用容量百分比等等。

[0111] 发作隔离队列 316 可具有任何所需的大小。在一个实施例中，隔离队列可存储约 3GB 的消息。

[0112] 2.4 生成病毒发作信息

[0113] 在一个实施例中，生成基于一个或多个消息特性指示出病毒发作的可能性的病毒发作信息。在一个实施例中，病毒发作信息包括数值，例如病毒得分值。病毒发作信息可与消息的一个或多个特性相关联，例如消息的附件类型、附件的大小、消息的内容（例如消息的主题行或者消息的正文的内容）、消息的发送者、消息的发送者的 IP 地址或域、消息的接收者、消息的发送者的 SenderBase 声望得分或者任何其他适当的消息特性。作为具体示例，病毒发作信息可将一个消息特性与一个病毒得分值关联起来，例如“EXE ＝ 4”，以指示出具有 EXE 类型附件的消息的病毒得分值“4”。

[0114] 在另一实施例中，病毒发作信息包括一个或多个规则，每个规则将病毒发作的可能性与一个或多个消息特性关联起来。作为具体示例，“如果为 EXE 并且大小＜ 50k，则为 4”这样形式的规则指示出对于带有 EXE 类型并且大小小于 50k 的附件的消息，病毒得分值为“4”。可以向消息传递网关提供一组规则，这组规则将被应用以确定传入的消息是否与规则的消息特性相匹配，从而指示出该规则可应用到该传入消息，因此应当基于相关联的病毒得分值来处理。对一组规则的使用在下文中参考图 5 来进一步描述。

[0115] 图 4 是根据一个实施例确定病毒得分值的方法的流程图。在一个实现方式中，图 4 的步骤可由病毒信息处理器 108 基于数据库 112 中从病毒信息源 104 和垃圾邮件陷阱 106 接收来的信息来执行。

[0116] 图 4 的步骤 401 指示出对于病毒信息处理器 108 可访问的每个不同的病毒信息源（例如病毒信息源 104 或垃圾邮件陷阱 106）执行某些计算步骤 402、404。

[0117] 步骤 402 包括利用给与更新近的先前病毒得分值以更大的权重的加权方法，通过

组合先前时间段的一个或多个先前病毒得分值来生成特定电子邮件文件附件类型的加权后当前平均病毒得分值。特定时间段的病毒得分值是指基于在特定来源处接收的具有可疑文件附件的消息的数目的得分值。消息在以下情况下被认为具有可疑附件：如果该附件满足一个或多个度量，例如特定的文件大小、文件类型等等，或者如果已知发送者的网络地址与先前的病毒发作相关联。该确定可以基于附件文件大小或文件类型或扩展名。

[0118] 病毒得分值的确定还可通过以下方式来进行：从消息中提取源网络地址，例如源 IP 地址值，并且向 SenderBase 服务发出查询以确定是否已知该源与垃圾邮件或病毒相关联。该确定还可以基于：(a) 直接附加到消息的文件附件的类型或扩展名、(b) 压缩文件、档案、.zip 文件或直接附加到消息的另一文件内包含的文件的类型或扩展名以及 (c) 从附件获得的数据指纹。对于在前述任何一种中找到的每个附件类型，可以生成和存储一个单独的病毒得分值。另外，可以基于消息中找到的最危险的附件类型来生成和存储病毒得分值。

[0119] 在一个实施例中，步骤 402 包括针对给定的文件附件类型，计算最后三个 15 分钟时段的病毒得分值的组合。另外，在一个实施例中，向 15 分钟时段的三个值应用加权值，其中最近的 15 分钟时间段的权重重于较早的 15 分钟时间段。例如，在一种加权方法中，向最老的 15 分钟时段 (30-45 分钟之前) 的病毒得分值应用 0.10 的乘数，向次老的 15 分钟时段 (15-30 分钟之前) 应用 0.25 的乘数，并且向 0-15 分钟之前时段的最近病毒得分值应用 0.65 的乘数。

[0120] 在步骤 404 中，通过将在步骤 402 确定的当前平均病毒得分值与长期平均病毒得分值相比较，来为特定的文件附件类型生成正常百分比病毒得分值。可以参考在 30 日时段内的所有 15 分钟时间段中该文件附件类型的 30 日平均值，来计算当前正常百分比级别。

[0121] 在步骤 405 中，所有来源 (例如病毒信息源 104 和垃圾邮件陷阱 106) 的所有正常百分比病毒得分值都被取平均，从而创建特定文件附件类型的总体正常百分比值。

[0122] 在步骤 406 中，总体正常百分比值被映射到特定文件附件类型的病毒得分值。在一个实施例中，病毒得分值是 0-5 之间的整数，并且总体正常百分比值被映射到病毒得分值。表 1 给出了病毒得分标度的示例。

[0123] 表 1- 示例性病毒得分标度

[0124]

正常百分比	得分	威胁级别
0 – 150	0	没有已知威胁/非常低的威胁
150 – 300	1	可能有威胁
300 – 900	2	小威胁
900 – 1500	3	中等威胁
> 1500	4	高威胁/极危险

[0125] 在其他实施例中，可以使用到 0 至 100、0 至 10、1 至 5 的得分值或任何其他所需的取值范围的映射。除了整数得分值外，也可使用非整数得分值。取代使用限定的取值范围，可以确定概率值，例如 0％至 100％范围内的概率，其中较高的概率指示出病毒发作的可能性较大，或者 0 至 1 范围内的概率，其中概率被表达为分数或小数，例如 0.543。

[0126] 作为一种优化，为了避免对于极低的 30 日计数可能发生的被零除 (division by zero) 问题，图 4 的过程可以将步骤 402 中计算的基线平均值加 1。实质上，加 1 通过减弱一些数据而有益地略微提高了值的噪声级别。

[0127] 表 2 给出了假设实施例中 EXE 文件类型的示例性数据：

[0128] 表 2-“.exe”文件类型的示例性数据

[0129]

来源	30 日平均值	当前“.exe”计数、40 分钟之前的、30 分钟之前的、15 分钟之前的	当前平均值	当前“.exe”的正常百分比
来源 1	3.6	21、40、3	14	382%
来源 2	15.4	50、48、7	21.6	140%
来源 3	1.7	1、1、15	10.1	600%
来源 4	1.3	15、15、15	15	1200%
平均正常百分比				581%
病毒得分				2

[0130] 在替换实施例中，图 2、图 3、图 4 的过程还可包括识别所报告的数据的趋势并识别病毒得分计算的异常的逻辑。

[0131] 由于大部分可执行文件是通过一类或另一类电子邮件附件来扩散的，因此这里的方法的策略集中于基于附件类型来作出方针判决。在替换实施例中，可通过考虑其他消息数据和元数据来形成病毒得分值，所述其他消息数据和元数据例如是消息中的通用资源定位符 (URL)、文件附件的名称、源网络地址等等。另外，在替换实施例中，可以为个体消息而不是文件附件类型分配病毒得分值。

[0132] 在另一实施例中，可以考虑其他度量来确定病毒得分值。例如，如果突然从以前从未向病毒信息处理器 108 或其信息来源发送过消息的新的主机接收到大量消息，则可以指示病毒。因此首次见到特定消息的日期较新近这一事实以及由病毒信息处理器 108 检测到的消息量峰值，可提供关于病毒发作的提早指示。

[0133] 2.5 使用病毒发作信息

[0134] 如上所述，病毒发作信息可简单地将病毒得分值与消息特性（例如附件类型）关联起来，或者病毒发作信息可包括一组规则，其中每个规则将病毒得分值与指示病毒的一个或多个消息特性关联起来。MGA 可向传入消息应用该组规则，以确定哪些规则与消息匹配。基于匹配传入消息的规则，MGA 可例如通过基于来自匹配规则的一个或多个病毒得分值确定病毒得分值，来确定消息包括病毒的可能性。

[0135] 例如，某一规则可以是“如果是‘exe’，则为 4”，以表示带有 EXE 附件的消息的病毒得分为 4。又例如，某一规则可以是“如果是‘exe’并且大小＜ 50k，则为 3”，以表示带有大小小于 50k 的 EXE 附件的消息的病毒得分为 3。又例如，某一规则可以是“如果 SBRS ＜ -5，

则为 4”，以表示如果 SenderBase 声望得分 (SBRS) 小于“-5”，则病毒得分为 4。又例如，某一规则可以是“如果是‘PIF’并且主题包含 FOOL，则为 5”，以表示如果消息具有 PIF 类型的附件并且消息的主题包括“FOOL”字符串，则病毒得分为 5。一般来说，规则可将任意数目的消息特性或可用于确定病毒发作的其他数据与关于匹配该消息特性或其他数据的消息包括病毒的可能性的指示符关联起来。

[0136] 另外，消息传递网关可应用例外，例如一个或多个隔离方针形式的例外，来确定否则即满足基于匹配规则而确定的（例如在图 3 的块 312 中确定的）基于病毒得分值的指定阈值的消息是否将要被放到发作隔离队列中，或者该消息是否要在不被放进发作隔离队列中的情况下被处理。MGA 可被配置为应用一个或多个应用规则的方针，例如总是允许消息被递送到一个电子邮件地址或一组电子邮件地址而不考虑病毒得分，或者总是递送带有指定类型的附件的消息，例如包含 PDF 文件的 ZIP 文件。

[0137] 一般来说，通过使病毒信息处理器提供规则而不是病毒得分值，每个 MGA 可以按由 MGA 的管理员所确定的方式来应用这些规则中的一些或全部，从而提供附加的灵活性来满足特定 MGA 的需要。结果，即使两个消息传递网关 107 使用同一组规则，每个 MGA 的管理员对规则应用进行配置的能力也意味着每个 MGA 可处理同一消息并且就所确定出的发生病毒攻击的可能性获得不同的结果，并且每个 MGA 可处理同一消息并且采取不同的动作，这取决于 MGA 的管理员所建立的配置。

[0138] 图 5 是根据一个实施例图示出用于管理病毒发作的一组规则的应用的流程图。图 5 所示的功能可由消息传递作为块 312 的一部分执行，或者在对传入消息的处理期间的任何其他适当的位置执行。

[0139] 在块 502 中，消息传递网关识别传入消息的消息特性。例如，消息传递网关 107 可确定消息是否具有附件，并且如果有的话，确定附件类型、附件大小以及附件名称。又例如，消息传递网关 107 可以基于发送方 IP 地址查询 SenderBase 服务，以获得 SenderBase 声望得分。为了描述图 5，假定消息具有 EXE 类型的大小为 35k 的附件，并且消息的发送主机具有 -2 的 SenderBase 声望得分。

[0140] 在块 504 中，消息传递网关基于消息的消息特性来确定规则组中的哪些规则匹配。例如，假定为了描述图 5，该规则组由以下五个将示例性的特性与所提供的假设病毒得分值关联起来的规则组成：

[0141] 规则 1：“如果是 EXE，则为 3”

[0142] 规则 2：“如果是 ZIP，则为 4”

[0143] 规则 3：“如果是 EXE 并且大小＞ 50k，则为 5”

[0144] 规则 4：“如果是 EXE 并且大小＜ 50k 且大小＞ 20k，则为 4”

[0145] 规则 5：“如果 SBRS ＜ -5，则为 4”

[0146] 在这些示例性规则中，规则 1 指示出 ZIP 附件比起 EXE 附件来更可能包括病毒，因为在规则 2 中病毒得分为 4，而在规则 1 中仅为 3。另外，上述示例性规则指示出大小大于 50k 的 EXE 附件更可能具有病毒，但大小小于 50k 但大于 20k 的 EXE 附件稍微不那么可能包括病毒，这也许是因为大多数带有 EXE 附件的可疑消息的大小都大于 50k。

[0147] 在消息具有 EXE 类型的大小为 35k 的附件并且与 -2 的 SenderBase 声望得分相关联的当前示例中，规则 1 和 4 匹配，而规则 2、3 和 5 不匹配。

[0148] 在块 506 中，消息传递网关基于来自匹配规则的病毒得分值来确定将要用于消息的病毒得分值。对将要用于消息的病毒得分值的确定可以基于多种方法中的任何一种来执行。所使用的特定方法可由消息传递网关的管理员来指定，并根据需要被修改。

[0149] 例如，可以使用在按列表顺序应用规则列表时首先匹配的规则，并忽略任何其他匹配的规则。从而，在本示例中，首先匹配的规则是规则 1，因此消息的病毒得分值是 3。

[0150] 又例如，使用具有最高病毒得分值的匹配规则。从而，在本示例中，规则 3 在匹配规则中具有最高病毒得分值，因此消息的病毒得分值是 5。

[0151] 又例如，使用具有最具体的一组消息特性的匹配规则。从而，在本示例中，规则 4 是最具体的匹配规则，因为规则 4 包括三个不同的标准，因此消息的病毒得分值是 4。

[0152] 又例如，来自匹配规则的病毒得分值可被组合以确定应用到消息的病毒得分值。作为具体示例，来自规则 1、3 和 4 的病毒得分可被取平均，以确定病毒得分值 4（例如 (3+4+5) ÷ 3 = 4）。又例如，可以使用匹配规则的病毒得分值的加权平均值，以向更具体的规则给予更大的权重。作为具体示例，每个病毒得分值的权重可以等于该规则中的标准数目（例如具有一个标准的规则 1 的权重为 1，而具有三个标准的规则 4 的权重为 3），从而规则 1、3 和 4 的加权平均产生病毒得分值 4.2（例如 (1*3+2*5+3*4) ÷ (1+2+3) = 4.2）。

[0153] 在块 508 中，消息传递网关使用在块 506 中确定的病毒得分值来确定指定的阈值病毒得分值是否得以满足。例如，假定在本示例中，阈值是病毒得分值 4。结果，在块 506 中由所有示例性方法确定出的病毒得分值都将满足该阈值，除了使用第一规则来匹配并且块 506 确定出的病毒得分值为 3 的第一示例之外。

[0154] 如果在块 508 中确定病毒得分值满足了指定的阈值，则在块 510 中，应用一个或多个隔离方针以确定是否将消息添加到发作隔离队列。例如，消息传递网关的管理员可以确定即使检测到了病毒发作，一个或多个用户或一组或多组用户的消息也应当永不被隔离。又例如，管理员可建立这样一条方针，即当病毒发作信息基于指定的阈值指示出病毒攻击时，具有某些特性的消息（例如带有大小至少为 75k 的 XLS 附件的消息）总是会被递送而不是被隔离。

[0155] 作为具体示例，机构的法律部门的成员可能频繁地接收到包含重要的法律文档的 ZIP 文件，这些文件不应当因为被放在发作隔离中而被延迟，即使消息传递网关确定正在发生病毒发作也是如此。从而，消息传递网关的邮件管理员可建立一条方针，以总是将带有 ZIP 附件的消息递送到法律部分，即使 ZIP 附件的病毒得分值达到或超过了指定的阈值。

[0156] 作为另一具体示例，邮件管理员可能希望使寻址到邮件管理员的电子邮件地址的消息总是被递送，因为这样的消息可提供用于对付病毒发作的信息。假定邮件管理员是老练的用户，则递送受病毒感染的消息的危险很低，因为邮件管理员很可能能够在病毒能够起作用之前识别并对付受感染的消息。

[0157] 对于在描述图 5 时使用的示例，假定邮件管理员已经建立了一条方针，即寻址到公司的高级工程经理的 EXE 附件总是被递送，即使这种消息的病毒得分值达到或超过阈值病毒得分值。从而，如果消息寻址到任何高级工程经理，则消息仍然被递送而不是被放到发作隔离中。但是，寻址到除高级工程经理外的其他人的消息则被隔离（除非被另一可应用方针所排除）。

