(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2010년01월21일
(11) 등록번호 10-0938072
(24) 등록일자 2010년01월13일

(51) Int. Cl.
G06Q 50/00 (2006.01)
(21) 출원번호 10-2004-0001266
(22) 출원일자 2004년01월08일
심사청구일자 2008년10월02일
(65) 공개번호 10-2004-0064232
(43) 공개일자 2004년07월16일
(30) 우선권주장
10/339,508 2003년01월09일 미국(US)
(56) 선행기술조사문헌
KR1020020063534 A*
US6161130 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
마이크로소프트 코포레이션
미국 워싱턴주 (우편번호 : 98052) 레드몬드 원 마이크로소프트 웨이
(72) 발명자
월래스앤드류제이.
미국98006워싱턴주벨레뷰175번플레이스에스이4427
쿨랜드닐케이.
미국98027워싱턴주이싸쿠아194번애비뉴에스이4562
(뒷면에 계속)
(74) 대리인
백만기, 이중희, 주성민

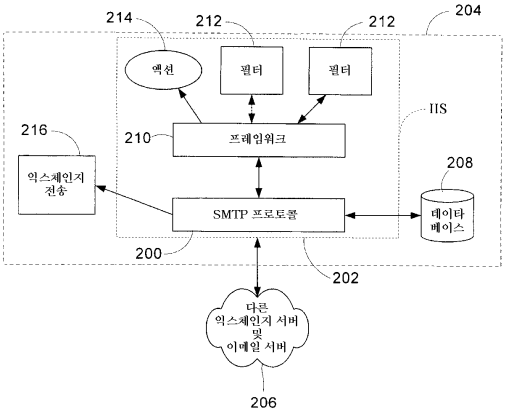
전체 청구항 수 : 총 24 항 심사관 : 계원호

(54) 안티-스팸 기술의 통합을 가능하게 하는 프레임워크

(57) 요 약

다수의 스팸 검출 솔루션들을 처리하기 쉽고 합리적인 방법으로 전개시켜 메시지가 스팸인지의 여부를 결정하는 방법이 제공된다. 프레임워크는 하나 이상의 안티-스팸 필터들을 기동하여(invoke) 메시지를 분석하고 메시지가 스팸인지의 여부의 신뢰 레벨을 리턴하고, 그 신뢰 레벨을 신뢰 레벨의 합산에 가산한다. 프레임워크는 규정된 임계값들의 세트에 대해 신뢰 레벨의 합산을 평가한다. 신뢰 레벨의 합산이 관리자에 의해 설정된 최상위 임계 세트보다 크면, 최상위 임계에 대한 특정된 액션이 취해진다. 만약 그렇지 않으면, 후속하는 필터들이 사용되어 최대 임계값을 초과하거나 또는 모든 필터들이 그 메시지를 평가할 때까지 메시지를 평가한다. 모든 필터들이 메시지를 평가한 후에, 신뢰 레벨의 합산은 모든 임계값들과 비교되어 매칭하는 임계값과 연관된 액션이 취해진다.

대 표 도 - 도2



(72) 발명자

왕콰이앙

미국98052워싱턴주레드몬드143번시티.엔이8426

닐리사무엘제이.

미국98074워싱턴주삼마미쉬엔이15번플레이스23327

애트웰시몬피.

미국98074워싱턴주삼마미쉬243번애비뉴엔이229

특허청구의 범위

청구항 1

복수의 안티-스팸 모듈을 구비하는 클라이언트 컴퓨팅 시스템에 메시지가 수신되었을 때 상기 메시지가 스팸인지의 여부를 결정하는 방법으로서,

a) 상기 복수의 안티-스팸 모듈 중 하나를 기동(invoke)하는 단계;

b) 상기 복수의 안티-스팸 모듈 중의 하나로부터 스팸 신뢰 레벨을 수신하는 단계;

c) 조정된 스팸 신뢰 레벨을 생성하기 위하여 상기 스팸 신뢰 레벨에 조정 인자(tuning factor)를 적용하는 단계 - 상기 적용된 조정 인자는 상기 복수의 안티-스팸 모듈 중 하나에 적용되는 사용자 정의 가중 인자(weighted factor)임 - ;

d) 상기 조정된 스팸 신뢰 레벨을 합산된 스팸 신뢰 레벨에 가산하는 단계 - 상기 사용자 정의 가중 인자는 상기 합산된 스팸 신뢰 레벨 내의 상기 조정된 스팸 신뢰 레벨에 대한 상대적 가중 인자를 정의함 - ;

e) 상기 합산된 스팸 신뢰 레벨을 적어도 하나의 임계값과 비교하는 단계 - 상기 적어도 하나의 임계값은 상부 임계값 및 하부 임계값을 포함함 - ;

f) 상기 합산된 스팸 신뢰 레벨이 상기 적어도 하나의 임계값보다 크면, 상기 적어도 하나의 임계값 중 어떤 것이 상기 상부 임계값에 가장 가까운지를 결정하고, 상기 상부 임계값에 가장 가까운 상기 적어도 하나의 임계값과 관련된 액션을 기동하는 단계; 및

g) 상기 합산된 스팸 신뢰 레벨이 상기 적어도 하나의 임계값보다 작으면 단계 a 내지 f를 반복하는 단계

를 포함하고,

상기 액션은

상기 메시지가 상기 클라이언트 컴퓨팅 시스템에 수신된 후에, 상기 합산된 스팸 신뢰 레벨이 상기 적어도 하나의 임계값 중 제1 임계 레벨을 초과하면, 접속을 해제(dropping);

상기 합산된 스팸 신뢰 레벨이 상기 적어도 하나의 임계값 중 제2 임계 레벨을 초과하고 상기 제1 임계 레벨 이하이면, 송신자에게 비전달 메시지(non-delivery message)를 반환; 및

상기 메시지가 상기 적어도 하나의 임계값 중 제3 임계 레벨을 초과하고 상기 제2 임계 레벨 이하이면, 상기 메시지를 정크 메일 폴더로 전달

중 하나를 포함하는,

방법.

청구항 2

삭제

청구항 3

제1항에 있어서,

상기 스팸 신뢰 레벨에 조정 인자를 적용하는 단계는 각각의 스팸 신뢰 레벨에 스케일링 인자를 적용하는 단계를 포함하는 방법.

청구항 4

제3항에 있어서,

상기 스케일링 인자를 적용하는 단계는 상기 스팸 신뢰 레벨을 1로 스케일링하는 단계를 포함하는 방법.

청구항 5

삭제

청구항 6

제1항에 있어서,

상기 스팸 신뢰 레벨에 조정 인자를 적용하는 단계는 비선형 신뢰 레벨 정규화를 이용하여 상기 스팸 신뢰 레벨을 정규화하는 단계를 포함하는 방법.

청구항 7

제1항에 있어서,

상기 액션을 기동하는 단계는, 상기 메시지를 삭제하는 단계, 비전달(non-delivery) 통지를 전송하는 단계, 및 상기 합산된 스팸 신뢰 레벨과 함께 상기 메시지를 클라이언트에게 전달하는 단계 중 하나를 기동하는 단계를 포함하는 방법.

청구항 8

제1항에 있어서,

상기 제1 임계는 99% 스팸 신뢰 레벨이고, 상기 제2 임계는 70% 스팸 신뢰 레벨이며, 상기 제3 임계 레벨은 40% 스팸 신뢰 레벨인, 방법.

청구항 9

제1항에 있어서,

단계 f는 상기 합산된 스팸 신뢰 레벨을 상기 메시지에 추가하는 단계를 포함하는 방법.

청구항 10

제1항에 있어서,

상기 복수의 안티-스팸 모듈은 블랙홀(blackhole) 리스트 및 터프(turf) 리스트 중 하나를 포함하는 방법.

청구항 11

복수의 안티-스팸 모듈의 스팸 신뢰 레벨을 통합하는 방법으로서,

메시지를 처리하기 위하여 상기 복수의 안티-스팸 모듈 중 적어도 하나를 기동하는 단계;

상기 복수의 안티-스팸 모듈 중 상기 적어도 하나로부터 상기 메시지의 스팸 신뢰 레벨을 수신하는 단계;

조정된 스팸 신뢰 레벨을 생성하기 위하여 상기 스팸 신뢰 레벨에 조정 인자를 적용하는 단계 - 상기 적용된 조정 인자는 상기 복수의 안티-스팸 모듈 중 하나에 적용되는 사용자 정의 가중 인자임 - ;

상기 조정된 스팸 신뢰 레벨을 합산된 스팸 신뢰 레벨에 가산하는 단계 - 상기 사용자 정의 가중 인자는 상기 합산된 스팸 신뢰 레벨 내의 상기 조정된 스팸 신뢰 레벨에 대한 상대적 가중 인자를 정의함 - ;

상기 합산된 스팸 신뢰 레벨을 적어도 하나의 임계값과 비교하는 단계 - 상기 적어도 하나의 임계값은 상부 임계값 및 하부 임계값을 포함함 - ; 및

상기 합산된 스팸 신뢰 레벨이 상기 적어도 하나의 임계값보다 크면, 상기 적어도 하나의 임계값 중 어떤 것이 상기 상부 임계값에 가장 가까운지를 결정하고, 상기 상부 임계값에 가장 가까운 상기 적어도 하나의 임계값과 관련된 액션을 기동하는 단계

를 포함하고,

상기 액션은

상기 합산된 스팸 신뢰 레벨이 상기 적어도 하나의 임계값 중 제1 임계 레벨을 초과하면, 접속을 해제;

상기 합산된 스팸 신뢰 레벨이 상기 적어도 하나의 임계값 중 제2 임계 레벨을 초과하고 상기 제1 임계 레벨 이하이면, 송신자에게 비전달 메시지를 반환; 및

상기 메시지가 상기 적어도 하나의 임계값 중 제3 임계 레벨을 초과하고 상기 제2 임계 레벨 이하이면,

상기 메시지를 정크 메일 폴더로 전달

중 하나를 포함하며,

상기 단계들은 상기 합산된 스팸 신뢰 레벨이 상기 임계값보다 크거나, 또는 상기 안티-스팸 모듈 모두가 기동될 때까지 반복되는,

방법.

