



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2018년06월08일
(11) 등록번호 10-1855539
(24) 등록일자 2018년04월30일

(51) 국제특허분류(Int. Cl.)
G06Q 50/30 (2012.01)
(21) 출원번호 10-2013-7004999
(22) 출원일자(국제) 2011년08월22일
심사청구일자 2016년07월20일
(85) 번역문제출일자 2013년02월27일
(65) 공개번호 10-2013-0103494
(43) 공개일자 2013년09월23일
(86) 국제출원번호 PCT/US2011/048664
(87) 국제공개번호 WO 2012/030576
국제공개일자 2012년03월08일
(30) 우선권주장
12/872,691 2010년08월31일 미국(US)
(56) 선행기술조사문헌
US06421709 B1*
Minh Tran, Grenville Armitage. End-users'
resource consumption of spam and a 3D
anti-spam evaluation framework. TENCON 2005
IEEE Region 10(2007.02.05.)*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
마이크로소프트 테크놀로지 라이선싱, 엘엘씨
미국 워싱턴주 (우편번호 : 98052) 레드몬드 원
마이크로소프트 웨이
(72) 발명자
안데르센 한스 크리스찬
미국 워싱턴주 98052-6399 레드몬드 원 마이크로
소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마
이크로소프트 코포레이션
파나슈크 아나톨리
미국 워싱턴주 98052-6399 레드몬드 원 마이크로
소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마
이크로소프트 코포레이션
(뒷면에 계속)
(74) 대리인
제일특허법인

전체 청구항 수 : 총 20 항

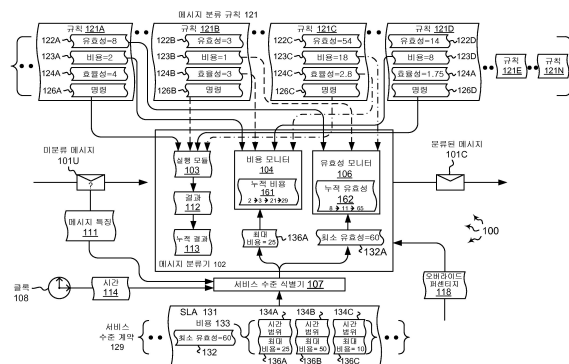
심사관 : 유원석

(54) 발명의 명칭 전자 메시지 스캐닝 규칙의 적응적 선택 기법

(57) 요약

본 발명은 전자 메시지 스캐닝 규칙을 적응적으로 선택하기 위한 방법, 시스템 및 컴퓨터 프로그램 제품에 이른다. 본 발명의 실시형태는 바람직하지 않은 메시지 내용(예를 들어, 스팸, 바이러스, 디지털 누설 등)에 대한 보호를 위해 동적으로 (그리고 잠재적으로는 예측불가능하게) 전자 메시지의 분류의 깊이(depth)/철저함(thoroughness)을 변화시키는 것에 관련된다. 최소 유효성은 유지되고, 가용 자원이 허용한다면, 증가된 보호를 제공하기 위해 초과될 수도 있다. 가용 메시지 분류 규칙의 최적 부분집합은 메시지 마다 선택될 수 있다. 규칙의 선택은 가용 시스템 자원, 최소 희망 유효성(예를 들어, SLA(Service Level Agreement)에 정의된) 및 규칙 특징에 기초한다. 피드백 루프가 선택된 분류 규칙 부분집합을 최적화하기 위해 사용될 수 있다.

대표도



(72) 발명자

리메니 벤카타 소마나드하 사르마

미국 워싱턴주 98052-6399 레드몬드 원 마이크로소프트 웨이 엘씨에이 - 인터내셔널 패이턴즈 마이크로소프트 코포레이션

커스 바트

미국 워싱턴주 98052-6399 레드몬드 원 마이크로소프트 웨이 엘씨에이 - 인터내셔널 패이턴즈 마이크로소프트 코포레이션

명세서

청구범위

청구항 1

하나 이상의 프로세서와 시스템 메모리를 포함하는 컴퓨터 시스템-상기 컴퓨터 시스템은 복수의 전자 메시지 분류 규칙을 포함-에서, 전자 메시지를 분류하는 데 사용되는 규칙을 적응적으로 선택하기 위한 방법으로서,

하나 이상의 전자 메시지를 수신하는 동작과,

상기 하나 이상의 전자 메시지 각각에 대해서, 전자 메시지 분류 규칙의 이전에 선택된 부분집합 내의 각각의 메시지 분류 규칙을 적용함으로써 상기 전자 메시지가 특정 메시지 특성을 가질 확률을 나타내는 결과를 계산하는 동작과,

상기 하나 이상의 전자 메시지 각각에 각각의 전자 메시지 분류 규칙을 적용하는 데 소비되는 자원의 양을 나타내는 자원 비용을 측정하는 동작과,

메시지 분류 규칙의 상기 이전에 선택된 부분집합 내의 각각의 메시지 분류 규칙에 대해, 상기 메시지 분류 규칙에 대한 측정된 자원 비용과 상기 계산된 결과로부터 효율성 메트릭(an efficiency metric)을 합성하는 동작-상기 효율성 메트릭은 전자 메시지 분류의 효율성을 정의하는 메트릭임-과,

상기 복수의 전자 메시지 분류 규칙에 포함되는 전자 메시지 분류 규칙에 대한 기존의 효율성 메트릭(existing efficiency metrics)과 상기 합성된 효율성 메트릭을 비교하는 동작과,

상기 합성된 효율성 메트릭을 기존의 효율성 메트릭과 비교한 결과에 적어도 부분적으로 기초하여 이후에 수신되는 전자 메시지를 분류하는 데 사용하기 위해 상기 복수의 전자 메시지 분류 규칙 중에서 전자 메시지 분류 규칙의 새로운 부분집합을 선택하는 동작을 포함하는

규칙 선택 방법.

청구항 2

제 1 항에 있어서,

상기 전자 메시지 분류 규칙의 새로운 부분집합을 선택하는 동작은 SLA(Service Level Agreement)에 따라 전자 메시지 분류 규칙의 새로운 부분집합을 선택하는 동작을 포함하는

규칙 선택 방법.

청구항 3

제 1 항에 있어서,

상기 전자 메시지 분류 규칙의 새로운 부분집합을 선택하는 동작은 효율성 점수에 기초하여 상기 복수의 전자 메시지 분류 규칙을 재정렬하는 동작을 포함하는

규칙 선택 방법.

청구항 4

제 1 항에 있어서,

상기 하나 이상의 전자 메시지를 수신하는 동작은 상기 하나 이상의 전자 메일 메시지를 수신하는 동작을 포함하는

규칙 선택 방법.

청구항 5

제 1 항에 있어서,

상기 하나 이상의 전자 메시지를 수신하는 동작은 상기 하나 이상의 SMS(Short Message Service) 메시지를 수신하는 동작을 포함하는

규칙 선택 방법.

청구항 6

제 1 항에 있어서,

상기 하나 이상의 전자 메시지를 수신하는 동작은 상기 하나 이상의 파일을 수신하는 동작을 포함하는

규칙 선택 방법.

청구항 7

제 1 항에 있어서,

상기 복수의 전자 메시지 분류 규칙은 전자 메시지를 스팸(SPAM)으로 또는 정당한 것(legitimate)으로 분류하는 데 사용되는

규칙 선택 방법.

청구항 8

제 1 항에 있어서,

상기 복수의 전자 메시지 분류 규칙은 전자 메시지를 악성코드(malware)를 포함하거나 악성코드를 포함하지 않는 것으로 분류하는 데 사용되는

규칙 선택 방법.

청구항 9

제 1 항에 있어서,

상기 복수의 전자 메시지 분류 규칙은 전자 메시지를 민감한 디지털 정보를 포함하거나 민감한 디지털 정보를 포함하지 않는 것으로 분류하는 데 사용되는

규칙 선택 방법.

청구항 10

하나 이상의 프로세서와 시스템 메모리를 포함하는 컴퓨터 시스템-상기 컴퓨터 시스템은 복수의 전자 메시지 분류 규칙을 포함-에서, 전자 메시지를 분류하는 데 사용되는 규칙을 적응적으로 선택하기 위한 방법으로서,

하나 이상의 전자 메시지를 수신하는 동작과,

상기 하나 이상의 전자 메시지 각각에 대하여:

전자 메시지 분류 규칙의 이전에 선택된 부분집합 내의 각각의 메시지 분류 규칙을 상기 전자 메시지에 적용하는 동작-상기 전자 메시지 분류 규칙의 이전에 선택된 부분집합은 상기 복수의 전자 메시지 분류 규칙의 부분집합임-과,

상기 전자 메시지 분류 규칙의 이전에 선택된 부분집합 내 각각의 전자 메시지 규칙에 대하여:

상기 전자 메시지 규칙이 상기 전자 메시지가 특정 메시지 특성을 가질 확률을 나타내는 결과를 계산하는 동작과,

상기 전자 메시지에 상기 전자 메시지 분류 규칙을 적용하는 데 소비되는 자원의 양을 나타내는 자원 비용을 측정하는 동작과,

각각의 전자 메시지에 각각의 전자 메시지 분류 규칙을 적용하는 것과 관련되는 상기 측정된 자원 비용과 상기 계산된 결과를 보유하는 동작과,

메시지 분류 규칙의 상기 이전에 선택된 부분집합 내의 각각의 메시지 분류 규칙에 대해, 상기 메시지 분류 규칙에 대한 상기 보유된 계산된 결과 및 측정된 자원 비용으로부터 효율성 메트릭을 합성하는 동작 -상기 효율성 메트릭은 전자 메시지 분류의 효율성을 정의하는 메트릭임- 과,

상기 복수의 전자 메시지 분류 규칙에 포함된 전자 메시지 분류 규칙에 대한 기존의 효율성 메트릭과 상기 합성된 효율성 메트릭을 비교하는 동작과,

상기 합성된 효율성 메트릭을 기존의 효율성 메트릭과 비교한 결과에 적어도 부분적으로 기초하여 이후에 수신되는 전자 메시지를 분류하는 데 사용하기 위해 상기 복수의 전자 메시지 분류 규칙 중에서 전자 메시지 분류 규칙의 새로운 부분집합을 선택하는 동작을 포함하는

규칙 선택 방법.

청구항 11

제 10 항에 있어서,

전자 메시지 분류 규칙의 이전에 선택된 부분집합 내의 각각의 메시지 분류 규칙을 적용하기 전에, 계산된 효율성 점수에 기초하여 전자 메시지 분류 규칙의 상기 선택된 부분집합을 선택하는 동작을 더 포함하는

규칙 선택 방법.

청구항 12

제 10 항에 있어서,

전자 메시지에 적어도 하나의 메시지 분류 규칙을 적용하는 것과 관련된 외부 피드백을 수신하는 동작과,

상기 전자 메시지에 상기 적어도 하나의 메시지 분류 규칙을 적용하는 것으로부터의 상기 계산된 결과에 상기 외부 피드백을 통합하는 동작을 더 포함하는

규칙 선택 방법.

청구항 13

제 12 항에 있어서,

상기 수신된 외부 피드백은 상기 전자 메시지에 상기 적어도 하나의 메시지 분류 규칙을 적용하는 것으로부터의 상기 계산된 결과가 오부정(false negative) 또는 오긍정(false positive) 중 하나임을 나타내는

규칙 선택 방법.

청구항 14

제 12 항에 있어서,

상기 외부 피드백을 통합하는 상기 계산된 결과에 기초하여 상기 적어도 하나의 메시지 분류 규칙에 대한 유효성 점수를 갱신하는 동작을 더 포함하는

규칙 선택 방법.

청구항 15

제 10 항에 있어서,

상기 전자 메시지 분류 규칙의 새로운 부분집합을 선택하는 동작은 SLA(Service Level Agreement)에 따라 전자 메시지 분류 규칙의 새로운 부분집합을 선택하는 동작을 포함하는

규칙 선택 방법.

