



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0117852  
(43) 공개일자 2021년09월29일

(51) 국제특허분류(Int. Cl.)  
G06F 40/279 (2020.01) H04L 29/06 (2006.01)  
(52) CPC특허분류  
G06F 40/279 (2020.01)  
G06F 21/56 (2013.01)  
(21) 출원번호 10-2020-0034699  
(22) 출원일자 2020년03월20일  
심사청구일자 2020년03월20일

(71) 출원인  
엘아이지넥스원 주식회사  
경기도 용인시 기흥구 마북로 207 (마북동)  
고려대학교 산학협력단  
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)  
(72) 발명자  
김휘강  
서울특별시 성북구 안암로 145, 로봇융합관 (안암동5가, 고려대학교안암캠퍼스)  
송현민  
서울특별시 성북구 안암로 145, 로봇융합관 (안암동5가, 고려대학교안암캠퍼스)  
(74) 대리인  
특허법인우인

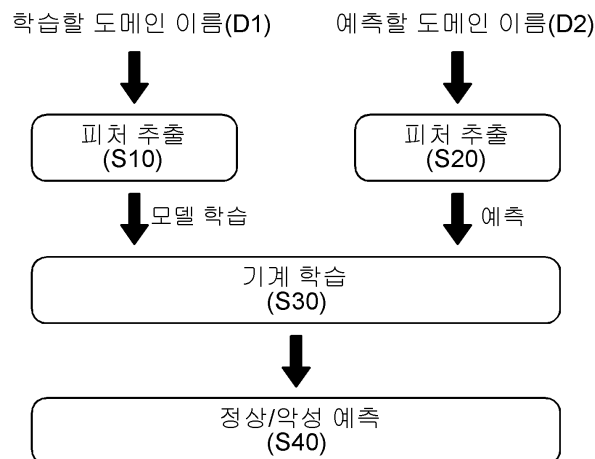
전체 청구항 수 : 총 13 항

(54) 발명의 명칭 악성 도메인 탐지 장치 및 방법

(57) 요약

본 발명은 문자 정보 피처를 이용한 악성 도메인의 탐지 방법 및 장치에 관한 것이다. 본 명세서의 일실시예에 따른 악성 도메인 탐지 방법은 학습할 도메인 이름을 입력받는 입력 단계; 상기 입력된 도메인 이름에서 복수의 피처를 추출하는 특징 추출 단계; 상기 추출된 피처를 기반으로 입력받은 도메인 이름이 정상 또는 악성인지 탐지할 수 있도록 학습하여 악성 도메인 분류 모델을 구축하는 학습 단계; 예측할 도메인 이름을 입력받는 단계; 상기 입력받은 도메인 이름에서 복수의 피처를 추출하고, 상기 학습 단계에서 구축된 악성 도메인 분류 모델로 전달하여 상기 입력받은 도메인 이름이 정상인지 또는 악성인지 예측하는 단계를 포함한다.

대표도 - 도1



(52) CPC특허분류

**H04L 63/0236** (2013.01)

(72) 발명자

**유정도**

서울특별시 성북구 안암로 145, 로봇융합관 (안암  
동5가, 고려대학교안암캠퍼스)

**박정호**

서울특별시 성북구 안암로 145, 로봇융합관 (안암  
동5가, 고려대학교안암캠퍼스)

**김광수**

경기도 성남시 분당구 판교로 333 (삼평동)

**조병모**

경기도 성남시 분당구 판교로 333 (삼평동)

**홍수연**

경기도 성남시 분당구 판교로 333 (삼평동)

## 명세서

### 청구범위

#### 청구항 1

학습할 도메인 이름을 입력받는 입력 단계;

상기 입력된 도메인 이름에서 복수의 피처를 추출하는 특징 추출 단계;

상기 추출된 피처를 기반으로 입력받은 도메인 이름이 정상 또는 악성인지 탐지할 수 있도록 학습하여 악성 도메인 분류 모델을 구축하는 학습 단계;

예측할 도메인 이름을 입력받는 단계; 및

상기 입력받은 도메인 이름에서 복수의 피처를 추출하고, 상기 학습 단계에서 구축된 악성 도메인 분류 모델로 전달하여 상기 입력받은 도메인 이름이 정상인지 또는 악성인지 예측하는 단계를 포함하는 악성 도메인 탐지 방법.

#### 청구항 2

제1항에 있어서,

상기 복수의 피처는 상기 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블의 길이 합을 나타내는 도메인 길이를 포함하는 것을 특징으로 하는 악성 도메인 탐지 방법.

#### 청구항 3

제1항에 있어서,

상기 복수의 피처는 상기 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블에서 특수문자의 개수를 나타내는 특수문자 발생빈도를 포함하는 것을 특징으로 하는 악성 도메인 탐지 방법.

#### 청구항 4

제1항에 있어서,

상기 복수의 피처는 상기 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블에서 특수문자의 비율을 나타내는 특수문자 발생비율을 포함하는 것을 특징으로 하는 악성 도메인 탐지 방법.

#### 청구항 5

제1항에 있어서,

상기 복수의 피처는 상기 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블에서 숫자의 개수를 나타내는 숫자 발생빈도를 포함하는 것을 특징으로 하는 악성 도메인 탐지 방법.

#### 청구항 6

제1항에 있어서,

상기 복수의 피처는 상기 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블에서 숫자의 비율을 나타내는 숫자 발생비율을 포함하는 것을 특징으로 하는 악성 도메인 탐지 방법.

#### 청구항 7

제1항에 있어서,

상기 복수의 피처는 상기 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블에서 영어 모음의 개수를 나타내는 모음 발생빈도를 포함하는 것을 특징으로 하는 악성 도메인 탐지 방법.

**청구항 8**

제1항에 있어서,

상기 복수의 피처는 상기 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블에서 영어 모음의 비율을 나타내는 모음 발생비율을 포함하는 것을 특징으로 하는 악성 도메인 탐지 방법.

**청구항 9**

제1항에 있어서,

상기 복수의 피처는 상기 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블에서 가장 긴 레이블의 길이를 나타내는 레이블 최대 길이를 포함하는 것을 특징으로 하는 악성 도메인 탐지 방법.

**청구항 10**

제1항에 있어서,

상기 복수의 피처는 상기 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블에서 연속된 숫자 또는 문자의 최대 길이를 포함하는 것을 특징으로 하는 악성 도메인 탐지 방법.

**청구항 11**

제1항에 있어서,

상기 복수의 피처는 상기 도메인 이름의 N-gram 비교값으로 판단하는 평판정보 또는 N-gram 희귀성 정보를 포함하며, 상기 N은 2 내지 5를 포함하는 것을 특징으로 하는 악성 도메인 탐지 방법.