청구항 12

제11항에 있어서,

상기 복수의 안티-스팸 모듈 중 적어도 하나를 기동하는 단계는, 상기 합산된 스팸 신뢰 레벨 중 하나가 상기 적어도 하나의 임계값보다 크고 상기 복수의 안티-스팸 모듈 모두가 기동되었을 때까지 상기 복수의 안티-스팸 모듈을 기동하는 단계를 포함하는 방법.

청구항 13

제11항에 있어서,

상기 제1 임계값은 99% 스팸 신뢰 레벨이고, 상기 제2 임계값은 70% 스팸 신뢰 레벨이며, 상기 제3 임계 레벨은 40% 스팸 신뢰 레벨인 방법.

청구항 14

제11항에 있어서,

상기 액션을 기동하는 단계는 상기 메시지에 대하여 취해진 상기 액션을 로그(logging)하는 단계를 포함하는 방법.

청구항 15

제11항에 있어서,

상기 스팸 신뢰 레벨에 조정 인자를 적용하는 단계는 각각의 스팸 신뢰 레벨에 스케일링 인자를 적용하는 단계를 포함하고, 상기 스케일링 인자를 적용하는 단계는 상기 스팸 신뢰 레벨을 1로 스케일링하는 단계를 포함하는 방법.

청구항 16

제11항에 있어서,

상기 스팸 신뢰 레벨에 조정 인자를 적용하는 단계는 각각의 스팸 신뢰 레벨에 스케일링 인자를 적용하는 단계를 포함하고, 상기 스케일링 인자를 적용하는 단계는 상기 복수의 안티-스팸 모듈 중 하나에 대한 신뢰 레벨로 상기 스팸 신뢰 레벨을 스케일링하는 단계를 포함하는 방법.

청구항 17

제11항에 있어서,

상기 스팸 신뢰 레벨에 조정 인자를 적용하는 단계는 비선형 신뢰 레벨 정규화를 이용하여 상기 스팸 신뢰 레벨을 정규화하는 단계를 포함하는 방법.

청구항 18

복수의 안티-스팸 모듈을 구비하는 클라이언트 컴퓨팅 시스템에 메시지가 수신되었을 때 상기 메시지가 스팸인지의 여부를 결정하기 위한 컴퓨터 실행가능한 명령들을 포함하는 컴퓨터 판독가능 기록 매체로서,

상기 명령들은,

a) 상기 복수의 안티-스팸 모듈 중 적어도 하나를 기동하는 단계;

b) 상기 복수의 안티-스팸 모듈 중 상기 적어도 하나로부터 상기 메시지의 스팸 신뢰 레벨을 수신하는 단계;

c) 정규화된 스팸 신뢰 레벨을 생성하기 위하여 상기 스팸 신뢰 레벨에 스케일링 인자를 적용하는 단계 - 상기 적용된 스케일링 인자는 상기 복수의 안티-스팸 모듈 중 하나에 적용되는 사용자 정의 가중 인자임 - ;

d) 상기 정규화된 스팸 신뢰 레벨을 합산된 스팸 신뢰 레벨에 가산하는 단계 - 상기 사용자 정의 가중 인자는 상기 합산된 스팸 신뢰 레벨 내의 상기 정규화된 스팸 신뢰 레벨에 대한 상대적 가중 인자를 정의함 - ;

e) 상기 합산된 스팸 신뢰 레벨을 적어도 하나의 임계값과 비교하는 단계 - 상기 적어도 하나의 임계값은 상부 임계값 및 하부 임계값을 포함함 - ;

f) 상기 합산된 스팸 신뢰 레벨이 상기 적어도 하나의 임계값보다 크면, 상기 적어도 하나의 임계값 중 어떤 것이 상기 상부 임계값에 가장 가까운지를 결정하고, 상기 상부 임계값에 가장 가까운 상기 적어도 하나의 임계값과 관련된 액션을 기동하는 단계 -

상기 액션은

상기 메시지가 상기 클라이언트 컴퓨팅 시스템에 수신된 후에, 상기 합산된 스팸 신뢰 레벨이 제1 임계 레벨을 초과하면, 접속을 해제;

상기 합산된 스팸 신뢰 레벨이 제2 임계 레벨을 초과하고 상기 제1 임계 레벨 이하이면, 송신자에게 비전달 메시지를 반환; 및

상기 메시지가 제3 임계 레벨을 초과하고 상기 제2 임계 레벨 이하이면, 상기 메시지를 정크 메일 폴더로 전달

중 하나를 포함함 - ; 및

g) 상기 합산된 스팸 신뢰 레벨이 상기 적어도 하나의 임계값보다 작으면, 상기 합산된 스팸 신뢰 레벨 중 하나가 상기 적어도 하나의 임계값보다 크고 상기 복수의 안티-스팸 모듈이 기동될 때까지, 단계 a 내지 f를 반복하는 단계

를 수행하기 위한 컴퓨터 판독가능 기록 매체.

청구항 19

삭제

청구항 20

제18항에 있어서,

상기 스팸 신뢰 레벨에 조정 인자를 적용하는 단계는 각각의 스팸 신뢰 레벨에 스케일링 인자를 적용하는 단계를 포함하는, 컴퓨터 판독가능 기록 매체.

청구항 21

제20항에 있어서,

상기 스케일링 인자를 적용하는 단계는 상기 스팸 신뢰 레벨을 1로 스케일링하는 단계를 포함하는, 컴퓨터 판독가능 기록 매체.

청구항 22

삭제

청구항 23

제18항에 있어서,

상기 스팸 신뢰 레벨에 조정 인자를 적용하는 단계는 비선형 신뢰 레벨 정규화를 이용하여 상기 스팸 신뢰 레벨을 정규화하는 단계를 포함하는, 컴퓨터 판독가능 기록 매체.

청구항 24

제18항에 있어서,

상기 액션을 기동하는 단계는, 상기 메시지를 삭제하는 단계, 비전달 통지를 전송하는 단계, 및 상기 합산된 스팸 신뢰 레벨과 함께 상기 메시지를 클라이언트에게 전달하는 단계 중 하나를 기동하는 단계를 포함하는, 컴퓨터 판독가능 기록 매체.

청구항 25

제18항에 있어서,

상기 제1 임계값은 99% 스팸 신뢰 레벨이고, 상기 제2 임계값은 70% 스팸 신뢰 레벨이며, 상기 제3 임계 레벨은 40% 스팸 신뢰 레벨인, 컴퓨터 판독가능 기록 매체.

청구항 26

제18항에 있어서,

상기 합산된 스팸 신뢰 레벨을 상기 메시지에 추가하는 단계를 수행하기 위한 컴퓨터 실행가능한 명령들을 더 포함하는 컴퓨터 판독가능 기록 매체.

청구항 27

제18항에 있어서,

상기 복수의 안티-스팸 모듈 중 적어도 하나를 기동하는 단계는 상기 메시지의 수신자 어드레스 리스트를 제공하는 단계를 포함하는, 컴퓨터 판독가능 기록 매체.

청구항 28

제18항에 있어서,

상기 메시지의 인코딩된 내용(content)을 크랙(crack)하는 단계를 포함하는 단계를 수행하기 위한 컴퓨터 실행가능한 명령들을 더 포함하는 컴퓨터 판독가능 기록 매체.

청구항 29

삭제

청구항 30

삭제

청구항 31

삭제

명 세 서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

<12> 본 발명은 일반적으로 전자 메시징에 관한 것으로, 특히 원치않는 전자 메일을 필터링하는 것에 관한 것이다.

<13> 전자 메시징, 특히 인터넷 상에서 전송되는 전자 메일("e-mail")은, 사회에서 급속히 퍼지고 있을 뿐 아니라, 그 약식(informality), 사용의 간편함 및 낮은 비용으로 인해, 많은 개인들 및 조직들이 선호하는 통신 방법이 되고 있다.