청구항 16

제 10 항에 있어서,

상기 전자 메시지 분류 규칙의 새로운 부분집합을 선택하는 동작은 효율성 점수에 기초하여 상기 복수의 전자 메시지 분류 규칙을 재정렬하는 동작을 포함하는

규칙 선택 방법.

청구항 17

스팸 검출 규칙을 적응적으로 선택하는 시스템에 있어서,

하나 이상의 프로세서와,

시스템 메모리와,

복수의 스팸 검출 규칙이 저장되고 메시지 분류기와 규칙 선택 및 재정렬 모듈을 나타내는 실행가능 명령이 저장되어 있는 하나 이상의 컴퓨터 저장 매체로서, 상기 메시지 분류기는,

하나 이상의 전자 메일 메시지를 수신하고,

상기 하나 이상의 전자 메일 메시지 각각에 대하여, 스팸 검출 규칙의 이전에 선택된 부분집합 내의 각각의 스팸 검출 규칙을 상기 전자 메일 메시지에 적용하고-상기 스팸 검출 규칙의 이전에 선택된 부분집합은 상기 복수의 스팸 검출 규칙의 부분집합임-,

상기 스팸 검출 규칙의 이전에 선택된 부분집합 내 각각의 스팸 검출 규칙에 대하여:

상기 전자 메일 메시지가 스팸일 확률을 나타내는 결과를 계산하고,

상기 하나 이상의 전자 메일 메시지 각각에 상기 스팸 검출 규칙을 적용하는 데 소비되는 자원의 양을 나타내는 자원 비용을 측정하고,

상기 스팸 검출 규칙에 대해 상기 계산된 결과 및 측정된 자원 비용으로부터 효율성 메트릭을 합성 -상기 효율성 메트릭은, 상기 계산된 결과를 측정된 자원 비용으로 나눈 몫에 기초하여, 전자 메시지를 스팸으로 분류하는 것의 효율성을 정의하는 메트릭임-

하도록 구성되고,

상기 규칙 선택 및 재정렬 모듈은,

상기 복수의 스팸 검출 규칙에 포함된 스팸 검출 규칙에 대한 기존의 효율성 메트릭과 상기 합성된 효

유효성 메트릭을 비교하고,

상기 합성된 효율성 메트릭을 기존 효율성 메트릭과 비교한 결과에 적어도 부분적으로 기초하여 이후에 수신되는 전자 메일 메시지를 분류하는 데 사용하기 위해 스팸 검출 규칙의 새로운 부분집합을 선택하도록 구성되는

규칙 선택 시스템.

청구항 18

제 17 항에 있어서,

스팸 검출 규칙의 새로운 부분집합을 선택하도록 구성되는 상기 규칙 선택 및 재정렬 모듈은, SLA(Service Level Agreement)에 따라 스팸 규칙의 새로운 부분집합을 선택하도록 구성되는 상기 규칙 선택 및 재정렬 모듈을 포함하는

규칙 선택 시스템.

청구항 19

제 17 항에 있어서,

스팸 검출 규칙의 새로운 부분집합을 선택하도록 구성되는 상기 규칙 선택 및 재정렬 모듈은, 효율성 점수에 기초하여 상기 복수의 스팸 분류 규칙을 재정렬하도록 구성되는 상기 규칙 선택 및 재정렬 모듈을 포함하는

규칙 선택 시스템.

청구항 20

제 17 항에 있어서,

하나 이상의 전자 메일 메시지를 수신하도록 구성되는 상기 메시지 분류기는, 인터넷으로부터 전자 메일 메시지를 수신하도록 구성되는 메시지 분류기를 포함하는

규칙 선택 시스템.

발명의 설명

기술 분야

배경 기술

[0001] 컴퓨터 시스템과 관련 기술은 사회의 많은 측면에 영향을 준다. 실제로, 컴퓨터 시스템의 정보 처리 능력은 우리가 살아가고 일하는 방식을 변화시켰다. 컴퓨터 시스템은 이제, 통상적으로 컴퓨터 시스템의 등장 전에는 수동으로 수행되던 수많은 태스크(예를 들어, 워드 프로세싱, 스케줄링, 회계 등)를 수행한다. 더 최근에는, 컴퓨터 시스템은 서로, 그리고 다른 전자 장치에 결합되어, 컴퓨터 시스템과 다른 전자 장치가 전자적 데이터를 전송할 수 있는 유선 및 무선의 컴퓨터 네트워크를 형성하였다. 따라서, 많은 연산 태스크의 수행이 다수의 상이한 컴퓨터 시스템 및/또는 다수의 상이한 연산 환경에 걸쳐 분산된다.

[0002] 많은 연산 환경에서, 예를 들어 이메일 메시지와 같은 전자적 메시지가 컴퓨터 시스템 사용자들 사이에서 정당하게 정보를 교환하는 데 사용된다. 그러나, 이들 연산 환경은 또한 사용자가, 종종 스팸(SPAM)이라고 지칭되는 요청하지 않은 및/또는 원하지 않는 전자 메시지를 겪게 한다. 많은 상이한 기술이 스팸을 검색하고 차단하

기 위해 개발되었다.

- [0003] 스팸 스캐닝 기술은 보통 한 세트의 메트릭을 절충하여야 하는데, 이는 유효성(effectiveness), 정확성(accuracy), 효율성(efficiency) 및 레이턴시(latency)를 포함한다. 유효성은 어떤 범위에서 스팸이 식별되고 차단될 수 있는지에 관련된다. 정확성은 어떤 범위에서 정당한 메시지가 스팸으로 잘못 식별되는지 (예를 들어, 오긍정(false positive)의 비율)에 관련된다. 효율성은 메시지를 스팸이나 정당한 것으로 식별하는 데 관련된 자원 소비에 관련된다. 레이턴시는 스캐닝의 결과로 각각의 개별 메시지가 전달에 있어 얼마나 지연되는지에 관련된다.
- [0004] 한 영역에서의 개선은 보통 하나 이상의 다른 영역에서의 열화를 의미하기 때문에, 이들 메트릭 사이의 균형은 상대적으로 복잡한 태스크일 수 있다. 예를 들어, 더 공격적인 안티-스팸 검출(유효성 증가)은 더 높은 오긍정(정확성 감소) 및/또는 더 복잡한 처리 알고리즘에 기인한 더 높은 CPU 부하(자원 소비 증가)를 유발할 수 있다.
- [0005] 또한, 이들 메트릭의 소정 조합은 종종 서비스 제공자가 지원하는 SLA(Service Level Agreement)로 매핑된다. 예를 들어, 안티-스팸 서비스 제공자는 X 이상의 유효성, Y 이상의 정확성 등을 지원하는 데 합의할 수 있다. SLA의 규정을 위반하는 것, 예를 들어, 유효성이 소정 시간 동안 X 미만이 되게 하는 것은 안티-스팸 서비스 제공자가 고객에서 소정의 금전적 환불을 하도록 할 수 있다.
- [0006] 그러나, 동시에, 안티-스팸 서비스는 통상 매우 다양한 시스템 부하를 경험한다. 예를 들어, 여하한 주어진 날 동안에, 주말에 그리고 계절에 따라 스팸의 양 및/또는 정당한 전자 메시지의 양이 변동할 수 있다. 불행하게도, 이는 서비스 제공자가 과제공(over provisioning)하도록 유발할 수 있다. 예를 들어, 공통의 디자인 패턴은 피크 부하 시간에 SLA를 보장하도록 충분한 전력을 갖는 스캐닝 서비스를 구축하는 것이고, 이는 평균 부하보다 3에서 5배 높을 수 있다.
- [0007] 실제로, 피크 부하를 위해 설계하는 것은 현저한 부분의 시간에 자원이 (잠재적으로 심각하게) 덜 사용되도록 한다. 통상적으로 스캐닝은, 존재한다면 제한된 가용 자원의 고려로써, 고정된 수의 단계 및/또는 고정된 수의 스캐닝 규칙의 사용을 포함한다. 그러므로, 비-피크 시간에, 자원이 추가의 스캐닝을 위해 사용가능하더라도 고정된 수의 단계 및/또는 규칙이 하나의 메시지를 스캐닝하는 데 사용된다. 이와 같이, 피크 부하를 위해 설계하는 것은 판매 제품의 비용의 관점에서 바람직하지 않지만, 그럼에도 SLA를 위해 요구된다.
- [0008] 예를 들어, 보통 고객, 프리미엄 고객, 저가 고객 등과 같은 다양한 수준의 서비스 각각을 통상 그들의 SLA에서 정의되는 상이한 메트릭으로 지원하는 때에 더욱 복잡해질 수 있다. 종종, 프리미엄 서비스 제안은, 서비스 제공자 측에서 더 많은 연산/처리 자원을 요구하는 더 높은 수준의 서비스(예를 들어, 높은 정확도, 낮은 레이턴시 등)를 보장하는 SLA와 함께 제공된다.
- [0009] 상이한 수준의 서비스를 다루기 위한 한 가지 설계 패턴은 모든 수준의 서비스에 대해 공통의 안티-스팸 서비스를 사용하는 것이다. 각각의 수준의 서비스는 요구되는 SLA에 의해 메시지가 거치는 처리 단계 및/또는 규칙의 수에 제한된다. 예를 들어, 프리미엄 고객의 이메일은 10개의 처리 단계를 거칠 수 있는 반면, 기본 고객의 이메일은 5개의 처리 단계만을 거칠 수 있다. 추가의 스캐닝을 위한 자원이 사용가능할 수 있어도 스캐닝의 품질을 낮추는 (예를 들어, 유효성 감소) 대가로 기본 고객을 서비스하는 비용이 감소된다. 더 낮은 스캐닝 품질에 추가하여, 기본 고객은 또한 기본 고객에게 제공되는 보호 수준에서의 약점 (시스템의 예측가능성)을 이용하는 타겟팅된 공격에 더 취약하다.
- [0010] 다른 통상의 패턴은 프리미엄 고객을 위한 하나와 기본 고객을 위한 다른 하나의 2개의 별개 시스템을 구축하는 것이다. 각각의 시스템은 고객의 유형에 따라 서비스의 품질과 서비스의 비용의 균형을 맞추도록 설계된다. 불행하게도, 이 시스템 유형은 중복되는 인프라스트럭처를 요구하여, 더 높은 전체 비용을 유발할 뿐만 아니라 피크 부하에서 SLA를 만족시키기 위한 과제공의 일반적 문제를 유발한다.

