디지털 포렌식 관점에서 패스워드 복구를 위한 사전 파일 구축 방안 연구

*임종민 *권혁돈 *최재민 *이상진

고려대학교 정보보호기술연구센터

*agezero@korea.ac.kr

A Study of Construct Dictionary File for Password Recovery in Digital Forensics Investigation

*Lim, Jong-Min *Kwon, Hyuk-Don *Choi, Jaemin *Lee, Sangjin

Center for Information Security Technologies(CIST), Korea University

요약

기술이 발전함에 따라 컴퓨터 범죄는 점차 증가하고 있으며, 용의자는 사건의 증거가 될 수 있는 파일들에 대해 패스워드 기능을 제공하는 응용프로그램을 활용하여 증거물에 대해 의도적인 접근을 막고 있다. 이로 인해 수사관은 암호화된 파일들에 대해 접근이 매우 어려운 상황이며, 해결 방안으로써 패스워드 복구 프로그램이 대안이 될 수 있다. 하지만 대다수의 패스워드 복구 프로그램들은 단순한 전수조사 공격 방식을 지원하거나 국가별 특징을 고려하지 않은 영문용 사전파일을 적용하여 복구하고 있기 때문에, 국내수사 환경에서 패스워드 검색에 한계가 따르고 있다. 따라서 수사관이 암호화된 파일에 대해 효율적으로 검색할 수 있는 방안이 필요하며, 이를 통해 빠른 시간 내에 증거물을 복구할 수 있는 방안이 강구되어야 한다. 본 논문에서는 최근 국내외 사전구축 사례 및 동향을 조사함으로써 효율적인 패스워드 사전 파일을 구축할 수 있는 방안을 제시하며, 이와 함께 용의자의 개인적인 정보를 이용하여 최적화된 사전파일을 생성할 수 있는 방안에 대해 설명한다.

1. 서론

현재 컴퓨터의 기술이 발전되고 컴퓨터의 보급이 늘어남에 따라 인터넷을 통한 여러 가지 범죄 기술이 널리 퍼지고 컴퓨터를 이용한 범죄 역시 증가하고 있는 추세이다[1]. 또한 범죄자는 자신의 증거가 될 수 있는 파일들에 대해 손쉽게 접근 제한을 할 수 있다. 대표적인 접근 제한 방법으로써 '패스워드 기반 암호 시스템(Password Based Encryption System)'을 사용할 수 있다. 패스워드 기반 암호 시스템을 이용하는 방법에는 여러 가지 방법이 있으며, 간단한 방법으로는 문서 편집 응용프로그램에서 지원하는 문서 파일의 암호화 방식과 파일 압축 응용프로그램에서 지원하는 파일의 암호화 방식이 있다. 이러한 방법으로 용의자가 임의로 주요 증거 파일에 패스워드 기반의 암호시스템을 사용하여 접근 제한을 했을 경우, 수사관은 패스워드 복구 프로그램을 이용하여 증거파일을 획득이 가능하다. 하지만 대다수의 패스워드 복구 프로그램은 단순한 전수조사 공격 방법을 지원하거나 공개된 사전파일을 적용하여 복구를 시도하고 있다. 전수조사 공격 방법은 사용가능한 패스워드의 모든 경우의 수를 대입 하는 방법으로써 패스워드의 길이가 길어지면 복구시간이 기하급수적으로 늘어나므로 패스워드 검색의 효율성이 떨어진다. 또한 공개된 사전파일을 사용한 공격방법은 한국인들의 패스워드를 고려하지 않은 방법이기 때문에 패스워드의 복구에 실패 확률이 높다. 따라서 수사관은 패스워드 기반 암호 시스템을 사용하여 암호화된 파일에 대해 효율적인 패스워드 복구를 할 수 있도록 사전파일을 구축하는 방안이 필요하다.