**청구항 12**

학습할 도메인 이름 및 예측할 도메인 이름을 입력받는 입력부;

상기 입력된 도메인 이름에서 복수의 피처를 추출하는 피처 추출부;

상기 피처 추출부에서 추출된 피처를 기반으로 입력받은 학습 도메인 이름이 정상 또는 악성인지 탐지할 수 있도록 학습하여 악성 도메인 분류 모델을 구축하는 기계학습부;

상기 입력부로부터 학습할 도메인 이름 리스트를 문자열로 전달받아, LSTM(Long Short-Term Memory) 모델을 사용하여 문자 패턴을 학습하고, 도메인 이름을 생성하는 도메인 생성부; 및

상기 피처 추출부에서 추출된 피처를 기반으로 입력받은 예측 도메인 이름이 정상 또는 악성인지 탐지하는 악성 도메인 탐지부를 포함하며,

상기 악성 도메인 탐지부는 상기 입력부를 통해서 예측할 도메인 이름이 입력되면, 상기 피처 추출부를 통해서 복수의 피처를 추출하고, 상기 추출된 복수의 피처를 전달받아서 상기 학습된 악성 도메인 분류 모델을 이용하여 상기 예측 도메인 이름이 정상 또는 악성인지 탐지하는 것을 특징으로 하는 악성 도메인 탐지 장치.

**청구항 13**

제12항에 있어서,

상기 복수의 피처는 상기 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블에 대해서 추출한 도메인 길이, 특수문자 발생빈도, 특수문자 발생비율, 숫자 발생빈도, 숫자 발생비율, 영어 모음 발생빈도, 영어 모음 발생비율, 레이블 최대 길이, 연속된 숫자의 길이, 연속된 문자의 길이, 평판 정보, N-gram 희귀성 정보 중 적어도 하나 이상을 포함하는 것을 특징으로 하는 악성 도메인 탐지 장치.

**발명의 설명****기술 분야**

본 발명은 문자 정보 피처를 이용한 악성 도메인의 탐지 방법 및 장치에 관한 것이다.

[0001]

## 배경 기술

- [0002] 많은 악성코드는 공격자가 구축한 C&C 서버와 통신하여 탈취한 정보를 전송하거나 공격 명령어를 수신한다. C&C(Command & Control) 서버란 일반적으로 감염된 좀비 PC가 해커가 원하는 공격을 수행하도록 원격지에서 명령을 내리거나 악성코드를 제어하는 서버를 말한다. 예를 들어 감염된 PC로 악성코드를 유포시키거나, 스팸 전송, 또는 DDoS 공격 명령 등을 전달하는 역할을 하며, 해당 C&C 서버의 인터넷 연결을 중단시키거나, 관련된 악성 도메인 접속을 차단하여 C&C 서버의 동작을 막을 수 있다.
- [0003] 악성코드는 도메인 이름(domain name)을 이용해 C&C 서버와 통신하는데 이를 위해서 악성코드에는 C&C 서버의 IP 주소 또는 도메인 주소를 내부에 가지고 있어야 하며, 이것을 hard-coding 되어 있다고 한다. 따라서, 악성코드가 hard-coding 방식으로 C&C 서버의 IP 주소나 도메인주소를 가지고 있다면, 보안 장비나 사법기관에서는 해당 IP 주소나 도메인 주소를 차단시킴으로써 악성코드에 감염된 디바이스가 C&C 서버에 접속하는 것을 방지할 수 있다.
- [0004] 악성코드 개발자는 이러한 접속 방지 기법을 우회하기 위하여, DGA(Domain Generation Algorithm)를 사용한다. 즉 악성코드는 먼저 hard-coding된 C&C 서버 주소로 접속을 시도를 하고, 접속이 안되면 DGA에서 생성한 도메인 주소로 접속을 시도하게 한다.
- [0005] DGA는 특정 시드값(seed)을 입력 받아 임의의 도메인 이름을 생성하는 알고리즘이다. 대부분의 경우, DGA에 의해 생성된 도메인 이름은 사람이 이해하기 힘든 형태의 문자열이며 그 형태가 매우 다양하므로 사람이 일일이 네트워크 블랙 리스트에 추가하여 규제하기 쉽지 않다. 따라서, 악성코드에 의한 피해를 최소화하기 위해서는 DGA로 생성되는 도메인 이름 문자열을 자동으로 식별하고 미리 차단할 방법이 필요하다.

## 발명의 내용

### 해결하려는 과제

- [0006] 본 발명은 도메인 이름 분류 모델의 학습시간이 비교적 짧으면서 악성 도메인의 탐지율이 높은 악성 도메인의 탐지 방법 및 장치를 제공하는 것을 목적으로 한다.
- [0007] 그러나 본 발명의 목적은 상기에 언급된 사항으로 제한되지 않으며, 언급되지 않은 또 다른 목적들은 아래의 기재로부터 당업자에게 명확하게 이해될 수 있을 것이다.

### 과제의 해결 수단

- [0008] 본 명세서의 일실시예에 따른 악성 도메인의 탐지 방법은 학습할 도메인 이름을 입력받는 입력 단계; 상기 입력된 도메인 이름에서 복수의 피처를 추출하는 특징 추출 단계; 상기 추출된 피처를 기반으로 입력받은 도메인 이름이 정상 또는 악성인지 탐지할 수 있도록 학습하여 악성 도메인 분류 모델을 구축하는 학습 단계; 예측할 도메인 이름을 입력받는 단계; 상기 입력받은 도메인 이름에서 복수의 피처를 추출하고, 상기 학습 단계에서 구축된 악성 도메인 분류 모델로 전달하여 상기 입력받은 도메인 이름이 정상인지 또는 악성인지 예측하는 단계를 포함한다.
- [0009] 또한, 상기 방법에서 복수의 피처는 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블에 대해서 추출한 도메인 길이, 특수문자 발생빈도, 특수문자 발생비율, 숫자 발생빈도, 숫자 발생비율, 영어 모음 발생빈도, 영어 모음 발생비율, 레이블 최대 길이, 연속된 숫자의 길이, 연속된 문자의 길이, 평판 정보 및 N-gram 희귀성 정보 중 적어도 하나 이상을 포함할 수 있다.
- [0010] 본 명세서의 일실시예에 따른 악성 도메인의 탐지 장치는 학습할 도메인 이름 및 예측할 도메인 이름을 입력받는 입력부; 상기 입력된 도메인 이름에서 복수의 피처를 추출하는 피처 추출부; 상기 피처 추출부에서 추출된 피처를 기반으로 입력받은 학습 도메인 이름이 정상 또는 악성인지 탐지할 수 있도록 학습하여 악성 도메인 분류 모델을 구축하는 기계학습부; 상기 입력부로부터 학습할 도메인 이름 리스트를 문자열로 전달받아, LSTM(Long Short-Term Memory) 모델을 사용하여 문자 패턴을 학습하고, 도메인 이름을 생성하는 도메인 생성부; 및 상기 피처 추출부에서 추출된 피처를 기반으로 입력받은 예측 도메인 이름이 정상 또는 악성인지 탐지하는 악성 도메인 탐지부를 포함하며, 상기 악성 도메인 탐지부는 상기 입력부를 통해서 예측할 도메인 이름이 입력되면, 상기 피처 추출부를 통해서 복수의 피처를 추출하고, 상기 추출된 복수의 피처를 전달받아서 상기 학습된 악성 도메인 분류 모델을 이용하여 상기 예측 도메인 이름이 정상 또는 악성인지 탐지하는 것을 특징으로 한다.