발명의 내용

해결하려는 과제

과제의 해결 수단

- [0011] 본 발명은 전자 메시지 스캐닝 규칙을 적응적으로 선택하기 위한 방법, 시스템 및 컴퓨터 프로그램 제품에 이른다. 일부 실시형태에서, 전자 메시지를 분류하는 데 사용되는 규칙은 적응적으로 선택된다. 하나 이상의 전자 메시지가 수신된다. 하나 이상의 전자 메시지 각각에 대해, 이전에 선택된 전자 메시지 분류 규칙의 부분집합 내의 각 메시지 분류 규칙이 전자 메시지에 적용된다. 이전에 선택된 전자 메시지 분류 규칙의 부분집합은 복수의 가용 전자 메시지 분류 규칙의 부분집합이다.
- [0012] 이전에 선택된 전자 메시지 분류 규칙의 부분집합 내의 각각의 전자 메시지 분류 규칙에 대해, 전자 메시지가 특정한 메시지 특성을 가질 확률을 나타내는 결과가 계산된다. 전자 메시지 분류 규칙을 전자 메시지에 적용하는 데 소비되는 자원의 양을 나타내는 자원 비용이 계산된다. 각 전자 메일 분류 규칙을 각각의 전자 메시지에 적용하는 것과 관련된 측정된 자원 비용과 계산 결과가 보유된다.
- [0013] 이전에 선택된 메시지 분류 규칙의 부분집합 내의 각각의 메시지 분류 규칙에 대해, 보유된 계산 결과와 메시지 분류 규칙에 대한 측정된 자원 비용으로부터 효율성 메트릭이 합성된다. 합성된 효율성 메트릭은 복수의 가용 전자 메시지 분류 규칙에 포함된 전자 메시지 분류 규칙에 대한 기존 효율성 메트릭과 비교된다. 합성된 효율성 메트릭을 기존 효율성 메트릭에 비교한 결과에 적어도 부분적으로 기초하여 복수의 가용 전자 메시지 분류 규칙 중에서 전자 메시지 분류 규칙의 새로운 부분집합이 선택된다. 전자 메시지 분류 규칙의 새로운 부분집합은 이후에 수신되는 전자 메시지를 분류하는 데 사용하기 위한 것이다. 따라서, 메시지 분류 규칙은 변화하는 메시지 내용 패턴에 적응하도록 사용되었다 사용되지 않았다 할 수 있다.
- [0014] 본 개요는 아래 상세한 설명에서 더 자세히 설명하는 개념 중 선택된 것을 단순화된 형태로 소개하기 위해 제공된다. 이 개요는 청구된 주제의 핵심 특징 또는 중요 특징을 식별하려는 것이 아니고, 청구된 주제의 범위를 결정하는 수단으로 사용되려는 것도 아니다.
- [0015] 본 발명의 추가적인 특징 및 장점은 아래에 따라오는 설명에 제시될 것이고, 부분적으로는 설명으로부터 명백할 것이며, 또는 본 발명의 실시예에 의해 이해될 수 있을 것이다. 본 발명의 특징과 장점은 첨부된 청구범위에서 구체적으로 특정된 수단 및 조합에 의해 실현되고 획득될 수 있을 것이다. 이들 및 기타의 본 발명 특징은 아래의 설명과 첨부된 청구범위로부터 더 완전히 명확하게 될 것이고, 또는 이후에 제시되는 본 발명의 실시예에 의해 이해될 수 있을 것이다.

도면의 간단한 설명

- [0016] 본 발명의 상술한 장점과 특징 및 다른 장점과 특징이 언어될 수 있는 방식을 설명하기 위해, 위에서 간략히 설명된 본 발명의 더 구체적인 설명이 첨부된 도면에 도시된 본 발명의 구체적 실시형태를 참조하여 이루어질 것이다. 이들 도면이 본 발명의 전형적인 실시형태만을 도시하고 따라서 본 발명의 범위를 제한하는 것으로 생각되어서는 안된다는 것을 이해한다면, 본 발명은 다음과 같은 첨부도면의 사용을 통해 더 구체적이고 상세하게 설명될 것이다.
- 도 1은 전자 메시지를 적응적으로 분류하는 것을 용이하게 하는 예시적인 컴퓨터 아키텍처를 도시한다.
- 도 2는 전자 메시지를 분류하는 데 사용되는 규칙을 적응적으로 선택하는 것을 용이하게 하는 예시적인 컴퓨터 아키텍처를 도시한다.
- 도 3은 전자 메시지를 적응적으로 분류하기 위한 예시적 방법의 흐름도를 도시한다.
- 도 4는 전자 메시지를 분류하는 데 사용되는 규칙을 적응적으로 선택하기 위한 예시적 방법의 흐름도를 도시한다.
- 도 5는 적응적인 전자 메시지 스캐닝 및 전자 메시지를 분류하는 데 사용되는 규칙을 적응적으로 선택하는 것을 용이하게 하는 다른 예시적 컴퓨터 아키텍처를 도시한다.

발명을 실시하기 위한 구체적인 내용

- [0017] 본 발명은 전자 메시지 스캐닝 규칙을 적응적으로 선택하기 위한 방법, 시스템 및 컴퓨터 프로그램 제품에 이른다. 일부 실시형태에서, 전자 메시지를 분류하는 데 사용되는 규칙은 적응적으로 선택된다. 하나 이상의 전자

메시지가 수신된다. 하나 이상의 전자 메시지 각각에 대해, 이전에 선택된 전자 메시지 분류 규칙의 부분집합 내의 각 메시지 분류 규칙이 전자 메시지에 적용된다. 이전에 선택된 전자 메시지 분류 규칙의 부분집합은 복수의 가용 전자 메시지 분류 규칙의 부분집합이다.

[0018] 이전에 선택된 전자 메시지 분류 규칙의 부분집합 내의 각각의 전자 메시지 분류 규칙에 대해, 전자 메시지가 특정한 메시지 특성을 가질 확률을 나타내는 결과가 계산된다. 전자 메시지 분류 규칙을 전자 메시지에 적용하는 데 소비되는 자원의 양을 나타내는 자원 비용이 계산된다. 각 전자 메일 분류 규칙을 각각의 전자 메시지에 적용하는 것과 관련된 측정된 자원 비용과 계산 결과가 보유된다.

[0019] 이전에 선택된 메시지 분류 규칙의 부분집합 내의 각각의 메시지 분류 규칙에 대해, 보유된 계산 결과와 메시지 분류 규칙에 대한 측정된 자원 비용으로부터 효율성 매트릭이 합성된다. 합성된 효율성 매트릭은 복수의 가용 전자 메시지 분류 규칙에 포함된 전자 메시지 분류 규칙에 대한 기존 효율성 매트릭과 비교된다. 합성된 효율성 매트릭을 기존 효율성 매트릭에 비교한 결과에 적어도 부분적으로 기초하여 복수의 가용 전자 메시지 분류 규칙 중에서 전자 메시지 분류 규칙의 새로운 부분집합이 선택된다. 전자 메시지 분류 규칙의 새로운 부분집합은 이후에 수신되는 전자 메시지를 분류하는 데 사용하기 위한 것이다. 따라서, 메시지 분류 규칙은 변화하는 메시지 내용 패턴에 적응하도록 서비스로 가져오거나 서비스에서 뺄 수 있다.

[0020] 다른 실시형태에서, 전자 메시지는 적응적으로 분류된다. 발신자로부터 수신자에게 전송된 전자 메시지는 특정한 시간에 수신된다. 수신된 전자 메시지에 적용가능한 서비스의 수준은 발신자와 수신자 중 하나 이상에 기초하여 식별된다.

[0021] 서비스의 수준은 적어도 최소 유효성 값 및 전자 메시지 스캐닝을 위한 최대 비용 값의 세트를 정의한다. 최소 유효성 값은, 메시지 분류 규칙의 조합이 서비스의 수준을 만족시키기 위해 가져야 하는 것인 최소 누적 총 유효성(minimum cumulative total effectiveness)을 나타낸다. 최대 비용 값의 세트 내의 각각의 최대 비용 값은 상이한 지정 시간 기간에 대응하고, 전자 메시지에 메시지 분류 규칙을 적용하는 데 사용될 수 있는 자원의 총량을 나타낸다. 최대 비용 값의 세트로부터의 최대 비용 값은, 선택된 최대 비용 값에 대한 지정된 시간 기간 내에 있는 특정 시간에 기초하여 수신된 전자 메시지를 스캐닝하는 때에 사용되기 위해 선택된다.

[0022] 하나 이상의 메시지 분류 규칙이 수신된 전자 메시지에 적용된다. 각각의 메시지 분류 규칙은 측정된 유효성, 측정된 자원 비용 및 측정된 자원 비용에 비추어 측정된 유효성에 기초한 계산된 효율성을 갖는다. 측정된 유효성은 전자 메시지가 특정한 메시지 특성을 가지는 것으로 적절하게 식별한 확률을 나타낸다. 하나 이상의 메시지 분류 규칙은 서비스 수준에서 정의된 최소 누적 총 유효성에 도달할 때까지 효율성의 순서로 적용된다.

[0023] 각각의 메시지 분류 규칙은, 전자 메시지가 특정 메시지 특성을 가질 확률을 나타내는 결과를 생성하도록 전자 메시지에 적용된다. 적용된 메시지 분류 규칙에 대해 측정된 자원 비용이 소비된 자원의 누적량에 더해진다. 소비된 자원의 누적량은 하나 이상의 메시지 분류 규칙 내에서 이전에 적용된 메시지 분류 규칙으로부터의 측정된 자원 비용을 더함으로써 계산된다.

[0024] 소비된 자원의 누적량이 선택된 최대 비용 값보다 작은지 여부가 판정된다. 판정에 기초하여 추가적인 메시지 분류 규칙이 전자 메시지에 적용된다. 소비된 자원의 양이 선택된 최대 비용 값보다 작으면, 수신된 전자 메시지에 더 많은 전자 메시지 규칙이 적용되어 서비스의 수준에서 정의된 것보다 높은 유효성을 초래한다. 소비된 자원의 양이 최소한 선택된 최대 비용 값과 같으면, 전자 메시지 규칙이 다른 전자 메시지에 적용된다.

[0025] 아래에서 자세히 설명하는 바와 같이, 본 발명의 실시형태는, 예를 들어, 하나 이상의 프로세서와 시스템 메모리와 같은 컴퓨터 하드웨어를 포함하는 특정 목적의 또는 범용의 컴퓨터를 포함하거나 이를 이용할 수 있다. 본 발명의 범위 내의 실시형태는 또한 컴퓨터 실행가능 명령 및/또는 데이터 구조를 운반하거나 저장하기 위한 물리적 및 기타의 컴퓨터 판독가능 매체를 포함한다. 이러한 컴퓨터 판독가능 매체는 범용 또는 특정 목적의 컴퓨터 시스템에 의해 액세스될 수 있는 여하한 가용 매체일 수 있다. 컴퓨터 실행가능 명령을 저장하는 컴퓨터 판독가능 매체는 물리적 저장매체이다. 컴퓨터 실행가능 명령을 운반하는 컴퓨터 판독가능 매체는 전송 매체(transmission media)이다. 그러므로, 제한이 아니라 예로서, 본 발명의 실시형태는 컴퓨터 저장 매체(장치) 및 전송 매체의 적어도 2가지 별개의 상이한 종류의 컴퓨터 판독가능 매체를 포함할 수 있다.

[0026] 컴퓨터 저장 매체(장치)는 RAM, ROM, EEPROM, CD-ROM 또는 기타 광 디스크 저장소, 자기 디스크 저장소 또는 기타 자기 저장 장치, 또는 회망 프로그램 코드 수단을 컴퓨터 실행가능 명령 또는 데이터 구조의 형태로 저장하는 데 사용될 수 있고 범용 또는 특정 목적의 컴퓨터에 의해 액세스될 수 있는 여하한 기타 매체를 포함한다.

[0027] "네트워크"는 컴퓨터 시스템 및/또는 모듈 및/또는 기타 전자 장치 사이에서 전자적 데이터의 전송을 가능하게

하는 하나 이상의 데이터 링크로 정의된다. 정보가 네트워크 또는 다른 통신 접속(유선, 무선 또는 유선과 무선의 조합)을 통해 컴퓨터로 전달되는 때에, 적당하게도 컴퓨터는 접속을 전송 매체로 본다. 전송 매체는 회망 프로그램 코드 수단을 컴퓨터 실행가능 명령 또는 데이터 구조의 형태로 운반하는 데 사용될 수 있고 범용 또는 특정 목적 컴퓨터에 의해 액세스될 수 있는 네트워크 및/또는 데이터 링크를 포함할 수 있다. 상기한 것들의 조합도 컴퓨터 판독가능 매체의 범위에 포함되어야 한다.