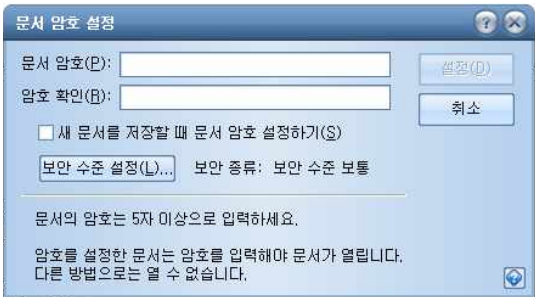
본 논문에서는 효율적으로 패스워드를 복구 할 수 있는 국내외 연구 사례 및 동향을 조사함으로써 빠르게 패스워드 복구가 가능한 사전파일을 구축하는 방안에 대해 연구하였다. 이와 함께 용의자의 개인적인 신상 정보를 활용하여 최적화된 사전파일을 생성할 수 있는 방안에 대해 제시한다.

2. 패스워드 기반 암호 시스템을 사용한 접근 제어 및 패스워드 복구 방법

가. 접근 제어 방법

용의자가 사용할 수 있는 패스워드 기반 접근 제어에는 여러 가지가 있을 수 있다. 용의자가 사용하는 응용프로그램에서 패스워드 기반의 암호시스템 기능을 제공한 경우에 암호화기능을 이용하여 접근 제어를 할 수 있다. 예를 들어 '한글과컴퓨터사'의 '한글'에 '문서 암호'라는 기능을 통해 문서자체에 암호를 사용하여 접근 제어를 할 수 있으며, '마이크로소프트사'의 엑셀, 워드, 파워포인트에서 제공하는 '문서 암호화'를 사용하여 각 문서별로 패스워드 기반의 암호화기능을 적용할 수 있다[2, 3]. 또한 'Adobe System사'의 'Acrobat PDF' 제품에도 '암호 보안' 기능을 사용하여 문서에 접근 제어 기능을 제공하고 있다[4].

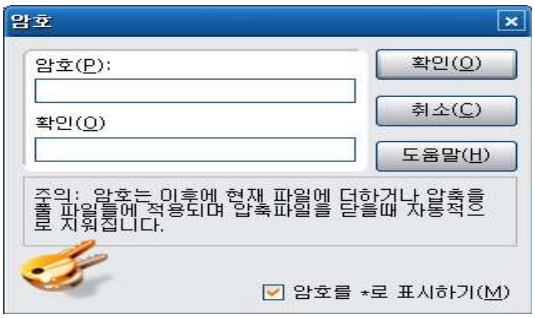
* 본 연구는 정보통신부 및 정보통신연구진흥원의 IT신성장동력핵심기술개발사업의 일환으로 수행하였음. [2007-S019-01, 정보투명성 보장형 디지털 포렌식 시스템 개발]



(그림 1) '한글' 프로그램의 암호설정 화면

[그림 1]은 '한글과컴퓨터사'의 '한글 2007'프로그램을 이용하여 패스워드 기반 접근 제어하는 그림이다. 위와 같이 용의자가 사용하는 문서 편집 응용프로그램에서 제공하는 패스워드 기반 암호화 기능을 이용하여 접근 제한을 하면, 수사관은 패스워드 복구 프로그램을 이용하여 패스워드를 복구하게 된다.

패스워드 기반 암호화를 사용하는 또 다른 방법으로써 용의자의 파일을 압축프로그램에 적용하는 방법이 존재한다. 파일 압축 프로그램도 패스워드를 사용하여 암호화가 가능하므로 사건현장에서 용의자가 여러 가지 압축 프로그램을 사용할 수가 있다. 예를 들면 [그림 2]와 같이 '알툴즈사'에서 개발한 '알집'을 이용한 압축을 통해 암호화를 사용할 수 있다[5].



(그림 2) '알집'의 암호화 적용화면

또한 'WinZip'과 'WinRAR'등 많은 압축 프로그램이 패스워드 기반 암호 시스템을 이용하여 암호화를 사용한다[6, 7].