[0158] 在一个实施例中，消息传递网关可被配置为处于两个状态之一中：“平静”和“紧

张”。如果没有消息被隔离，则平静状态适用。但是，当病毒发作信息被更新并且指示出指定的阈值被超过时，状态从平静变化到紧张，而不论消息传递网关所接收到的任何消息是否正被隔离。紧张状态一直持续，直到病毒发作信息被更新并且指示出指定阈值不再被超过为止。

[0159] 在一些实现方式中，每当发生系统状态变化（例如从平静到紧张或者从紧张到平静）时，警告消息就被发送到操作者或管理员。此外，当先前不满足阈值的低病毒得分值现在达到或超过阈值时，警告可被发出，即使系统的整体状态没有变化（例如，系统先前从平静变化到紧张，并且在处于紧张状态中时，从病毒信息处理器接收到也达到或超过阈值的另一病毒得分）。类似地，当先前满足阈值的高病毒得分已经下降并且现在小于指定的阈值时，警告可被发出。

[0160] 警告消息可包括一类或多类信息，包括（但不限于）以下信息：为其改变病毒发作信息的附件类型、当前病毒得分、先前病毒得分、当前阈值以及病毒发作信息的最近更新发生的时间。

[0161] 2.6 附加特征

[0162] 除了上述特征之外，在特定实现方式中可使用以下附加特征中的一个或多个。

[0163] 一个附加特征是获得特别设计来帮助识别病毒威胁的基于发送者的数据。例如，当 MGA 查询例如 SenderBase 这样的服务以获得连接 IP 地址的 SenderBase 声望得分时，SenderBase 可提供特定于该连接 IP 地址的病毒威胁数据。病毒威胁数据是基于 SenderBase 为该 IP 地址收集的数据的，并且反映了该 IP 地址就以下方面而言历史：在源自该 IP 地址或与该 IP 地址相关联的公司的消息中检测到病毒的频率如何。这使得 MGA 能够仅仅基于消息的发送者而从 SenderBase 获得病毒得分，而不需要关于来自发送方 IP 地址的特定消息的内容的任何信息或知识。关于发送者的病毒威胁的数据可被用来取代或附加于如上所述确定的病毒得分，或者关于发送者的病毒威胁的数据可被考虑在病毒得分的计算中。例如，MGA 可以基于 SE 听病毒威胁数据来增大或减小特定的病毒得分值。

[0164] 另一特征是在动态或拨号主机直接连接到外部 SMTP 服务器时，使用动态或拨号黑名单来识别很有可能被病毒感染的消息。正常情况下，预期连接到因特网的动态和拨号主机会通过主机的本地 SMTP 服务器发送传出消息。但是，如果主机被病毒感染，则病毒可能导致主机直接连接到外部 SMTP 服务器，例如 MGA。在这种情况下，该主机被导致主机建立到外部 SMTP 服务器的直接连接的病毒所感染的可能性较高。示例包括垃圾邮件和开放转发阻止系统 (SORBS) 动态主机以及非另一伪列表 (not justanother bogus list，缩写为 NJABL) 动态主机。

[0165] 但是，在一些情况下，直接连接不是病毒发起的，例如当新手用户进行直接连接时或者当连接来自诸如 DSL 或者线缆调制解调器这样的非动态宽带主机时。然而，这种从拨号或动态主机到外部 SMTP 服务器的直接连接可能导致确定出高病毒得分或者增加已经确定的病毒得分，以反映直接连接是由病毒引起的可能性的增大。

[0166] 另一个特征是使用一个跟踪过去已被病毒所开拓的主机的已开拓主机黑名单来作为病毒信息源。主机可能在服务器是开放转发器、开放代理或者具有另一个允许任何人向任何地方递送电子邮件的弱点时被开拓。已开拓主机黑名单利用以下技术之一来跟踪已被开拓的主机：感染主机的内容正在发送和定位已经经由连接时间扫描而感染的主机。示

例包括使用来自复合阻止列表 (CBL) 和开放代理监视器 (OPM) 的数据的开拓黑名单 (XBL) 和分布式服务器抵制列表 (DSBL)。

[0167] 另一个特征是供病毒信息处理器形成具有过去发送病毒历史的发送者和网络的黑名单。例如,最高的病毒得分可被分配给已知仅发送病毒的个体 IP 地址。中等病毒得分可与已知既发送病毒又发送未被病毒的合法消息的个体 IP 地址相关联。中到低病毒得分可被分配给包含一个或多个个体的受感染主机的网络。

[0168] 另一个特征是除了以上所述之外,并入更宽的一组测试来识别可疑消息,例如识别附件特性。例如,通用头部测试可用于测试任何通用消息头部,以查找固定的字符串或者正则表达式,例如在下面的示例中 :

[0169] head X_MIME_FOO X-Mime =~ /foo/

[0170] head SUBJECT_YOUR Subject =~ /your document/

[0171] 又例如,通用正文测试可用于通过搜索固定字符串或正则表达式来测试消息正文,例如在下面的示例中 :

[0172] body HEY_PAL/hey pal|long time,no see/

[0173] body ZIP_PASSWORD A.zip password is/i

[0174] 又例如,函数测试可用于制作定制的测试,以测试消息的非常具体的方面,例如在下面的示例中 :

[0175] eval EXTENSION_EXE message_attachment_ext(" .exe")

[0176] eval MIME_BOUND_FOO mime_boundary(" -/d/d/d/d[a-fj"]

[0177] eval XBLJP connecting_ip(exploited host)

[0178] 又例如,元测试可用于构建在多个特征 (如以上那些) 上以创建规则的元规则,例如在下面的示例中 :

[0179] meta VIRUS_FOO((SUBJECT_FOO1||SUBJECT_FOO2)&&BODY_FOO)

[0180] meta VIRUS_BAR(SIZE_BAR+SUBJECT_BAR+BODY_BAR > 2)

[0181] 另一个可使用的特征是将上述病毒得分确定方法扩展到一个或多个机器学习技术,以便不需要运行所有的规则并且通过使假阳性和假阴性达到最低限度来提供精确的分类。例如,可以采用以下方法中的一种或多种 :判决树,以提供离散的答案 ;感知,以提供加性得分 ;以及类贝叶斯分析,以将概率映射到得分。

[0182] 另一个特征是基于病毒的后果将来自病毒发作的威胁的严重性考虑到病毒得分确定中。例如,如果病毒导致受感染的计算机的硬盘驱动器的所有内容被删除,则病毒得分可被增大,而仅仅显示一条消息的病毒可以使得病毒得分不变或者甚至被减小。

[0183] 另一个附加特征是扩展用于处理可疑消息的选项。例如,可疑消息可被标记以指示出该消息是可疑的,其方式例如是通过向消息 (例如在主题或正文中) 添加病毒得分以便可警告用户为该消息确定出的病毒危险级别。又例如,可以创建新的消息,以警告接收者有人尝试向其发送受病毒感染的消息,或者创建包括消息的未受病毒感染部分的新的未受感染的消息。

[0184] 2.7 示例性使用情形

[0185] 下面的假设描述提供了关于这里描述的方法如何可用来管理病毒发作的示例。

[0186] 作为第一使用情形,假定一条新的名为“Sprosts.ky”病毒通过嵌入在 Microsoft

Excel 中的 Visual Basic 宏而扩散。就在病毒命中后不久，对于 .xls 附件病毒得分从 1 变到 3，并且这里的方法的用户即 Big Company 开始延迟 Excel 文件的递送。Big Company 的网络管理员接收到声明 .xls 文件现在被隔离的电子邮件。Sophos 随后在一小时之后发出警告，声明新的更新文件可用于阻止病毒。网络管理员随后确认其 IronPort C60 安装了最新的更新文件。虽然网络管理员已经为隔离队列设置了 5 小时的延时时段，但是 Excel 文件对于该公司是很重要的，因此管理员无法承受再等四个小时。因此，管理员访问 IronPort C60 并手工地冲刷队列，在附加的 Excel 文件经过 Sophos 防病毒检查的情况下发送所有消息。管理员发现这些消息中有 249 个是病毒阳性的，并且有 1 未被 Sophos 抓住，因为它未被感染。这些消息在总共延迟 1-1/2 小时后被递送。

[0187] 作为第二使用情形，假定“Clegg.P”病毒通过加密的 .zip 文件而扩散。Big Company 的网络管理员接收到警告病毒得分值已经猛增的电子邮件，但是管理员忽略了该警告，而是依赖于这里提供的自动处理。经过一晚上的六个小时之后，管理员接收到第二页，警告他隔离队列已经达到了 75％的容量。在管理员去上班之前，Clegg.P 已经充满了 Big Company 的隔离队列。幸运的是，网络管理员已在 IronPort C60 上设置了方针以在隔离队列溢出时像通常那样递送消息，并且在隔离队列溢出之前，Sophos 经过一夜已经发表出新的更新。在病毒得分值触发隔离队列之前只有两个用户被感染，因此管理员仅仅面临着充满的隔离队列。在所有消息都是病毒的假定下，管理员将消息从队列中冲刷出，自动删除它们以节余出 IronPortC60 上的负载。作为一种预防性方法，网络管理员开始在将来指定的一段时间中阻止所有加密的 .zip 文件。

[0188] 3.0 用于阻止垃圾邮件消息的方法

[0189] 图 7 是可用在用于阻止“垃圾邮件”消息的方法中或用于其他种类的电子邮件扫描过程的系统的框图。在此上下文中，术语“垃圾邮件 (spamn)”是指任何不请自来的电子邮件，而术语“正常邮件 (ham)”是指合法的群发电子邮件 (bulk email)。术语“TI”是指威胁识别，也就是确定正在发生病毒发作或垃圾邮件通信。

[0190] 在服务提供者 700 内，一个或多个 TI 开发计算机 702 耦合到全集 (corpus) 服务器集群 706，该全集服务器集群 706 容宿着威胁识别规则的全集或主库，并且在评估基础上向消息应用威胁识别规则以使得生成得分值。服务提供者 700 的邮件服务器 704 向全集服务器集群 706 贡献正常电子邮件。一个或多个垃圾邮件陷阱 716 向全集贡献垃圾电子邮件。垃圾邮件陷阱 716 是这样的电子邮件地址 :这些地址是建立并播种给垃圾邮件发送者的，以便这些地址只接收垃圾电子邮件。在垃圾邮件陷阱 716 处接收的消息可被变换成存储在全集服务器集群 706 中的消息签名或校验和。一个或多个化身 (avatar)714 向全集贡献未分类电子邮件以供评估。

[0191] 由全集服务器集群 706 创建的得分被耦合到规则 /URL 服务器 707，该规则 /URL 服务器 707 向位于服务提供者 700 的客户处的一个或多个消息传递网关 107 公布与病毒、垃圾邮件和其他电子邮件威胁相关联的规则和 URL。消息传递网关 107 周期性地通过 HTTPS 传送检索新规则。威胁操作中心 (TOC)708 可生成并向全集服务器集群 706 发送试验性的规则以供测试。威胁操作中心 708 是指在确定病毒威胁并对其作出响应时涉及到的人员、工具、数据和设备。TOC 708 还向规则 /URL 服务器 707 公布被许可用于生产的规则，并且向规则 -URL 服务器发送已知不与垃圾邮件、病毒或其他威胁相关联的列入白名单的 URL。TI

小组 710 可手工地创建其他规则并将它们提供到规则 /URL 服务器。

[0192] 为了图示出清楚的示例，图 7 示出了一个消息传递网关 107。但是，在各种实施例和商业实现方式中，服务提供者 700 耦合到各种客户或客户站点处的大量现场部署的消息传递网关 107。消息传递网关 107、化身 714 和垃圾邮件陷阱 716 通过例如因特网这样的公共网络连接到服务提供者 700。

[0193] 根据一个实施例，客户消息传递网关 107 中的每一个维护一个本地 DNS URL 黑名单模块 718，该模块包括可执行逻辑和 DNS 黑名单。DNS 黑名单的结构可包括多个 DNS 类型 A 记录，这些记录将例如 IP 地址这样的网络地址映射到与该 IP 地址相关联的声望得分值。IP 地址可代表垃圾邮件消息的发送者的 IP 地址，或者与在垃圾邮件消息中找到的或者已知与诸如钓鱼式攻击或病毒之类的威胁相关联的 URL 的根域相关联的服务器地址。

[0194] 从而，每个消息传递网关 107 维护其自己的 IP 地址 DNS 黑名单。相反，在现有方法中，DNS 信息是维护在必须通过网络通信接收所有查询的全局位置中的。本方法提高了性能，因为 MGA 生成的 DNS 查询无需穿越网络才能到达中央 DNS 服务器。该方法还易于更新；中央服务器周期性地向消息传递网关 107 发送递进的更新。为了过滤垃圾邮件消息，消息传递网关 107 中的其他逻辑可以从被测消息中提取一个或多个 URL，向黑名单模块 718 提供（URL、位屏蔽）列表形式的输入，并且接收黑名单 IP 地址命中列表形式的输出。如果指示了命中，则消息传递网关 107 可阻止电子邮件的递送、隔离电子邮件或者应用其他方针，例如在递送之前将 URL 从消息中剥离。

[0195] 在一个实施例中，黑名单模块 718 还测试电子邮件中的 URL 毒害 (URL poisoning)。URL 毒害是指一种被垃圾邮件发送者使用的技术，该技术将恶意的或破坏性的 URL 放在也包含非恶意 URL 的不请自来的电子邮件消息内，从而使得点击 URL 的未存怀疑的用户可能无意间触发了恶意的本地动作、显示广告等等。“良好”URL 的存在旨在防止垃圾邮件检测软件将消息标记为垃圾邮件。在一个实施例中，黑名单模块 718 可确定作为输入提供的恶意和良好 URL 的特定组合何时表示垃圾邮件消息。

[0196] 一个实施例提供了一个系统，该系统用于取得 DNS 数据并且将其移到散列型本地数据库中，该数据库可以接受若干数据库查询并且随后接收 DNS 响应。

[0197] 前述方法可实现在构造为 SpamAssassin 开放源项目的插件的计算机程序中。SpamAssassin 由一组 Perl 模块构成，该组模块可与和 SpamAssassin 一起发运的核心程序一起使用，该核心程序提供用于执行消息检查的网络协议，例如“spamd”。SpamAssassin 的插件体系结构可通过应用编程接口来扩展；程序员可添加新的检查试探和其他功能，而无需改变核心代码。插件是在配置文件中标识的，并且是在运行时加载的并成为 SpamAssassin 的功能部分。API 定义试探的格式（检测垃圾邮件中常用的单词或短语的规则）和消息检查规则。在一个实施例中，试探是基于单词的字典的，并且消息传递网关 107 支持一个用户接口，该用户接口使得管理员能够编辑字典的内容以添加或删除不许可的单词或已知的良好单词。在一个实施例中，管理员可配置防垃圾邮件逻辑 119 以在执行其他防垃圾邮件扫描之前对照特定于企业的内容字典来扫描消息。该方法使得消息能够在包含特定于企业的术语或行业标准术语的情况下首先收到较低的得分，而不经历其他计算上昂贵的垃圾邮件扫描。

[0198] 另外，就更宽的意义上来说，前述方法使得垃圾邮件检查引擎接收并使用形成了

声望确定的基础但尚未在垃圾邮件检查中直接使用的信息。该信息可被用于修改垃圾邮件检查器的权重值和其他试探。因此，垃圾邮件检查器可更精确地确定新接收到的消息是否是垃圾邮件。另外，垃圾邮件检查器能够得知全集中的大量信息，这也提高了精度。

[0199] 3.1 从垃圾邮件扫描中提早退出

[0200] 防垃圾邮件逻辑 119 通常以完整的方式作用于每个消息，这意味着每个消息的每个元素都被完整地解析，然后每个注册的测试都被执行。这给出了关于一份消息是正常邮件还是垃圾邮件的非常精确的总评价。但是，一旦消息足够地“垃圾邮件化”，它就被标记为并视为垃圾邮件。不需要附加的信息来促成对消息的二元处置。当一个实施例实现垃圾邮件和正常邮件的阈值时，防垃圾邮件逻辑 119 的性能通过以下方式而提高 ：一旦该逻辑确定消息足够地“垃圾邮件化”以至于能够确信它就是垃圾邮件，就从消息扫描功能中退出。在本描述中，这种方法被称为从防垃圾邮件解析或扫描中提早退出。