[0028] 또한, 다양한 컴퓨터 시스템 컴포넌트에 도달할 때에, 컴퓨터 실행가능 명령 또는 데이터 구조의 형태의 프로그램 코드 수단은 전송 매체로부터 컴퓨터 저장 매체 (장치)로 자동으로 전달될 수 있다 (또는 그 역이다). 예를 들어, 네트워크 또는 데이터 링크를 통해 수신된 컴퓨터 실행가능 명령 또는 데이터 구조는 네트워크 인터페이스 모듈(예를 들어, "NIC") 내의 RAM에서 버퍼링되고 최종적으로 컴퓨터 시스템 RAM 및/또는 컴퓨터 시스템의 덜 휘발성인 컴퓨터 저장 매체로 전달될 수 있다. 그러므로, 컴퓨터 저장 매체 (장치)는, 송신 매체도 (또는 주로 송신 매체를) 사용하는 컴퓨터 시스템 컴포넌트에 포함될 수 있음을 이해하여야 한다.

[0029] 컴퓨터 실행가능 명령은, 예를 들어, 프로세서에서 실행되는 때에 범용 컴퓨터, 특수 목적 컴퓨터 또는 특수 목적 처리 장치가 특정 기능 또는 기능의 그룹을 수행하도록 하는 명령과 데이터를 포함한다. 컴퓨터 실행가능 명령은, 예를 들어, 바이너리, 어셈블리 언어와 같은 중간 포맷 명령 또는 심지어 소스 코드일 수 있다. 구조적 특성 및/또는 방법적 동작에 특정한 언어로 주제가 설명되었지만, 첨부된 청구범위에 정의된 주제는 반드시 상술한 특성이나 동작에 제한되어야 하는 것은 아님을 이해하여야 한다. 오히려, 설명된 특성 및 동작은 청구범위를 구현하는 예시적인 형태로서 개시된다.

[0030] 당업자는 본 발명이 네트워크 연산 환경에서, 퍼스널 컴퓨터, 데스크톱 컴퓨터, 랩톱 컴퓨터, 메시지 프로세서, 핸드헬드 장치, 멀티프로세서 시스템, 마이크로프로세서 기반 또는 프로그램가능 컴퓨터 전자기기, 네트워크 PC, 미니컴퓨터, 메인프레임 컴퓨터, 이동 전화, PDA, 페이지, 라우터, 스위치 등을 포함하는 많은 유형의 컴퓨터 시스템 구성으로 실시될 수 있음을 인식할 것이다. 본 발명은 또한 네트워크를 통해 (유선 데이터 링크, 무선 데이터 링크 또는 유선과 무선 데이터 링크의 조합에 의해) 링크된 로컬 및 원격 컴퓨터 시스템이 모두 태스크를 수행하는 분산 시스템 환경에서 실시될 수도 있다. 분산 시스템 환경에서, 프로그램 모듈은 로컬 및 원격 메모리 저장 장치 모두에 위치될 수 있다.

[0031] 일반적으로, 본 발명의 실시형태는 바람직하지 않은 메시지 내용(예를 들어, 스팸, 바이러스, 디지털 누설(leakage) 등)에 대한 보호를 위해 동적으로 (그리고 잠재적으로는 예측불가능하게) 전자 메시지의 분류의 깊이(depth)/철저함(thoroughness)을 변화시키는 것에 관련된다. 최소 유효성은 유지되고, 가용 자원이 허용한다면, 증가된 보호를 제공하기 위해 초과될 수도 있다. 가용 메시지 분류 규칙의 최적 부분집합은 메시지마다 선택될 수 있다. 규칙의 선택은 가용 시스템 자원, 최소 회망 유효성(예를 들어, SLA(Service Level Agreement)에 정의된) 및 규칙 특징에 기초한다. 피드백 루프가 분류 규칙 부분집합을 최적화하기 위해 사용될 수 있다.

[0032] 이와 같이, 명세서와 다음의 청구범위에서, "메시지 분류"는 전자 메시지(예를 들어, 전자 메일 메시지, SMS(Short Message Service) 메시지, 파일 등)을, 예를 들어, 내용, 메시지 크기, 첨부, 비즈니스 대 소비자 도메인, 보내진 영역, 발신자, 수신자, 시간, 날짜 등과 같은 메시지 (또는 파일) 특성에 기초하여 상이한 "분류(class)"로 분류하는 것을 포함한다.

[0033] 일부 실시형태에서, 전자 메시지는 전자 메시지에 대응하는 서비스 수준(예를 들어, SLA에 따라)을 결정하기 위해 분류된다. 서비스의 수준은 전자 메시지에 대한 메시지 분류 규칙의 추가적 적용을 정의한다. 서비스의 수준은 전자 메시지에 어떠한 유형의 다른 메시지 분류 규칙이 얼마나 많이 적용되어야 하는지 정의할 수 있다. 예를 들어, 다른 국가에서 메시지를 분류하는 데 매우 효과적인 분류 규칙이 다른 나라에서 메시지를 분류하는 데 덜 효과적일 수 있고, 그 역일 수 있다.

[0034] 일부 실시형태에서, 추가의 분류는 전자 메시지가 원치않는 것인지 및/또는 요청하지 않은 전자 메시지(예를 들어, 스팸)인지 여부, 전자 메시지가 악성코드(malware)를 포함하거나 또는 달리 감염되거나 및/또는 위험한지 (예를 들어, 바이러스, 스파이웨어, 트로이 목마 등) 여부, 민감한 정보가 전자 메시지에서 누설되는지 여부 등을 판정하는 것에 관련된다. 예를 들어, 디지털 누설 방지(digital leakage prevention)("DLP") 시스템은 전자 메시지는 민감한 정보를 포함하는지 여부를 판정하기 위해 규칙을 이용할 수 있다.

[0035] 도 1은 적응적으로 전자 메시지를 분류하는 것을 용이하게 하는 예시적인 컴퓨터 아키텍처(100)를 도시한다. 도 1을 참조하면, 컴퓨터 아키텍처(100)는 메시지 분류기(102), 서비스 수준 식별기(107), 클록(108), 오버라이

드 퍼센티지(118), 메시지 분류 규칙(121) 및 서비스 수준 계약(131)을 포함한다. 각각의 게시된 컴포넌트는 예를 들어 LAN(Local Area Network), WAN(Wide Area Network) 및 심지어 인터넷과 같은 네트워크를 통해 서로 접속된다(또는 그 일부이다). 따라서, 각각의 도시된 컴포넌트뿐만 아니라 여하한 다른 접속된 컴퓨터 시스템과 그들의 컴포넌트는 메시지 관련 데이터를 생성하고, 네트워크를 통해 메시지 관련 데이터(예를 들어, IP(Internet Protocol) 데이터그램 및 기타 TCP(Transmission Control Protocol), HTTP(Hypertext Transfer Protocol), SMTP(Simple Mail Transfer Protocol) 등과 같은 IP 데이터그램을 이용하는 더 높은 계층 프로토콜)을 교환한다.

[0036] 규칙(121)은, 예를 들어, 전자 메시지를 분류하는 데 사용될 수 있는 규칙(121A 내지 121N)과 같은 복수의 메시지 분류 규칙을 포함한다. 각각의 규칙은 유효성, 비용, 효율성을 나타낼 수 있고, 명령을 포함할 수 있다. 유효성은 사용되고 있는 스캐닝의 유형에 기초하여 규칙이 메시지를 어떻게는 바람직하지 않은 것으로 얼마나 정확하게 식별하는지를 나타낸다. 예를 들어, 스팸을 검출하기 위한 규칙의 유효성은 규칙이 오공정이 없이 스팸을 검출하는 확률을 나타낼 수 있다. 비용은 실행 모듈이 규칙의 명령을 실행하는 때에 소비되는 (예를 들어, 추정된) 시스템 자원의 양을 나타낸다. 효율성은 자원 소비의 측면에서 유효성에 기초하여 규칙이 얼마나 효율적인지를 나타낸다. 일부 실시형태에서, 효율성은 유효성을 비용으로 나눈 몫이다. 명령은 전자 메시지를 분류하는 데 관련된 결과를 생성하도록 (예를 들어, 전자 메시지가 스팸인지, 악성코드를 포함하는지, 민감한 정보를 포함하는지 등을 판정하도록) 실행된다.

[0037] 일반적으로, 메시지 분류기(102)는 전자 메시지 특성에 기초하여 전자 메시지를 분류하도록 구성된다. 도시된 바와 같이, 메시지 분류기(102)는 실행 모듈(103), 비용 모니터(104) 및 유효성 모니터(106)를 포함한다. 실행 모듈(103)은 수신된 규칙에 포함된 명령(예를 들어, 스크립트 또는 기타 실행가능 코드)를 실행하도록 구성된다. 명령은 전자 메시지를 분류하기 위한 데이터 포인트로 사용될 수 있는 개별 결과를 산출한다. 예를 들어, 개별 결과는 전자 메시지가 원치않는 및/또는 요구하지 않은 전자 메시지(예를 들어, 스팸)인지 여부, 감염되거나 위험한지 여부, 민감한 정보를 포함하는지 여부 등을 나타낼 수 있다. 실행 모듈(103)은 다수의 상이한 규칙을 실행하는 것으로부터 개별 결과를 누적할 수 있다. 그러면 메시지 분류기(102)가 메시지를 분류하기 위해 누적된 개별 결과를 사용할 수 있다.

[0038] 비용 모니터(104)는 전자 메시지를 스캐닝하는 것과 관련된 계속되는 자원 비용을 추적하도록 구성된다. 규칙이 실행되는 때에, 비용 모니터(104)는 전자 메시지에 대해 실행되는 여하한 규칙에 대한 총 자원 비용을 유지한다. 일부 실시형태에서, 각각의 규칙이 실행되는 때에, 그 규칙에 대한 비용이 이전에 실행되는 여하한 규칙에 대한 자원 비용에 더해진다.

[0039] 유효성 모니터(106)는 전자 메시지를 스캐닝하는 것의 계속되는 유효성을 추적하도록 구성된다. 규칙이 실행되는 때에, 유효성 모니터(106)는 전자 메시지에 대해 실행되는 여하한 규칙에 대한 총 유효성을 유지한다. 일부 실시형태에서, 각각의 규칙이 실행되는 때에, 그 규칙에 대한 유효성이 이전에 실행된 여하한 규칙에 대한 유효성에 더해진다.

[0040] 서비스 수준 계약(129)은 SLA(131)를 포함하는 복수의 SLA를 포함한다. 각각의 SLA는 최소 유효성과 하나 이상의 비용을 포함한다. 각각의 비용은 특정한 날짜/시간 범위에 적용가능하다. 최소 유효성은 메시지를 스캐닝하는 때에 (자원 소비가 초과되더라도) 달성될 누적 유효성(즉, 복수의 분류 규칙에 대한 유효성의 합)을 나타낸다. 표 1은 고객 유형에 기초한 SLA 당 유효성의 예이다.

표 1

고객 유형	최소 유효성
기본 고객	75
프리미엄 고객	100

[0041]

[0042] 표 1은 최소 유효성(즉, 복수의 분류 규칙의 적용으로부터 나오는 누적 유효성)이 기본 고객에 대해 75이고 프리미엄 고객에 대해 100임을 나타낸다. SLA에서 최소 유효성을 할당하는 때에 다른 요인도 고려될 수 있다.

[0043] 하나 이상의 비용은 각각 시간 범위와 최대 비용을 포함한다. 각각의 시간 범위/최대 비용 쌍은 규칙을 적용하기 위한 최대 자원 비용이 메시지가 시간 범위 내에서 수신되는 때에 메시지에게 고려되어야 함을 나타낸다.

시간 범위/최대 비용 쌍은 상이한 서비스 수준에 대해 달라지거나 동일할 수 있다. 일부 실시형태에서, 시간 범위/최대 비용 쌍은, 시간 범위/최대 비용 쌍이 많은 SLA에 대해 동일하도록 공통으로 액세스가능한 표에 할당된다. 다른 실시형태에서, 시간 범위/최대 비용 쌍은, 예를 들어, SLA 내에 포함되는 것을 통하는 등으로, SLA마다 할당될 수 있다. 표 2는 시간 범위/최대 비용 쌍의 예이다.