나. 패스워드 복구 방법

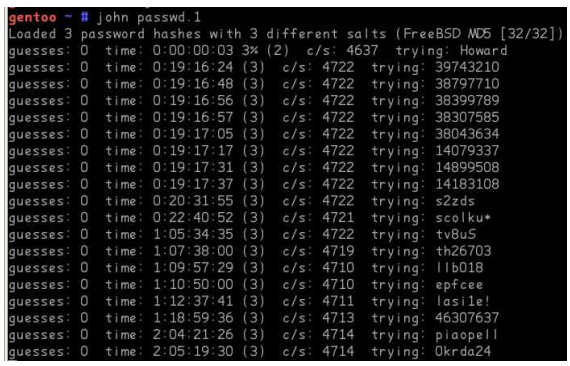
용의자는 위에서 제시한 응용프로그램들을 활용하여, 법정에서 자신에게 불리하게 사용될 수 있는 증거파일에 패스워드 기반 암호화를 적용할 수 있다. 수사관은 압수된 하드디스크 드라이브에서 증거파일이라고 의심되는 암호화된 파일을 입수하게 되며, 입수한 증거 파일을 패스워드 복구 프로그램에 적용하게 된다. 일반적으로 많이 활용되고 있는 패스워드 복구프로그램은 'Elcomsoft사'의 프로그램들이다[8]. 아래의 [그림 3]은 'Elcomsoft사'의 패스워드 복구 프로그램 중 하나인 'ARPR 1.50'의 그림으로 패스워드 복구 방법에는 '전수조사 공격(Brute force Attack)'과 '사전 공격(Dictionary Attack)'을 이용하여 패스워드를 복구한다. 대부분의 패스워드 복구 프로그램들도 전수조사 공격과 사전 공격을 이용하여 패스워드를 복구한다. 그러나 전수조사 공격을 이용한 패스워드 복구는 복구 성공률이 좋지만, 시간이 너무 많이 걸리기 때문에 5자 이하의 패스워드에 대해서만 활용 해야만 한다.

(그림 3) 'WinRAR' 패스워드 복구 프로그램

효율적인 대응방안으로 사전파일을 이용한 방법이 사용 되고 있으나, 사용되는 사전파일에 따라 패스워드 복구시간과 패스워드 복구 성공 여부가 가변적이다. 따라서 패스워드에 자주 사용되는 사전파일을 이용하여 패스워드 복구 프로그램을 실행시키면 패스워드 복구시간을 단축시키고 패스워드 복구 성공률을 높일 수 있다.

3. 사전구축 사례 및 동향

국외의 경우에 패스워드 기반 암호화를 이용하는 사용자들의 패스워드 패턴을 활용하여 자동으로 검사 및 패스워드 복구를 해주는 프로그램이 존재하고 있다. 대표적인 패스워드 복구 프로그램으로는 'John the Ripper' 프로그램이 있다[9]. 'John the Ripper'는 여러 가지 플랫폼을 지원하여 Uxix, DOS, Windows의 패스워드의 복구가 가능하고, 다양한 암호 알고리즘 적용이 가능한 것이 특징이다. 예를 들면 DES, MD5, LM-hash 등 다양한 알고리즘에 적용이 가능하다. 따라서 여러 가지 플랫폼과 알고리즘에 적용이 가능한 것은 사람들의 패스워드 생성 규칙을 사전에 분석하고, 자주 사용되는 패턴에 맞게 패스워드 후보를 생성해주는 알고리즘이 있기 때문이다. 다음의 [그림 4]는 리눅스에서 'John the Ripper 버전 1.7.2'를 이용하여 실제로 패스워드 복구를 실행하는 화면이다.



(그림 4) John the Ripper 버전 1.7.2의 실행 화면

최근의 국내연구에서는 '개인키 보호용 패스워드 공격에 관한 연구'가 이뤄졌으며, 한글 패스워드 특징을 분석 하고 여러 암호 알고리즘을 사용하여 패스워드 복구를 시도 하였다[10]. 연구의 주요 특징으

로는 'John the Ripper' 소스 분석과 함께 한글 패스워드 패턴 조사 및 한글 단어 수집을 통해 'John the Ripper'에 한글 패스워드 처리 모듈을 추가한 것이다. 한글 패스워드 패턴을 분석한 결과, 717개의 패스워드 중 한글 패스워드가 20%가 존재하며 한글 패스워드 중에는 각 사용자별 이름이 많이 활용되었다.