[0201] 利用提早退出，通过不评估仅仅进一步确认消息是垃圾邮件的数百条规则，可以节省很多时间。由于通常存在很少的负得分规则，因此一旦达到某个阈值，逻辑 119 就可肯定地确定消息是垃圾邮件。利用被称为规则排序和执行 (Rule Ordering and Execution) 以及按需解析 (Parse onDemand) 的机制，也实现了两个进一步的性能增进。

[0202] 规则排序和执行是使用指示符来允许迅速地达到确定的机制。规则被排序并被放到测试群组中。在每个群组被执行并且当前的得分被检查之后，判决消息是否足够地“垃圾邮件化”。如果是，则逻辑 119 中断规则处理并宣布消息是垃圾邮件的裁决。

[0203] 按需解析只在需要时才执行作为防垃圾邮件逻辑 119 的一部分的消息解析。例如，如果仅解析消息头部就导致确定消息是垃圾邮件，则其他的解析操作不再被执行。具体而言，可应用到消息头部的规则可能是非常良好的垃圾邮件指示符 ；如果防垃圾邮件逻辑 119 基于头部规则确定消息是垃圾邮件，则正文不被解析。结果，防垃圾邮件逻辑 119 的性能提高，因为解析头部在计算上比解析消息正文更昂贵。

[0204] 又例如，如果应用到非 HTML 正文元素的规则导致垃圾邮件裁决，则消息正文被解析，但 HTML 元素被排除。解析 HTML 或测试 URI 黑名单（在下文中进一步描述）只在需要时执行。

[0205] 图 11 是利用提早退出方法执行消息威胁扫描的过程的流程图。在步骤 1102 中，接收到多个规则。规则指定指示出与消息相关联的威胁的电子消息特性。从而，当规则匹配消息元素时，该消息很可能具有威胁或者是垃圾邮件。每个规则具有一个优先级值，并且每个规则与一个消息元素类型相关联。

[0206] 在步骤 1104 中，接收到电子邮件消息，该消息具有针对接收者账户的目的地地址。该消息包括多个消息元素。这些元素通常包括头部、原始正文和 HTML 正文元素。

[0207] 在步骤 1106 中，提取下一个消息元素。如块 1106A 中所指示的，步骤 1106 可包括提取头部、原始正文或者 HTML 正文元素。作为示例，假定在步骤 1106 只提取了消息头部。提取通常包括暂时拷贝到数据结构中。

[0208] 在步骤 1108 中，基于规则的优先级顺序，在针对同一元素类型的一组规则中选择下一规则。从而，步骤 1108 反映了对于在步骤 1106 提取的当前消息元素，只有针对该元素类型的规则被考虑，并且规则是根据其优先级顺序被匹配的。例如，如果在步骤 1106 提取了消息头部，则只有头部规则被匹配。与过去的方法不同，不同时考虑整个消息，并且不同

时考虑所有规则。

[0209] 在步骤 1109 中，通过仅将当前消息元素仅匹配到当前规则来确定消息的威胁得分值。或者，步骤 1108 和 1109 可包括选择对应于当前消息元素类型的所有规则并且将所有这样的规则匹配到当前消息元素。从而，图 11 包括通过在每个规则之后进行测试来执行提早退出，或者匹配针对特定消息元素类型的所有规则并且随后确定是否可能提早退出。

[0210] 当在步骤 1110 中测试出威胁得分值大于指定的阈值时，在步骤 1112 执行从扫描、解析和匹配的退出，并且在步骤 1114 输出该威胁得分值。结果，当在扫描、提取和规则匹配过程中阈值很早就被超过时，完成了从扫描过程的提早退出并且可以快得多地输出威胁得分值。具体而言，如果头部规则产生超过阈值的威胁得分值，则可以跳过呈现出 HTML 消息元素并将规则匹配到它们这一计算上昂贵的过程。

[0211] 但是，如果在步骤 1110 威胁得分值不大于阈值，则在步骤 1111 执行测试以确定是否已经匹配了针对当前消息元素的所有规则。在上述在步骤 1110 的测试之前匹配针对消息元素的所有规则的替换方案中，步骤 1111 是不必要的。如果对于同一消息元素类型存在其他规则，则控制返回到步骤 1108 以匹配这些规则。如果针对同一消息元素类型的所有规则已经被匹配，则控制返回到步骤 1106 以考虑下一消息元素。

[0212] 图 11 的过程可实现在防垃圾邮件扫描引擎中、防病毒扫描器中或者能够识别多种不同种类的威胁的通用威胁扫描引擎中。威胁可包括病毒、垃圾邮件或钓鱼式攻击中的任何一种。

[0213] 因此，在一个实施例中，一旦已经达到关于消息处置的确定，执行防垃圾邮件、防病毒或其他消息扫描操作的逻辑引擎就不对消息执行测试或操作。引擎将规则分组成优先级组，以便最有效并且最不昂贵的测试被提早执行。引擎被逻辑上排序以便直到特定规则或规则群组要求解析为止才进行解析。

[0214] 在一个实施例中，规则优先级值被分配给规则并且允许规则在执行时被排序。例如，优先级为 -4 的规则在优先级为 0 的规则之前运行，并且优先级为 0 的规则在优先级为 1000 的规则之前运行。在一个实施例中，规则优先级值是在规则组被创建时由管理员分配的。示例性的规则优先级包括 -4、-3、-2、-1、BOTH、VOF，并且是基于规则的功效、规则类型以及规则的成型花销来分配的。例如，非常有效并且是简单的正则表达式比较的头部规则可以是 -4（首先运行）优先级。BOTH 指示出规则对于检测垃圾邮件和病毒都是有效的。VOF 指示出规则是执行来检测病毒发作的。

[0215] 在一个实施例中，威胁识别小组 710（图 7）确定规则分组和排序并分配优先级。TI 小组 710 还可不断地评估规则的统计有效性，以确定如何对它们的执行进行排序，其中包括分配不同的优先级。

[0216] 在一个实施例中，首先消息头部被解析并且头部规则运行。接下来，消息正文解码被执行并且原始正文规则被运行。最后，HTML 元素被呈现，并且正文规则和 URI 规则被运行。在每个解析步骤之后，执行测试以确定当前的垃圾邮件得分是否大于垃圾邮件阳性阈值。如果是，则解析器退出并且后续的步骤不被执行。作为附加或替换，测试是在每个规则被运行之后执行的。

[0217] 表 3 是一个矩阵，该矩阵表述了在提早退出的实现中，防垃圾邮件逻辑 119 内的事件的示例性操作顺序。HEAD 行指示出消息 HEAD 被解析，并且头部测试被运行，并且这样的

测试支持提早退出，并且被允许具有完全的优先级范围（-4..VOF）。

[0218] 表 3- 用于提早退出的示例性操作顺序

[0219]

解析	测试（按顺序）	EE	允许的优先级
HEAD	header	提早退出	-4、-3、-2、-1、BOTH
	header_eval	提早退出	
解码	rawbody	提早退出	-3、-2、-1、BOTH
	rawbody_eval	提早退出	
呈现	body	提早退出	-2、-1、BOTH
	body_uri	提早退出	
	body_eval	提早退出	
	meta	提早退出	BOTH
VOF	VOF	无	VOF（将运行 BOTH 规则）

[0220] 3.2 垃圾邮件扫描裁决缓存

[0221] 某些垃圾邮件消息可能导致防垃圾邮件逻辑 119 需要很长的时间来确定关于消息是否是垃圾邮件的裁决。从而，垃圾邮件发送者可使用“有毒消息”攻击，这种攻击反复地发送这种困难消息，尝试强制系统管理员禁用防垃圾邮件逻辑 119。为了解决这个问题和提高性能，在一个实施例中，防垃圾邮件逻辑 119 生成的消息防垃圾邮件裁决被存储在消息传递网关 107 中的裁决缓存 115，并且防垃圾邮件逻辑 119 重复使用缓存的裁决来处理具有相同的正文的消息。

[0222] 在一种有效的实现方式中，当从缓存检索的裁决与实际扫描可能返回的裁决相同时，该裁决被称为“真实裁决”。来自缓存的不与来自扫描的裁决相匹配的裁决被称为“虚假裁决”。在一种有效的实现方式中，一些性能增进被牺牲来确保可靠性。例如，在一个实施例中，消息“主题”行的摘要被包括作为缓存的密钥的一部分，这降低了缓存命中率，但也降低了虚假裁决的机率。

[0223] 垃圾邮件发送者可能通过在其他内容都相同的正文连续消息中包括非打印的无效 URL 标签来使裁决缓存的使用无效。在消息正文内使用这种标签将会导致对于这种连续的消息，正文的消息摘要不同。在一个实施例中，可以使用一种模糊摘要生成算法，其中解析 HTML 元素并且从摘要算法的输入中去除未显示的字节。

[0224] 在一个实施例中，裁决缓存 115 被实现为来自防垃圾邮件逻辑 119 的裁决的 Python 字典。缓存的密钥是消息摘要。在一个实施例中，防垃圾邮件逻辑 119 包括 Brightmail 软件并且缓存密钥包括 DCC “fuz2”消息摘要。Fuz2 是消息正文的那些有意义地唯一的部分的 MD5 散列或摘要。Fuz2 解析 HTML 并且跳过消息中不影响用户在查看消息时看到的内容的那些字节。Fuz2 还尝试跳过消息中被垃圾邮件发送者频繁改变的部分。例如，在摘要的输入中排除以“Dear”开始的主题行。

[0225] 在一个实施例中，当防垃圾邮件逻辑 119 开始处理符合垃圾邮件或病毒扫描条件

的消息时，消息摘要被创建和存储。如果创建消息摘要失败或者如果裁决缓存 115 的使用被禁止，则摘要被设置为“无”。摘要被用作密钥以在裁决缓存 115 中执行查找，以确定是否已为具有相同的消息正文的消息存储了先前计算的裁决。术语“相同”是指消息中被读者认为对于判定消息是否是垃圾邮件有意义的那些部分相同。如果在缓存中发生命中，则缓存的裁决被检索并且进一步的消息扫描不被执行。如果在缓存中不存在摘要，则消息被用防垃圾邮件逻辑 119 扫描。

[0226] 在一个实施例中，裁决缓存 115 具有大小限度。如果达到了大小限度，则从缓存中删除最久以前使用的条目。在一个实施例中，每个缓存条目在可配置的条目寿命结束时期满。寿命的默认值是 600 秒。大小限度被设置为 100 乘以条目寿命。因此，缓存需要相对较少量的约 6MB 的存储器。在一个实施例中，缓存中的每个值是一个三元组，该三元组包括输入的时间、裁决以及防垃圾邮件逻辑 119 完成最初扫描所花的时间。

[0227] 在一个实施例中，如果在缓存中存在所请求的缓存密钥，则该值的输入时间被与当前时间相比较。如果条目仍是新近的，则缓存中的项目的值作为裁决被返回。如果条目已经期满，则其被从缓存中删除。

[0228] 在一个实施例中，可以进行若干尝试以在裁决被缓存之前计算消息摘要。例如，如果 fuz2 可用则 fuz2 被使用，否则如果 fuz1 可用则 fuz1 被使用，否则如果“所有 mime 部分”可用则其被用作摘要，否则不创建缓存条目。在一个实施例中，“所有 mime 部分”摘要包括消息的 MIME 部分的摘要的串联。如果没有 MIME 部分，则使用整个消息正文的摘要。在一个实施例中，“所有 mime 部分”摘要仅在防垃圾邮件逻辑 119 由于某种其他原因而执行消息正文扫描时被计算。正文扫描提取 MIME 部分，并且计算摘要的少量成本是可忽略的；因此操作可被高效地组合起来。

[0229] 在一个实施例中，每当消息传递网关 107 接收到来自规则 -URL 服务器 707(L 科 7) 的规则更新时，裁决缓存就被冲刷。在一个实施例中，每当防垃圾邮件逻辑 119 的配置发生变化时（例如由于管理动作或者由于加载新的配置文件），裁决缓存就被冲刷。

[0230] 在一个实施例中，防垃圾邮件逻辑 119 可并行地扫描多个消息。因此，两个或多个相同的消息可被同时扫描，从而导致缓存错失 (cachemiss)，因为裁决缓存尚未基于消息之一被更新。在一个实施例中，仅当消息的一个拷贝被完全扫描之后，裁决才被缓存。同一消息的当前正被扫描的其他拷贝是缓存错失。

[0231] 在一个实施例中，防垃圾邮件逻辑 119 周期性地扫描整个裁决缓存并且删除期满的裁决缓存条目。在这种情况下，防垃圾邮件逻辑 119 在日志文件 113 中写入日志条目，该日志条目报告缓存命中、错失、期满和添加的计数。防垃圾邮件逻辑 119 或裁决缓存 115 可维护计数器变量，以便执行日志记录或性能报告。

[0232] 在其他实施例中，缓存的摘要可被用于消息过滤器或者防病毒裁决。在一个实施例中，多个校验和被用于创建更丰富的密钥，该更丰富密钥既提供更高的命中率又提供更低的虚假裁决率。另外，可在裁决缓存中存储其他信息，例如扫描较长的消息以找出垃圾邮件所需的时间量。

[0233] 可引入优化以满足特定防垃圾邮件软件或逻辑的特定要求。例如，Brightmail 创建跟踪字符串并将该跟踪字符串与消息裁决一起返回；跟踪字符串可被添加到消息作为 X-Brightmail-Tracker 头部。跟踪字符串可被 Brightmail 对 Microsoft Outlook 的插件

用来实现语言识别。跟踪字符串也在插件报告假阳性时被发送回 Brightmail。

[0234] 对于具有相同正文的消息，裁决和跟踪字符串都可以是不同的。在一些情况下，正文不是垃圾邮件，但是垃圾邮件被编码在了主题中。在一种方法中，消息主题行被与消息正文包括在一起，作为消息摘要算法的输入。但是，当消息的正文很明显是垃圾邮件或者很明显是两者病毒时，主题行可能是不同的。例如，两个消息可能包含相同的病毒，但是主题头部可能是不同的。每个消息可能具有不同于其他消息的简短的文本附件，并且可能具有不同的名称。附件中的文件的名称可能是不同的。但是，当两个消息都被扫描时，将会得出相同的裁决。

[0235] 在一个实施例中，利用病毒阳性规则来提高缓存命中率。如果附件的摘要匹配病毒阳性裁决和垃圾邮件阳性裁决，则先前的垃圾邮件裁决被重复使用，即使主题和序言不同。

[0236] 在一些相似的消息中，不同的 From（发件人）值和不同的消息 ID 行会导致生成不同的跟踪字符串。垃圾邮件裁决相同，但是明显虚假的“From”值和明显虚假的消息 ID 将会导致更快地找到裁决并且在跟踪字符串中报告其他规则。在一个实施例中，From 头部和消息 ID 头部被从第二消息中删除并且消息被重新扫描，并且跟踪字符串与针对第一消息的相同。

[0237] 4.0 基于消息试探、发送者信息、动态隔离操作和细颗粒规则的病毒检测方法

[0238] 4.1 利用消息试探进行检测

[0239] 根据一种方法，提供了利用试探方法检测病毒。用于检测病毒发作的基本方法在 2004 年 12 月 6 日递交的 Michael Olivier 等人的题为“Methodand apparatus for managing computer virus outbreaks”的共同未决申请 No.11/006,209 中有所描述。