[0011] 마찬가지로, 상기 장치에서 복수의 피처는 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블에 대해서 추출한 도메인 길이, 특수문자 발생빈도, 특수문자 발생비율, 숫자 발생빈도, 숫자 발생비율, 영어 모음 발생빈도, 영어 모음 발생비율, 레이블 최대 길이, 연속된 숫자의 길이, 연속된 문자의 길이, 평판 정보 및 N-gram 회귀성 정보 중 적어도 하나 이상을 포함할 수 있다.

### 발명의 효과

[0012] 본 발명의 일실시예에 따르면, 도메인 이름 분류 모델의 학습시간이 비교적 짧으면서 악성 도메인의 탐지율이 높은 악성 도메인의 탐지 방법 및 장치가 제공되는 효과가 있다.

### 도면의 간단한 설명

[0013] 도 1은 본 발명의 일실시예에 따른 악성 도메인 예측 과정을 나타낸 흐름도이다.  
 도 2는 본 발명의 일실시예에 따른 문자 정보 피처를 설명하기 위한 참고도이다.  
 도 3은 본 발명의 일실시예에 따른 도메인 이름 생성 모델의 구조를 도시한 도면이다.  
 도 4는 도메인 이름 생성 모델의 입력과 출력을 나타낸 참고도이다.  
 도 5는 본 발명의 일실시예에 따른 악성 도메인 탐지 장치를 개략적으로 도시한 블록도이다.  
 도 6은 본 발명의 일실시예에 따른 도메인 이름 생성 모델의 입출력 방식을 설명하기 위한 참고도이다.

### 발명을 실시하기 위한 구체적인 내용

[0014] 이하, 본 발명의 바람직한 실시예를 첨부된 도면들을 참조하여 상세히 설명한다. 우선 각 도면의 구성요소들에 참조 부호를 부가함에 있어서, 동일한 구성요소들에 대해서는 비록 다른 도면상에 표시되더라도 가능한 한 동일한 부호를 가지도록 하고 있음에 유의해야 한다. 또한, 본 발명을 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다. 또한, 이하에서 본 발명의 바람직한 실시예를 설명할 것이나, 본 발명의 기술적 사상은 이에 한정하거나 제한되지 않고 당업자에 의해 변형되어 다양하게 실시될 수 있음은 물론이다.

[0015] 도 1은 본 발명의 일실시예에 따른 악성 도메인 예측 과정을 나타낸 흐름도이다.

[0016] 우선, 학습할 도메인 이름(D1)을 입력받고, 입력된 도메인 이름에서 복수의 피처를 추출한다(S10).

[0017] 본 명세서의 일실시예에 따르면, 복수의 피처(feature)는 다음 표 1과 같이 도메인의 문자정보와 관련된 14가지 요소로 구성될 수 있다.

표 1

| No. | 피처 명칭     | 내용                                 |
|-----|-----------|------------------------------------|
| 1   | 도메인 길이    | TLD를 제외한 도메인 이름의 길이                |
| 2   | 특수문자 발생빈도 | TLD를 제외한 도메인 이름 내 특수문자의 개수         |
| 3   | 특수문자 발생비율 | TLD를 제외한 도메인 이름 내 특수문자의 비율         |
| 4   | 숫자 발생빈도   | TLD를 제외한 도메인 이름 내 숫자의 개수           |
| 5   | 숫자 발생비율   | TLD를 제외한 도메인 이름 내 숫자의 비율           |
| 6   | 모음 발생빈도   | TLD를 제외한 도메인 이름 내 영어 모음의 개수        |
| 7   | 모음 발생비율   | TLD를 제외한 도메인 이름 내 영어 모음의 비율        |
| 8   | 레이블 최대 길이 | TLD를 제외한 도메인 이름의 레이블 중 가장 긴 길이     |
| 9   | 연속된 숫자 길이 | TLD를 제외한 도메인 이름 내 연속된 숫자 중 가장 긴 길이 |
| 10  | 연속된 문자 길이 | TLD를 제외한 도메인 이름 내 연속된 문자 중 가장 긴 길이 |
| 11  | 평판정보(N=2) | 2-gram에 대한 평판 정보                   |
| 12  | 평판정보(N=3) | 3-gram에 대한 평판 정보                   |
| 13  | 평판정보(N=4) | 4-gram에 대한 평판 정보                   |
| 14  | 평판정보(N=5) | 5-gram에 대한 평판 정보                   |

[0018]