<14> 불행하게도, 이메일 수신자는 불필요한 대량의 메일을 증가적으로 받고 있다. 인터넷 기반의 상거래가 성장함에 따라, 폭넓게 성장하는 다양한 전자 상인들이 증가하고 있는 분야의 이메일 수신자에게 그들의 상품 및 서비

스를 광고하는 불필요한 메일을 반복적으로 보내고 있다. 인터넷을 통해 제품을 주문하거나 그렇지 않으면 상인과 거래하는 대부분의 소비자들은 이들 상인들로부터 이러한 유혹을 받는 것을 예상하고 있으며, 사실상 정기적으로 받고 있다.

<15> 그러나, 전자 메일러(mailer)들은 증가하고 있는 다수의 수신자와 접촉하기 위해 그들의 배급 리스트를 지속적으로 확장하고 있다. 예를 들면, 다양한 웹사이트에 의해 생성되는 방문자 정보 등의 아마도 악의없어 보이는 요청에 응답하여 단지 그들의 이메일 어드레스를 제공하는 수신자들은 종종 불필요한 메일을 수신하고, 불만스럽게도, 그들은 그들이 전자 배급 리스트 상에 포함되어 있음을 알게 된다. 이는 수신자의 동의는 말할 것도 없이, 수신자가 알지 못하는 사이에 발생한다. 게다가, 전자 메일러는 종종 그 배급 리스트를, 판매, 대여, 또는 다른 식으로, 그 사용을 위한 또 다른 메일러, 및 후속 메일러 등에게 퍼트릴 것이다. 따라서, 시간이 지남에 따라, 이메일 수신자는 폭넓게 증가하는 각종 다량의 메일러가 보유하는 별도의 배급 리스트의 결과로서 종종 불필요한 메일이 연달아 쏟아지는 것을 발견하게 된다. 개개인은 일년에 걸쳐 수백, 심지의 수천개의 불필요한 이메일을 쉽게 수신할 수 있다. 이메일 배급 리스트 상의 개인들은 더 짧은 기간에 걸쳐 상당히 많은 수의 불필요한 메시지를 받을 것으로 예상하고 있다.

<16> 또한, 할인 사무소 또는 컴퓨터 공급자에 대한 제공, 저당율 시세, 또는 한 종류 또는 다른 종류의 회의에 참석하는 초대와 같은, 많은 불필요한 이메일 메시지는 양호하지만, 포르노, 선동적이고 욕설의 내용과 같은 다른 것들은 그 수신자에게 불쾌하다. 이 불필요한 메시지들은 "정크(junk)" 메일 또는 "스팸(spam)"이라고 알려져 있다. 스팸으로부터의 이메일 부하는 합법적인 이메일로부터 발생되는 부하와 동일할 수 있다.

<17> 정크 우송 메일을 처리하는 작업과 유사하게, 이메일 수신자는 스팸을 제거하기 위해 그가 받고있는 메일을 선별해야만 한다. 컴퓨터 산업은 이 문제점을 인식하였고 스팸의 제거를 자동화하기 위한 기술을 개발해오고 있다. 예를 들면, 한가지 기술은 터프(turf) 리스트이다. 이메일 수신자들은 규정된 룰에 기초한 특성 세트를 사용하여 메일 수신을 확인하거나 거절하는, 터프 리스트를 신청한다. 불행히도, 주어진 이메일 메시지가 스팸인 지의 여부를 선택하는 것은 특정 수신자 및 메시지의 실제 내용에 달려있다. 어떤 수신자에게 스팸일 수 있는 것이 다른 사람에게는 스팸이 아닐 수 있기 때문에, 이는 터프 리스트의 기능을 제한한다. 또한, 전자 메일러(즉, 스팸 발생자)는 메시지의 실제 내용이 제목 라인으로부터 명확하지 않고 메시지의 본문을 읽는 것에 의해서만 식별되도록 메시지를 준비할 것이다.

<18> 개발된 다른 기술은 블랙홀(black hole) 리스트로서 알려져 있다. 블랙홀 리스트는 스팸이 보내지는 공지된 스팸 어드레스의 리스트이다. 이메일 송신자의 어드레스가 블랙홀 리스트에 대하여 체크된다. 그 어드레스가 리스트 상에 있으면, 이메일은 승인되지 않는다. 스팸 발생자는 이 기술을 피하기 위해 단순히 그 어드레스를 변경한다. 다른 기술도 개발되고 있다. 어떤 기술도 100% 효과적이지는 않다. 스팸을 방지하기 위한 이메일 서버의 혁신은 이 혁신을 극복하기 위한 스팸 생성자들의 혁신에 직면한다.

발명이 이루고자 하는 기술적 과제

<19> 본 발명은 다수의 스팸 검출 솔루션들을 처리하기 쉽고 합리적인 방법으로 함께 작용하도록 전개될 수 있고, 새로운 혁신들이 급속한 전개 모델 하에서 생성 및 전개될 수 있게 하는 프레임워크를 제공한다.

<20> 안티-스팸(anti-spam) 모듈을 이용하여 이메일 메시지가 스팸인지를 결정하는 방법이 개시된다. 이 방법은 안티-스팸 모듈 중 하나를 기동하고 안티-스팸 모듈로부터 스팸 신뢰 레벨을 수신한다. 스팸 신뢰 레벨에 조정 인자(tuning factor)를 제공하여 조정된 스팸 신뢰 레벨을 생성한다. 조정된 스팸 신뢰 레벨은 합산된 스팸 신뢰 레벨에 부가되고, 합산된 스팸 신뢰 레벨은 적어도 하나의 임계값과 비교된다. 합산된 스팸 신뢰 레벨이 임계값보다 크면, 적어도 하나의 임계값와 관련된 액션이 기동된다. 이 프로세스는, 합산된 스팸 신뢰 레벨이 임계값보다 크거나 모든 안티-스팸 모듈들이 기동될 때까지 반복된다.

<21> 일 실시예에서, 상부 임계값을 포함하는 복수의 임계값들이 사용되고, 합산된 스팸 신뢰 레벨은 각 임계값과 비교된다. 합산된 스팸 신뢰 레벨이 하나 이상의 임계값보다 크면, 상부 임계값과 가장 가까운, 초과된 임계값과 관련된 액션이 기동된다.

<22> 조정 인자는 스팸 신뢰 레벨에 1을 곱하거나, 스팸 신뢰 레벨을 제공한 안티-스팸 모듈의 사용자 신뢰 레벨에 의해 스팸 신뢰 레벨을 스케일링하는 등의 단순한 스케일링 인자로부터, 비선형 신뢰 레벨 정규화를 이용하여 스팸 신뢰 레벨을 정규화하는 복잡한 조정 인자까지 변할 수 있다.

<23> 기동된 액션들은 합산된 스팸 신뢰 레벨이 제1 임계 레벨을 초과하면 접속을 해제(dropping)하고, 합산된 스팸

신뢰 레벨이 제2 임계 레벨을 초과하고 제1 임계 레벨 이하이면 비전달 메시지를 송신자에게 반환하며, 메시지가 제3 임계 레벨을 초과하고 제2 임계 레벨 이하이면 정크 메일 폴더로 메시지를 전달하고, 합산된 스팸 신뢰 레벨을 클라이언트에게 보내어 클라이언트가 사용자마다의 주문형 액션을 수행가능하게 한다.

<24> 본 발명의 다른 특징 및 이점은 하기 첨부 도면을 참조하여 진행되는 예시적인 실시예들의 상세한 설명으로부터 명백해질 것이다.

발명의 구성 및 작용

<25> 첨부된 클레임은 특별히 본 발명의 특징을 개시하지만, 본 발명은, 그 목적 및 이점과 함께, 첨부 도면을 참조한 이하의 상세한 설명으로부터 가장 잘 이해될 것이다.

<26> 도면을 참조하면, 유사한 참조 부호는 유사한 구성요소를 지칭하고, 본 발명은 적합한 컴퓨팅 환경에서 구현되는 것으로서 설명된다. 요구되지는 않지만, 본 발명은 퍼스널 컴퓨터에 의해 실행되는 컴퓨터 모듈과 같은 컴퓨터 실행가능한 명령들의 일반적인 문맥으로 기술될 것이다. 일반적으로, 프로그램 모듈은 특정 태스크를 수행하거나 특정의 추상적 데이타 유형을 구현하는 루틴, 프로그램, 오브젝트, 컴포넌트, 데이타 구조 등을 포함한다. 또한, 당업자들은 본 발명이 핸드 헬드 장치, 멀티 프로세서 시스템, 마이크로 프로세서 기반 또는 프로그램가능한 소비 전자기기, 네트워크 PC, 미니컴퓨터, 메인프레임 컴퓨터 등을 포함하는 다른 컴퓨터 시스템 구성에 의해 실행될 수 있음을 이해할 것이다. 본 발명은 또한 통신 네트워크를 통해 링크되는 원격 프로세싱 장치에 의해 태스크가 수행되는 분산 컴퓨팅 환경에서도 실행될 수 있다. 분산 컴퓨팅 환경에서, 프로그램 모듈은 로컬 및 원격 메모리 저장 장치에 배치될 수 있다.

