신경망을 이용한 유연한 프로액티브 패스워드 체킹 방법

박신혜[0] 김원일 김동규
아주대학교 정보 통신 전문 대학원
(sinne[0], wikim, dkkim)@ajou.ac.kr

The Flexible Proactive Password Checking Methods using Neural Network

Sihn-hye Park[0] Wonil Kim Dong-Kyoo Kim
Dept. of Information Communication Engineering GSIC AJOU

요 약

다중 사용자 환경에서 컴퓨터 시스템 보안을 위한 사용자 인증(user authentication)은 패스워드(password), 토큰(token), 스마트 카드(smart card), 지문(fingerprint), 음성(voiceprint) 등 다양한 정보들을 통하여 시스템에 접근하는 사용자의 신원을 확인하고, 인증되지 않은 사용자의 접근을 제한한다. 이들 중 가장 보편적으로 사용되는 패스워드 기반 사용자 인증은 구현이 쉽고, 관리 비용이 적게 든다는 장점이 있다. 패스워드 기반 사용자 인증에서 패스워드의 선택은 시스템의 보안을 위하여 매우 중요하다. 따라서 시스템 관리 차원에서 사용자의 패스워드를 검사할 필요가 있다. 본 논문에서는 패스워드를 추측하기 쉬운 패스워드와 추측하기 어려운 패스워드로 분류하는 근거가 되는 여러 가지 패스워드들에 대한 특징들 중, 패스워드에 대한 언어적인 정보를 구별할 수 있는 특징을 제안한다. 또한 이를 신경망(neural network)을 사용하여 구현함으로써 보안 시스템의 특성에 따라 패스워드의 적합성 여부를 유연하게 조정할 수 있는 프로액티브 패스워드 체킹 방법을 제안한다.

1. 서 론

다중 사용자 환경에서 컴퓨터 시스템 보안을 위한 사용자 인증(user authentication)은 패스워드(password), 토큰(token), 스마트 카드(smart card), 지문(fingerprint), 음성(voiceprint) 등 다양한 정보들을 통하여 시스템에 접근하는 사용자의 신원을 확인하고, 인증되지 않은 사용자의 접근을 제한한다[1][3]. 이들 중 패스워드 기반 사용자 인증은 가장 보편적으로 사용되는 인증 방법으로, 구현이 쉽고 관리 비용이 적게 든다는 이점이 있다[1][2].

패스워드 기반 사용자 인증에 있어서 패스워드의 선택은 컴퓨터 시스템의 보안을 위하여 매우 중요하다. 일반적으로 사용자는 추측하기 쉬운 패스워드(전화번호, 생일, 가족 또는 친구 이름, 일반 단어 등)를 선택하는 경향이 있다. 이와 같은 사용자의 특성으로 패스워드 기반 사용자 인증을 사용하는 시스템은 패스워드 크래킹(password cracking)에 의해 공격 받을 수 있다는 취약성을 가지고 있다. 따라서 시스템 관리 차원에서 사용자의 패스워드를 검사할 필요가 있다[4].

시스템 측면에서 패스워드를 체크하는 방법에는 리액티브 패스워드 체킹 방법(reactive password checking)과 프로액티브 패스워드 체킹 방법(proactive password checking)이 있다[4]. 리액티브 패스워드 체킹은 시스템 관리자가 Crack[5]과 같은 프로그램을 주기적으로 실행하여 추측하기 쉬운 패스워드를 찾는 방법이다. 패스워드를 체크한 후 추측하기 쉬운 패스워드가 발견되면 사용자에게 패스워드의 수정을 요구한다. 이 방법은 사용자가 패스워드 수정 요구를 받는 즉시 패스워드를 재설정 하지 않으면, 해당 시스템은 취약한 패스워드가 수정되기 전까지 취약성을 가지게 된다. 이러한 문제점을 해결할 수 있는 방법으로 프로액티브 패스워드 체킹 방법이 있다. 프로액티브 패스워드 체킹 방법은 패스워드의 설정 또는 수정 시에 사용자가 설정하고자 하는 패스워드가 적합한 지의 여부를 검사하는 방법이다. 사용자가 패스워드를 설정 또는 수정하고자 할 때, 패스워드를 체크한 후 설정하고자 하는 패스워드가 추측하기 쉬운 패스워드임이 판명되면 사용자는 패스워드를 수정해야 한다. 사용자는 패스워드가 추측하기 어려운 패스워드로 판명될 때까지 시스템에 패스워드를 설정할 수 없다. 프로액티브 패스워드 체킹은 미리 선택되어진 패스워드의 특징에 따라 시스템에 대한 패스워드 설정 가능 여부가 정해진다.