- [0019] 이후, 추출된 피처를 기반으로 입력받은 도메인 이름이 정상 또는 악성인지 탐지할 수 있도록 기계 학습 또는 딥러닝 기반의 탐지 모델에 전달하여 악성 도메인을 분류할 수 있는 학습 모델을 구축한다(S30).
- [0020] 정상 도메인 이름 및 악성 도메인 이름을 학습 데이터로서 학습하는 학습 모델은 다양한 종류의 기계 학습 모델을 포함할 수 있으며, 예를 들면, 장단기 기억 학습 모델(Long Short-Term Memory, LSTM)을 이용할 수도 있다.
- [0021] 장단기 기억 학습 모델(Long Short-Term Memory)이란, 순환 신경망(Recurrent Neural Network)의 변형 모델로서, 순환 신경망의 경우에는 모든 데이터를 순환하며 모든 데이터를 기억하며 학습하는 것이지만, 장단기 기억 학습 모델의 경우에는 순환 신경망과 동일한 과정을 진행하면서도 순환 신경망과는 달리 모든 데이터를 기억하며 학습하는 것이 아닌, 모델 스스로 장기 기억과 단기 기억을 구분하여 필요 없는 기억을 지우는 방법으로 학습을 진행하는 것이다.
- [0022] 구체적으로, 장단기 기억 학습 모델(LSTM)은 여러 개의 게이트(gate)가 붙어있는 셀(cell)로 이루어져 있으며, 셀의 정보를 새로 저장하거나, 셀의 정보를 불러오거나 셀의 정보를 유지하는 기능을 포함하고 있다. 셀은 셀에 연결된 게이트의 값을 보고 무엇을 저장할지, 언제 정보를 내보낼지, 언제 쓰고 언제 지울지를 결정하고, 게이트는 0에서 1사이의 값을 가져 게이트의 값에 비례하여 여러 가지 작동을 한다. 즉, 장단기 기억 학습 모델은 이전 프레임의 신호를 현재 프레임에 반영하여 시간적으로 연속하는 데이터를 처리하는 기법으로서, 셀에 이전의 정보들을 저장하고, 게이트들을 통하여 셀로 흐르는 정보의 양을 게이트 값에 비례하여 여러 가지 작동을 함으로써 조절하는 것이다.
- [0023] 이후, 예측할 도메인 이름(D2)을 입력받고, 입력받은 도메인 이름에서 복수의 피처를 추출한다(S20).
- [0024] 그리고, 추출된 피처를 학습 단계에서 구축된 악성 도메인 분류 모델로 전달하여(S30), 입력받은 도메인 이름이 정상인지 또는 악성인지 예측한다(S40).
- [0025] 이하에서는 표 1 및 도 2를 참조하여 본 명세서의 일실시예에 따른 문자 정보 피처들을 설명한다.
- [0026] 도 2는 본 발명의 일실시예에 따른 문자 정보 피처를 설명하기 위한 참고도이다.
- [0027] 첫번째 피처 정보는 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블의 길이 합을 나타내는 도메인 길이 정보이다. 도 2(a)에 도시된 바와 같이, 도메인 길이 정보는 TLD인 'com'을 제외한 레이블인 'example-domain1' 길이인 15의 값을 나타낸다.
- [0028] 두번째 피처 정보는 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블에서 특수문자의 개수를 나타내는 특수문자 발생빈도 정보이다. 즉, 도 2(b)에 도시된 바와 같이, TLD인 'com'을 제외한 레이블인 'example-domain-2'에서 특수문자인 '\_' 및 '-'가 각각 2번씩 총 4개 발생하였으며 이를 나타낸 4의 값이 특수문자 발생빈도 정보가 된다.
- [0029] 또한, 세번째 피처 정보는 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블에서 특수문자의 비율을 나타내는 특수문자 발생비율 정보이며, 이는 하기 수학적 식 1과 같이 구할 수 있다. 즉, 도 2(b)를 참조하면, 특수문자인 '\_' 및 '-'가 각각 2번씩 총 4번 발생한 빈도인 4를 전체 도메인 길이인 18로 나누면 특수문자 발생비율은 0.2222 값을 나타낸다.

## 수학적 식 1

$$(\text{특수문자 발생비율}) = \frac{(\text{특수문자 발생빈도})}{(\text{도메인 길이})}$$

- [0030]
- [0031] 네번째 피처 정보는 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블에서 숫자의 개수를 나타내는 숫자 발생빈도 정보이다. 즉, 도 2(c)에 도시된 바와 같이, TLD인 'com'을 제외한 레이블인 'example-d0main94'에서 숫자인 '1', '0', '1', '9' 및 '4'가 나타난 횟수인 숫자 발생빈도 5를 확인할 수 있다.
- [0032] 또한, 다섯번째 피처 정보는 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블에서 숫자의 비율을 나타내는 숫자 발생비율 정보이며, 이는 하기 수학적 식 2와 같이 구할 수 있다. 즉, 도 2(c)에 도시된 바와 같이, TLD인 'com'을 제외한 레이블인 'example-d0main94'에서 숫자인 '1', '0', '1', '9' 및 '4'가 나타난 횟수인 숫자

발생빈도 5를 확인할 수 있으며, 숫자 발생빈도 5를 전체 도메인 길이인 16으로 나누면 숫자 발생비율은 0.3125 값을 나타낸다.

## 수학식 2

$$(\text{숫자발생비율}) = \frac{(\text{숫자발생빈도})}{(\text{도메인길이})}$$

[0033]

[0034]

여섯번째 피쳐 정보는 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블에서 영어 모음의 개수를 나타내는 모음 발생빈도 정보이다. 즉, 도 2(d)에 도시된 바와 같이, TLD인 'com'을 제외한 레이블인 'example-domain6'에서 영어모음인 'e', 'a', 'e', 'o', 'a' 및 'i'가 나타난 횟수인 모음 발생빈도 6을 확인할 수 있다.

[0035]

또한, 일곱번째 피쳐 정보는 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블에서 영어 모음의 비율을 나타내는 모음 발생비율 정보이며, 이는 하기 수학식 3과 같이 구할 수 있다. 다. 즉, 도 2(d)에 도시된 바와 같이, TLD인 'com'을 제외한 레이블인 'example-domain6'에서 영어모음인 'e', 'a', 'e', 'o', 'a' 및 'i'가 나타난 횟수인 모음 발생빈도 6을 전체 도메인 길이인 15로 나누면 모음 발생비율은 0.4 값을 나타낸다.

## 수학식 3

$$(\text{모음발생비율}) = \frac{(\text{모음발생빈도})}{(\text{도메인길이})}$$

[0036]

[0037]

여덟번째 피쳐 정보는 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블에서 가장 긴 레이블의 길이를 나타내는 레이블 최대 길이 정보이다. 즉, 도 2(e)에 도시된 바와 같이, TLD인 'com'을 제외한 레이블들인 'exam', 'ple-doma', 'in8' 중에서 가운데 레이블인 'ple-doma'의 레이블이 가장 긴 길이인 8의 값을 가지고 있으며 이를 나타낸 8이 레이블 최대 길이 정보임을 확인할 수 있다.

[0038]

아홉번째 피쳐 정보는 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블에서 연속된 숫자의 최대 길이 정보이다. 즉, 도 2(f)에 도시된 바와 같이, TLD인 'com'을 제외한 레이블인 'examp1274-d0m61n9'에서 연속된 숫자 '1274'의 길이가 가장 긴 길이를 가지고 있으며, 이를 나타내는 길이 값 4가 연속된 숫자의 최대 길이 정보임을 확인할 수 있다.

[0039]

열번째 피쳐 정보는 도메인 이름에서 TLD(Top Level Domain)를 제외한 레이블에서 연속된 문자의 최대 길이 정보이다. 즉, 도 2(g)에 도시된 바와 같이, TLD인 'com'을 제외한 레이블인 'example-d0main10'에서 연속된 문자 'examp'의 길이가 가장 긴 길이를 가지고 있으며, 이를 나타내는 길이 값 5가 연속된 문자의 최대 길이 정보임을 확인할 수 있다.

[0040]

열한번째 내지 열네번째 피쳐 정보는 도메인 이름의 N-gram 비교값으로 판단하는 도메인 이름 평판정보이다.