<27> 도 1은 본 발명이 구현될 수 있는 적합한 컴퓨팅 시스템 환경(100)의 예를 도시한다. 컴퓨팅 시스템 환경(100)은 단지 적합한 컴퓨팅 환경의 일례이고 본 발명의 사용 또는 기능의 범위에 대해 임의의 제한을 가하는 것을 의도하지 않는다. 컴퓨팅 환경(100)은 예시적인 운영 환경(100)에서 기술되는 컴퓨넌트들 중 어느 하나 또는 조합과 관련하여 임의의 의존성 또는 필요성을 갖는 것으로 해석되어서는 안된다.

<28> 본 발명은 다수의 다른 범용 또는 특정 목적의 컴퓨팅 시스템 환경 또는 구성으로 동작한다. 본 발명에 사용되기에 적합할 수 있는 공지된 컴퓨팅 시스템, 환경, 및/또는 구성의 예는 퍼스널 컴퓨터, 서버 컴퓨터, 핸드 헬드 또는 랩톱 장치, 멀티프로세서 시스템, 멀티프로세서 기반의 시스템, 셋톱 박스, 프로그램가능한 소비 전자기기, 네트워크 PC, 미니컴퓨터, 메인프레임 컴퓨터, 상기 시스템 또는 장치 중 임의의 것을 포함하는 분산 컴퓨팅 환경 등을 포함하지만, 이에 제한되지 않는다.

<29> 본 발명은 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터 실행가능한 명령들의 일반적인 문맥으로 기술될 수 있다. 일반적으로, 프로그램 모듈은 특정 태스크를 수행하거나 특정의 추상적 데이타 유형을 구현하는 루틴, 프로그램, 오브젝트, 컴포넌트, 데이타 구조 등을 포함한다. 본 발명은 또한 통신 네트워크를 통해 링크되는 원격 프로세싱 장치에 의해 태스크가 수행되는 분산 컴퓨팅 환경에서 실행될 수도 있다. 분산 컴퓨팅 환경에서, 프로그램 모듈들은 메모리 저장 장치를 포함하는 로컬 및 원격 컴퓨터 저장 매체 모두에 배치될 수 있다.

<30> 도 1을 참조하면, 본 발명을 구현하기 위한 예시적인 시스템은 컴퓨터(110)의 형태의 범용 컴퓨팅 장치를 포함한다. 컴퓨터(110)의 컴포넌트들은 프로세싱 유닛(120), 시스템 메모리(130), 및 시스템 메모리를 포함하는 각종 시스템 컴포넌트를 프로세싱 유닛(120)과 결합시키는 시스템 버스(121)를 포함하지만, 이에 제한되지 않는다. 시스템 버스(121)는 메모리 버스 또는 메모리 제어기, 주변회로 버스, 및 각종 버스 아키텍처 중 임의의 것을 사용하는 로컬 버스를 포함하는 여러가지 유형의 버스 구조들 중 임의의 것일 수 있다. 제한되지 않는 예로서, 이러한 아키텍처는 ISA(Industry Standard Architecture) 버스, MCA(Micro Channel Architecture) 버스, EISA(Enhanced ISA) 버스, VESA(Video Electronics Standards Associate) 로컬 버스, 및 메자닌(Mezzanine) 버스로도 알려져 있는 PCI(Peripheral Component Interconnect) 버스를 포함한다.

<31> 컴퓨터(110)는 일반적으로 다양한 컴퓨터 판독가능한 매체를 포함한다. 컴퓨터 판독가능한 매체는 컴퓨터(110)에 의해 액세스될 수 있는 임의의 이용가능한 매체일 수 있고, 휘발성 및 불휘발성 매체, 제거가능 및 제거불가능 매체 모두를 포함한다. 제한되지 않는 예로서, 컴퓨터 판독가능한 매체는 컴퓨터 저장 매체 및 통신 매체를 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능한 명령, 데이타 구조, 프로그램 모듈 또는 다른 데이타와 같은 정보의 저장을 위한 임의의 방법 및 기술에서 구현되는, 휘발성 및 불휘발성, 제거가능 및 제거불가능 매체 모두를 포함한다. 컴퓨터 저장 매체는 RAM, ROM, EEPROM, 플래시 메모리 또는 다른 메모리 기술, CD-ROM, DVD(digital versatile disks) 또는 다른 광디스크 저장 장치, 자기 카세트, 자기 테이프, 자기 디스

크 저장 장치 또는 다른 자기 저장 장치, 또는 원하는 정보를 저장하는데 사용될 수 있고 컴퓨터(110)에 의해 액세스될 수 있는 임의의 다른 매체를 포함하지만, 이에 제한되지 않는다. 통신 매체는 일반적으로 컴퓨터 판독가능한 명령, 데이타 구조, 프로그램 모듈 또는 반송파와 같은 변조된 데이타 신호 또는 다른 전송 메카니즘의 다른 데이타를 구현하며, 임의의 정보 전달 매체를 포함한다. 용어 "변조된 데이타 신호(modulated data signal)"는, 신호의 정보를 인코딩하는 방식으로 설정되거나 변경된 하나 이상의 특징을 갖는 신호를 의미한다. 제한되지 않는 예로서, 통신 매체는 유선 네트워크 또는 직접 유선 접속과 같은 유선 매체, 및 음향, RF, 적외선 및 다른 무선 매체 등의 무선 매체를 포함한다. 상술한 것들의 임의의 조합도 컴퓨터 판독가능한 매체의 범위 내에 포함되어야 한다.

<32> 시스템 메모리(130)는 ROM(131), RAM(132)과 같은 휘발성 및/또는 불휘발성 메모리 형태의 컴퓨터 저장 매체를 포함한다. 시동(start-up) 중과 같이, 컴퓨터(110) 내의 소자들 사이에서 정보를 전달하도록 돕는 기본 루틴을 포함하는 BIOS(basic input/output system)(133)는 통상적으로 ROM(131)에 저장된다. RAM(132)은 통상적으로 프로세싱 유닛(120)에 의해 현재 작동되거나 및/또는 즉시 액세스가능한 프로그램 모듈 및/또는 데이타를 포함한다. 제한되지 않는 예로서, 도 1은 운영 시스템(134), 응용 프로그램(135), 다른 프로그램 모듈(136), 및 프로그램 데이타(137)를 도시한다.

<33> 컴퓨터(110)는 또한 다른 제거가능/제거불가능, 휘발성/불휘발성 컴퓨터 저장 매체를 포함할 수 있다. 단지 예로서, 도 1은 제거불가능한 불휘발성의 자기 매체로부터 판독하거나 또는 그것에 기입하는 하드 디스크 드라이브(141), 제거가능한 불휘발성의 자기 디스크(152)로부터 판독하거나 그것에 기입하는 자기 디스크 드라이브(151), 및 CD ROM 또는 다른 광 매체와 같은 제거가능한 불휘발성의 광디스크(156)로부터 판독하거나 그것에 기입하는 광디스크 드라이브(155)를 포함한다. 예시적인 운영 환경에서 사용될 수 있는 다른 제거가능/제거불가능, 휘발성/불휘발성 컴퓨터 저장 매체는 자기 테이프 카세트, 플래시 메모리 카드, DVD, 디지털 비디오 테이프, 고상 RAM, 고상 ROM, 등을 포함하지만 이에 제한되는 것은 아니다. 하드 디스크 드라이브(141)는 통상적으로 인터페이스(140) 등의 제거불가능한 메모리 인터페이스를 통해 시스템 버스(121)에 접속되고, 자기 디스크 드라이브(151) 및 광디스크 드라이브(155)는 통상적으로 인터페이스(150) 등의 제거가능한 메모리 인터페이스에 의해 시스템 버스(121)에 접속된다.

<34> 상술되고 도 1에 도시된 드라이브 및 그에 관련된 컴퓨터 저장 매체는 컴퓨터 (110)에 대해 컴퓨터 판독가능한 명령, 데이타 구조, 프로그램 모듈 및 다른 데이타의 저장을 제공한다. 도 1에서, 예를 들면, 하드 디스크 드라이브(141)는 운영 시스템(144), 응용 프로그램(145), 다른 프로그램 모듈(146), 및 프로그램 데이타(147)를 저장하는 것으로 도시된다. 이들 컴포넌트들이 운영 시스템(134), 응용 프로그램(135, 다른 프로그램 모듈(136), 및 프로그램 데이타(137)와 동일할 수도 있고 다를 수도 있음을 유의한다. 운영 시스템(144), 응용 프로그램(145), 다른 프로그램 모듈(146), 및 프로그램 데이타(147)는 최소한 그들이 다른 카피라는 것을 도시하기 위해 다른 번호로 주어진다. 사용자는 키보드(162) 및 통상 마우스, 트랙볼 또는 터치 패드라 칭해지는 포인팅 장치(161) 등의 입력 장치를 통해 커맨드 및 정보를 컴퓨터(110)에 입력할 수 있다. 다른 입력 장치들(도시하지 않음)은 마이크로폰, 조이스틱, 게임 패드, 위성 안테나(satellite dish), 스캐너 등을 포함할 수 있다. 이들 및 다른 입력 장치들은 종종 시스템 버스에 결합되는 사용자 입력 인터페이스(160)를 통해 프로세싱 유닛(120)에 접속되지만, 병렬 포트, 게임 포트 또는 유니버셜 직렬 버스(USB) 등의 다른 인터페이스 및 버스 구조에 의해 접속될 수 있다. 모니터(191) 또는 다른 유형의 표시 장치도 비디오 인터페이스(190) 등의 인터페이스를 통해 시스템 버스(121)에 접속된다. 모니터 외에, 컴퓨터는 또한 출력 주변 인터페이스(195)를 통해 접속될 수 있는, 스피커(197) 및 프린터(196) 등의 다른 주변 출력 장치를 포함할 수도 있다.