[0240] 在此上下文中，消息试探是指在没有关于消息的签名信息可用时用于确定消息是病毒的可能性的一组因素。试探可包括检测垃圾邮件中常用的单词或短语的规则。试探可根据消息文本中使用的语言而不同。在一个实施例中，管理用户可选择在防垃圾邮件扫描中使用哪种语言试探。消息试探可用于确定 VSV 值。消息的试探可由执行基本防垃圾邮件扫描和防病毒扫描的扫描引擎确定。

[0241] 基于试探操作的结果（取代或者附加于病毒发作的定义），消息可被放在隔离存储中，因为它可能包含病毒。这种定义在以上引用的 Olivier 等人的申请中有所描述。从而，全集服务器集群 706 包含过去的病毒历史，并且如果试探结果是消息匹配该过去历史中的模式，则消息可被隔离，而不管它是否匹配病毒发作的定义。这种提早隔离在 TOC 准备病毒发作定义的同时提供了消息处理的有益延迟。

[0242] 图 8 是在假设的示例性病毒发作中时间与受感染的机器数目之间的关系图。在图 8 中，水平轴 814 表示时间，垂直轴 812 表示受感染机器的数目。点 806 表示以下时间：防病毒软件卖家（例如 Sophos）公布更新后的病毒定义，该更新后的病毒定义将会检测带有病毒的消息并且防止受正在使用该防病毒软件的消息传递网关 107 保护的网络中的机器上的进一步感染。点 808 表示 TOC708 公布识别同一病毒的病毒发作的规则的时间。曲线 804 如图 8 所示变化，使得受感染的机器数目随着时间增长，但是增长率在点 808 之后下降，然后受感染机器总数在点 806 之后最终进一步大大下降。这里所述的基于试探的提早隔离在点 810 处被应用，以帮助减少被覆盖在曲线 804 的区域 816 内的机器的数目。

[0243] 在一个实施例中使用可变的隔离时间。当试探指示消息包含病毒的可能性较高时，可以增加隔离时间。这为 TOC 或者防病毒卖家提供了最大限度的时间来准备规则或定义，同时向不那么可能包含病毒的消息应用最低限度的隔离延迟。这样，隔离时间被耦合到消息包含病毒的概率，从而导致对隔离缓冲器空间的最优使用，并且使得隔离非病毒消息的时间达到最低限度。

[0244] 4.2 基于发送者的病毒检测

[0245] 根据一种方法，病毒得分被确定并且与消息发送者的 IP 地址值相关联地存储在数据库中。从而得分指示出源自关联地址的消息包含病毒的可能性。前提是发送一个病毒的机器很可能被另一病毒感染或者再次被同一病毒或更新后的病毒所感染，这是因为这些机器没有受到很好的保护。另外，如果机器正在发送垃圾邮件，则它更可能正在发送病毒。

[0246] IP 地址可指定远程机器，或者可以指定处于消息传递网关 107 正在保护的公司网络内的机器。例如，IP 地址可指定公司网络内无意中被病毒所感染的机器。这种被感染的机器很可能发送其他包含病毒的消息。

[0247] 在相关方法中，在整体消息处理中，病毒发作检测检查可与消息传递网关 107 内的垃圾邮件检查同时执行。从而，病毒发作检测可在消息被解析并经历垃圾邮件检测的同时执行。在一个实施例中，一个线程以有序的串行方式执行前述操作。另外，某些试探操作的结果可被用于通知防垃圾邮件检测操作和防病毒检测操作两者。

[0248] 在一个实施例中，VSV 值是基于以下之中的任何一个或多个来确定的：文件扩展名；按本地的、按全局的、针对每个发送者者识别和针对每个内容识别的消息量的峰值；基于附件内容，例如 Microsoft 可执行文件；以及基于发送者的威胁识别信息。在各种实施例中，使用各种基于发送者的威胁识别信息。示例包括动态或拨号主机黑名单、已开拓主机黑名单和病毒热区。

[0249] 连接到因特网的动态和拨号主机一般通过本地 SMTP 服务器发送传出邮件。当主机直接连接到外部 SMTP 服务器（例如消息传递网关 107）时，主机很可能已受到危害并且正在发送垃圾邮件消息或电子邮件病毒。在一个实施例中，消息传递网关 107 包括维护着过去已经以前述方式操作的动态主机的黑名单的逻辑，或者到动态主机黑名单的连接可在外部来源（例如 NJABL 动态主机列表和 SORBS 动态主机列表）处获得。

[0250] 在该实施例中，图 5 的步骤 502 处对传入消息的消息特性的识别还包括确定消息的发送者是否处于动态主机黑名单中。如果是，则确定或分配更高的 VSV 值。

[0251] 步骤 502 还可包括连接到或管理已开拓主机黑名单并且确定消息的发送者是否在已开拓主机黑名单上。已开拓主机黑名单基于受感染主机发送的内容来跟踪已知被病毒感染的主机或者已知发送垃圾邮件的主机并且通过连接时间扫描来定位已被感染的主机。示例包括 XBL(CBL 和 OPM) 和 DSBL。

[0252] 在另一实施例中，服务提供者 700 基于从客户消息传递网关 107 接收的发送者信息来创建并存储具有过去发送病毒历史的发送者和网络的内部黑名单。在一个实施例中，客户消息传递网关 107 周期性地发起到全集服务器集群 706 的网络通信，并且报告消息传递网关 107 的内部逻辑确定为垃圾邮件或与病毒或其他威胁相关联的消息的发送者的网络地址（例如 IP 地址）。服务提供者 700 处的逻辑可周期性地扫描内部黑名单并且确定是否已知任何网络地址只发送病毒或垃圾邮件。如果是，则该逻辑可将较高的威胁级别值或

VSV 与这些地址关联起来存储。中等威胁级别值可与已知既发送病毒也发送合法电子邮件的网络地址关联起来存储。中或低威胁级别值可与包含一个或多个个体受感染主机的网络相关联。

[0253] 对照黑名单的测试可利用上述类型的规则来发起。例如，下面的规则可发起黑名单测试：

[0254] eval DYNAMIC_IP connecting_ip(dynamic)

[0255] eval HOTZONE_NETWORK connecting_ip(hotzone)

[0256] eval XBL_IP connecting_ip(exploited host)

[0257] 4.3 包括重新扫描的动态隔离操作

[0258] 在现在方法中，消息是以先进先出顺序从隔离中释放的。或者，在另一实施例中可使用首先退出算法。在该方法中，当隔离缓冲器已满时，排序机制确定哪些消息应当首先被释放。在一个实施例中，被认为危险性最小的消息被首先释放。例如，由于试探的结果而被隔离的消息首先被释放，由于匹配病毒发作测试的结果而被隔离的消息其次被释放。为了支持该机制，每个隔离的消息与指示出隔离原因的信息相关联地被存储在消息传递网关 107 的隔离中。然后，消息传递网关 107 中的过程可以基于原因而释放消息。

[0259] 通过在由消息传递网关 107 处理的配置文件中指定顺序，可以数据驱动方式来配置排序。从而，从服务提供者向客户消息传递网关 107 公布包含排序的新配置文件自动地致使这些消息传递网关 107 采用新的排序。

[0260] 类似地，基于在被隔离的消息离开隔离时与消息相关联的威胁级别，可在消息离开隔离时对其采取不同的动作。例如，看起来极为有威胁性但是可能由于溢出而离开隔离的消息可经历剥离并递送操作，其中附件被剥离并且消息在无附件的情况下被递送到接收者。或者，具有较低威胁级别的消息被像正常情况那样递送。

[0261] 在另一替换方案中，X 头部可被添加到较低威胁级别的消息。该替换方案在以下情况时适用：客户端电子邮件程序（例如 Eudora，MicrosoftOutlook）被配置以这样一条规则，该规则识别 X 头部并将具有 X 头部的消息放在特殊文件夹（例如“潜在危险消息”）中。在另一替换方案中，具有特定威胁级别的消息的文件附件被重命名（消息被“拔去尖刀”），从而要求接收方用户确实地再次重命名文件附件才能使其用于应用。该方法旨在使用户在重命名并打开文件之前小心地检查它。消息可被转发到管理员以便评估。在一个实施例中可结合这些替换方案中的任何一个。

[0262] 图 9 是用于重新扫描可能包含病毒的消息的方法的流程图。根据一个实施例，当 TI 小组 710 向消息传递网关 107 发布新的威胁规则时，每个消息传递网关对照新的规则重新扫描其隔离中的消息。该方案提供了以下优点，即消息可以更早地从隔离中释放，因为在后面阶段的处理中将会利用新规则检测到消息包含病毒。在此上下文中，“释放”是指将消息从隔离中去除并将其发送到防病毒扫描过程。

[0263] 或者，重新扫描可减少或增加消息的隔离时间。这使得隔离中的消息数目最小化，并且降低了释放受感染的消息的可能性。这种非有意释放可能发生在例如以下情况下：隔离具有固定的释放时间，并且在防病毒卖家或其他来源发布将会捕获释放的消息的病毒定义之前该固定的释放定时器就已期满。在这种场景下，恶意的消息将会被自动释放，并且下游处理不会捕获到它。

[0264] 在一个实施例中，若干事件之中的任何一种将会触发对消息隔离中的消息进行重新扫描。另外，图 9 的方法适用于处理由于病毒、垃圾邮件或者消息的其他威胁或不合需要的特性的结果而在隔离中的消息。在步骤 902 中，重新扫描定时器被启动并且运行直到期满，并且在期满时在步骤 906 中触发对隔离队列中的所有消息的重新扫描。

[0265] 作为附加或替换，在步骤 904 中，消息传递网关 107 接收来自规则 -URL 服务器 707 的一个或多个新的病毒威胁规则、防垃圾邮件规则、URL、得分或其他消息分类信息。接收这种信息也可触发步骤 906 处的重新扫描。新的规则、得分和其他信息被用在重新扫描步骤中，以为隔离中的每个消息生成新的 VSV。例如，TOC 服务器 708 可通过规则 -URL 服务器 707 公布最初较宽的一组针对病毒发作的规则，然后随着获知关于发作的更多信息而缩小规则的范围。结果，与较早的规则组相匹配的消息可能不与修改后的规则相匹配，并且变成已知的假阳性。这里的方法尝试响应于规则更新而自动释放已知假阳性，而无需消息传递网关 107 的管理员干预。

[0266] 在一个实施例中，隔离队列 316 中的每个消息具有指示出消息何时进入隔离的存储时间值，并且步骤 906 处的重新扫描是按隔离进入时间的顺序执行的，最老的消息最先。

[0267] 在步骤 908 中，像图 3 的步骤 312 中那样，执行测试以确定消息的新 VSV 是否大于或等于特定的阈值。VSV 阈值是由消息传递网关 107 的管理员设置的，用于确定隔离消息的容限。如果 VSV 低于阈值，则消息很可能能够从隔离中释放。因此，控制传递到步骤 910，在该步骤中应用正常隔离退出递送方针。

[0268] 可选地，在一个实施例中，消息传递网关 107 可实现单独的报告阈值。当在步骤 907 测试出消息具有超过报告阈值的 VSV 时，消息传递网关 107 在步骤 909 通知服务提供者 700 并且继续处理消息。这种通知可为新病毒发作的确定提供重要的输入。在某些实施例中，这种报告是“SenderBase 网络参与”(SBNP) 的一个方面，并且可由管理员利用配置设置来有选择地启用。

[0269] 在步骤 910 应用递送方针可包括立即将消息排队以便以未修改的形式递送到接收者，或者剥离附件，或者执行内容过滤，或者对消息执行其他检查。应用递送方针可包括向消息添加指示出病毒扫描结果的 X 头部。所有可应用的 X 头部可按动作的发生的顺序被添加到消息。应用递送方针可包括修改消息的主题行以指示出可能存在病毒、垃圾邮件或其他威胁。应用递送方针可包括将消息重定向到备用接收者，并且存储消息的归档拷贝以便其他逻辑、系统或个人以后分析。

[0270] 在一个实施例中，在步骤 910 应用递送方针包括在消息处于若干隔离中的任何一个之中并且一个隔离确定剥离附件是正确动作时在递送消息之前从消息中剥离所有附件。例如，消息传递网关 107 可支持病毒发作隔离队列 316 和单独的隔离队列，该单独隔离队列保存着看起来违犯了网关的方针的消息，例如存在不允许的单词。假定病毒发作隔离队列 316 被配置为在递送之前在溢出时剥离附件。假定消息既在病毒发作隔离队列 316 中又在单独方针隔离队列中，并且正好使病毒发作隔离队列 316 溢出。如果管理员随后手工从该方针隔离队列中释放同一消息，则附件在递送之前再次被剥离。

[0271] 在步骤 912，消息被递送。

[0272] 如果步骤 909 的测试为真，则消息有问题并且很可能需要被保持在隔离中。

[0273] 可选地，每个消息可被分配一个期满时间值，并且该期满时间值与隔离队列 316

相关联地被存储在消息传递网关 107 的数据库中。在一个实施例中，期满时间值等于消息进入隔离队列 316 的时间和指定的保持时间。期满时间值可以基于消息内容或消息的试探可不同。

[0274] 在步骤 914 中，执行测试以确定消息期满时间是否已经期满。如果是，则将消息从隔离中去除，但此时的消息去除被认为是异常或提早退出，因此在步骤 918 应用异常退出递送方针。然后，消息可在步骤 912 中被递送，并且经历步骤 918 的递送方针。在步骤 918 应用的递送方针可不同于在步骤 910 应用的方针。例如，步骤 910 的方针可提供不受限的递送，而在步骤 918 可能要求去除附件（用于递送这样的消息：可疑，但在隔离中的时间已经长于期满时间）。

[0275] 如果在步骤 914 消息时间尚未期满，则消息被保持在隔离中，如步骤 916 所示。如果使 VSV 超过阈值的规则变化，则规则名称和描述在消息数据库中被更新。

[0276] 在各种实施例中，图 9 的不同步骤可以使消息传递网关 107 发送一个或多个警告消息给管理员或者指定的用户账户或群组。例如，在步骤 904、912 或 916 可生成警告。示例性的警告事件包括达到指定的隔离充满级别或空间限度；隔离溢出；接收到新的发作规则，例如这样一条规则：如果匹配则将 VSV 设置为高于在消息传递网关中配置的隔离阈值；接收到去除发作规则的信息；以及在消息传递网关中更新新规则的尝试失败。去除发作规则的信息可包括接收到将特定类型的消息的威胁级别降低到在消息传递网关中配置的隔离阈值之下的新规则。

[0277] 另外，图 9 的不同步骤可以使消息传递网关 107 在日志文件 113 中写入一个或多个描述执行的动作的日志条目。例如，日志文件条目可在消息被异常地释放或在提早退出时被写入。警告或日志条目可在隔离在指定级别下充满时被发送或写入。例如，警告或日志条目在隔离达到 5％满、50％满、75％满等等时被发送或写入。日志条目可包括隔离接收时间、隔离退出时间、隔离退出标准、隔离退出动作、隔离中的消息数目等等。

[0278] 在其他实施例中，警告消息可指示出扫描警告更新失败；规则更新失败；在指定时间段中接收规则更新失败；拒绝指定百分比的消息；拒绝指定数目的消息；等等。

[0279] 图 10 是实现上述逻辑的消息传递网关中的消息流模型的框图。消息试探 1002 和病毒发作规则 1004 被提供给扫描引擎，例如防病毒检查器 116，该扫描引擎生成 VSV 值或者病毒威胁级别 (VTL) 值 1005。如果 VSV 值超过指定的阈值，则消息进入隔离 316。

[0280] 多个退出标准 1006 可以使消息能够离开隔离 316。示例性的退出标准 1006 包括时间限度 1008 的期满、溢出 1010、手工释放 1012 或者规则更新 1014。当退出标准 1006 得到满足时，一个或多个退出动作 1018 随后发生。示例性的退出动作 1018 包括剥离和递送 1020、删除 1022、正常递送 1024、利用关键字（例如 [SPAM]）标记消息主题 2016 以及添加 X 头部 1028。在另一实施例中，退出动作可包括警告消息的指定接收者。