[0041]

여기서, N-gram이란 입력한 문자열을 N개의 기준 단위로 절단하는 방법이다. 예를 들어 "Here is a dog" 라는 문장을 문자 절단 단위의 3-gram으로 만든다면, "Her", "ere", "re\_", "e\_i", "\_is", "is\_", "s\_a", "\_a\_", "a\_d", "\_do", "dog" 로 만들 수 있다. 같은 작업을 단어 단위의 2-gram으로 만든다면, "Here is", "is a", "a dog" 가 된다. 이와 같이 문자열을 소정 단위로 절단하여 재 구성하면 나오는 빈도를 분석할 수도 있고, 검색어 엔진에서 키워드를 뽑아내는 용도로도 쓰일 수 있으며, 음성 인식에서도 활용할 수 있다.

[0042]

평판정보는 특정 도메인 이름을 정상 도메인 이름과 N-gram에 대해 비교하였을 때, 어떤 차이가 있는지 계산한 피쳐 정보이다. 여기서, 평판정보를 K라고 할 때, N-gram에 대한 평판정보 값은 정상 도메인에서 나타나는 N-gram 개수를  $C_{N-Gram}(i)$ 라고 하면, 다음 수학식 4를 통해서 구할 수 있다.

#### 수학식 4

$$K = \sum_{i=1}^{N\text{-Gram 개수}} \log_2 \left( \frac{C_{N\text{-Gram}}(i)}{N} \right)$$

상기 수학식 4를 사용하기 위해서는 먼저 정상 도메인 이름에 대한 N-gram의 개수를 모두 계산하여야 하며, 예를 들어 N=3일 때 'domain'이라는 도메인 이름의 평판정보를 K(3)이라고 하면 다음과 같이 구해질 수 있다.

$$K(3) = W_{dom} + W_{oma} + W_{mai} + W_{ain} = \log_2 \left( \frac{C_{dom}}{3} \right) + \log_2 \left( \frac{C_{oma}}{3} \right) + \log_2 \left( \frac{C_{mai}}{3} \right) + \log_2 \left( \frac{C_{ain}}{3} \right)$$

본 명세서의 바람직한 일실시예에 따르면, 도메인 이름의 평판정보는 N=2, 3, 4, 5일 때 각각에 대한 평판정보를 모두 피쳐로 사용한다.

본 발명의 다른 일실시예에 따르면, 상기 열번째 내지 열네번째 피쳐 정보는 도메인 이름 내 N-gram의 최소한 정도를 나타내는 N-gram 회귀성 정보가 될 수도 있다. N-gram 회귀성 정보를 R이라고 할 때, 정상 도메인에서 나타나는 N-gram 개수를  $C_{N\text{-Gram}}(i)$ 라고 하면, 다음 수학식 5와 같이 정의될 수 있다.

#### 수학식 5

$$W_{N\text{-Gram}}(i) = \log_2 \left( \frac{C_{N\text{-Gram}}(i)}{N} \right)$$

$$R = \frac{\sum_i^{N\text{-Gram 개수}} W_{N\text{-Gram}}(i)}{N\text{-Gram 개수}}$$

수학식 5는, 해당 N-gram이 전체 N-gram에서 나올 확률을 의미하며, 확률이 가장 큰 경우와 가장 작은 경우의 값 크기 차이가 크기 때문에 log를 취해 크기를 보정한다. 확률에 log를 취하면 모든 값이 음수가 나오므로 마이너스를 취해 양수로 만들며, 이 값이 해당 N-gram의 가중치(weight)가 된다. 가중치의 크기가 클수록 해당 N-gram이 나올 확률은 작은 것을 의미하며, 크기가 작을수록 N-gram이 나올 확률은 큰 것을 의미한다. 또한, N-gram 회귀성은 해당 도메인 이름이 가진 모든 N-gram의 가중치의 평균을 의미하며, 정상 도메인 이름의 N-gram 분포와 확연한 차이를 보이는 N-gram으로 이루어진 악성 도메인 이름을 보다 높은 확률로 탐지할 수 있게 된다.

한편, 악성 도메인을 탐지하여 분류하는 모델은 특정 도메인 이름 목록에 대해 학습을 수행한 후, 학습한 내용을 기반으로 전달된 도메인 이름이 정상인지 악성인지 예측을 수행한다. 그런데, 만약 학습 단계에서는 없었던 새로운 패턴의 악성 도메인 이름이 전달될 경우에는 해당 악성 도메인을 제대로 판별하지 못 할 가능성이 존재한다. 본 발명에서는 이와 같은 문제점을 해결하기 위해서 새로운 패턴의 도메인 이름을 사전에 생성할 수 있다.

이하에서는 도 3 및 도 4를 참조하여, 기존에 없었던 도메인 이름을 생성하는 실시예를 설명하며, 이를 통해서 잠재적 악성 도메인의 예측률을 더욱 높일 수 있는 방안을 제시한다.

도 3은 본 발명의 일실시예에 따른 도메인 이름 생성 모델의 구조를 도시한 도면이다.

도메인 이름 생성 모델에는 문자수준 텍스트 생성(character-level text generation)을 적용한다. 문자수준 텍스트 생성을 이용하면 도메인 이름을 한꺼번에 생성하지 않고 어떤 문자열이 있을 때 해당 문자열 바로 다음에 올 문자를 하나씩 예측하여 도메인 이름을 생성한다. 이때, 생성 모델에서 점 "."을 포함한 각 문자는 특정한 실수 벡터로 매핑되어 계산에 이용된다.

또한, 생성 모델은 도 3에 도시된 바와 같이, 장단기 기억 학습 모델(Long Short-Term Memory, LSTM)을 사용한

다. LSTM은 문자열과 같은 연속적인 데이터를 다루는데 주로 사용되며 이전 상태를 기억하는 성질이 있다. 즉, 이전의 입력 값을 바탕으로 현재의 출력 값을 학습한다. 이를 문자열에 적용하면 문자의 패턴을 학습할 수 있으며, 이러한 성질을 이용해 도메인 이름의 패턴을 학습시킨 뒤 도메인 이름을 생성하도록 만든다.

[0055] 도시된 바와 같이, LSTM 셀이 출력한 값 중 하나는 다음 LSTM 셀로 직접 전달하며, 다른 하나는 완전연결 계층(fully-connected layer)로 전달한다. 완전연결 계층은 LSTM으로부터 전달 받은 값이 어떤 실수 벡터에 해당하는지 학습한다. 완전연결 계층에 연결된 소프트 맥스 함수는 해당 실수 벡터와 가장 가까운 문자 벡터를 계산하여 어떤 문자에 해당하는 벡터인지 예측한다. 예측된 문자 벡터는 다음 LSTM 셀에 전달하여 도메인 이름 생성 모델이 계속해서 문자를 예측하고 생성할 수 있도록 한다.

[0056] 도 4는 도메인 이름 생성 모델의 입력과 출력을 나타낸 참고도이다.