본 논문에서는 패스워드를 추측하기 쉬운 패스워드와 추측하기 어려운 패스워드로 분류함에 있어서 패스워드에 대한 언어적인 정보를 구별할 수 있는 특징을 제안한다. 또한 이를 신경망(neural network)을 사용하여 구현함으로써 보안 시스템의 특성에 따라 패스워드 설정 가능 여부를 유연하게 조정할 수 있는 프로액티브 패스워드 체킹 방법을 제안한다.

2. 기존의 프로액티브 패스워드 체킹 방법

프로액티브 패스워드 체커(proactive password checker)는 간단한 프로그램으로 구현되어 질 수 있다[8]. 기본적으로 체커(checker)는 추측하기 쉬운 패스워드 목록을 가지고 있고, 사용자가 설정하고자 하는 패스워드가 패스워드 목록에 포함되어 있는 지를 체크한다. 패스워드가 목록에 포함되어 있으면, 패스워드는 시스템에 설정되어 질 수 없다[4][8]. 이 방법은 시스템에

패스워드 목록을 저장하고, 이를 검색하기 위해 많은 시간과 공간이 필요하므로, 일반적으로 사용하기에 적합하지 않다. 따라서 이러한 문제들을 해결하기 위하여 많은 연구가 이루어졌다. 예를 들어 Spafford는 OPUS 시스템에서 Bloom filters를 사용하였고, Davies와 Gaunesan은 BApasswd 시스템에서 Markov 모델과 삼중자(trigrams)를 사용하였다. 그리고 Francesco는 decision tree를 이용하여 ProCheck를 구현하였고, Carlo는 뉴럴 네트워크를 이용하는 새로운 방법을 제안하였다[4][7]. 이 중 Carlo가 제안한 뉴럴 네트워크를 이용한 프로액티브 패스워드 체킹 방법은 선형 분리 공간(linearly separable space) 상에서 패스워드를 분류하였다[7][8]. 또한 패스워드의 분류를 위한 4가지 특징, 즉, 클래스, 특수 문자의 수, 대소문자의 차, 다이그램(digrams)을 제안하였다[8]. 이 방법은 같은 문자들로 구성되었으나 배열이 다른 단어 또는 문장을 구분할 수 없다. 예를 들어 직관적으로, 사용자는 " Iloveyou" 문장은 추측하기 쉬운 패스워드로 " yIvleoo" 은 추측하기 어려운 패스워드로 판단할 수 있으나, 이 시스템은 이 둘을 이와 같이 분류할 수 없다.

3. 제안된 프로액티브 패스워드 체킹 방법

본 논문에서는 패스워드를 추측하기 쉬운 패스워드와 추측하기 어려운 패스워드로 분류함에 있어서 패스워드에 대한 언어적인 정보를 구별할 수 있는 특징을 제안한다. 또한 이를 신경망(neural network)를 사용하여 구현함으로써 보안 시스템의 특성에 따라 패스워드의 설정 가능 여부를 유연하게 조정할 수 있는 프로액티브 패스워드 체킹 방법을 제안한다.

3.1 패스워드의 5가지 특징과 11가지 요소

표 1은 패스워드를 추측하기 쉬운 패스워드와 추측하기 어려운 패스워드로 분류하기 위하여 제안한 패스워드의 5가지 특징들이다.