[0281] 在一个实施例中，消息传递网关 107 维护一个数据结构，该数据结构对于与消息相关联的每个发送方主机，定义用于作用于接收自该主机的消息的方针。例如，主机访问表包括布尔属性值，该值指示出是否对该主机执行这里针对图 3、图 9 描述的病毒发作扫描。

[0282] 另外，在消息传递网关 107 中处理的每个消息可被存储在一个数据结构中，该数据结构承载着指示出在消息传递网关内执行什么消息处理的元数据。元数据的示例包括：消息的 VSV 值；导致该 VSV 值的规则的名称和相应的规则描述；消息隔离时间和溢出优先

级；指定是否执行防垃圾邮件和防病毒扫描和病毒发作扫描的标志；以及使得能够绕开内容过滤器的标志。

[0283] 在一个实施例中，存储在消息传递网关 107 中的一组配置信息为来自网关的消息的每个潜在接收者的病毒发作扫描指定附加的程序行为。由于消息传递网关 107 通常将消息流量控制到一组有限的用户，例如雇员、立约人或者企业专用网络中的其他用户，因此这种配置信息可针对所有潜在接收者被管理。例如，每接收者配置值可指定这里描述的扫描不考虑的消息附件文件扩展名类型（“.doc”、“.ppt”等等）的列表，以及指示出消息不应当被隔离的值。在一个实施例中，配置信息可为每个接收者包括特定的阈值。从而，取决于相关联的阈值，步骤 312 和步骤 908 的测试对于不同的接收者可具有不同的结果。

[0284] 消息传递网关 107 还可管理对已经利用图 9、图 9 的技术过滤的消息进行计数的数据库表、这种消息的 VSV 以及发送到消息隔离 316 的消息的计数。

[0285] 在一个实施例中，每个消息隔离 316 具有多个相关联的程序化动作，这些动作控制消息如何退出隔离。再次参考图 3，退出动作可包括基于操作者判决 318 将消息从消息隔离 316 中手工释放。退出动作可包括像图 9 中那样在期满定时器期满时将消息从消息隔离 316 中自动释放。作为溢出方针 322 的实现，退出动作可包括当隔离充满时从消息隔离 316 中提早退出。“提早退出”是指基于例如队列溢出这样的资源限制，在与消息相关联的期满时间值结束之前提前释放消息。

[0286] 正常消息退出动作和提早退出动作可被组织为具有以上针对递送方针步骤 910 所描述的类型的主动作和次动作。主动作可包括反弹、删除、剥离附件并递送以及递送。次动作可包括主题标记、X 头部、重定向或归档。次动作不与主动作删除相关联。在一个实施例中，次动作重定向使得消息能够被发送到在全集服务器集群 706 处或者服务提供者 700 内的另一元素处而不是消息传递网关 107 上容宿的次级“盒外”隔离队列。该方法使得 TI 小组 710 能够检查隔离的消息。

[0287] 在一个实施例中，由隔离队列溢出导致的从隔离的提早退出动作可包括主动作中的任何一个，其中包括剥离附件并递送。次动作中的任何一个可被用于这种提早退出。消息传递网关 107 的管理员可通过利用命令接口或 GUI 向消息传递网关发出配置命令，来选择在提早退出时使用的主动作和次动作。作为附加或替换，由于执行防病毒扫描或其他消息扫描而确定出的消息试探可导致执行不同的提早退出动作作为响应。

[0288] 在一个实施例中，消息传递网关 107 中的本地数据库存储消息隔离 316 中的接收消息的文件附件的名称，以及文件附件的大小。

[0289] 步骤 906 处的重新扫描可响应于消息传递网关 107 的其他动作而对特定消息发生。在一个实施例中，消息传递网关 107 实现可根据一个或多个规则而改变接收到的消息的内容的内容过滤器。如果内容过滤器改变先前被进行过病毒扫描的接收消息的内容，则该消息的 VSV 值可能在重新扫描时改变。例如，如果内容过滤器从消息中剥离附件，并且病毒在附件中，则剥离的消息可能不再有病毒威胁。因此，在一个实施例中，当内容过滤器改变接收到的消息的内容时，步骤 906 处的重新扫描被执行。

[0290] 在一个实施例中，消息传递网关 107 的管理员可利用控制台命令或其他用户接口命令来搜索隔离 316 的内容。在一个实施例中，搜索可以基于附件名称、附件类型、附件大小和其他消息属性来执行。在一个实施例中，按文件类型的搜索可以仅对处于隔离 316 中

但不处于方针隔离或其他隔离中的消息执行，因为这种搜索要求对消息正文进行扫描，这可能负面地映射性能。在一个实施例中，管理员可根据前述属性中的任何一种按分类的顺序显示病毒发作的内容。

[0291] 在一个实施例中，当消息通过图 3 或图 9 的过程而被放在隔离 316 中时，消息传递网关 107 自动地显示病毒发作隔离的视图。在一个实施例中，该视图对于隔离中的每个消息包括以下属性值：发作标识符或规则名称；发送者名称；发送者域；接收者名称；接收者域；主题名称；附件名称；附件类型；附件大小；VSV；隔离进入时间；隔离剩余时间。

[0292] 在一个实施例中，消息传递网关 107 存储重新插入密钥，该重新插入密钥包括可与已从隔离 316 中手工释放的消息相关联的可选的唯一文本串。当释放的消息具有与之相关联的重新插入密钥时，在递送之前，在消息传递网关 107 中的后续处理期间，释放的消息不能再次被隔离。

[0293] 4.4 细颗粒规则

[0294] 消息规则是抽象的陈述，这些抽象陈述如果在防垃圾邮件逻辑 119 中与消息相比匹配则导致较高的垃圾邮件得分。规则可具有规则类型。示例性的规则类型包括受危害主机、可疑垃圾邮件源、头部特性、正文特性、URI 和学习。在一个实施例中，可以应用特定的发作规则。例如，病毒发作检测机制可确定具有 20kb 大小的 ZIP 文件附件的某种类型的消息代表病毒。该机制可创建一种规则，根据这种规则，客户消息传递网关 107 将会隔离具有 20kb ZIP 附件的消息，但不会隔离具有 1MB ZIP 附件的消息。结果，更少的阳隔离操作会发生。

[0295] 在一个实施例中，病毒信息逻辑 114 包括支持建立关于消息头部和消息正文的规则或测试以识别固定字符串或正则表达式的逻辑。例如，一个实施例允许定义下面的规则：

[0296] head X_MIME_FOO X-Mime =～ /foo/

[0297] head SUBJECT_YOUR Subject = -/your document/

[0298] body HEY_PAL/hey pal|long time,no see/

[0299] body ZIP_PASSWORD ^.zip password is/i

[0300] 在一个实施例中，函数测试可测试消息的特定方面。每个函数执行定制代码以检查消息、已经捕捉的关于消息的信息，等等。测试无法利用通用头部或正文测试的简单逻辑组合来形成。例如，用于在不检查文件内容的情况下匹配病毒的有效测试是将“文件名”或“名称”MIME 字段的扩展名与声称的 MIME 内容类型相比较。如果扩展名是“doc”并且内容类型既不是 application/octet-stream 也不是 application/.*word，则内容是可疑的。对于 PowerPoint、Excel、图像文件、文本文件和可执行文件可执行类似地比较。

[0301] 测试的其他示例包括：测试 base64 型内容的第一行是否匹配指示出 Microsoft 可执行文件的正则表达式 /^TV[nopqr]/；测试电子邮件优先级是否被设置到高，但是却没有 X 寄信人或用户代理头部；测试消息是否是多部分 / 替换的，但替换部分的内容非常不同；测试消息是否是多部分的，但只包含 HTML 文本；查找特定的 MIME 边界格式以找出新的发作。

[0302] 在一个实施例中，病毒信息逻辑 114 包括支持建立包括多个链接的规则的元规则的逻辑。示例包括：

[0303] meta VIRUS_FOO ((SUBJECT_FOO1||SUBJECTJFOO2)&&BODY_FOO)

[0304] meta VIRUS_BAR(SIZE_BAR+SUBJECTJBAR+BODY_BAR > 2)

[0305] 在一个实施例中，病毒信息逻辑 114 包括这样的逻辑，该逻辑支持对照基于文件附件大小、文件名称关键字、加密文件、消息 URL 和防病毒逻辑版本值的规则来建立和测试消息。在一个实施例中，与文件附件大小相关的规则是基于离散值而不是每个可能的大小值创建的；例如，规则可按以下增量来指定文件大小：对于 0-5K 的文件指定 1K 增量；对于从 5K 到 1MB 大小的文件指定 5K 增量；以及指定 1MB 增量。

[0306] 文件名关键字规则在消息的文件附件具有包括规则中的一个或多个关键字的名称时在消息上匹配。加密文件规则测试文件附件是否被加密。这种规则可用于隔离有加密容器（例如加密的 ZIP 文件）作为消息附件的消息。消息 URL 规则在消息正文包含规则中指定的一个或多个 URL 时在消息上匹配。在一个实施例中，除非在系统中安装了至少一个消息 URL，否则不会扫描消息以识别 URL。

[0307] 基于防病毒逻辑版本值的规则在消息传递网关 107 正在运行具有匹配版本的防病毒逻辑时匹配消息。例如，规则可指定“7.3.1”的 AV 签名版本，并且如果消息传递网关正在利用具有该版本号的签名文件运行 AV 软件则该规则在消息上匹配。

[0308] 在一个实施例中，在接收到对于一组消息来说比先前接收的规则更具体的一个新规则时，消息传递网关 107 自动地降低消息的存储 VSV。例如，假定 TOC 708 最初分发了一个规则，即任何具有 .ZIP 文件附件的消息都被分配以 VSV “3”。TOC 708 随后分发了一个规则，既 30KB 到 35KB 之间的 .ZIP 文件附件具有 VSV“3”。作为响应，消息传递网关 107 将具有不同文件大小的 .ZIP 附件的所有消息的 VSV 降低到默认 VSV，例如“1”。

[0309] 在一个实施例中，防垃圾邮件逻辑 119 可以基于诸如接收者地址、接收者域和常用单词或短语之类的传出消息特性来学习识别特定于某个组织的合法电子邮件。在此上下文中，传出消息是由与专用网络 110 上的计算机 120A、120B、120C 相关联的用户账户所编写的、通过消息传递网关 107 被导向逻辑上在消息传递网关外部的接收者账户的消息。这种接收者账户通常在连接到公共网络 102 的计算机上。由于所有的传出消息在递送到网络 102 中之前都经过消息传递网关 107，因此这种传出消息几乎永远不会是垃圾邮件，并且消息传递网关可扫描这种消息并自动生成与非垃圾邮件消息相关联的试探或者规则。在一个实施例中，学习是通过以下方式完成的：在传出消息的文本上训练防垃圾邮件逻辑 119 中的贝叶斯过滤器，然后利用该贝叶斯过滤器来测试传入消息。如果被训练的贝叶斯过滤器返回高概率，则根据传出消息不是垃圾邮件的概率，传入消息很可能不是垃圾邮件。

[0310] 在一个实施例中，消息传递网关 107 周期性地轮询规则 -URL 服务器 707 以请求任何可用的规则更新。HTTPS 可被用于递送规则更新。在一个实施例中，消息传递网关 107 的管理员可通过输入规则更新的 URL 并利用浏览器和代理服务器或者固定地址连接到规则 -URL 服务器 707，来访问和检查规则更新。管理员随后可以将更新递送到被管理网络内的所选消息传递网关 107。接收规则更新可包括在消息传递网关 107 的接口中显示用户通知，或者在日志文件 113 中写入条目，声明接收到规则更新或者消息传递网关成功地连接到了规则 -URL 服务器 707。

[0311] 4.5 与服务提供者的通信

[0312] 图 1 中的客户消息传递网关 107 可实现“给家打电话”或者“SenderBase 网络参

与”服务，其中消息传递网关 107 可打开与服务提供者 700 的连接并且提供关于消息传递网关 107 已处理的消息的信息，以便这种来自现场的信息可被添加到全集或者在服务提供者处以其他方式使用，以提高计分、发作检测和试探。

[0313] 在一个实施例中，树形数据结构和处理算法被用于提供从消息传递网关 107 到服务提供者的高效数据通信。

[0314] 作为防垃圾邮件和防病毒检查的一部分而生成的来自服务提供者的数据被发送到现场的消息传递网关 107。结果，服务提供者创建描述其希望消息传递网关 107 向其返回什么数据的元数据。消息传递网关 107 在一段时间（例如 5 分钟）中收集匹配元数据的数据。消息传递网关 107 随后连接回服务提供者，并根据元数据的规范提供现场数据。

[0315] 在该方法中，在不同的时间定义和递送不同的元数据到消息传递网关 107 使得服务提供者能够指示现场的消息传递网关 107 将不同的数据递送回服务提供者。从而，“打电话回家”服务在服务提供者的指导下变得可扩展。不需要更新 MGA 处的软件。

[0316] 在一种实现方式中，树被实现为散列的散列。存在嵌套散列（或 Python 中的字典）到树的标准映射。某些节点以这样的方式被命名：从 MGA 返回关于哪些事物是哪些的数据。通过为树中的节点命名，而不是仅基于其位置来描述事物，MGA 不需要知道服务提供者将会如何处理数据。MGA 只需要按名称定位正确的数据，并将数据的拷贝发送回服务提供者。MGA 唯一需要知道的是数据的类型，即数据是数值还是字符串。MGA 不需要对数据执行计算或变换以适应服务提供者。

[0317] 对于数据结构施加了约束。规则是树的端点总是两者之一。如果目标数据是数字，则叶节点是计数器。当 MGA 看到到达的下一消息，则其递增或递减该节点的计数器。如果目标数据是字符串，则叶节点被用该字符串值覆写。

[0318] 利用计数器方法，任何形式的数据都可被传输。例如，如果 MGA 需要将平均得分值传输回服务提供者，而不是让服务提供者通知 MGA 其希望 MGA 返回特定值作为平均得分，则使用两个计数器，一个用于顶部值，一个用于底部值。MGA 不需要知道哪个是哪个。它只是对规定的值进行计数并返回它们。服务提供者处的逻辑知道从 MGA 接收的值是计数器并且需要被取平均和存储。

[0319] 从而，该方法提供了一种用于对数据进行透明的核对和传送的方法，其中传送数据的设备不知道数据的具体用途，但是可核对并提供数据。另外，服务提供者可更新其软件以请求来自消息传递网关 107 的附加的值，但是不需要对 MGA 软件进行更新。这使得服务提供者能够收集数据而无需改变现场的数百或数千的消息传递网关 107。

[0320] 可从消息传递网关 107 传输到服务提供者 700 的示例性的数据包括 X 头部值，其包含在特定消息上匹配并且导致垃圾邮件裁决的扰乱规则。

[0321] 4.7 传出白名单模块

[0322] 在图 3 的配置中，客户消息传递网关 107 可被部署在客户网络中，从而使得它们接收和处理传入和传出消息流量。因此，消息传递网关 107 可被配置以传出消息白名单。在该方法中，离开消息传递网关 107 的指定消息的目的地网络地址和权重值一起被放在传出消息白名单中。当接收到传入消息时参考传出消息白名单，并且如果权重值适当则递送具有传出白名单中的源网络地址的传入消息。也就是说，在确定是否应当递送消息时考虑了权重值；传出白名单中地址的存在不一定要求递送。其原理是从传出白名单中的实体接收