[0057] 도시된 바와 같이, 도메인 이름을 입력하는 경우 도메인 이름을 하나씩 생성 모델에 전달하는 것이 아니라 도메인 이름 리스트를 하나의 문자열로 전달하며, 생성 모델 역시 도메인 이름을 하나씩 생성하지 않고 도메인 이름 리스트를 문자열로 출력한다. 이때 입력 도메인 이름과 출력 도메인 이름 내의 각 도메인 이름은 개행문자 ' '로 구분될 수 있다.

[0058] 도메인 생성 모델은 학습할 도메인 이름 데이터를 받아 해당 데이터의 문자열 시퀀스(character sequence)의 패턴을 학습한다. 학습 후 도메인 생성 모델은 입력 받은 데이터에 존재하지 않는 새로운 도메인 이름을 생성할 수 있다. 생성된 도메인 이름 중에는 이미 등록되어 있는 도메인도 있으며, 도메인 이름 규칙에 어긋나는 유효하지 않은 도메인도 존재한다. 유효하지 않은 도메인은 제외하기 위해 생성 모델 내에서 자체적으로 필터링 규칙을 적용한다.

[0059] 표 2는 유효하지 않은 도메인 이름 규칙을 예시적으로 나타낸 표이다.

표 2

| No | 유효하지 않은 도메인 이름 규칙                        | 예시                                                                           |
|----|------------------------------------------|------------------------------------------------------------------------------|
| 1  | 레이블이 하나밖에 존재하지 않는 도메인 이름                 | example-domain                                                               |
| 2  | 역 슬래시 기호(₩)가 존재하는 도메인 이름                 | example₩domain.com                                                           |
| 3  | 점(.)이 연속으로 오는 도메인 이름                     | example-domain..com                                                          |
| 4  | TLD의 길이가 2 미만인 도메인 이름                    | example-domain.c                                                             |
| 5  | 특수문자('-', '_')로 시작하거나 끝나는 레이블이 있는 도메인 이름 | _example-domain.com<br>example-domain-.com                                   |
| 6  | 길이가 63을 초과하는 레이블이 있는 도메인 이름              | abcdefghijklmnopqrstuvwxyz<br>abcdefghijklmnopqrstuvwxyz0<br>12345678901.com |
| 7  | 점(.)을 제외하고 총 길이가 3 미만인 도메인 이름            | co                                                                           |

[0060]

[0061] 표 2를 참조하면, 레이블이 하나밖에 존재하지 않는 도메인 이름은 도메인 이름 규칙에서 벗어나는 도메인이 생성된 것으로 판단하여 필터링된다.

[0062] 즉, 도메인 이름 생성 모델은 TLD를 포함한 도메인 이름 목록을 전달받아 해당 도메인 이름들의 패턴을 학습한다. 그럼에도 불구하고 Top Level Domain, Second Level Domain 등 없이 레이블이 하나밖에 존재하지 않는 도메인 이름이 생성되는 경우가 있어 이를 제외하도록 하였다.

[0063] 또한, 역 슬래시 기호(₩)가 존재하는 도메인 이름은 도메인 이름 규칙에서 벗어나는 도메인이 생성된 것으로 판단하여 필터링된다. 도메인 이름 생성 모델은 도메인 이름을 하나씩 전달받지 않고 여러 도메인 이름을 한꺼번에 전달받으며, 각 도메인 이름은 개행 문자(

)로 구분된다. 즉, 생성 모델이 개행 문자까지 학습하며 완전하지 않은 제어 문자(역 슬래시 기호를 포함)가 생성되는 경우가 있어 이를 제외하도록 하였다.

[0064] 또한, 점(.)이 연속으로 오는 도메인 이름은 도메인 이름 규칙에서 벗어나는 도메인이 생성된 것으로 판단하여 필터링된다. 즉, 도메인 이름 규칙상 점이 연속으로 올 수 없기 때문에 점이 연속으로 오는 도메인 이름은 제외하도록 한다.

[0065] 또한, TLD의 길이가 2미만인 도메인 이름은 도메인 이름 규칙에서 벗어나는 도메인이 생성된 것으로 판단하여 필터링된다. 즉, 오늘날 모든 TLD의 길이는 최소 2이상이기 때문에 TLD 길이가 2미만인 도메인 이름은 제외하도록 하였다.

[0066] 또한, 특수문자('-', '\_')로 시작하거나 끝나는 레이블이 있는 도메인 이름은 도메인 이름 규칙에서 벗어나는 도메인이 생성된 것으로 판단하여 필터링된다. 즉, 도메인 이름 규칙상 하이픈(-)과 언더바(\_)는 레이블의 맨 앞이나 맨 끝에 위치할 수 없어 이에 해당하는 도메인 이름은 제외하도록 하였다.

[0067] 또한, 길이가 63을 초과하는 레이블이 있는 도메인 이름은 도메인 이름 규칙에서 벗어나는 도메인이 생성된 것으로 판단하여 필터링된다. 즉, 도메인 이름 규칙상 한 레이블의 길이가 63을 초과할 수 없어 길이가 63을 초과하는 레이블을 포함하는 도메인 이름은 제외하도록 하였다.

[0068] 또한, 점(.)을 제외하고 총 길이가 3미만인 도메인 이름은 도메인 이름 규칙에서 벗어나는 도메인이 생성된 것으로 판단하여 필터링된다. 즉, 오늘날 사용되는 도메인 이름의 길이는 최소 3이므로 이보다 짧은 길이의 도메인 이름은 제외하였다.

[0069] 한편, 도메인 이름을 다루는 기존의 모델들은 고정된 길이(최대 길이)를 정한 뒤 도메인 이름이 그 길이에 못 미칠 경우 도메인 이름 뒤에 0과 같은 문자를 더해 길이를 일정하게 맞추는 패딩(padding)을 적용하였다. 그러나, 도메인 이름의 최대 길이가 길수록 0패딩 역시 길어질 수 있는데 0패딩이 길어질수록 이것을 다루는 기계학습이 될 중요한 부분(패딩된 부분)을 학습하는데 시간을 소요하므로 학습시간 관점에서 효율적이지 못한 단점이 있다. 따라서, 본 발명의 일실시예에 따르면, 도메인 이름 생성 모델에서는 도 6(a)에 도시된 바와 같은 패딩을 적용하지 않는 대신, 도 6(b)에 도시된 바와 같이 개행 문자(

)가 포함된 문자열을 입력 및 출력으로써 다룰 수 있다. 즉, 0과 같은 문자 패딩 없이 개행 문자가 포함된 문자열을 그대로 학습에 사용한다면 학습시간 관점에서 보다 효율적인 모델을 구현할 수 있다.

[0070] 도 5는 본 발명의 일실시예에 따른 악성 도메인 탐지 장치를 개략적으로 도시한 블록도이다.

[0071] 도시된 바와 같이, 악성 도메인 탐지 장치는 도메인 데이터 입력부(10), 피쳐 추출부(20), 기계학습부(30), 도메인 생성부(40) 및 악성 도메인 탐지부(50)를 포함하여 구성된다.