표 1. 패스워드의 5가지 특징과 11가지 요소

5 특징	11 요소
Configuration	모음의 개수 : n_v
	자음의 개수 : n_c
	숫자의 개수 : n_d
	특수 문자의 개수 : n_s
Difference	대문자의 개수와 소문자의 개수 차이 : n_{uld}
Adjacency	인접하는 문자의 개수 : n_{ac}
	인접하는 숫자의 개수 : n_{ad}
Repetition	반복하는 문자의 개수 : n_{rc}
	반복하는 숫자의 개수 : n_{rd}
Alternation	모음 문자와 자음 문자의 교대의 개수 : n_{a1}
	문자, 숫자, 특수 문자의 교대의 개수 : n_{a2}

위와 같은 5가지 특징들 중 Configuration, Difference, Adjacency, Repetition은 패스워드에 대한 문자들의 구성에 관한 정보들을 통하여 얻을 수 있는 반면, Alternation은 언어적인 정보들을 통하여 얻을 수 있다. 이 특징들은 표 2와 같이 11가지 요소들로 나뉘어 진다. 이 11가지 요소들 중 모음에서 자음으로, 자음에서 모음으로 바뀌는 회수는 패스워드 분류의 언어적 요소로 매우 중요하다. 일반적으로 사전에 있는 단어와 문장들은 모음과 자음의 적절한 배합으로 이루어져 있기 때문이다. 본 논문에서 제안된 시스템은 이러한 언어적인 정보를 사용하여 패스워드를 추측하기 쉬운 패스워드와 추측하기 어려운 패스워드를 분류한다.

3.2 네트워크 구조

본 논문에서 제안된 시스템은 다중 계층 퍼셉트론(perceptrons)을 사용한다. 다중 계층 퍼셉트론은 입력 계층(input layer), 출력 계층(output layer), 숨김 계층(hidden layer)으로 이루어진 지각 단위(sensory unit)로 구성된 신경망(neural network)이다[9]. 표 2는 제안된 시스템의 입력 계층을 구성하는 11개의 입력 노드이다. 입력 노드의 값은 각각 패스워드의 길이로 나뉘어진다. 출력 계층은 S자형 함수(sigmoid function)를 사용하는 결정 노드(decision node)이다. 출력 노드의 값은 0과 1사이가 되며, 추측하기 쉬운 패스워드 일수록 1에, 추측하기 어려운 패스워드 일수록 0에 가까운 값을 얻는다. 숨김 계층은 30개의 숨김 노드들로 구성된다.

표 2. 11가지 입력 노드 요소 값 (n_p : 패스워드 길이)

특징	요소	입력 노드에 대한 요소 값
Configuration	n_v	n_v / n_p
	n_c	n_c / n_p
	n_d	$1 - n_d / n_p$
	n_s	$1 - n_s / n_p$
Difference	n_{uld}	n_{uld} / n_p
Adjacency	n_{ac}	n_{ac} / n_p
	n_{ad}	n_{ad} / n_p
Repetition	n_{rc}	n_{rc} / n_p
	n_{rd}	n_{rd} / n_p
Alternation	n_{a1}	n_{a1} / n_p
	n_{a2}	$1 - n_{a2} / n_p$

3.3 트레이닝과 테스팅

본 논문에서 제안된 뉴럴 네트워크는 시스템이 추측하기 쉬운 패스워드와 추측하기 어려운 패스워드의 차이점을 배우도록 트레이닝된다. 트레이닝 데이터는 패스워드 체킹 라이브러리(CrackLib v2.7)[5]에서 임의로 뽑은 10,000개의 추측하기 쉬운 단어와 임의적으로 생성한 10,000개의 추측하기 어려운 단어이고, 테스팅 데이터는 트레이닝 데이터에 속하지 않는 추측하기 쉬운 5,000개의 단어와 추측하기 어려운 5,000개의 단어이다. 트레이닝은 2,500번 반복하였고, 에러율은 0.003 이하이다.

3.4 결과

제안된 시스템은 99.9% 성공률로 단어들을 분류하였다. 이 결과는 시스템이 추측하기 쉬운 패스워드와 추측하기 어려운 패스워드를 효과적으로 구분한다는 것을 보여준다. 특히 이 시스템은 패스워드의 언어적인 정보를 구분한다. 예를 들면, 아래의 표 3

과 같이 " Iloveyou" 와 " yIvleoou" 를 구분할 수 있다. 표 3은 배열이 다른 문자 모음에 대한 테스트 결과로, 단어를 구성하고 있는 문자들의 배열에 따라 결과값이 다르게 나타난다. 표 4, 5, 6은 추측하기 쉬운 패스워드와 추측하기 어려운 패스워드에 대한 테스트 결과로 추측하기 쉬운 패스워드 일수록 1에, 추측하기 어려운 패스워드 일수록 0에 가까운 값임을 볼 수 있다. 즉, 추측하기 쉬운 패스워드와 추측하기 어려운 패스워드의 경계점(threshold)을 어떤 값으로 설정하느냐에 따라 패스워드의 선택 범위가 달라진다. 따라서 보안 시스템의 특성에 따라 패스워드의 적합성 여부를 유연하게 조정할 수 있다.