4. 패스워드 사전파일 방안 연구

가. 패스워드 패턴 분석을 위한 분석 항목 분류 기준

패스워드는 일반적으로 키보드를 통해 입력 가능하므로, 본 논문에서는 검색의 범위를 숫자, 문자, 특수문자로 정의하였다. 따라서 숫자, 문자, 특수문자를 분류하여 분석했다.

문자	◈ 문자열의 길이 ◈ 문자열의 의미 분석
숫자	◈ 숫자 길이 ◈ 같은 숫자, 순차적인 숫자 사용 유무 ◈ 숫자 반복 사용 유무
특수문자	◈ 빈도수
사용자명 영타입력	◈ 이름을 영문타자로 친 경우
이름 연관성	◈ 로마자표기법 ◈ 로마자표기 단축 사용 유무
생년월일 연관성	◈ 생년월일 사용유무

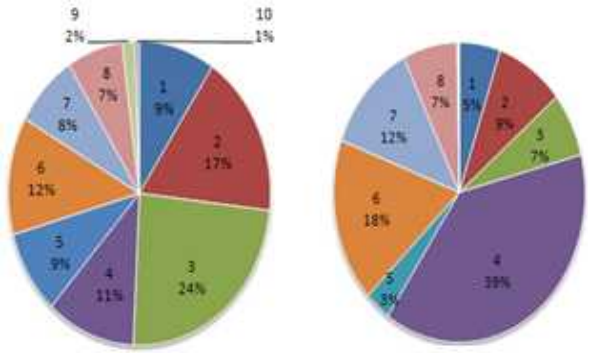
[표 1] 패스워드 패턴 분석을 위한 주요 항목

위의 표는 한글 패스워드를 사용하는 사람들의 패스워드 생성 패턴에 대해 분석 기준을 마련한 표이다. 문자열의 길이와 문자가 가지고 있는 의미에 대해 분석하였다. 문자가 가지고 있는 의미로 예를 들면 영문 자판을 한글로 변환하였을 때 유의미한 단어인지 분석하고, 자주 사용되는 단어들을 분석하였다. 숫자의 경우는 숫자가 사용되는 길이와 같은 숫자, 연속적인 숫자 사용 유무 등에 대하여 분석하였다. 특수문자의 경우에는 자주 사용되는 문자의 빈도수를 조사하였다. 패스워드 생성 패턴 중에서 사용자명이 가장 많이 사용되었으며[10], 본 논문에서는 이를 기반으로 여러 항목별로 분류하였다. 단순히 사용자명을 영문타자로 패스워드를 생성한 경우와 로마자표기로 변환하여 패스워드를 생성한 경우, 그리고 로마자표기를 단축하여 사용한 경우를 따로 분석하였다. 또한 패스워드를 생성하는 패턴 중에 자주 사용되는 패턴이 자신의 생년월일을 사용하는 경우가 있다. 따라서 생년월일 중에 어떠한 정보가 가장 많이 사용되는지 조사하였다.

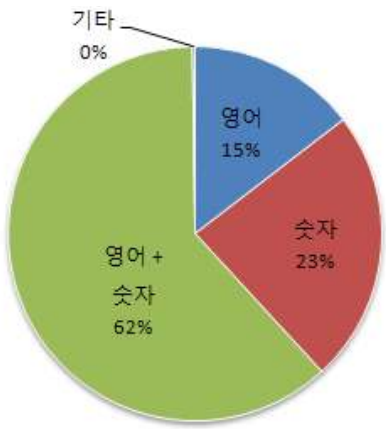
나. 분석 항목별 분석 결과

각 분석 항목에 대해 한글 패스워드 생성 패턴을 분석하였다. 패스워드의 길이는 조사자의 91%가 6~8자를 사용하여 패스워드를 생성한다고 했다. 그리고 5자 이하의 패스워드는 조사자의 0.4%만이 사용하는 것으로 조사 됐다. 따라서 사전파일 생성에 6자 이상을 생성하는 것이 패스워드 복구에 시간을 단축시킬 수 있다. 또한 5글자 이하의 패스워드는 사전 공격을 종료한 후 전수조사 공격을 시행하는 것이 패스워드 복구에 확률을 높일 수 있다.