的消息不应当是垃圾邮件或者有威胁的，因为向该实体发送消息就暗示着信任。传出白名单可维护在服务提供者处，以便分发到其他客户消息传递网关 107。

[0323] 可利用若干种方法来执行权重值的确定。例如，可以利用声望计分系统来处理目的地地址，并且可以基于所得到的声望得分来选择权重值。可以跟踪并比较消息标识符以确定传入消息实际上是否是在回复过去发送的先前消息。可以使用消息标识符的缓存。从而，如果 Reply-To（回复给...）头部包含先前由同一消息传递网关 107 发送的消息的消息标识符，则很可能该回复不是垃圾邮件或威胁。

[0324] 5.0 实现机构 - 硬件概述

[0325] 这里描述的用于管理计算机病毒发作的方法可以用多种方式来实现，并且本发明不限于任何特定实现方式。该方法可被集成到电子邮件系统或邮件网关装置或其他合适的设备中，或者可以实现为独立的机构。另外，该方法可用计算机软件、硬件或其组合来实现。

[0326] 图 6 是示出可以实现本发明的实施例的计算机系统 600 的框图。计算机系统 600 包括用于传输信息的总线 602 或其他通信机构和与总线 602 相耦合用于处理信息的处理器 604。计算机系统 600 还包括诸如随机存取存储器 (RAM) 或其他动态存储设备之类的主存储器 606，其耦合到总线 602，用于存储信息和处理器 604 要执行的指令。主存储器 606 还可用于存储在处理器 604 执行指令期间的临时变量或其他中间信息。计算机系统 600 还包括只读存储器 (ROM)608 或其他静态存储设备，其耦合到总线 602，用于存储静态信息和处理器 604 的指令。提供了诸如磁盘或光盘之类的存储设备 610，其耦合到总线 602，用于存储信息和指令。

[0327] 计算机系统 600 可以经由总线 602 耦合到显示器 612，例如阴极射线管（“CRT”），用于向计算机用户显示信息。包括字母数字和其他键的输入设备 614 被耦合到总线 602，用于向处理器 604 传输信息和命令选择。另一类用户输入设备是光标控制装置 616，例如鼠标、跟踪球、触笔或光标方向键，用于向处理器 604 传输方向信息和命令选择，并用于控制显示器 612 上的光标移动。该输入设备一般具有两个轴（第一轴（例如 x）和第二轴（例如 y））上的两个自由度，其允许设备指定平面中的位置。

[0328] 本发明涉及使用计算机系统 600 来向消息内容应用试探测试、管理动态威胁隔离队列以及进行具有从解析和扫描的提早退出的消息扫描。根据本发明的一个实施例，向消息内容应用试探测试、管理动态威胁隔离队列以及具有从解析和扫描的提早退出的消息扫描由计算机系统 600 响应于处理器 604 执行包含在主存储器 606 中的一条或多条指令的一个或多个序列而提供。这种指令可以被从另一计算机可读介质（如存储设备 610）读取到主存储器 606 中。包含在主存储器 606 中的指令序列的执行使得处理器 604 执行这里描述的过程步骤。在替换实施例中，可以使用硬线电路来替代软件指令或与软件指令相组合以实现本发明。从而，本发明的实施例并不限于硬件电路和软件的任何特定组合。

[0329] 这里所用的术语“计算机可读介质”指参与向处理器 604 提供指令以供执行的任何介质。这种介质可以采取许多形式，包括但不限于：非易失性介质、易失性介质和传输介质。非易失性介质例如包括光盘或磁盘，如存储设备 610。易失性介质包括动态存储器，如主存储器 606。传输介质包括同轴电缆、铜线和光纤，包括含总线 602 的线路。传输介质也可以采取声波或光波的形式，例如在无线电波和红外数据通信期间生成的声波或光波。

[0330] 计算机可读介质的常见形式例如包括软盘、柔性盘、硬盘、磁带或任何其他磁介

质，CD-ROM、任何其他光介质，穿孔卡、纸带、任何其他具有孔图案的物理介质，RAM、PROM 和 EPROM、FLASH-EPROM、任何其他存储器芯片或磁带盒 (cartridge)，下文中描述的载波，或者计算机可以读取的任何其他介质。

[0331] 计算机可读介质的各种形式可用于将一条或多条指令的一个或多个序列传输到处理器 604 以供执行。例如，指令可以首先承载在远程计算机的磁盘上。远程计算机可以将指令加载到其动态存储器中，并利用调制解调器经由电话线发送指令。计算机系统 600 本地的调制解调器可以接收电话线上的数据，并使用红外发送器来将数据转换为红外信号。红外检测器可以接收在红外信号中携带的数据，并且适当的电路可以将数据置于总线 602 上。总线 602 将数据传输到主存储器 606，处理器 604 从主存储器 606 检索指令并执行指令。主存储器 606 接收的指令可以可选地在处理器 604 执行之前或之后存储到存储设备 610 上。

[0332] 计算机系统 600 还包括耦合到总线 602 的通信接口 618。通信接口 618 提供到连接到本地网络 622 的网络链路 620 的双向数据通信耦合。例如，通信接口 618 可以是综合业务数字网络 (ISDN) 卡或调制解调器，以提供到相应类型电话线的数字通信连接。又例如，通信接口 618 可以是局域网 (LAN) 卡，以提供到兼容 LAN 的数据通信连接。也可以实现无线链路。在任何这种实现方式中，通信接口 618 发送并接收电的、电磁的或光信号，这些信号携带了代表各种类型信息的数字数据流。

[0333] 网络链路 620 一般经过一个或多个网络提供到其他数据设备的数据通信。例如，网络链路 620 可以经过本地网络 622 提供到主机计算机 624 或由因特网服务供应商 (ISP)626 操作的数据设备的连接。ISP 626 又经过全球分组数据通信网络（现在通常称为“因特网”628) 提供数据通信服务。本地网络 622 和因特网 628 都使用携带数字数据流的电的、电磁的或光信号。经过各种网络的信号和在网络链路 620 上并经过通信接口 618 的信号（这些信号携带去往和来自计算机系统 600 的数字数据）是传输信息的载波的示例性形式。

[0334] 计算机系统 600 可以经过网络、网络链路 620 和通信接口 618 发送消息并接收数据，包括程序代码。在因特网示例中，服务器 630 可以经过因特网 628、ISP 626、本地网络 622 和通信接口 618 发送针对应用程序的请求代码。根据本发明，一个这种下载的应用程序提供了如这里所述的向消息内容应用试探测试、管理动态威胁隔离队列以及进行具有从解析和扫描的提早退出的消息扫描。

[0335] 接收到的代码可以在接收时被处理器 604 执行，和 / 或被存储在存储设备 610 或其他非易失性存储介质中以供后续执行。以这种方式，计算机系统 600 可以获得载波形式的应用代码。

[0336] 6.0 扩展和替换

[0337] 在前述说明书中，已参考具体实施例描述了本发明。但是，应当清楚，在不脱离本发明更宽广的精神和范围的前提下，可以进行各种修改和改变。因此，说明书和附图都应当认为是示例性的，而非限制性的。本发明包括其他上下文和应用，其中这里描述的机制和过程可用于其他机制、方法、程序和过程。

[0338] 此外，在本说明书中，某些过程步骤是以特定的顺序来阐述的，并且使用了字母和字母数字标注来标识某些步骤。除非本公开中具体声明，否则本发明的实施例不限于任何

特定的执行这种步骤的顺序。具体而言，这些标注的使用只是为了便于标识步骤，而不想要暗示、指定或要求执行这种步骤的特定顺序。另外，其他实施例可以使用比这里论述的更多或更少的步骤。

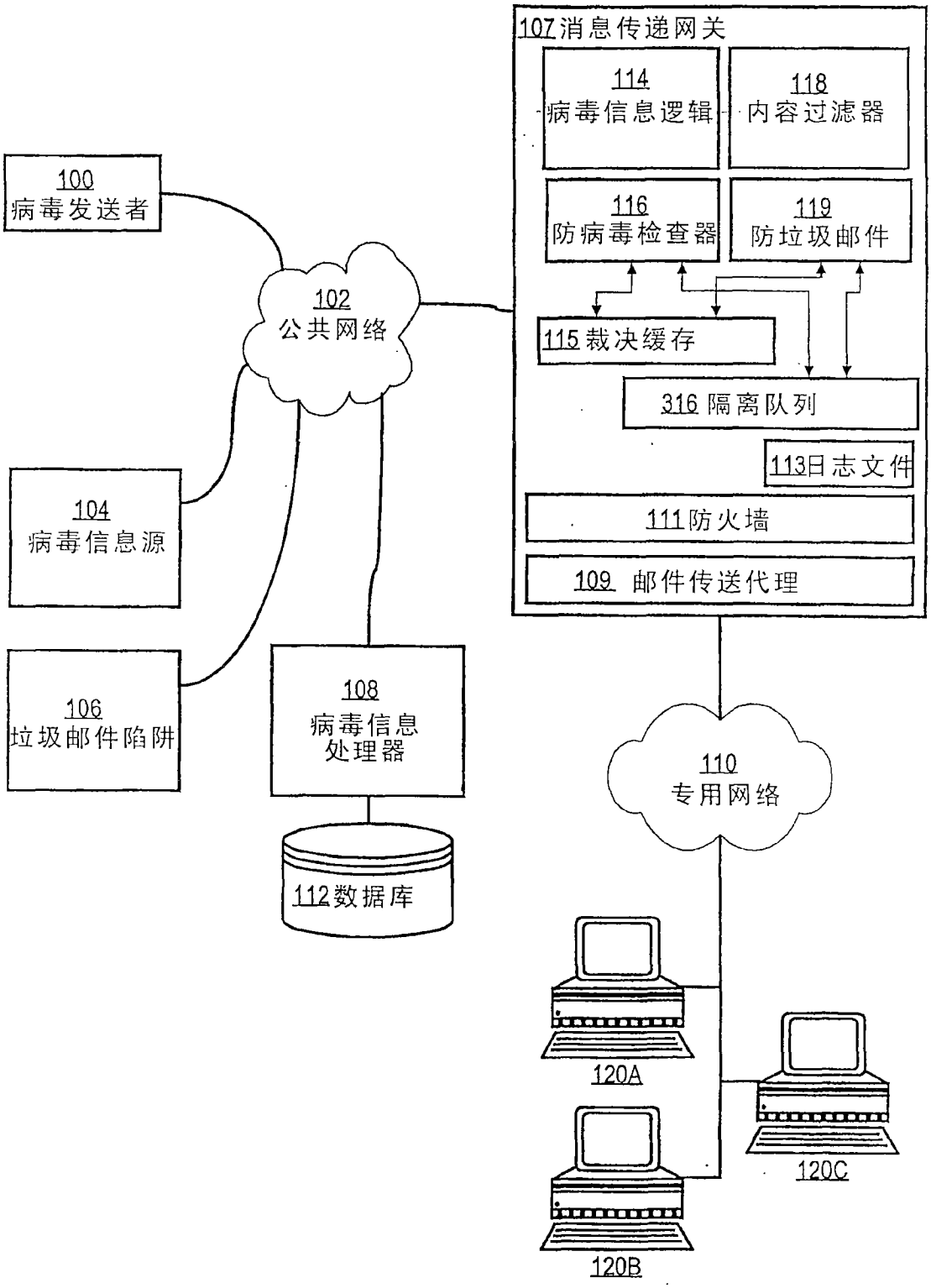

107消息传递网关
114
病毒信息逻辑
118
内容过滤器
116
防病毒检查器
119
防垃圾邮件
115裁决缓存
316隔离队列
113日志文件
111防火墙
109 邮件传送代理
100
病毒发送者
102
公共网络
104
病毒信息源
106
垃圾邮件陷阱
108
病毒信息
处理器
112数据库
110
专用网络
120A
120C
120B

图1

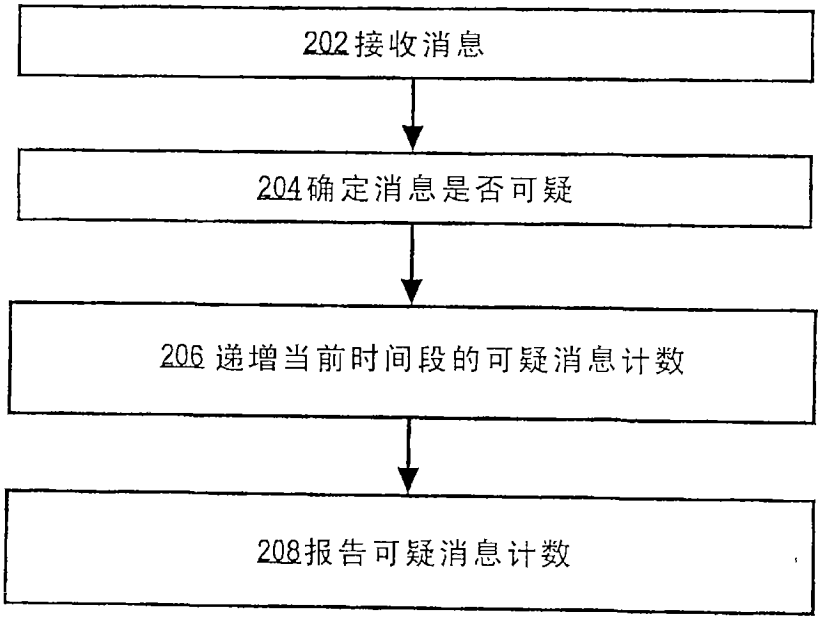


图2

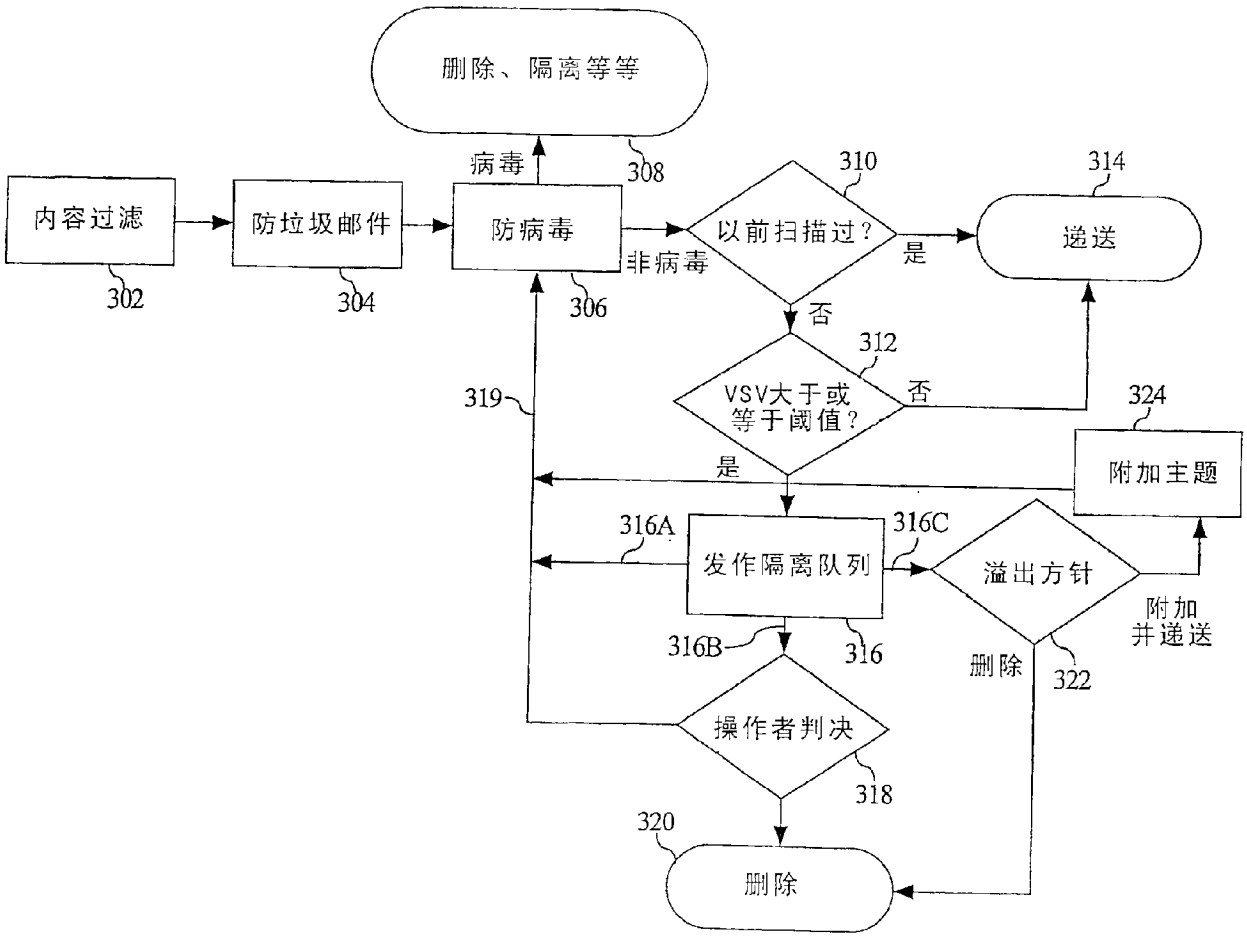

删除、隔离等等
病毒
308
内容过滤
防垃圾邮件
防病毒
非病毒
以前扫描过？
310
是
314
递送
否
302
304
306
312
VSV大于或
等于阈值？
否
319
324
附加主题
是
316A
发作隔离队列
316C
溢出方针
附加
并递送
316B
316
删除
322
操作者判决
318
320
删除