[0072] 입력부(10)는 학습할 도메인 이름 및 예측할 도메인 이름을 입력받는다.

[0073] 피쳐 추출부(20)는 입력된 도메인 이름에서 복수의 피쳐를 추출한다.

[0074] 기계학습부(30)는 피쳐 추출부(20)에서 추출된 피쳐를 기반으로 입력받은 학습 도메인 이름이 정상 또는 악성인지 탐지할 수 있도록 학습하여 악성 도메인 분류 모델을 구축한다.

[0075] 도메인 생성부(40)는 입력부(10)로부터 학습할 도메인 이름 리스트를 문자열로 전달받아, LSTM(Long Short-Term Memory) 모델을 사용하여 문자 패턴을 학습하고, 도메인 이름을 생성한다.

[0076] 악성 도메인 탐지부(50)는 피쳐 추출부(20)에서 추출된 피쳐를 기반으로 입력받은 예측 도메인 이름이 정상 또는 악성인지 탐지한다. 이때, 악성 도메인 탐지부(50)는 입력부(10)를 통해서 예측할 도메인 이름이 입력되면, 피쳐 추출부(20)를 통해서 복수의 피쳐를 추출하고, 상기 추출된 복수의 피쳐를 전달받아서 상기 학습된 악성 도메인 분류 모델을 이용하여 상기 예측 도메인 이름이 정상 또는 악성인지 탐지할 수 있다.

[0077] 이상에서 설명한 본 발명의 실시예를 구성하는 모든 구성요소들이 하나로 결합하거나 결합하여 동작하는 것으로 기재되어 있다고 해서, 본 발명이 반드시 이러한 실시예에 한정되는 것은 아니다. 즉, 본 발명의 목적 범위 안에서라면, 그 모든 구성요소들이 하나 이상으로 선택적으로 결합하여 동작할 수도 있다. 또한, 그 모든 구성요소들이 각각 하나의 독립적인 하드웨어로 구현될 수 있지만, 각 구성요소들의 그 일부 또는 전부가 선택적으로 조합되어 하나 또는 복수개의 하드웨어에서 조합된 일부 또는 전부의 기능을 수행하는 프로그램 모듈을 갖는 컴퓨터 프로그램으로서 구현될 수도 있다. 또한, 이와 같은 컴퓨터 프로그램은 USB 메모리, CD 디스크, 플래쉬 메

모리 등과 같은 컴퓨터가 읽을 수 있는 기록매체(Computer Readable Media)에 저장되어 컴퓨터에 의하여 읽혀지고 실행됨으로써, 본 발명의 실시예를 구현할 수 있다. 컴퓨터 프로그램의 기록매체로서는 자기 기록매체, 광 기록매체, 캐리어 웨이브 매체 등이 포함될 수 있다.

[0078] 또한, 기술적이거나 과학적인 용어를 포함한 모든 용어들은, 상세한 설명에서 다르게 정의되지 않는 한, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 갖는다. 사전에 정의된 용어와 같이 일반적으로 사용되는 용어들은 관련 기술의 문맥상의 의미와 일치하는 것으로 해석되어야 하며, 본 발명에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

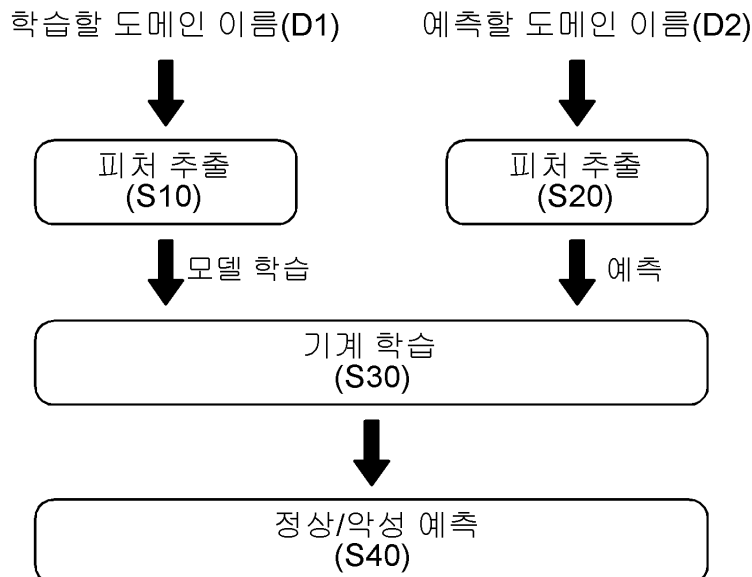
[0079] 이상의 설명은 본 발명의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 본질적인 특성에서 벗어나지 않는 범위 내에서 다양한 수정, 변경 및 치환이 가능할 것이다. 따라서, 본 발명에 개시된 실시예 및 첨부된 도면들은 본 발명의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예 및 첨부된 도면에 의하여 본 발명의 기술 사상의 범위가 한정되는 것은 아니다. 본 발명의 보호 범위는 아래의 청구 범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 권리 범위에 포함되는 것으로 해석되어야 할 것이다.