표 2

시간(time of day)	최대 비용
피크 시간	50
보통 시간	75
한산한(off-peak) 시간	100

[0044]

[0045]

표 2는 분류 규칙을 적용하기 위한 최대 자원 비용이 피크 시간 동안에는 50, 보통 시간에는 75, 한산한 시간에는 100임을 나타낸다. 다른 인자도 고려될 수 있다.

[0046]

최대 비용은 시간에 따라 변할 수 있다. 메시지 분류 서버가 추가의 하드웨어를 추가하고, 그러므로, 더 높은 연산 능력을 가지면, 최대 비용 숫자(figure)가 상승할 수 있다. 반면에, 서비스가 추가의 고객을 추가하거나 부하가 갑자기 증가하면, 최대 비용 숫자는 감소할 수 있다.

[0047]

일부 실시형태에서, 최소 유효성은 최대 비용에 비해 더 높은 중요성으로 고려된다. 이들 실시형태에서, 최소 유효성이 달성되는 것을 보장하기 위해 최대 비용을 초과하는 자원이 소비될 수 있다. 최대 비용보다 적은 자원을 이용하여 최소 유효성이 달성되면, 최대 비용에 도달하거나 이를 초과할 때까지 유효성을 증가시키도록 추가의 분류 규칙이 적용될 수 있다.

[0048]

서비스 수준 식별기(107)는 수신된 전자 메시지에 대응하는 서비스 수준을 식별하도록 구성된다. 메시지 특성과 시간/날짜에 기초하여, 서비스 수준 식별기(107)는 서비스 수준 계약(131)으로부터 적당한 SLA를 식별할 수 있다. 클록(108)은 날짜와 시간을 유지하고 그 정보를 전자 메시지가 수신되는 때에 서비스 수준 식별기(107)에 전송할 수 있다. 서비스 수준 식별기는 메시지에 대한 최대 비용과 최소 유효성을 메시지 분류기(102)에 보낼 수 있다. 메시지 분류 규칙마다, 수신된 메시지에 어떤 분류 규칙을 얼마나 많이 적용할지를 결정하기 위해, 메시지 분류기(102)는 누적 유효성을 최대 유효성에 비교하고 누적 비용을 최대 비용에 비교할 수 있다.

[0049]

오버라이드 퍼센티지(118)는 최소 유효성이 이미 달성되고 최대 비용이 이미 도달되거나 초과된 때에도 추가 분류 규칙이 전자 메시지에 적용되는 소정의 퍼센티지를 정의한다. 오버라이드 퍼센티지(118)는 (예를 들어, 효율성 때문에) 그렇지 않았으면 간과되었을 분류 규칙이 가끔 실행될 수 있도록 한다. 일부 실시형태에서, 오버라이드 퍼센티지(118)는 메시지 분류 규칙(121) 내의 모든 규칙이 전자 메시지에 전자 메시지에 적용되는 퍼센티지를 나타낸다.

[0050]

도 3은 전자 메시지를 적응적으로 분류하기 위한 예시적 방법(300)의 흐름도를 도시한다. 방법(300)은 컴퓨터 아키텍처(100)의 데이터와 컴포넌트의 관점에서 설명될 것이다.

[0051]

방법(300)은 발신자로부터 수신자에게 전송된 전자 메시지를 특정 시간에 수신하는 동작을 포함한다(동작 301). 예를 들어, 메시지 분류기(102)는 시간(114)(클록(108)에 의해 표시)에 메시지(101U)를 수신할 수 있다. 메시지(101U)는 발신자 주소와 수신자 주소를 포함하는 메시지 특성(111)을 포함할 수 있다.

[0052]

방법(300)은 다음 중 하나 이상에 기초하여 수신된 전자 메시지에 적용가능한 서비스 수준을 식별하는 동작을 포함한다: 발신자와 수신자, 적어도 최소 유효성 값과 최대 비용 값의 세트를 정의하는 서비스 수준, 메시지 분류 규칙의 조합이 서비스 수준을 만족시키기 위해 가져야하는 최소 누적 총 유효성을 나타내는 최소 유효성 값, 상이한 지정 시간 기간에 대응하는 최대 비용 값의 세트 내의 각각의 최대 비용 값, 메시지 분류 규칙을 전자 메시지에 적용하는 데 사용될 수 있는 자원의 총량을 나타내는 각각의 최대 비용 값(동작 302). 예를 들어, 서비스 수준 식별기(107)는 메시지 특성(111)과 시간(114)을 수신할 수 있다. 메시지 특성(111)(예를 들어, 발신자 및/또는 수신자 주소)에 기초하여, 서비스 수준 식별기(107)는 SLA(131)를 메시지(101U)를 분류하는 데 적용할 수 있다고 식별할 수 있다.

- [0053] 도시된 바와 같이, SLA(131)는 최소 유효성(132)과 비용(133)을 정의한다. 비용(133)은 시간 범위/최대 비용 쌍을 포함하는데, 이는 시간 범위(134A)/최대 비용(136A), 시간 범위(134B)/최대 비용(136B), 시간 범위(134C)/최대 비용(136C) 등을 포함한다.
- [0054] 방법(300)은 선택된 최대 비용 값에 대해 지정된 시간 기간 내에 있는 특정 시간에 기초하여 수신된 전자 메시지를 스캐닝하는 때에 사용될 최대 비용 값을 최대 비용 값의 세트로부터 선택하는 동작을 포함한다(동작 303). 예를 들어, 서비스 수준 식별기(107)는 시간이 시간 범위(134A) 내에 있다고 판정할 수 있다. 응답으로, 서비스 수준 식별기(107)는 미분류 메시지(101U)를 스캐닝하는 때에 사용되도록 최대 비용(136A)을 선택할 수 있다.
- [0055] 서비스 수준 식별기(107)는 메시지 분류기(102)에 최소 유효성(132)과 최대 비용(136A)을 보낼 수 있다. 메시지 분류기(102)는 미분류 메시지(101U)에 메시지 분류 규칙을 적용하는 것이 언제 중단되어야 하는지를 결정하기 위해 최소 유효성(132)과 최대 비용(136A)을 사용할 수 있다.
- [0056] 방법(300)은 하나 이상의 메시지 분류 규칙을 수신된 전자 메시지에 적용하는 동작을 포함하는데, 각각의 메시지 분류 규칙은 측정된 유효성, 측정된 자원 비용, 측정된 자원 비용에 비추어 측정된 유효성에 기초하여 계산된 효율성을 갖고, 측정된 유효성은 전자 메시지를 특정 메시지 특성을 갖는 것으로 적절하게 분류할 확률을 나타내고, 하나 이상의 메시지 분류규칙은 서비스 수준에서 정의된 최소 누적 총 유효성이 달성될 때까지 효율성의 순서로 적용된다(동작 304). 예를 들어, 메시지 분류기는 최소 유효성(132)(즉, 60)이 달성될 때까지 효율성의 순서로 규칙(121)로부터의 규칙을 적용할 수 있다.
- [0057] 개시된 규칙 중에서, 규칙(121) 내의 규칙에 대해 효율성(124A)(즉, 4)이 최고일 수 있다. 그러므로, 규칙(121A)은 미분류 메시지(101U)에 적용되는 첫 번째 규칙이다. 규칙(121A)을 적용하는 때에, 누적 유효성(162)은 유효성(122A)과 동일한 8이다. 메시지 분류기(102)는 8이 60보다 작다고 판정하고 따라서 최소 유효성(132A)을 달성하기 위해 추가의 분류 규칙이 적용된다.
- [0058] 효율성(124B)(즉, 3)이 규칙(121) 내의 규칙에 대해 다음으로 높을 수 있다. 그러므로, 규칙(121B)이 미분류 메시지(101U)에 적용되는 다음 규칙이다. 규칙(121B)을 적용하는 때에, 누적 유효성(162)은 11이고, 유효성(122A)과 유효성(122B)의 합과 같다. 메시지 분류기(102)는 11이 60보다 작다고 판정하고 따라서 최소 유효성(132A)을 달성하기 위해 추가의 분류 규칙이 적용된다.
- [0059] 효율성(124C)(즉, 2.8)이 규칙(121) 내의 규칙에 대해 다음으로 높을 수 있다. 그러므로, 규칙(121C)이 미분류 메시지(101U)에 적용되는 다음 규칙이다. 규칙(121C)을 적용하는 때에, 누적 유효성(162)은 65이고, 유효성(122A)과 유효성(122B)과 유효성(122C)의 합과 같다. 메시지 분류기(102)는 65가 60보다 크다고 판정하고 따라서 SLA(131)를 만족하기 위해 추가의 분류 규칙이 요구되지 않는다.
- [0060] 각각의 적용된 하나 이상의 메시지 분류 규칙에 대해, 방법(300)은 전자 메시지가 특정 메시지 특성을 갖는 확률을 나타내는 결과를 생성하기 위해 전자 메시지에 메시지 분류 규칙을 적용하는 동작을 포함한다(동작 305). 예를 들어, 실행 모듈(103)은 결과(112)를 생성하기 위해 미분류 메시지(101U)에 대해 명령(126A)을 실행할 수 있다. 결과(112)는 미분류 메시지(101U)가 원하지 않는 및/또는 요구하지 않은 전자 메시지가거나, 감염된 또는 위험한 메시지가거나, 민감한 정보를 포함하는 등의 확률을 나타낸다. 실행 모듈(103)은 누적 결과(113)에 결과(112)를 저장할 수 있다. 명령(126B 및 126C)은 결과를 생성하기 위해 미분류 메시지(101U)에 대해 유사하게 실행될 수 있다. 이들 결과도 누적 결과(113)에 저장될 수 있다.
- [0061] 각각의 적용된 하나 이상의 메시지 분류 규칙에 대해, 방법(300)은 소비되는 자원의 누적량에, 적용되는 메시지 분류 규칙에 대해 측정된 자원 비용을 더하는 동작을 포함하는데, 소비되는 자원의 누적량은 하나 이상의 메시지 분류 규칙 내의 이전에 적용된 메시지 분류 규칙으로부터의 측정된 자원 비용을 합함으로써 계산된다(동작 306). 예를 들어, 규칙(121A, 121B, 121C)을 적용하는 때에, 누적 비용(161)은 21이고, 이는 비용(123A)과 비용(123B)과 비용(123C)의 합이다.
- [0062] 방법(300)은 선택된 최대 비용 값보다 소비된 자원의 누적량이 작은지 여부를 판정하는 동작을 포함한다(동작 307). 예를 들어, 비용 모니터(104)는 누적 비용(161)이 최대 비용(136A)보다 작은지 여부를 판정할 수 있다. 방법(300)은 판정에 기초하여 전자 메시지에 추가의 메시지 분류 규칙을 적용하는 동작을 포함한다(동작 308). 예를 들어, 메시지 분류기(102)는 누적 비용(161)이 최대 비용(136A)보다 작은지 여부에 기초하여 전자 메시지에 추가의 메시지 분류 규칙을 적용할 수 있다.
- [0063] 컴퓨터 아키텍처(100)에 도시되는 바와 같이, 최소 유효성(132A)을 달성하는 때에, 누적 비용(161)(즉, 21)은

최대 비용(136A)(즉, 25)보다 작다. 그러므로, 미분류 메시지(101U)를 분류하는 유효성을 증가시키기 위해 추가의 분류 규칙이 미분류 메시지(101U)에 적용될 수 있다.