<35> 컴퓨터(110)는 원격 컴퓨터(180) 등의 하나 이상의 원격 컴퓨터들과의 논리 접속을 사용하여 네트워크 환경에서 동작할 수 있다. 원격 컴퓨터(180)는 다른 퍼스널 컴퓨터, 서버, 라우터, 네트워크 PC, 피어 장치 또는 다른 공통 네트워크 노드일 수 있으며, 도 1에 단지 메모리 저장 장치(181)만이 도시되었지만, 통상적으로 퍼스널 컴퓨터(110)에 관련하여 상술한 모든 소자들 또는 다수의 소자들을 포함한다. 도 1에 도시된 논리 접속은 근거리 통신망(LAN : 171), 광역 통신망(WAN : 173)을 포함하지만 다른 네트워크를 포함할 수도 있다. 이러한 네트워킹 환경은 오피스, 기업형 컴퓨터 네트워크, 인트라넷 및 인터넷에서 흔하다.

<36> LAN 네트워킹 환경에서 사용되는 경우, 퍼스널 컴퓨터(110)는 네트워크 인터페이스 또는 어댑터(170)를 통해 LAN(171)에 접속된다. WAN 네트워킹 환경에서 사용되는 경우, 컴퓨터(110)는 통상적으로 모뎀(172), 또는 인터넷 등의 WAN(173)을 통해 통신을 수립하기 위한 다른 수단들을 포함한다. 내부 또는 외부에 있을 수 있는 모뎀(172)은 사용자 입력 인터페이스(160), 또는 다른 적합한 메카니즘을 통해 시스템 버스(121)에 접속될 수 있다. 네트워크 환경에서, 퍼스널 컴퓨터(110)에 대하여 도시된 프로그램 모듈, 또는 그 부분들은 원격 메모리 저장

장치에 저장될 수 있다. 제한되지 않는 예로서, 도 1은 메모리 장치(181)에 상주하는 원격 응용 프로그램(185)을 도시한다. 도시된 네트워크 접속은 예시적이며 컴퓨터 사이의 통신 링크를 수립하는 다른 수단들이 사용될 수 있음을 이해할 것이다.

<37> 이어지는 설명에서, 본 발명은 다르게 도시되지 않는다면, 하나 이상의 컴퓨터에 의해 수행되는 동작들의 처리 및 심볼 표시를 참조하여 설명될 것이다. 이러한 것으로서, 때때로 컴퓨터에 의해 실행되는 것으로 참조되는 이러한 작용들 및 동작들은 컴퓨터의 프로세싱 유닛에 의해 데이타를 구조화된 형태로 나타내는 전자 신호의 조작을 포함함을 이해할 것이다. 이 조작은 데이타를 변환하거나 그것을 컴퓨터의 메모리 시스템의 위치들에서 유지하며, 이는 당업자들이 이해할 수 있는 방식으로 컴퓨터의 동작을 재구성 또는 그렇지 않으면 변경한다. 데이타가 유지되는 데이타 구조는 데이타의 포맷에 의해 정의되는 특별한 특성을 갖는 메모리의 물리적 위치이다. 그러나, 본 발명은 상기 문맥으로 기술되지만, 당업자들은 후술되는 각종 작용 및 동작이 하드웨어에서도 구현될 수 있음을 이해할 수 있으며 이에 제한되는 것을 의미하지 않는다.

<38> 익스체인지(Exchange) 서버를 갖는 SMTP(Simple Mail Transfer Protocol)는 본 발명을 설명하는데 사용될 것이다. 익스체인지는 마이크로소프트 코포레이션에 의해 생산되는 이메일 서버이다. SMTP는 인터넷 상에서 사용되는 유력한 이메일 프로토콜이다. SMTP 및 익스체인지가 사용되지만, 본 발명에는 다른 전송 프로토콜 및 메일 서버가 사용될 수 있다. SMTP는 인터넷을 통해 익스체인지와 같은 하나의 이메일 서버로부터 다른 이메일 서버로 메일을 전송하는데 사용되는 메시지 포맷을 정의하는 TCP/IP(Transmission Control Protocol/Internet Protocol) 통신 프로토콜이다. SMTP에 의하면, 이메일 메시지는 통상적으로 다음 방식으로 전송된다. 사용자는 이메일 메시지를 생성하기 위해 이메일 프로그램을 실행하고, 이메일 프로그램은 그 메시지 텍스트 및 제어 정보를 인출 메시지의 큐(queue)에 배치한다. 큐는 통상적으로 이메일 서버에 액세스가능한 파일의 집합으로서 구현된다.

<39> 익스체인지 서버는 목적지 이메일 서버 상의 예비 SMTP 포트와 TCP(Transmission Control Protocol) 접속을 수립하고 인터넷을 통해 메시지를 전송하기 위해 SMTP를 사용한다. 송신 및 수신 서버들 사이의 SMTP 세션에 의해 메시지가 송신 호스트 상의 큐로부터 스테이지의 수신 호스트 상의 큐로 전송되는 결과를 야기시킨다. 스테이지는 수립되고 있는 접속의 IP 어드레스를 제공하는 송신 서버로부터 모든 메시지 헤더 및 메시지 내용의 수신까지 변화한다. 메시지 전송이 완료되면, 수신 서버는 SMTP에 의해 사용되는 TCP 접속을 닫고, 송신 호스트는 그 메일 큐로부터 메시지를 제거하며, 수신자는 그의 구성된 이메일 프로그램을 사용하여 메일 큐의 메시지를 판독한다.

<40> 이제 도 2를 참조하면, SMTP 스택(200)은 IIS(Internet information server : 202) 내에서 운영되며, 이것은 서버(204) 상에 설치되며, 마이크로소프트 코포레이션에 의해 판매되는 웹서버 소프트웨어이다. IIS(202)는 SMTP를 통해 인터넷 상의 다른 익스체인지 서버(206) 또는 SMTP 서버(도시하지 않음)와 통신한다. IIS(202)는 인출 또는 인입 메시지를 저장하는데 사용되는 데이타베이스(208)를 갖는다. 인입하는 메시지에 대한 SMTP 프로토콜(200)과의 접속이 수립되면, 이벤트가 시작되어 프레임워크(210)에 의해 수신된다. 프레임워크(210)가 메시지를 가로채어 그것을 하나 이상의 필터(212)로 보낸다. 필터(212)는 그 메시지를 분석하고, 필터(212)가 갖고 있으며 그 메시지가 스팸이라는 신뢰 레벨을 결정하며, 그 신뢰 레벨을 프레임워크(210)로 보낸다. 프레임워크(210)는 신뢰 레벨에 기초하여 그것이 다른 필터(212) 또는 액션(214)을 기동하기를 원하는지의 여부를 결정한다. 액션(214)은 접속을 해제하고, 메시지를 익스체인지 전송(216)으로 송신하고, 메시지를 삭제한다. 익스체인지 전송(216)은 메시지를 라우팅한다. 이것은 메시지가 서버(204) 상의 메일박스로 전달되어야 하는지 또는 SMTP(200)를 통해 다른 서버(206)로 갈 필요가 있는지를 결정한다.

<41> 이제 도 3을 참조하면, 필터(210)는 여러 유형의 안티-스팸 검출 기술들을 포함한다. 예를 들면, 필터(210)의 유형은 실시간 블랙홀 리스트 모듈(300), 비선형 모듈(302), 안티 바이러스 모듈(306)이 익스체인지 서버(204)와 통신하는데 사용되는 안티 바이러스 API(304), 터프 리스트 모듈(308), 및 메시지가 스팸인지를 결정하기 위해 그 자신의 룰을 이용하는 다른 필터들(310)일 수 있다. 예를 들면, 다른 필터(310)는 텍스트 분류, 키워드 매칭 등일 수 있다. 이와 같은 여러 유형의 안티-스팸 검출 기술들은 메시지를 분석하여 특정 메시지의 신뢰 레벨을 결정할 수 있는데, 이러한 신뢰 레벨은 그 특정 메시지가 스팸인지 여부를 나타낸다.