图3

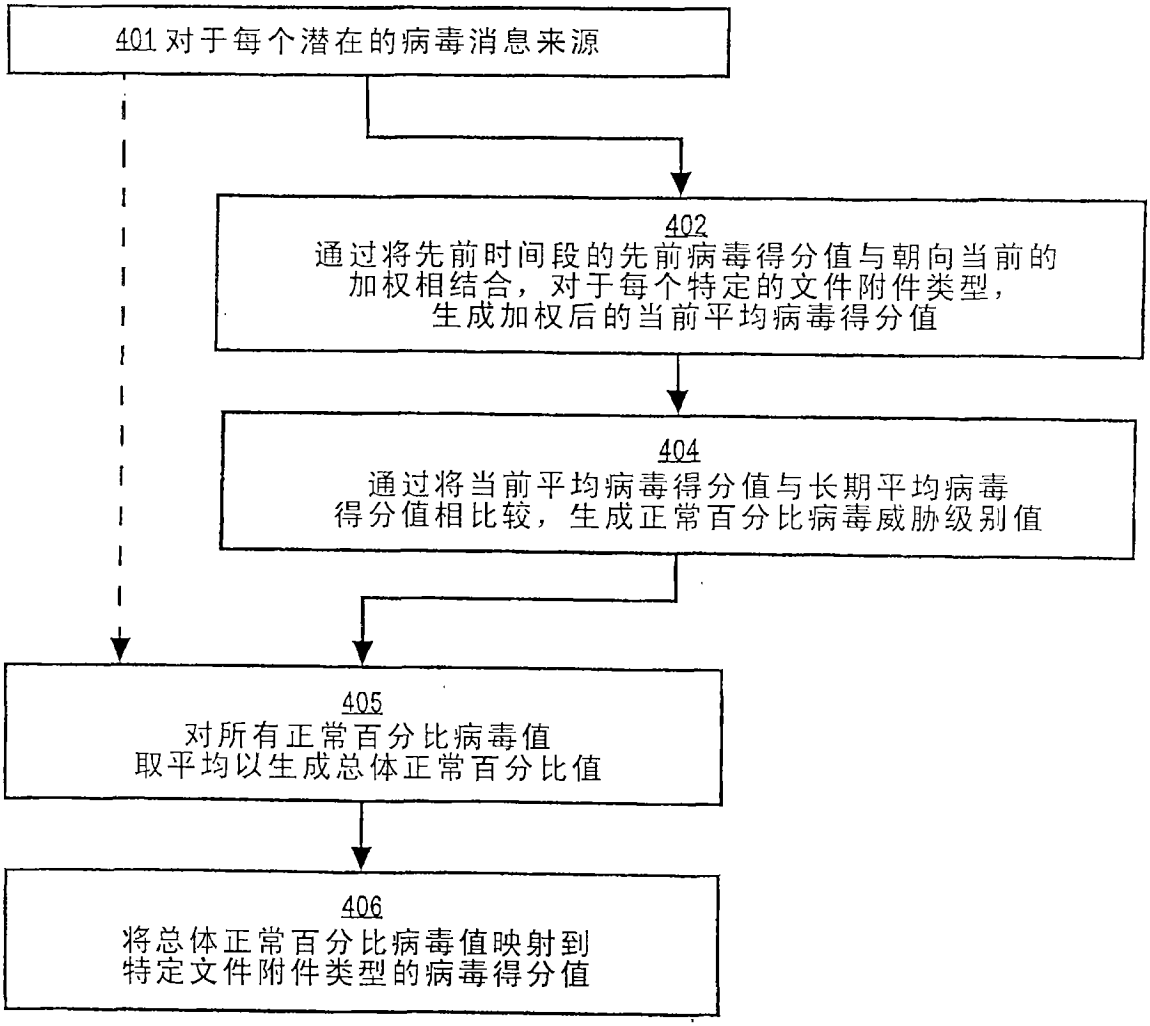

401 对于每个潜在的病毒消息来源
402
通过将先前时间段的先前病毒得分值与朝向当前的加权相结合，对于每个特定的文件附件类型，生成加权后的当前平均病毒得分值
404
通过将当前平均病毒得分值与长期平均病毒得分值相比较，生成正常百分比病毒威胁级别值
405
对所有正常百分比病毒值取平均以生成总体正常百分比值
406
将总体正常百分比病毒值映射到特定文件附件类型的病毒得分值

图4

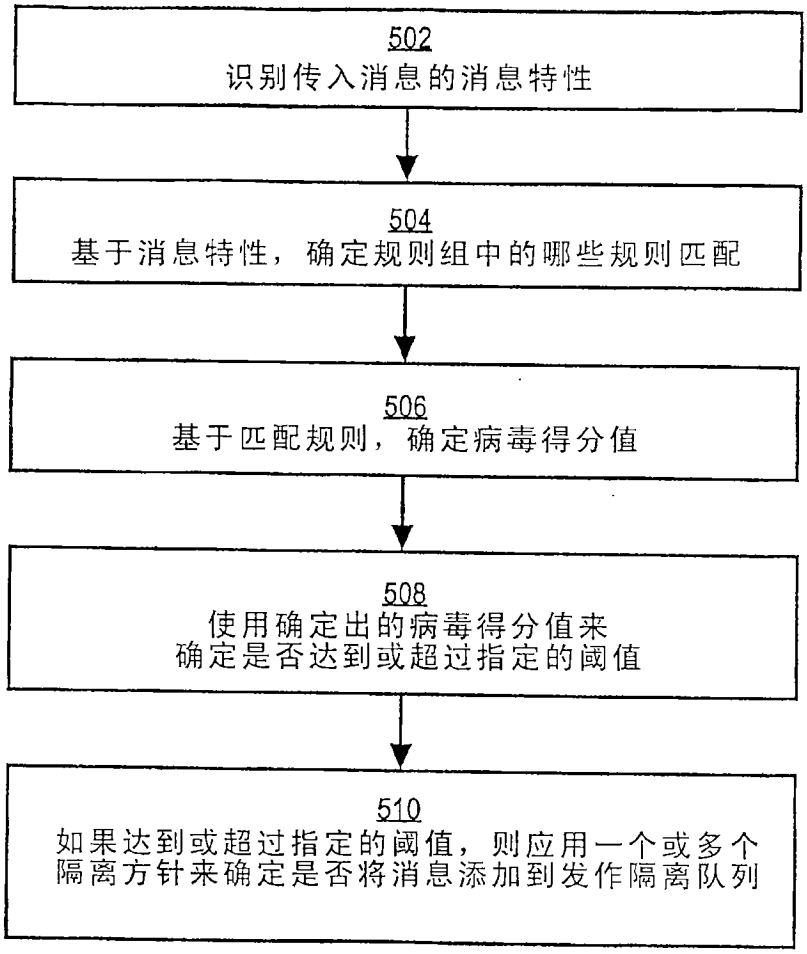

502
识别传入消息的消息特性
504
基于消息特性，确定规则组中的哪些规则匹配
506
基于匹配规则，确定病毒得分值
508
使用确定出的病毒得分值来
确定是否达到或超过指定的阈值
510
如果达到或超过指定的阈值，则应用一个或多个
隔离方针来确定是否将消息添加到发作隔离队列

图5

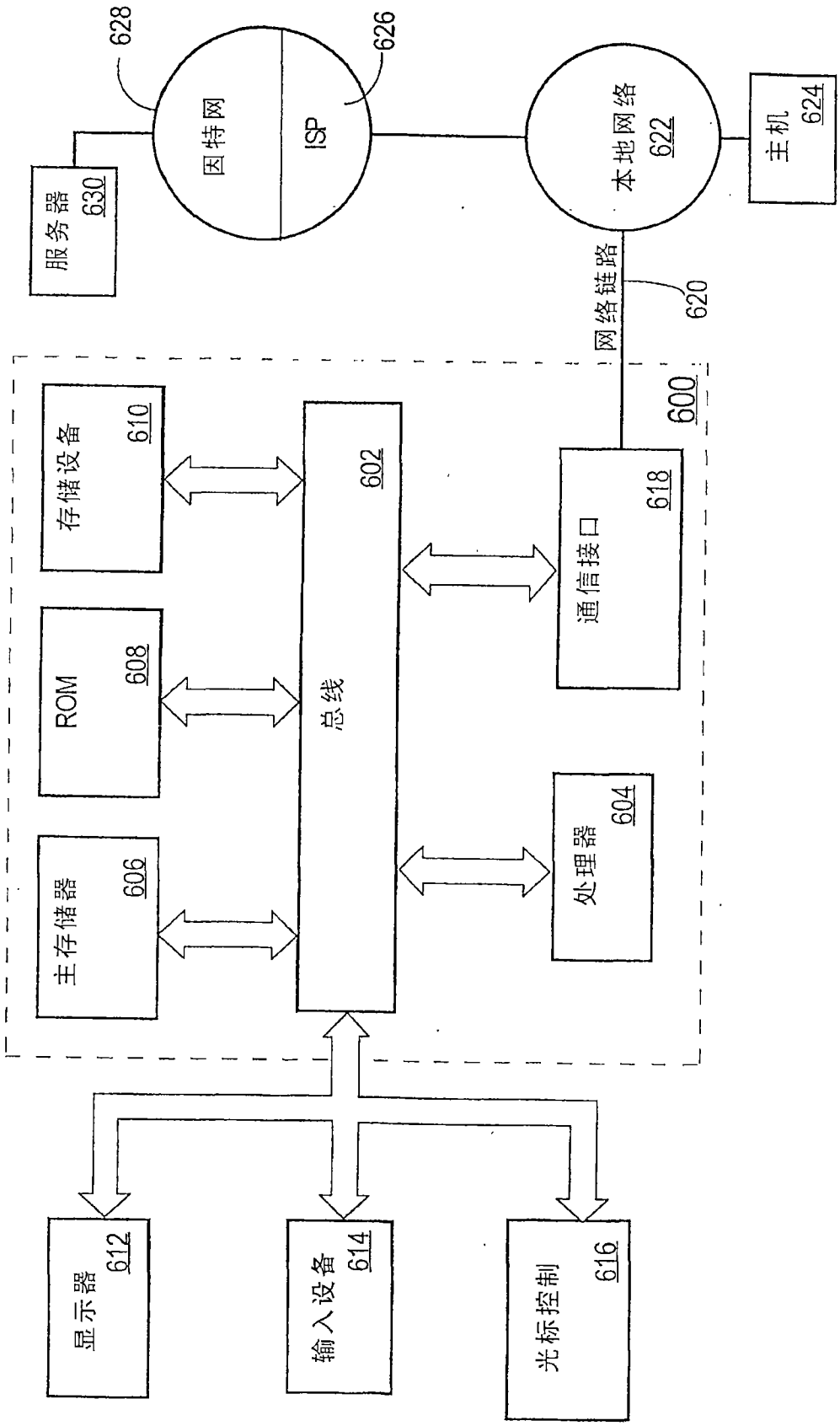

628
因特网
ISP
626
服务器
630
本地网络
622
主机
624
网络链路
620
600
存储设备
610
602
通信接口
618
ROM
608
总线
主存储器
606
处理器
604
显示器
612
输入设备
614
光标控制
616

图6

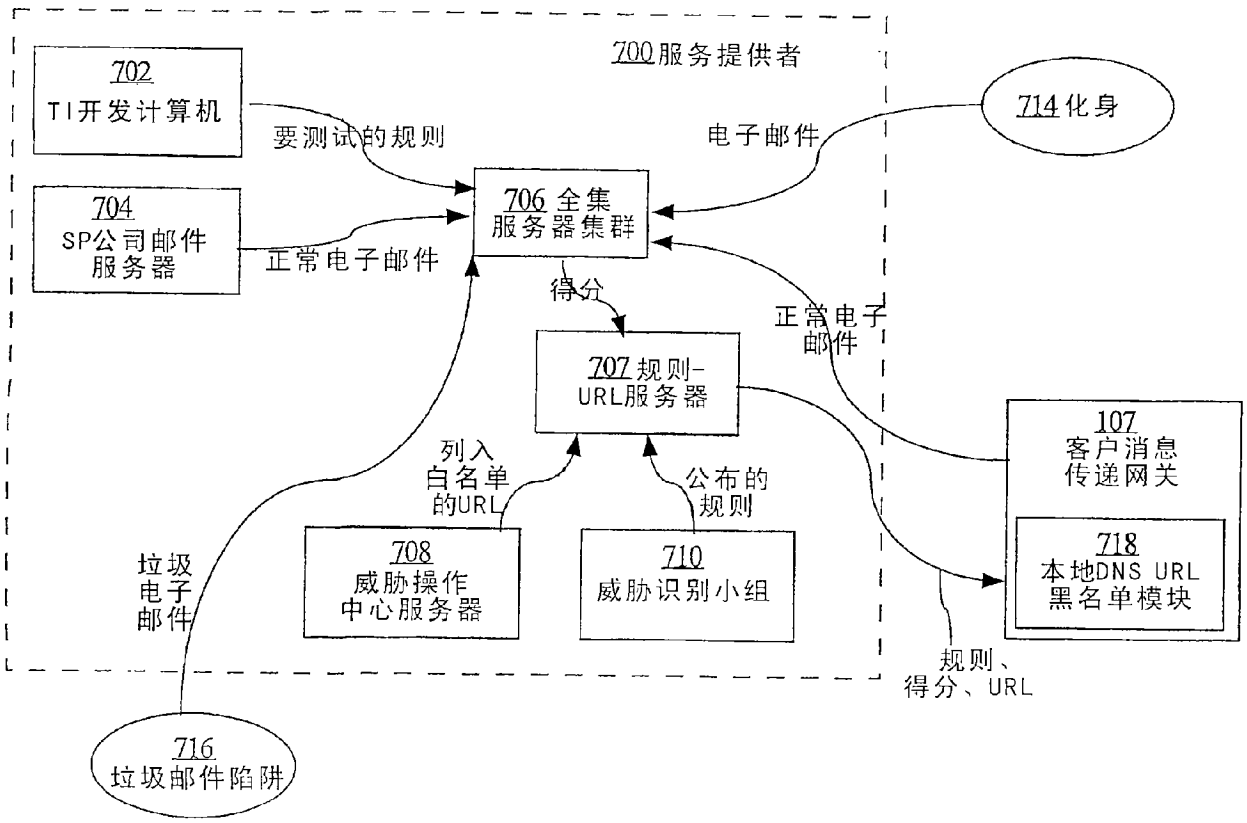

700服务提供者
702
TI开发计算机
要测试的规则
704
SP公司邮件
服务器
正常电子邮件
706 全集
服务器集群
得分
707 规则-
URL服务器
714化身
电子邮件
正常电子
邮件
107
客户消息
传递网关
718
本地DNS URL
黑名单模块
列入
白名单
的URL
公布的
规则
708
威胁操作
中心服务器
710
威胁识别小组
垃圾
电子
邮件
规则、
得分、URL
716
垃圾邮件陷阱