[0064] 예를 들어, 효율성(124D)(1.75)이 규칙(121) 내의 규칙에 대해 다음으로 높을 수 있다. 그러므로, 규칙(121D)이 미분류 메시지(101U)에 다음으로 적용되는 규칙이다. 이와 같이, 실행 모듈(103)은 결과를 생성하고 누적 결과(113)에 결과를 저장하기 위해 미분류 메시지(101U)에 대해 명령(126D)을 실행할 수 있다. 규칙(121D)을 적용하는 때에, 누적 비용(161)은 29로 전환되는데, 이는 비용(123A)과 비용(123B)과 비용(123C)과 비용(123D)의 합이다 (최소 유효성(123A)이 이미 달성되었기 때문에 유효성(122D)은 실질적으로 무시된다). 누적 비용(161)(즉, 29)이 최대 비용(136A)(즉, 25)을 초과하였으므로, 추가의 규칙이 미분류 메시지(101U)에 적용되지 않는다.

[0065] 다르게는, 규칙(121C)의 적용 후에 최소 유효성(132A)을 달성하는 때에 누적 비용(161)이 최대 비용(136A)(즉, 25)과 같거나 더 크면, 미분류 메시지(101U)에 추가의 규칙이 적용되지 않는다. 그러나, 최소 유효성(132A)은 여전히 달성된다.

[0066] 최소 유효성(132A)이 달성되고 최대 비용(136A)에 도달하거나 이를 초과하는 때에, 메시지 분류기(102)는 오버라이드 퍼센티지(118)를 참조할 수 있다. 메시지 분류기(102)는 미분류 메시지(101U)에 추가의 분류 규칙이 적용될지 여부를 결정하기 위해 오버라이드 퍼센티지(118)를 사용할 수 있다. 만약 그렇다면, 메시지 분류기(102)는, 예를 들어 규칙(121E)와 같은, 메시지 분류 규칙(121) 중의 하나 이상의 (또는 남아 있는 모든) 규칙을 미분류 메시지(101U)에 적용한다. 오버라이드 퍼센티지(118)의 사용은 사용되지 않았거나 사용이 제한되었을 메시지 분류 규칙의 성능(예를 들어, 유효성 및 비용)이 평가되고 적절하게 변경될 수 있도록 한다. 변경에 기초하여, 이들 분류 규칙의 사용의 빈도가 증가될 수 있다. 예를 들어, 더 오래된 규칙의 성능의 평가는 그 더 오래된 규칙이 스팸 패턴의 변경 때문에 이제는 더 효율적이라는 것을 밝힐 수 있다.

[0067] 미분류 메시지(101U)에 추가의 규칙이 적용되지 않는 때에, 메시지 분류기(102)는 미분류 메시지(101U)를 분류하기 위해 누적 결과(113)를 사용할 수 있다. 예를 들어, 누적 결과(113)로부터 메시지 분류기(102)가 미분류 메시지(101U)를 적절한 메시지, 또는 원하지 않는 및/또는 요청되지 않은 메시지(예를 들어, 스팸), 악성코드를 포함 또는 미포함하는 것, 민감한 정보를 포함하거나 미포함하는 것 등으로 분류할 수 있다. 메시지 분류기(102)는 분류를 표시하기 위해 분류된 메시지(101C)를 출력할 수 있다.

[0068] 분류된 메시지(101C)를 출력하는 때에, 메시지 분류기(102)는 다음 전자 메시지를 분류하는 것으로 전이할 수 있다.

[0069] 도 2는 전자 메시지를 분류하기 위해 사용되는 적응적인 규칙 선택을 용이하게 하는 예시적인 컴퓨터 아키텍처(200)를 도시한다. 도 2를 참조하면, 컴퓨터 아키텍처(200)는 메시지 분류기(202), 메시지 분류 규칙(221) 및 규칙 선택 및 재정렬 모듈(216)을 포함한다. 도시된 컴포넌트 각각은, 예를 들어, LAN(Local Area Network), WAN(Wide Area Network) 및 심지어 인터넷과 같은 네트워크를 통해 서로에게 접속된다 (또는 그 일부이다). 따라서, 도시된 컴포넌트의 각각 뿐만 아니라 여하한 다른 접속된 컴퓨터 시스템 및 그들의 컴포넌트가 메시지 관련 데이터를 생성하고 메시지 관련 데이터(예를 들어, IP(Internet Protocol) 데이터그램 및 기타 TCP(Transmission Control Protocol), HTTP(Hypertext Transfer Protocol), SMTP(Simple Mail Transfer Protocol) 등과 같은 IP 데이터그램을 이용하는 더 높은 계층 프로토콜)를 교환할 수 있다.

[0070] 규칙(221)은, 전자 메시지를 분류하는 데 사용될 수 있는 예를 들어 규칙(221A 내지 221N)과 같은 복수의 메시지 분류 규칙을 포함한다. 규칙(121)과 유사하게, 규칙(121) 내의 각각의 규칙은 유효성, 비용, 효율성을 나타낼 수 있고 명령을 포함할 수 있다.

[0071] 일반적으로, 메시지 분류기(202)는 전자 메시지 특성에 기초하여 전자 메시지를 분류하도록 구성된다. 예를 들어, 메시지 분류기는 미분류 메시지(201U)를 입력으로 수신하고 분류된 메시지(201C)를 출력으로 생성할 수 있다. 분류된 메시지(201C) 내의 각각의 메시지는, 예를 들어, 메시지가 스팸인지, 악성코드를 포함하는지, 민감한 정보를 포함하는지 여부 등을 나타내도록 분류될 수 있다.

[0072] 도시된 바와 같이, 메시지 분류기(202)는 자원 모니터(213)를 포함하는 실행 모듈(203) 및 효율성 합성기(214)를 포함한다. 실행 모듈(203)은 수신된 규칙 내에 포함된 명령(예를 들어, 스크립트나 기타 실행가능 코드)를 실행하도록 구성된다. 명령은 전자 메시지를 분류하는 데이터 포인트로서 사용될 수 있는 개별 결과(잠재적으로 외부 사용자 피드백을 겪음)를 산출한다. 리소스 모니터(213)는 규칙 실행 동안 다양한 소비되는 자원(예를 들어, 시스템 메모리, 프로세서, 네트워크 대역폭 등)의 양을 (예를 들어, 본질적으로 실시간으로) 모니터링할

수 있다.

- [0073] 효율성 합성기(214)는 소비된 자원의 표시와 결과를 수신하고 적용된 규칙에 대한 갱신된 효율성을 합성할 수 있다. 적용된 규칙에 대한 소비된 자원과 결과는 합성된 효율성과의 일관성을 위해 규칙에 대한 유효성 및/또는 비용을 갱신하는 데 사용될 수도 있다.
- [0074] 이와 같이, 각 분류 규칙의 비용과 유효성은 측정된 값일 수 있고, 특정 시점(point-in-time)에 (예를 들어, 적용되는 때) 측정될 수 있고, 시간에 따라 변할 수 있다. 스팸 패턴과 내용이 진화하면서, 분류 규칙은 더 또는 덜 효과적이게 될 수 있다. 특정 역사적인 스팸 캠페인이 양적으로 부활하는 경우, 과거의 규칙이 갑자기 더 효율적이게 된다. 또한, 소프트웨어가 업그레이드되고 최적화되면서, 규칙의 비용은 감소할 수 있다.
- [0075] 규칙 선택 및 재정렬 모듈(216)은 전자 메시지에 적용하기 위한 규칙을 (예를 들어, 효율성에 기초하여) 메시지 분류 규칙(221)으로부터 선택할 수 있다. 규칙 선택 및 재정렬 모듈(216)은 또한 메시지 분류 규칙(221)을 (예를 들어, 효율성에 기초하여) 정렬할 수도 있다.
- [0076] 도 4는 전자 메시지를 분류하기 위해 사용되는 규칙을 적응적으로 선택하기 위한 예시적인 방법(400)의 흐름도를 도시한다. 방법(400)은 컴퓨터 아키텍처(200)의 데이터 및 컴포넌트의 관점에서 설명될 것이다.
- [0077] 방법(400)은 하나 이상의 전자 메시지를 수신하는 동작을 포함한다(동작 401). 예를 들어, 메시지 분류기(202)는 미분류 메시지(201U)를 수신할 수 있다.
- [0078] 하나 이상의 전자 메시지 각각에 대해, 방법(400)은 전자 메시지 분류 규칙의 이전에 선택된 부분집합 내 각각의 메시지 분류 규칙을 전자 메시지에 적용하는 동작을 포함하는데, 전자 메시지 분류 규칙의 이전에 선택된 부분집합은 복수의 전자 메시지 분류 규칙의 부분집합이다(동작 402). 예를 들어, 메시지 분류기(202)는 (예를 들어, SLA 및 가능하게는 오버라이드 퍼센티지 내의 최소 유효성 및 최대 비용에 기초하여) 미분류 메시지(201U) 내 각각의 메시지에 규칙(221A-221C)을 적용할 수 있다.
- [0079] 전자 메시지 분류 규칙의 이전에 선택된 부분집합 내 각각의 전자 메시지 분류 규칙에 대하여, 방법(400)은 전자 메시지가 특정 메시지 특성을 가질 확률을 나타내는 결과를 전자 메시지 규칙이 계산하는 동작을 포함한다(동작 403). 예를 들어, 실행 모듈(203)은 결과(212)를 생성하기 위해 201U 내의 미분류 메시지에 대해 명령(226A)을 실행할 수 있다. 결과(212)는 (예를 들어, 메시지의 지정된 수신자에 기초하여) 201U 내의 메시지가 원치않는 전자 메시지이거나, 감염되거나 위험한 전자 메시지이거나, 민감한 정보를 포함하는 등의 확률을 나타낼 수 있다. 규칙(221B 및 221C)에 대한 결과도 계산될 수 있다.
- [0080] 외부 피드백(예를 들어, 사용자로부터)이 계산된 결과로 통합(incorporated)될 수 있다. 예를 들어, 외부 피드백(261)은 결과(212)로 통합될 수 있다. 외부 피드백은 사용자의 유효성 인식에 기초하여 계산된 유효성을 높이거나 낮출 수 있다. 적당한 경우에, 외부 피드백도 규칙(221B와 221C)에 대한 계산된 결과로 통합될 수 있다.
- [0081] 일부 실시형태에서, 잡히지 않은 스팸, 악성코드 또는 민감한 정보를 포함하는 전자 메시지(오부정(false negative))뿐만 아니라 스팸, 악성코드 또는 민감한 정보를 포함하는 것으로 분류된 정당한 메시지(오긍정)가 추가의 분석을 위해 제출된다. 이 유형의 피드백은 유효성 점수를 튜닝하는 데에도 사용될 수 있다.
- [0082] 이전에 선택된 전자 메시지 분류 규칙의 부분집합 내의 각각의 전자 메시지 분류 규칙에 대해, 방법(400)은 전자 메시지 분류 규칙을 전자 메시지에 적용하는 데 소비되는 자원의 양을 나타내는 자원 비용을 측정하는 동작을 포함한다(동작 404). 예를 들어, 자원 모니터(213)는 201U로부터의 메시지에 대해 명령(226A)을 실행함으로써 소비되는 소비 자원(231)의 양을 나타내는 자원 비용을 측정할 수 있다. 규칙(221B 및 221C)에 대한 자원 소비 비용도 측정될 수 있다.
- [0083] 방법(400)은 각각의 전자 메시지에 각각의 전자 메시지 분류 규칙을 적용하는 것과 관련된 측정된 자원 비용과 계산된 결과를 보유하는 동작을 포함한다(동작 405). 예를 들어, 메시지 분류기(202)는 미분류 메시지(201U) 내의 다른 메시지에 대해 규칙(221A)을 실행하는 데 대한 자원 비용과 함께 소비된 자원(231)과 결과(212)를 보유할 수 있다. 미분류 메시지(201U)의 메시지에 대해 규칙(221B와 221C)을 실행하기 위한 자원 비용과 결과도 보유될 수 있다.
- [0084] 메시지 분류 규칙의 이전에 선택된 부분집합 내의 각각의 메시지 분류 규칙에 대해, 방법(400)은 메시지 분류 규칙에 대한 측정된 자원 비용과 보유된 계산 결과로부터 효율성 메트릭을 합성하는 동작을 포함한다(동작 406). 그러므로, 규칙(221A, 221B, 221C) 각각에 대해, 효율성 합성기(214)는 측정된 자원 비용과 보유된 계산

결과로부터 효율성 메트릭을 합성할 수 있다. 예를 들어, 규칙(221A)에 대해, 효율성 합성기(214)는 결과(212)와 소비 자원(231) 뿐만 아니라 규칙(221A)을 미분류 메시지(201U) 내의 다른 메시지에 적용하는 것으로부터의 측정된 자원 비용과 계산된 결과로부터 합성된 효율성(232)을 합성할 수 있다. 효율성은 규칙(221B과 221C)에 대해서도 합성될 수 있다.