<42> 실시간 블랙홀 리스트 모듈(300)은 메시지 송신자의 IP 어드레스를 공지된 스팸 어드레스들의 리스트와 비교한다. IP 어드레스가 공지된 리스트 상에 있으면, 익스체인지 서버(204)는 메일을 받아들이지 않는다. 비선형 모듈(302)은 s-형상의 곡선과 같은 함수들, 스팸과 합법적인 메시지 사이를 강제로 분리시키는 베이지안(Bayesian) 함수 등을 이용하여 필터(210)의 신뢰 레벨을 정규화한다. 예를 들면, 필터(210)가 95%의 신뢰 레

벨을 리턴하면, 비선형 모듈(302)은 신뢰 레벨을 96%까지 스케일링하지만 40%의 신뢰 레벨은 30%의 신뢰 레벨까지 스케일링될 수 있다. 터프 리스트 모듈(308)은 송신자의 메일 어드레스 및/또는 도메인, 메일의 타겟 수신자(들), 및 메시지 id, 날짜, 제목, 첨부 형태 및 이름과 같은 실제 메시지 본문의 특징을 포함하는, 이용가능한 정보에 기초하여 송신자와 익스체인지 서버 사이의 SMTP 프로토콜 교환 중에 메일을 거부한다.

<43> 프레임워크(210)는 하나 이상의 안티-스팸 필터(210)의 기동을 관리하고, 각 기동의 결과들을 정규화하며, 정규화된 결과를 평가하고, 그 결과에 대한 액션을 취한다. 프레임워크(210)는 가장 통상적으로 네트워크의 에지에서(즉, 인터넷으로부터 처음으로 이메일을 수신하는 메일 서버) 서버(204)에 대해 전개된다. 텍스트 분류와 같이, 사용되는 기술들 중 일부는 메시지의 중요도 또는 민감성을 식별하는 등의 다른 용도로 사용될 수 있다. 이 결과로서, 프레임워크는 또한 내부 서버 상에 유용하게 전개될 수 있다. 프레임워크(210)는 독립형 구현으로부터 이동을 돕기 위해 기존의 독립형 스팸 검출 구현에 의해 기동되는 유틸리티의 라이브러리로서, 또는 보다 바람직하게는 안티-스팸 필터(210)를 호출하는데 사용되는 잠재적인 이벤팅 메카니즘(이하에 기술함)으로부터 추상적 개념을 제공하는 래퍼(wrapper)로서 단독으로 사용될 수 있다. 래퍼 실시예는 이메일용으로 개발된 안티-스팸 필터(210)가 인스턴트 메시징 등의, 고민스러운 메시지를 판정하는 다른 메시징 솔루션들에도 사용될 수 있게 한다. 이들 경우에 있어서, 프레임워크는 구축 또는 동작 시에 안티-스팸 기술로 링크되는 라이브러리로서 전달된다.

<44> 익스체인지의 SMTP 스택(200)의 아키텍쳐는, 이벤트가 통상적으로 COM 오브젝트로서 구현되는 이벤트 싱크(event sink)에 대한 스택(200)에 의해(즉, 이로부터 기원함) 시작된다는 것이다. 새로운 안티-스팸 기술들이 전개되면, 그것은 인스톨시 프로토콜 이벤트 시스템으로 등록하는 COM 오브젝트를 구현한다. 등록 코드는 프레임워크(210)에 의해 전달된다. 프레임워크(210)의 설치는 해당 서버 상에 소프트웨어를 설치하는 것, 이벤트 싱크를 등록하는 것, 시스템 관리자 콘솔을 통해 특정 서버에 대해 특정 기술을 가능하게 또는 불가능하게 하는 것, 및 스팸이 수신될 때 이어지는 평가 및 액션 전략을 수립하는 것을 포함한다. 특정 기술의 가능/불가능은 네트워크의 모든 서버들이 동일한 소프트웨어 바이너리를 포함하게 함으로써 프레임워크(210)의 처리능력을 증진시킨다.

<45> 이제 도 4를 참조하면, 안티-스팸 모듈들(212)을 통합하고 메시지가 스팸인지의 여부를 결정하는 프로세스가 도시된다. 실행 시에, 접속이 SMTP 스택(200)(및 그후에는 포인트들)에 오픈되면, 이벤트가 시동된다(단계 400). 이벤트 디스패치 시스템은 그것의 등록 리스트를 점검하고 대응하는 오브젝트를 기동한다. 그것이 래퍼(wrapper)로서 작용하는 경우에는 직접적으로, 또는 그것이 라이브러리 함수로서 호출되고 있는 경우에는 간접적으로, 프레임워크(210)를 기동하게 한다(단계 402). 양 경우에 있어서, 프레임워크(210)는 자신의 구성을 조사하여 등록되었던 안티-스팸 기술들(300-310) 중 어느 것이 시스템 관리자에 의해 "인에이블" 또는 "디스에이블" 되었는지를 결정한다. 인에이블되었던 안티-스팸 기술들이 계산되고, 스팸 신뢰 레벨의 합산은 제로로 설정된다(단계 404).

<46> 특정 안티-스팸 필터(212)가 "인에이블"되면, 프레임워크(210)는 조사용의 안티-스팸 필터(212)에 이용가능한 어떠한 정보라도 획득하고 그 정보를 필터(212)로 전송한다(단계 406). 이용가능한 정보의 양 및 유형은 필터(212)가 기동되는 프로토콜의 단계에 따라 변경될 것이다. 예를 들어, 여기서 기동된 제1 시기에는 수립되고 있는 접속의 IP 어드레스에 대한 정보만 일 수 있다. 시스템이 메시지를 승인하기 전의 최종 호출에서, 모든 메시지 헤더들과 내용(content)은 이용가능하다. 프레임워크(210)가 메시지의 인코딩된 내용을 크랙(cracking)할 수 있으면, 프레임워크(210)는 메시지의 인코딩된 내용을 안티-스팸 필터(212)가 더욱 용이하게 사용할 수 있는 형태로 크랙할 것이다. 프레임워크가 래퍼로서 구현되는 경우, 이 정보는 자동적으로 이용가능해진다. 라이브러리로서 구현되는 경우에는, 정보는 안티-스팸 필터(212)가 특정적으로 요청하는 경우에만 이용가능할 것이다. 프레임워크의 형태(즉, 래퍼 또는 라이브러리)에 관계 없이, 메시지 내용을 크랙하는 것과 같은 유틸리티 기능들이 안티-스팸 필터에 의해 수동적으로 기동되어 CPU 부하를 감소시킨다. 일 실시예에서, 프레임워크(210)는 또한 필터(212)가 스팸과 같은 한개의 메일의 평가의 부분으로서 사용할 수 있는 메시지의 수신 어드레스의 룩업(lookup)을 제공한다.

<47> 그 평가를 완료하면, 필터(212)는 리턴값 혹은 레퍼런스 중의 하나에 의해(또는 프레임워크 라이브러리를 호출함으로써) 프레임워크(210)로 되돌아가고, 솔루션이 갖는 신뢰의 평가는 특정 메일 메시지가 스팸이라는 것이다(단계 408). 프레임워크(210)는 통상적으로 0-100% 범위의 답을 기대하며, 여기서 0%는 명백하게 스팸이 아님을 나타내며, 100%는 명백하게 스팸임을 나타낸다. 일 실시예에 있어서, 퍼센티지는 0과 1000 사이의 숫자로서 표현된다. 상이한 측정들을 사용하는 다양한 안티-스팸 기술들(300-310)을 수용하기 위하여, 프레임워크(210)는 기동되는 각각의 개별적인 필터(212)로부터의 결과들에 적용되는 스케일링 또는 조정 인자를 제공하여 정규

화되거나 또는 조정된 스팸 신뢰 레벨을 생성한다(단계 410). 스케일링 또는 조정 인자는 서버(204)의 관리자에 의해 구성된다. 이것은 상이한 필터들(210)의 결과를 비교하기 위하여 프레임워크가 수행하여야만 하는 정규화를 구성한다. 이 정규화된 숫자는 스팸 신뢰 레벨로서 참조될 것이다. 스팸 신뢰 레벨은 스팸 신뢰 레벨의 합산에 가산된다(단계 412). 프레임워크는 계산된 스팸 신뢰 레벨을, 구동 합산으로써 지속성을 위해 메시지 자체에 및/또는 성능을 위해 메모리에 저장한다. 연속적인 솔루션의 평가 결과들은 스팸 신뢰 레벨의 합산에 가산된다.