표 3. 배열이 다른 문자 모음에 대한 테스트 결과

문자모음	패스워드	결과	패스워드	결과
{g, o, o, d, j, o, b, !}	goodjob!	0.986738	gbjd!ooo	0.184618
{I, l, o, v, e, y, o, u}	Iloveyou	0.958713	yIvleoou	0.392575
{K, y, u, n, g, g, i, d, o}	Kyunggido	0.873008	nggKyduio	0.320702
{P, a, l, d, a, l, g, u}	Paldalgu	0.887745	aauPlldg	0.301937
{W, o, n, c, h, u, n}	Wonchun	0.755111	nchWuno	0.235934

표 4. 추측하기 쉬운 패스워드에 대한 테스트 결과

Passwords	Output
paperers	0.994223
papering	0.993600
papillae	0.990678
papistry	0.976735
papooses	0.994774

표 5. 문자, 숫자, 특수 문자의 조합으로 이루어진 추측하기 어려운 패스워드에 대한 테스트 결과

Passwords	Output
%Ri9S7D_	0.000059
ga#cU{3a	0.000861
<nq8tB>9	0.000044
$sPg^ Jul	0.000349
XaC;"E37	0.000990

표 6. 문자로만 이루어진 추측하기 어려운 패스워드에 대한 테스트 결과

Passwords	Output
nqBHGKmM	0.000372
OhMCkrud	0.005539
qfmnWdgZ	0.005969
EymjKrXI	0.000194
lKITVCeY	0.166722

4. 결 론

본 논문에서는 프로액티브 패스워드 체킹에 대하여 추측하기 쉬운 패스워드와 추측하기 어려운 패스워드를 분류하는 근거가 되는 여러 가지 특징들 중, 단어 또는 문장을 언어적으로 인식할 수 있는 특징을 제안하였다. 또한 이를 다중 계층 퍼셉트론을 사용하는 신경망으로 구현함으로써, 다양한 보안 시스템 특성에 따라 시스템에 대한 패스워드의 적합성 여부를 유연하게 조정할 수 있는 방법을 제안하였다.

향후 본 논문에서 제안된 패스워드의 언어적 정보에 관한 특징을 확장하고, 이에 대한 타당성을 검증할 필요가 있다. 또한 패스워드의 각 특징들이 패스워드 분류에 기여하는 정도를 연구하여, 이 특징들에 적절한 가중치를 부여하는 방법을 구현할 예정이다.

5. 참고 문헌

[1] Ravi Sandhu, Pierangela Samarati, " Authentication, Access Control, and Audit" , ACM Computing Surveys, Vol.28, No.1, March 1996.

[2] Fabian Monrose, Michael K. Reiter, Susanne Wetzel, " Password Hardening Based on Keystroke Dynamics, CCS' 99 ACM 1999.

[3] Charles P.Pfleeger, " Security Computing" , Prentice-Hall International, Inc., 1997.

[4] Jeff Yan, " A Note on Proactive Password Checking" , ACM New Security Paradigms Workshop, New Mexico, USA, September 2001.

[5] http://www.users.dircon.co.uk/~crypto/

[6] M.Bishop, " Improving System Security via Proactive Password Checking" , Computers and Security, Vol. 14, No. 3, pp. 233-249, 1995.

[7] F.Bergadano, B.Crispo, and G.Ruffo, " High Dictionary Compression for Proactive Password Checking" , ACM Transactions on Information and System Security, Vol. 1, No. 1, pp. 3-25, November 1998.

[8] C.Blundo, P.D' Arco, A.De Santis, C.Galdi, " A Novel Approach to Proactive Password Checking" , INFRASEC 2002, Bristol (UK), October 1-3.

[9] Simon Haykin, " Neural Networks A Comprehensive Foundation" , 1997.