- [0085] 그러면 메시지 분류기(202)는 224A를 합성된 효율성(232)으로 대체할 수 있다. 유효성(222A)과 비용(223A)도 합성된 효율성(232)과의 일관성을 위해 적당하게 갱신될 수 있다. 규칙(221B과 221C)에 대한 효율성, 유효성 및 비용도 적당하게 갱신될 수 있다.
- [0086] 방법(400)은 복수의 전자 메시지 분류 규칙에 포함된 전자 메시지 분류 규칙에 대한 기존 효율성 메트릭과 합성된 효율성 메트릭을 비교하는 동작을 포함한다(동작 407). 예를 들어, 합성된 효율성(232)은 메시지 분류 규칙(221) 중 다른 것에 포함된 효율성에 비교될 수 있다. 또한 규칙(221B과 221C)에 대한 합성된 효율성은 다른 메시지 분류 규칙(221)에 포함된 효율성과 비교될 수 있다.
- [0087] 방법(400)은 합성된 효율성 메트릭을 기존 효율성 메트릭에 비교한 결과에 적어도 부분적으로 기초하여 이후에 수신되는 전자 메시지를 분류하는 데 사용하기 위해 복수의 전자 메시지 분류 규칙 중에서 전자 메시지 분류 규칙의 새로운 부분집합을 선택하는 동작을 포함한다(동작 408). 예를 들어, 합성된 효율성에 기초하여, 규칙(221A, 221B 및 221C)가 서로에 대해서뿐만 아니라 메시지 분류 규칙(221) 중 다른 것에 대해서도 더 또는 덜 효율적이게 될 수 있다. 이와 같이, 전자 메시지를 분류하기 위해 (예를 들어, SLA에 기초하여) 새로운 규칙의 부분집합이 선택되는 때에 하나 이상의 규칙(221A, 221B 및 221C) 중 하나 이상은 탈락할 수 있다.
- [0088] 일부 실시형태에서, 메시지를 분류(예를 들어, 스팸, 악성코드, 민감한 정보 등을 포착)하는 데 있어서의 각각의 규칙의 상대적인 유효성 및 실제 관찰되는 규칙 실행 비용에 대한 라이브 측정이 이루어지면서, 메시지 분류 규칙에 대한 비용 및 효율성은 (예를 들어, 하루 중에) 계속하여 재측정된다. 더 효율적인 규칙에 대해, 그것이 더 많은 메시지에 대해 실행되므로 규칙의 유효성과 비용에 대한 더 많은 데이터가 존재한다. 덜 효율적인 규칙에 대해, 예를 들어 1%와 같은 오버라이드 퍼센티지(또는 랜덤 확률)는 적어도 기본량의 갱신 비용 및 유효성 정보가 수집됨을 나타낸다. 비용과 효율성이 재계산되면서, 규칙을 순서지우는데 사용되는 효율성 점수도 재계산된다. 이후의 전자 메시지는 갱신된 점수에 기초하여 선택되는 메시지 분류 규칙을 이용하여 분류된다.
- [0089] (예를 들어, 새로운 유형의 스팸이나 악성코드를 잡기 위해) 새로운 규칙이 기록되면서, 규칙집의 크기가 성장한다. 새로이 소개된 규칙은 유효성 점수 0과 비용 점수 1로 소개될 수 있는데, 이는 효율성 점수가 0이 되게 하고 규칙을 목록의 맨 마지막에 놓는다. 시간에 따라, 규칙이 오버라이드 퍼센티지에 따라 메시지에 적용되면서, 충분한 실제 데이터가 결국 누적되어 새로운 규칙에 대한 비용과 유효성에 대한 더 실제적인 값 및 그에 의한 더 적당한 효율성 점수를 계산할 수 있다. 효율성 점수가 재계산되면서, 새로운 규칙은 자동으로 목록 내에서 그의 최적 순서로 이주할 것이다.
- [0090] 시간에 따라, 규칙집은 너무 크게 성장하여 현재의 오버라이드 퍼센티지(예를 들어, 1%)에 기초하여 적당하게 실행할 수 없게 될 수 있다. 이와 같이, 다른 tier가 추가될 수 있는데, 여기서 아마도 0.1 미만의 효율성 점수를 갖는 규칙이 예를 들어, 0.1%와 같은 감소된 오버라이드 퍼센티지에 기초하여 실행된다. 이러한 낮은 효율성 규칙의 결과는 메시지를 분류하는 데 사용되지도 않고, 대신에 갱신된 비용과 유효성 정보를 생성하는 데에만 사용될 수 있다.
- [0091] 도 5는 적응적 전자 메시지 스캐닝 및 전자 메시지를 분류하는 데 사용되는 규칙의 적응적 선택을 용이하게 하는 예시적인 컴퓨터 아키텍처(500)를 도시한다.
- [0092] 미분류 메일(501U)이 수신된다. 발신자/수신자 정보가 고객 분류(531)로 전송된다. 고객 분류(531)는 미분류 메시지(501U)를 분류하는 때에 사용할 최소 유효성을 식별한다. 스테이지 541에서, 고객 분류(531)에 대한 최소 유효성이 규칙(511A, 511B 및 511C)을 실행하는 것으로부터 달성된다. 스테이지 542에서, 규칙(511D)을 포함하는 하나 이상의 추가 규칙이 자원 가용성(504)이 허용하는 바에 따라 기회주의적으로(opportunisticly) 실행된다. 스테이지 543에서, 랜덤 확률(518)에 기초하여 규칙(511N)까지의 하나 이상의 다른 규칙을 실행하지 않기로 하는 결정이 오버라이드되고, 이들 규칙(511N)까지의 하나 또는 다른 규칙이 실행된다. 규칙(511A 내지 511N)의 결과에 기초하여 분류된 메일(501C)이 (예를 들어, 스팸 또는 정당한 것으로) 출력된다.
- [0093] 규칙(511A 내지 511N)에 대해 각각의 규칙(503)의 런타임에 의해 성능 데이터가 수집된다. 갱신된 비용 점수가 규칙(511A 내지 511N)으로 다시 기록된다. 규칙(511A 내지 511N)에 대해 각각의 규칙(512)의 결과가 긍정(예를 들어, 스팸임) 또는 부정(예를 들어, 정당한)으로 결정된다. 오긍정 및 오부정을 식별하기 위해 외부 피드백

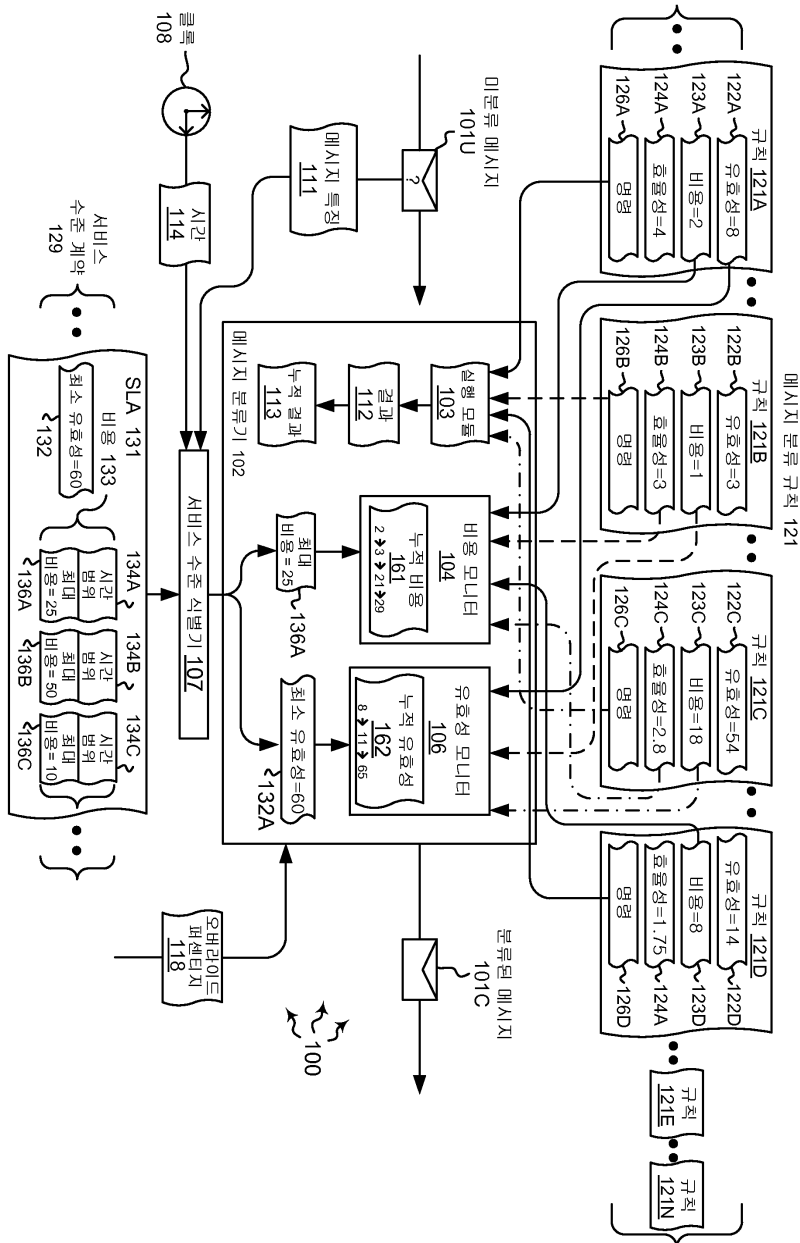
(561)이 결과에 통합된다. 갱신된 유효성 점수는 규칙(511A 내지 511N)으로 다시 기록된다. 효율성이 재계산되고 재계산된 효율성에 기초하여 규칙이 재정렬된다.