<48> 정규화(즉, 스케일링 인자의 적용)는 다양한 방법으로 구현될 수 있다. 예를 들면, 결과들을 정규화하기 위한 한가지 방법은 각 필터의 결과들을 동등하게 신뢰하여 그 결과들을 간단하게 합산하는 것이다(예를 들면, 0.5+0.7+0.8+...=합산). 다른 방법으로는 각각의 것에 스케일을 적용하는 것이다. 예를 들면, 특정 필터가 스팸을 검출하는 방법을 관리자가 원하는 경우, 관리자는 다소 높은 숫자(예를 들어, 0.9)에 의해 그 필터의 스팸 신뢰 레벨을 스케일링한다. 마찬가지로, 관리자가 그다지 신뢰하지 않는 필터가 있으면, 관리자는 다소 낮은 숫자(예를 들어, 0.3)에 의해 그 필터의 스팸 신뢰 레벨을 스케일링한다. 스케일링 인자의 다른 예는 s-형태의 곡선과 같은 가중 곡선(weighted curve)을 사용하여 비선형 신뢰 레벨 정규화를 사용하는 것이다. 예를 들면, 필터(210)가 95%의 스팸 신뢰 레벨을 리턴하면, 스케일링 인자는 그것을 더 높게, 예를 들어 96%로 스케일링한다. 스팸 신뢰 레벨이 범위의 중간(예를 들어, 50-55%)에 있으면, 좀 더 극단적인 스케일링이 적용되어 그것을 더 낮게 스케일링한다(예를 들어, 30-35%).

<49> 프레임워크(210)는 관리자에게, 스팸 신뢰 레벨의 합산이 초과하는 최대 임계값에 기초하여 메시지와 함께 취해지는 다양한 액션들을 관리자가 정의할 수 있게 하는 다수의 임계값을 설정하는 기능을 제공한다. 액션들은 메시지가 스팸인지의 여부에 대한 더 좋은 아이디어를 알 때까지 메시지가 전달되는 것을 방지하고, 접속을 해제하고, 비전달 메시지를 송신자에게 전송하고, 메시지를 삭제하고, 스팸 신뢰 레벨의 합산에 기초하여 그것을 다른 필터(210)로 전달하고, 메시지를 수신자에게 전송 등을 한다. 임계값들 및 대응하는 액션들의 디폴트 세트는 프레임워크(210)에 제공된다.

<50> 스케일링된 스팸 신뢰 레벨의 합산은 관리자에 의해 설정된 상부 임계 세트와 비교된다(단계 414). 스팸 신뢰 레벨의 합산이 상부 임계값을 초과하면, 상부 임계값에 대해 구성된 액션이 취해진다(단계 416). 스팸 신뢰 레벨의 합산이 상부 임계 레벨을 초과하지 않고 메지시를 평가하는데 이용될 수 있는 더 많은 필터들(210)이 있으면, 상부 임계값을 초과하거나 또는 메시지의 끝을 수신하거나 할 때까지 단계 404 내지 단계 416이 반복된다(단계 420). 메시지의 끝이 수신되지 않으면, SMTP 스택(200)은 다음 메시지 승인 상태로 이동하고(422), 다음 메시지 승인 상태에 대해 단계 406 내지 420이 반복된다. 메시지의 끝을 수신하고 모든 인에이블된 필터들이 메시지를 분석하고 나면, 스팸 신뢰 레벨의 합산은 스팸 신뢰 레벨의 합산이 임계값을 초과할 때까지 상부 임계값으로부터 하부 임계값으로의 순서대로 나머지 임계값들과 비교된다(단계 426). 스팸 신뢰 레벨의 합산이 임계값을 초과하면, 그 임계값에 대해 구성된 액션이 취해진다.

<51> 요약하면, 필터(212)가 그것을 분석하는 것을 완료한 후에, 프레임워크(210)는 관리자에 의해 규정된 임계값의 세트에 대해 스팸 신뢰 레벨의 합산을 평가한다. 스팸 신뢰 레벨의 합산이 관리자가 설정한 최상위 임계값보다 크면, 최상위 임계값에 대해 특정된 액션을 메시지와 함께 취한다. 그렇지 않으면, 후속하는 필터들이 사용되어 최대 임계값을 초과하거나 또는 모든 필터들이 메시지를 평가할 때까지 메시지를 평가한다. 모든 필터들이 메시지를 평가한 후에, 스팸 신뢰 레벨의 합산은 모든 임계값에 대해 비교되어, 매칭 임계값이 선택된다. 그 임계값과 연관된 액션이 그 다음에 취해진다. 예를 들면, 스팸 신뢰 레벨의 합산이 99% 신뢰 레벨 임계값을 초과하면, 메시지 접속은 조용하게 해제될 수 있다. 스팸 신뢰 레벨의 합산이 70% 신뢰 레벨 임계값을 초과하면, 비전달 통지가 송신자에게 리턴될 수 있다. 스팸 신뢰 레벨의 합산이 40% 신뢰 레벨 임계값을 초과하면, 메시지는 사용자의 메일박스에 있는 "정크 메일(junk mail)" 폴더로 전달된다. 스팸 신뢰 레벨의 합산이 임의의 임계값들을 초과하지 않으면, 메시지는 적법한 것으로 여겨져서 사용자의 받은편지함(inbox)으로 전달된다.

<52> 모든 점에 있어서, 특정 메시지와 함께 취해진 액션들은 관리자가 기록하기로 선택한 정보의 레벨에 따라, 메시지 추적 테이블을 로그하거나 그에 추가될 수 있다. 디폴트 세트의 액션들은 프레임워크(210)와 함께 관리자에게 이용가능하다. 새로운 안티-스팸 필터들을 전개하는데 이용되었던것과 유사한 방식으로 추가의 액션 실행 코드를 전달함으로써 추가의 액션들이 부가될 수 있다.

<53> 프레임워크(210)에 의해 메일박스로의 전달을 위해 승인된 임의의 메시지는 공지된 속성으로 그것에 저장된 메시지의 스팸 신뢰 레벨의 합산을 가진다. 메시지를 처리하는 전달 에이전트는 그 자신의 논리의 일부로서 이 속성을 평가하기 위해 선택할 수 있다. 메시지 또는 이러한 메시지들의 테이블을 보는 클라리언트는, 잘못 계

산될 수 있는 메시지들을 식별하기 위하여 보조로써, 스팸 신뢰 레벨의 오름차순 또는 내림차순으로 메시지들을 나열하도록 선택할 수 있다.

발명의 효과

<54> 다양한 기존과 미래의 안티-스팸 필터들 및 기술들을 사용하여 네트워크 경계의 에지에서 스팸과 바이러스들이 검출될 수 있고 처리될 수 있게 하기 위한 플랫폼이 서술되었음을 알 수 있다. 이 플랫폼은 솔루션 및 기술들이 합리적인 방법으로 상호작용하고 관리될 수 있게 함에 따라, 스팸 검출의 도전적인 환경에서 서버 측에 급속한 혁신을 전개하는 기능을 제공할 수 있게 한다.

<55> 본 발명의 원리가 적용될 수 있는 많은 가능한 실시예들의 견지에서, 도면에 대하여 본 명세서에 서술된 실시예들은 단지 예시적인 것을 의미하는 것이며 본 발명의 범위를 한정하는 것으로 취급되어서는 안된다. 예를 들어, 당업자들은 소프트웨어로 나타내어진 예시적인 실시예들의 구성요소들이 하드웨어로 구현될 수 있으며 또 그 역도 가능하고, 또는 예시된 실시예들은 구성 및 상세함에 있어서 본 발명의 사상으로부터 벗어나지 않으면서 변형될 수 있다. 따라서, 본 명세서에 서술된 발명은 첨부하는 특허청구범위 및 그 등가물의 범위 내에 속할 수 있는 모든 실시예들을 예상할 수 있다.

도면의 간단한 설명

<1> 도 1은 본 발명이 상주하는 예시적인 컴퓨터 시스템을 일반적으로 도시하는 블록도.

<2> 도 2는 SMTP 프로토콜 스택을 사용하는 시스템에서의 본 발명의 프레임워크를 일반적으로 도시하는 블록도.

<3> 도 3은 본 발명에 따라 사용된 안티-스팸 모듈들의 예를 도시하는 블록도.

<4> 도 4는 안티-스팸 모듈을 통합하여 메시지가 스팸인지를 결정하는 프로세스를 도시하는 플로우챠트.