### 부호의 설명

[0080] 10: 입력부  
20: 피처 추출부  
30: 기계 학습부  
40: 도메인 생성부  
50: 악성 도메인 탐지부

### 도면

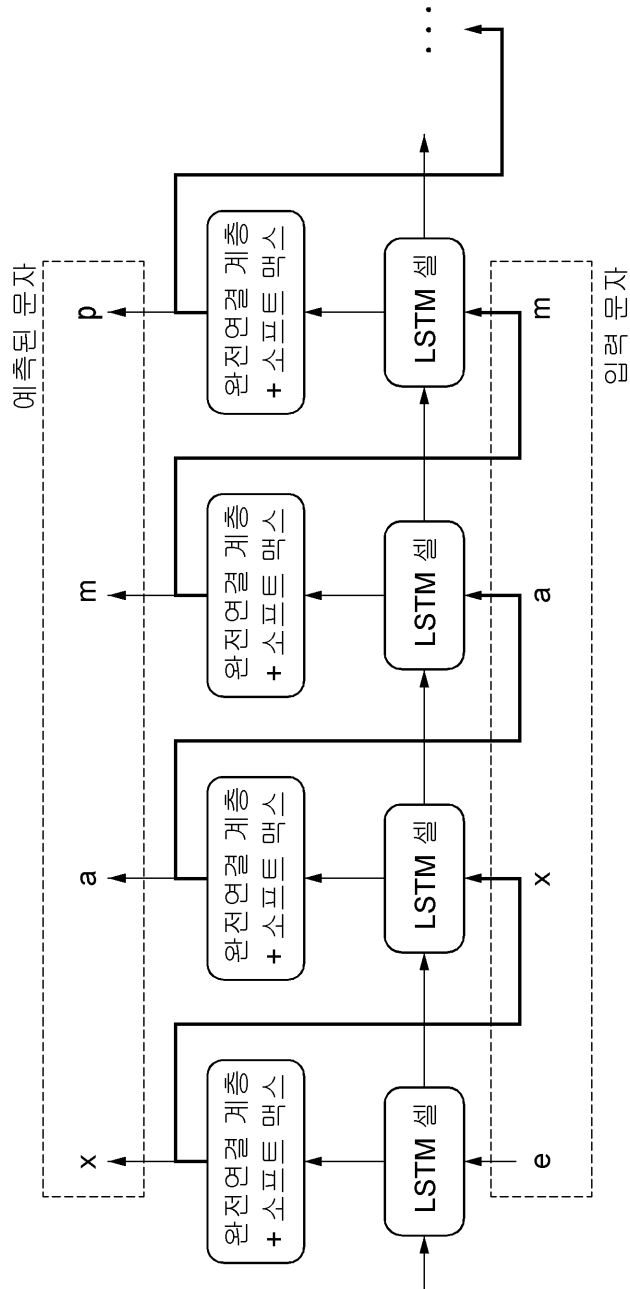
#### 도면1



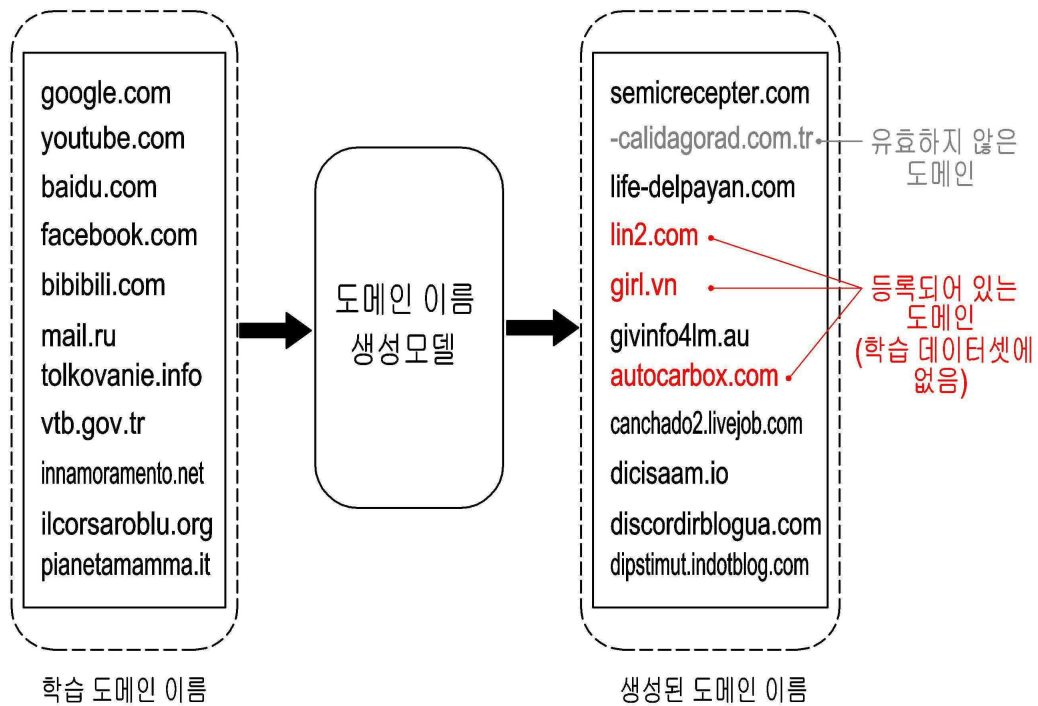
도면2

- (a) example-domain1.com  
 도메인 길이 = 15 TLD
- (b) ex\_ample-do-main-2.com  
 특수문자 발생빈도 = 4 TLD  
 특수문자 발생비율 =  $4 / 18 = 0.2222$
- (c) example-d0ma1n94.com  
 숫자 발생빈도 = 5 TLD  
 숫자 발생비율 =  $5 / 16 = 0.3125$
- (d) example-domain6.com  
 모음 발생빈도 = 6 TLD  
 모음 발생비율 =  $6 / 15 = 0.4$
- (e) exam.p1e-doma.in8.com  
 레이블 길이 = 4 레이블 길이 = 8 레이블 길이 = 3 TLD  
 레이블 최대 길이 = 8
- (f) examp1274-d0m61n9.com  
 연속된 숫자의 최대 길이 = 4 TLD
- (g) examp1e-d0main10.com  
 연속된 문자의 최대 길이 = 5 TLD

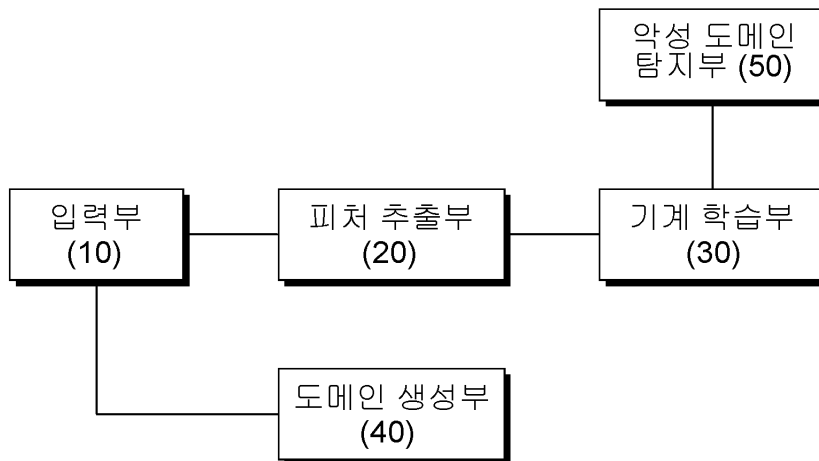
도면3



도면4



도면5



도면6

|   |   |   |   |   |   |   |   |    |   |    |    |
|---|---|---|---|---|---|---|---|----|---|----|----|
| g | o | o | g | i | e | . | c | o  | m | r  | io |
| y | o | u | t | u | b | e | . | c  | o | m  | io |
| b | a | i | d | u | . | c | o | m  | r | in |    |
| f | a | c | e | b | o | o | k | .  | c | o  | m  |
| b | i | i | i | b | i | i | i | .  | c | o  | m  |
| m | a | i | i | . | r | u | r | in |   |    |    |

(b)

|   |   |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|---|
| g | o | o | g | l | e | . | c | o | m | o | o |
| y | o | u | t | u | b | e | . | c | o | m | o |
| b | a | i | d | u | . | c | o | m | o | o | o |
| f | a | c | e | b | o | o | k | . | c | o | m |
| b | i | l | i | b | i | l | i | . | c | o | m |
| m | a | i | l | . | r | u | o | o | o | o | o |

(a)