[0094]

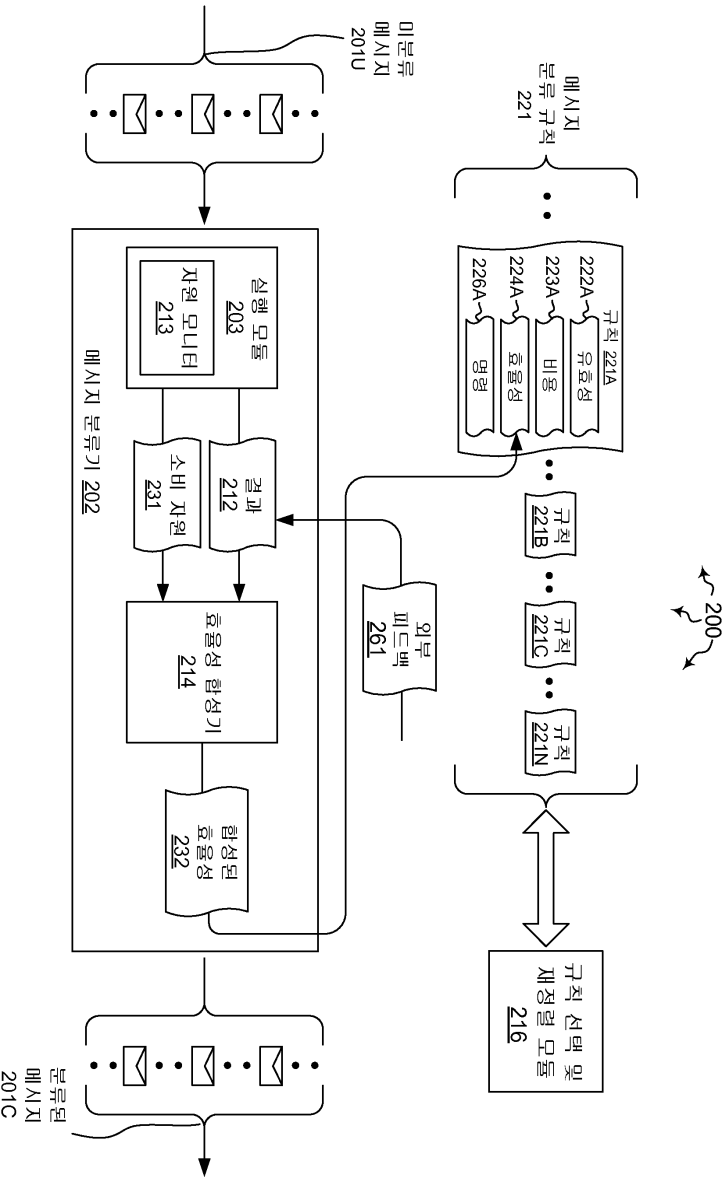
본 발명은 그 사상 또는 본질적 특징으로부터 벗어남이 없이 다른 구체적인 형태로 구현될 수 있다. 설명된 실시형태는 모든 측면에서 예시적인 것이고 제한적이지 않은 것으로 생각되어야 한다. 그러므로 본 발명의 범위는 위의 설명이 아니라 첨부된 청구범위에 의해 표시된다. 청구범위의 균등물의 의미와 범위 내의 모든 변경은 청구범위에 포함된다.

도면

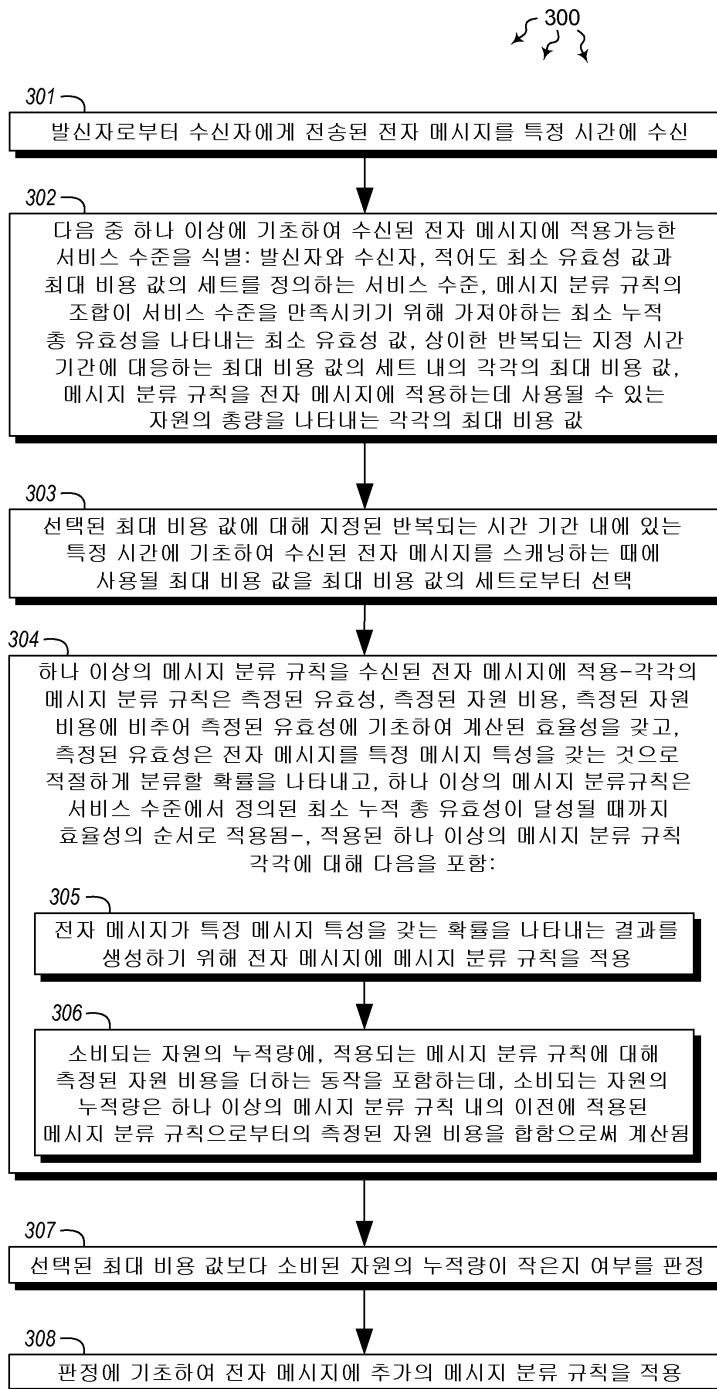
도면1



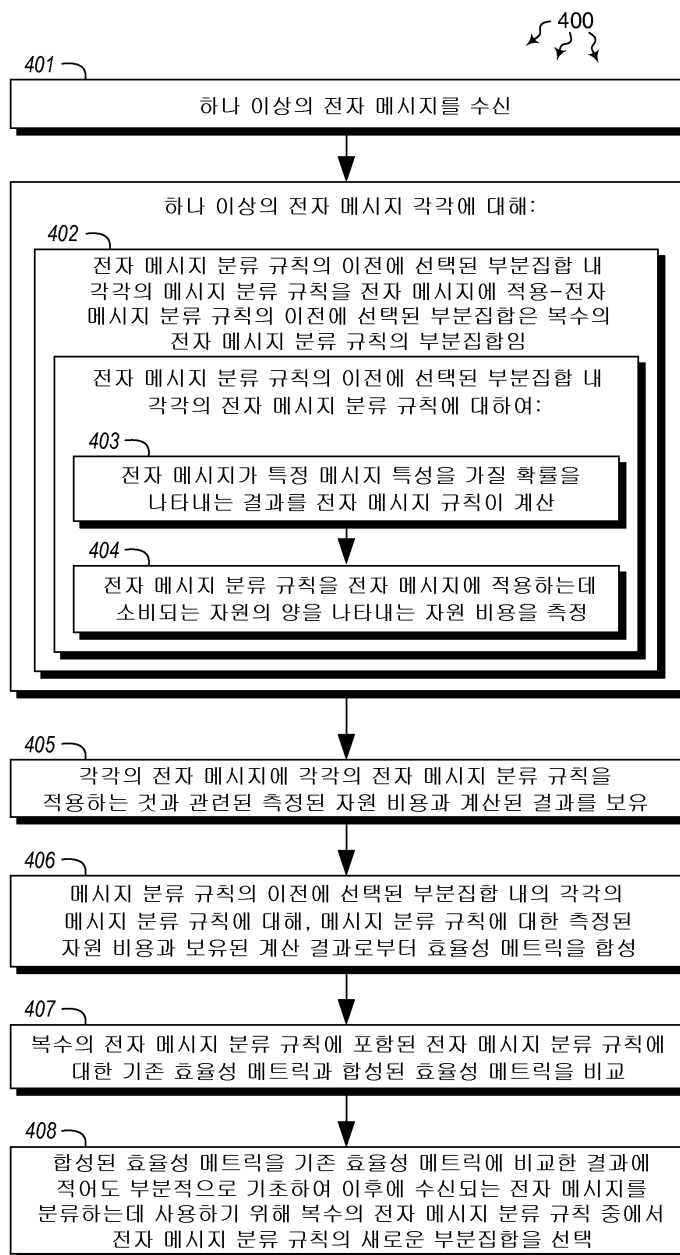
도면2



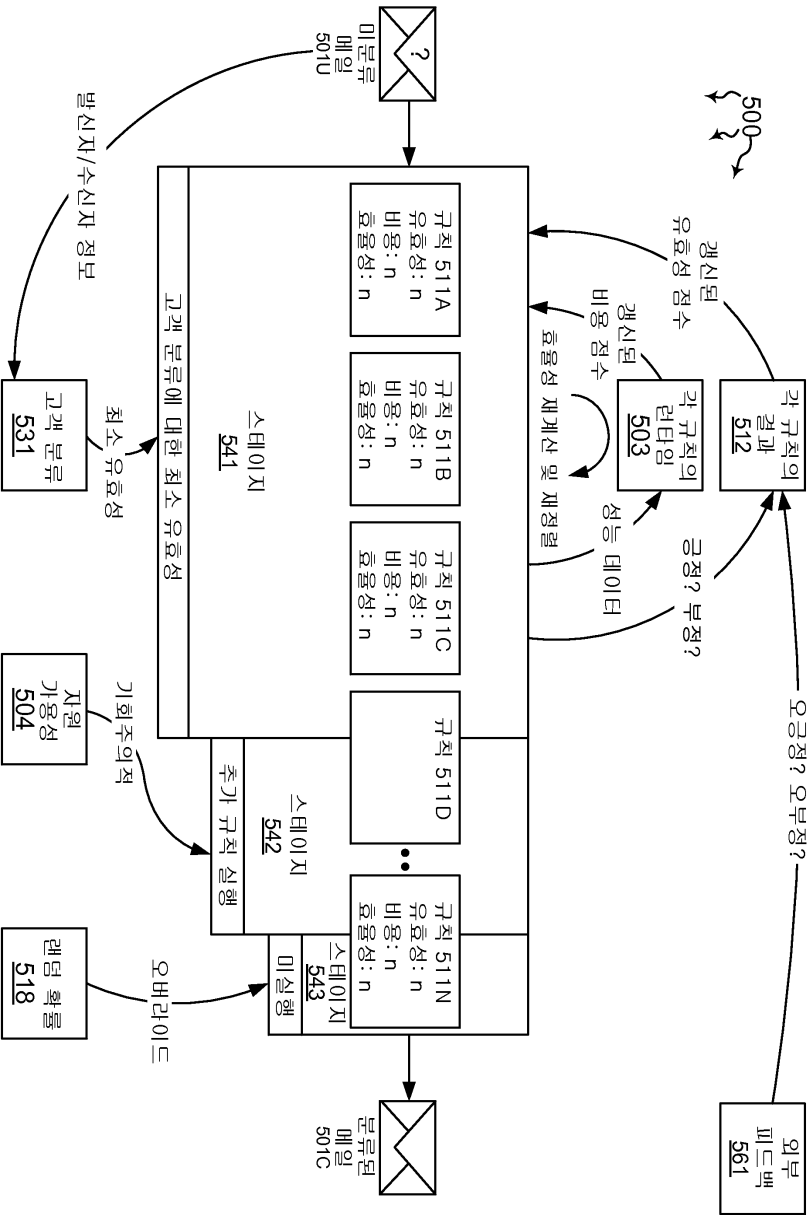
도면3



도면4



도면5



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 제4항의 2줄

【변경전】

하나 이상의

【변경후】

상기 하나 이상의

【직권보정 2】

【보정항목】 청구범위

【보정세부항목】 제6항의 2줄

【변경전】

하나 이상의

【변경후】

상기 하나 이상의

【식권보정 3】

【보정항목】 청구범위

【보정세부항목】 제5항의 2줄

【변경전】

하나 이상의

【변경후】

상기 하나 이상의