图7

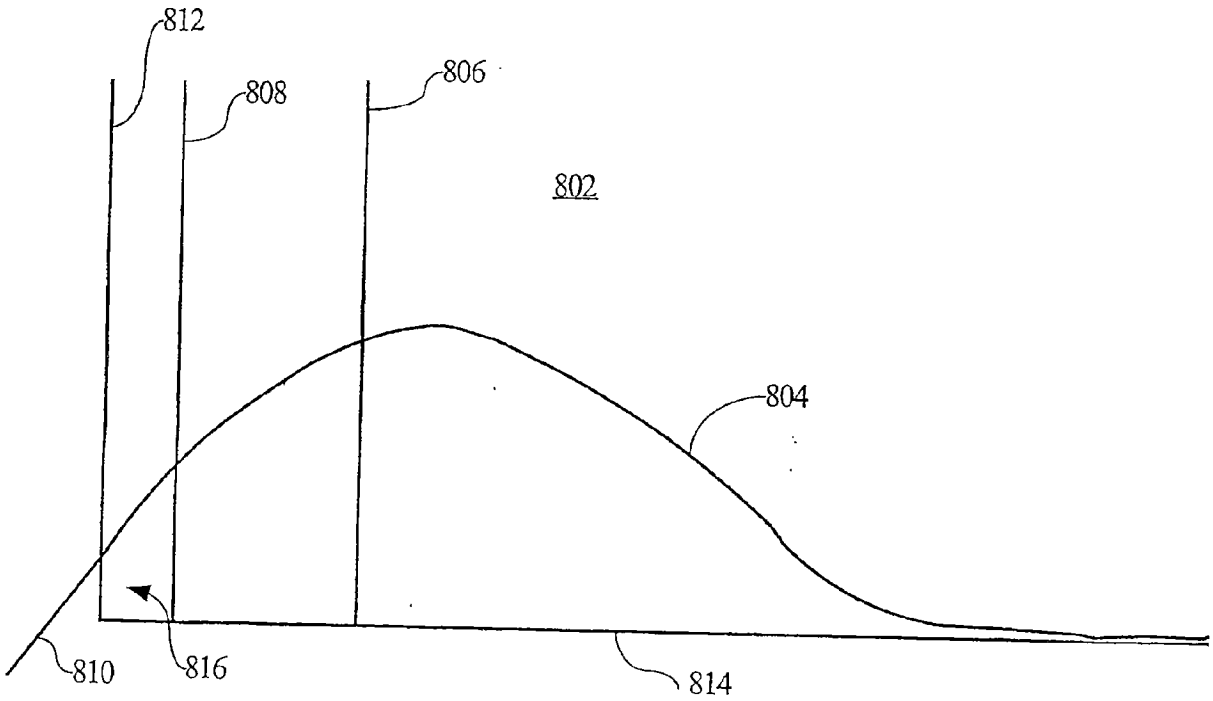


图8

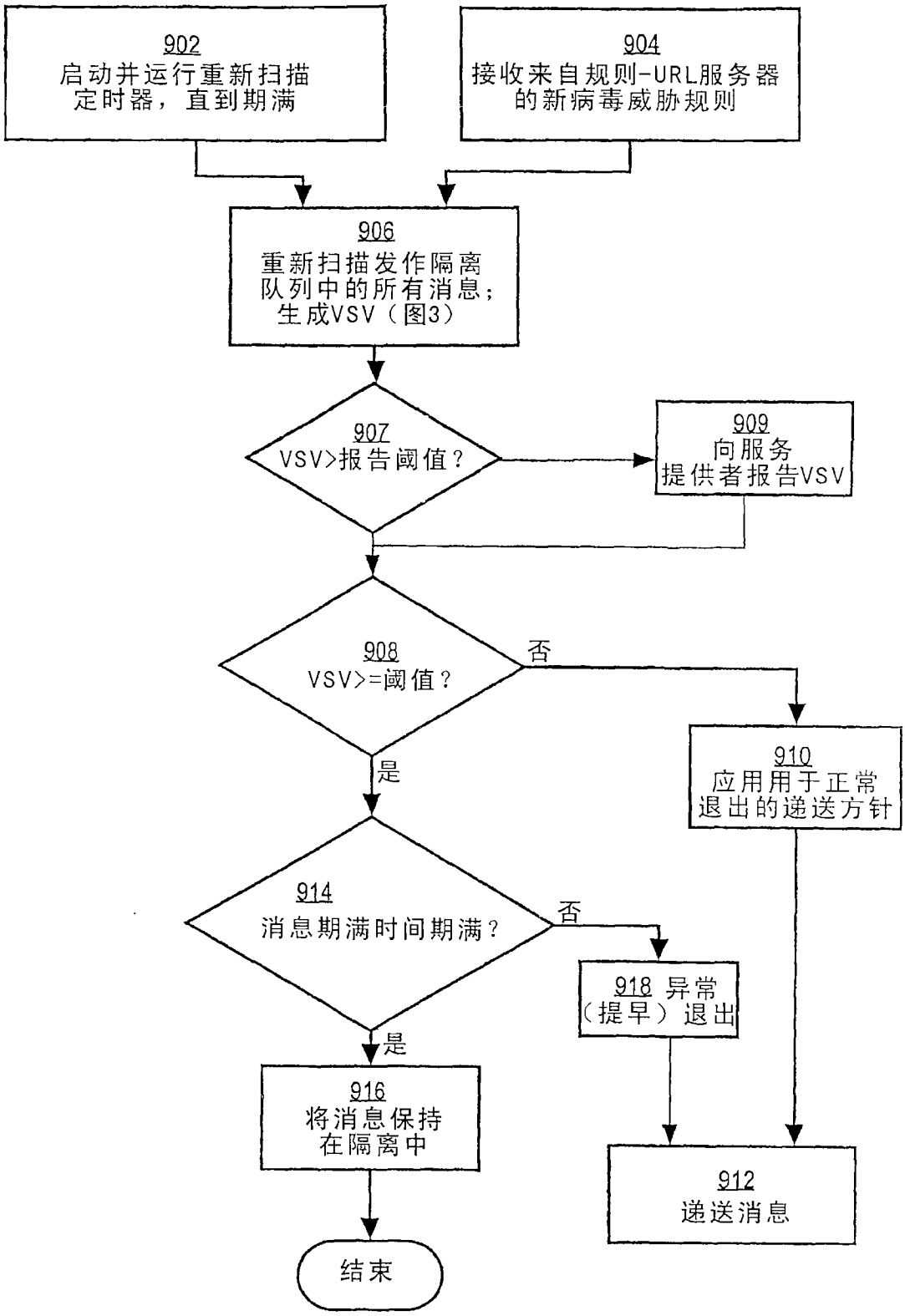

902
启动并运行重新扫描定时器，直到期满
904
接收来自规则-URL服务器的新病毒威胁规则
906
重新扫描发作隔离队列中的所有消息；生成VSV（图3）
907
VSV>报告阈值？
909
向服务提供者报告VSV
908
VSV>=阈值？
否
是
910
应用用于正常退出的递送方针
914
消息期满时间期满？
否
是
918 异常（提早）退出
916
将消息保持在隔离中
912
递送消息
结束

图9

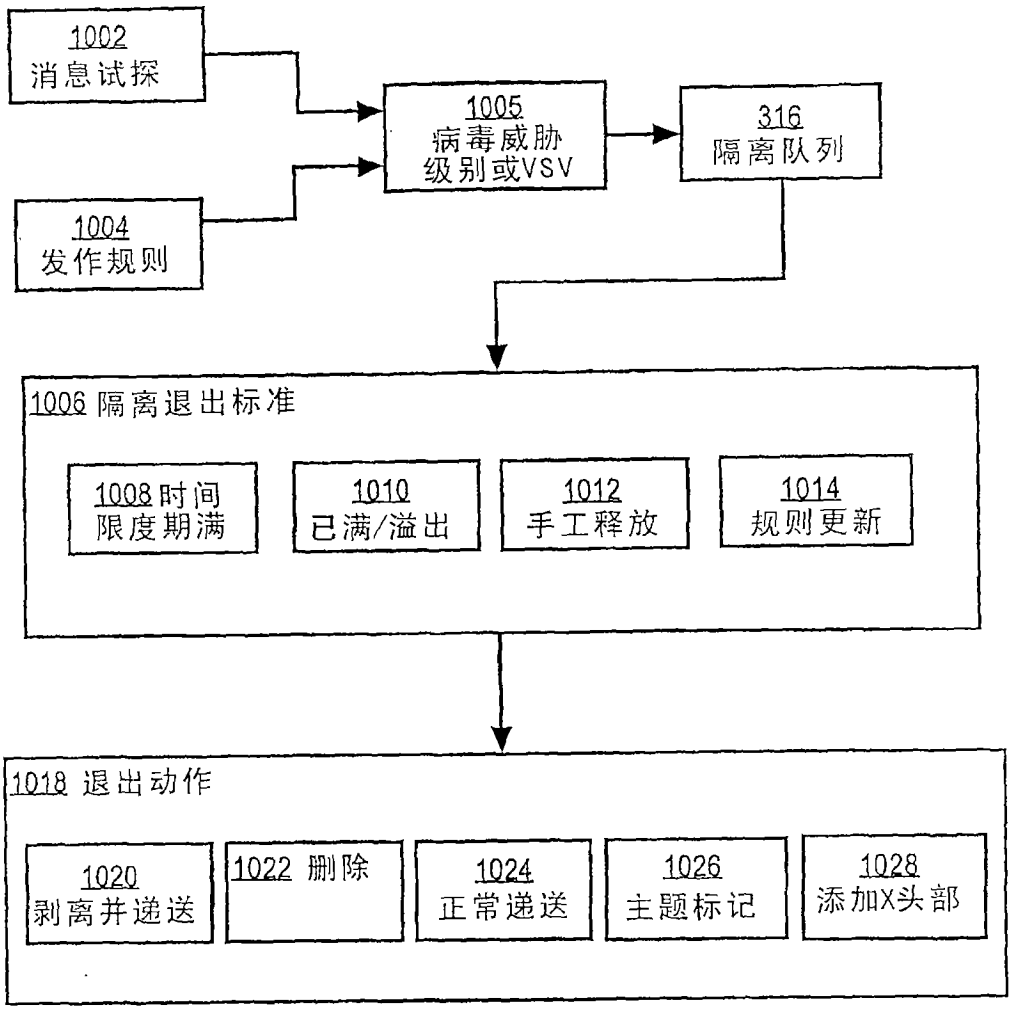

1002
消息试探
1004
发作规则
1005
病毒威胁
级别或VSV
316
隔离队列
1006 隔离退出标准
1008 时间
限度期满
1010
已满/溢出
1012
手工释放
1014
规则更新
1018 退出动作
1020
剥离并递送
1022 删除
1024
正常递送
1026
主题标记
1028
添加X头部

图10

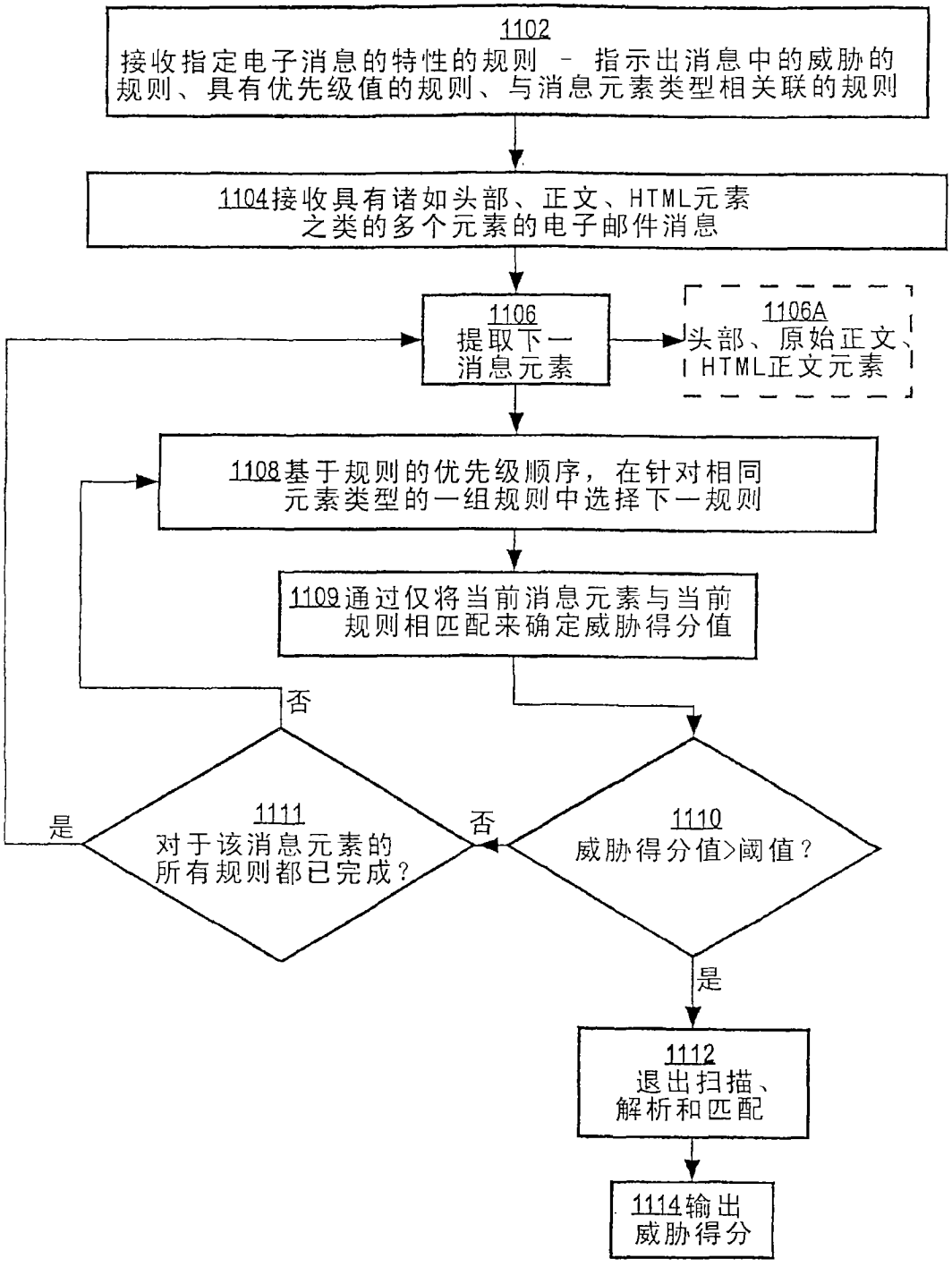

1102
接收指定电子消息的特性的规则 - 指示出消息中的威胁的规则、具有优先级值的规则、与消息元素类型相关联的规则
1104接收具有诸如头部、正文、HTML元素之类的多个元素的电子邮件消息
1106
提取下一消息元素
1106A
头部、原始正文、HTML正文元素
1108基于规则的优先级顺序，在针对相同元素类型的一组规则中选择下一规则
1109通过仅将当前消息元素与当前规则相匹配来确定威胁得分值
否
1111
对于该消息元素的所有规则都已完成？
是
否
1110
威胁得分值>阈值？
是
1112
退出扫描、解析和匹配
1114输出威胁得分

图11