<5> <도면의 주요 부분에 대한 부호의 설명>

<6> 200: SMTP 프로토콜

<7> 208: 데이타베이스

<8> 210: 프레임워크

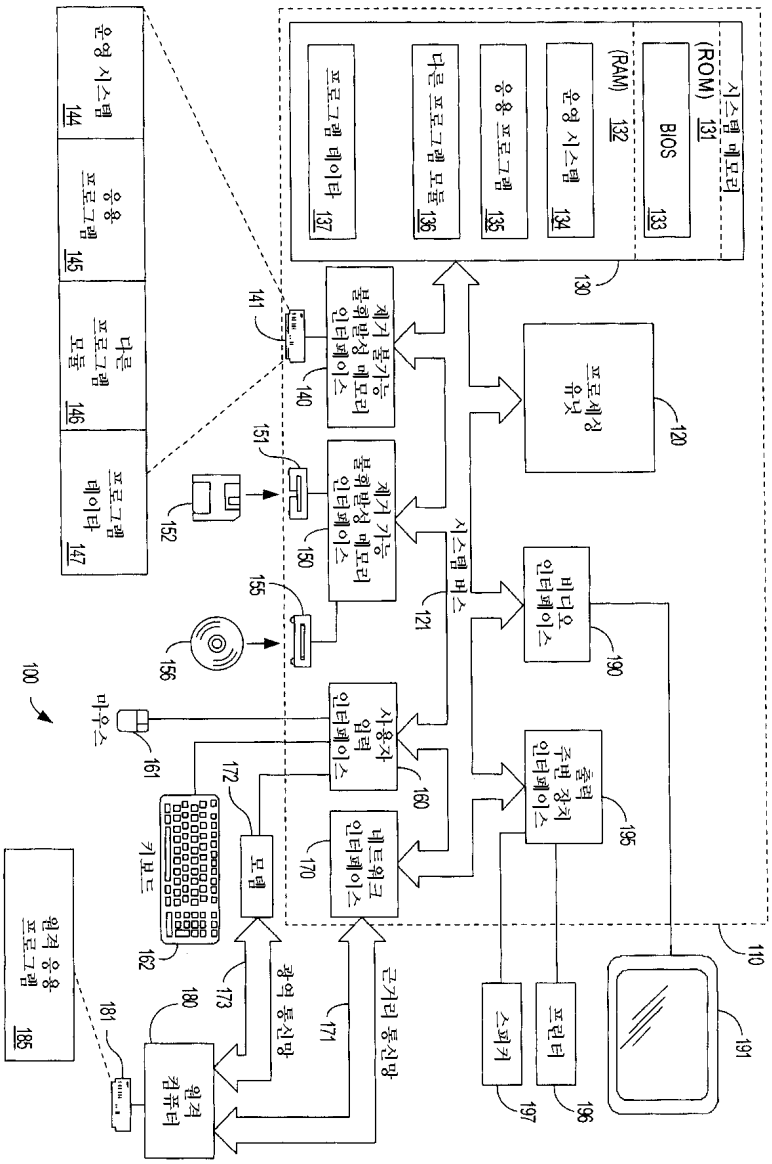
<9> 212: 필터

<10> 214: 액션

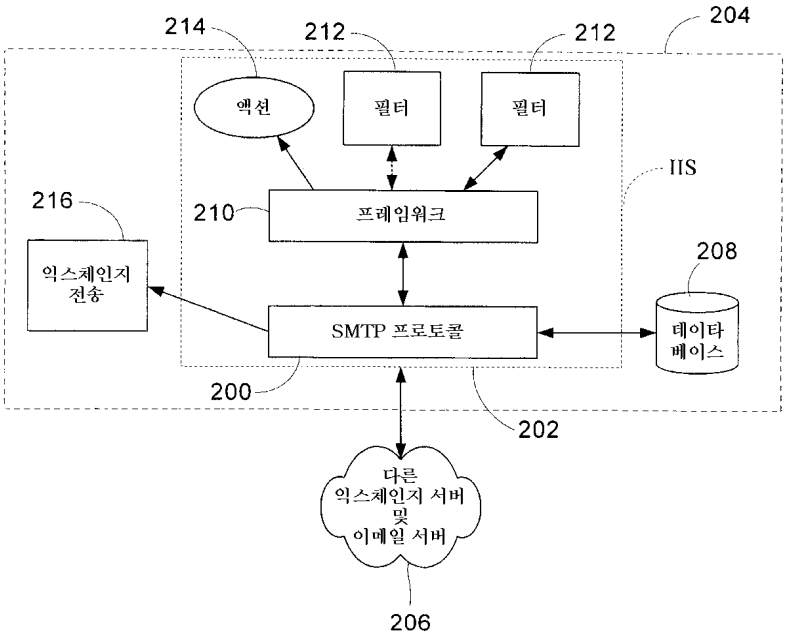
<11> 216: 익스체인지 전송

도면

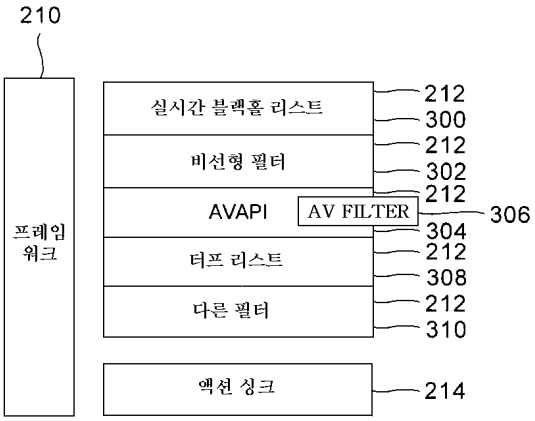
도면1


시스템 메모리
(ROM) 131
BIOS 133
(RAM) 132
운영 시스템 134
응용 프로그램 135
다른 프로그램 모듈 136
프로그램 데이터 137
130
프로세싱 유닛 120
시스템 버스 121
비디오 인터페이스 190
출력 주변 장치 인터페이스 195
사용자 입력 인터페이스 160
네트워크 인터페이스 170
제거 불가능 불휘발성 메모리 인터페이스 140
제거 가능 불휘발성 메모리 인터페이스 150
141
151
155
마우스 161
키보드 162
모뎀 172
원격 컴퓨터 180
181
원격 응용 프로그램 185
광역 통신망 173
근거리 통신망 171
모니터 191
프린터 196
스피커 197
110
100
152
156
운영 시스템 144
응용 프로그램 145
다른 프로그램 모듈 146
프로그램 데이터 147

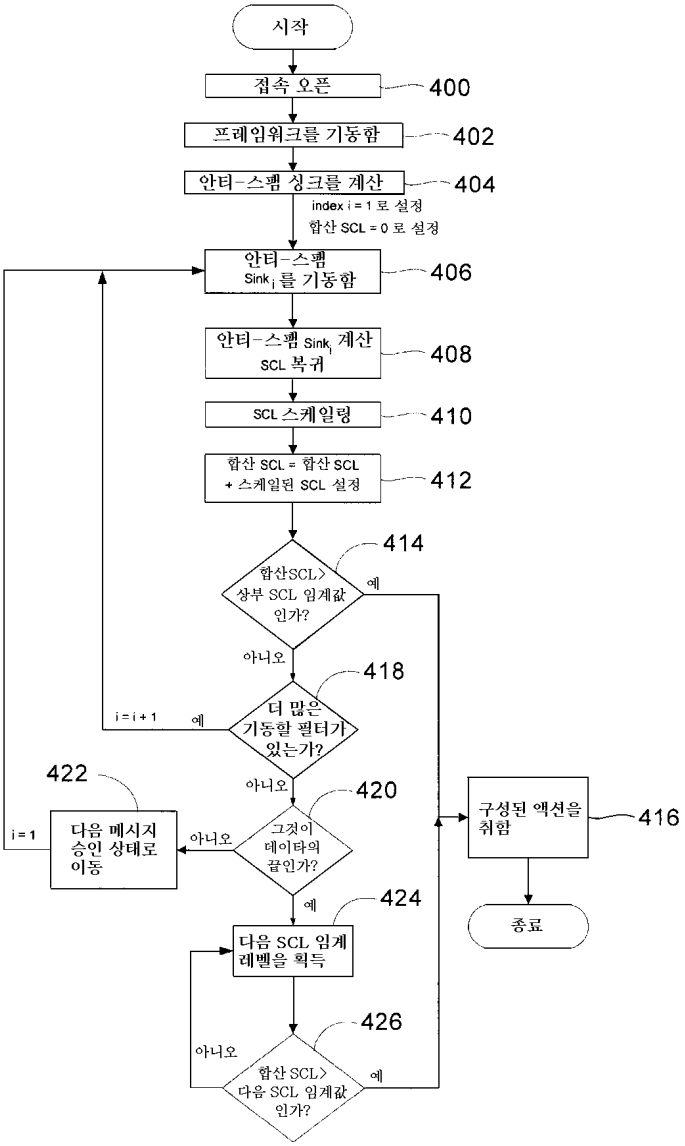
도면2


214
212
212
204
액션
필터
필터
IIS
216
210
프레임워크
익스체인지
전송
208
SMTP 프로토콜
데이타
베이스
200
202
다른
익스체인지 서버
및
이메일 서버
206

도면3


210
프레임
워크
실시간 블랙홀 리스트
212
300
비선형 필터
212
302
AVAPI
AV FILTER
212
306
304
터프 리스트
212
308
다른 필터
212
310
액션 싱크
214

도면4


시작
접속 오픈
400
프레임워크를 기동함
402
안티-스팸 싱크를 계산
404
index i = 1 로 설정
합산 SCL = 0 로 설정
안티-스팸 Sink_i 를 기동함
406
안티-스팸 Sink_i 계산 SCL 복귀
408
SCL 스케일링
410
합산 SCL = 합산 SCL + 스케일된 SCL 설정
412
414
합산SCL> 상부 SCL 임계값 인가?
예
아니오
418
더 많은 기동할 필터가 있는가?
i = i + 1
예
아니오
422
420
다음 메시지 승인 상태로 이동
i = 1
아니오
그것이 데이타의 끝인가?
구성된 액션을 취함
416
예
종료
424
다음 SCL 임계 레벨을 획득
426
아니오
합산 SCL> 다음 SCL 임계값 인가?
예