(1) 문자열의 길이



(그림 5) 영문자 길이(좌)와, 숫자의 길이(우)



(그림 6) 문자의 조합 패턴

(2) 문자열의 의미

패스워드에 사용되는 문자열의 특징을 분석한 결과, 키보드의 왼쪽 자리에 위치한 패스워드를 많이 사용하였으며, 이름의 로마자표기 단축형태가 많이 보였다. 또한 키보드 위치 순서대로를 패스워드를 사용한 패턴도 분석되었고, 한글 단어 중 '사랑', '사랑해', 'LOVE', 사람 이름, 은어, 인터넷용어 등의 패턴이 분석되었다. 또한 'abcd'와 같은 연속된 영문자도 사용되었고, 'aaa'와 같은 동일한 문자의 연속의 패턴도 나타났다.

(3) 숫자의 특징

숫자의 경우를 살펴보면, 동일한 숫자를 연속으로 사용한 패턴이 숫자를 순차적으로 사용하는 패턴보다 더 많은 것을 분석을 통해 확인할 수 있었다.

(4) 특수문자 빈도수

패스워드 생성 패턴 중 문자열 내에서 특수문자를 활용한 경우를 살펴보면, 1자리를 포함하는 패턴이 가장 많았으며, 2자리를 포함하는 패턴도 다수 존재하였다. 즉 특수문자를 사용한 패턴 중 1자리와 2자리를 합친 패턴이 91%로 통계적조사결과가 나타났다. 따라서 패스워드 사전파일 생성 시에 특수문자의 경우 1자리와 2자리를 생성하는 패턴은 검색 시에 매우 중요한 요소라 할 수 있다.

(5) 사용자명 영타입력

한글 이름을 영타로 사용해서 패스워드를 생성한 패턴을 분석한

결과 이름을 포함하는 패스워드를 가장 많이 생성 하고, 성을 포함하는 패스워드 생성 패턴이 뒤를 이었다. 또한 성과 이름을 그대로 영타를 사용하여 패스워드를 생성한 패턴도 분석되었다.

(6) 이름의 연관성

한글 이름을 로마자로 변형하여 패스워드에 사용하는 패턴에 대해 분석한 결과, 이니셜을 사용한 패스워드가 가장 많은 패턴이 나타났다.

(7) 생년월일 연관성

생년월일을 포함하는 패턴을 살펴보면 '월/일'을 입력하는 것이 가장 많이 나타났으며, '연/월/일'을 사용하는 패턴이 뒤를 이었다.

다. 패스워드 생성 패턴 종합 분석 결과

패스워드 길이	6~8자가 전체의 91%
문자 조합 타입	영어 + 숫자 > 숫자 > 영어 > 숫자 + 특수문자
영어와 숫자의 순서	영어가 먼저 사용됨
한글이름의 영타 표기	이름 포함 > 성 포함 > 이름 동일 > 성명 포함, 성명 동일
한글이름의 로마자표기	단축 포함 > 단축 동일 > 성 > 이름
영타의 한글 변환	'사랑해', '사랑', 'LOVE', 은어, 인터넷 용어 등
생년월일 사용 빈도	MMDD > YYMMDD > YY
숫자의 연속성	조사 대상의 11%가 숫자를 연속으로 사용
자주 사용하는 특수문자	특수문자 사용자의 93% 1~2자리 사용

[표 2] 패스워드 생성 패턴의 종합 분석표

위의 [표 2]는 한글 패스워드를 생성하는 패턴을 종합적으로 분석한 것이다. 제시한 내용을 토대로 각 개인에 맞는 패스워드 사전파일을 자동으로 생성해주는 알고리즘을 설계가 추가적으로 필요하다.

5. 결론 및 향후 연구방향

컴퓨터 범죄의 용의자는 법정에서 자신에게 불리한 증거가 될 만한 파일들을 패스워드 기반 암호 시스템을 활용해 접근 제어를 하고 있다. 주로 사용하는 응용프로그램에는 다양하게 존재 하며, 이에 대해 수사관은 패스워드 복구 프로그램들을 통해 패스워드를 복구한다. 패스워드 복구프로그램은 전수조사 공격방식과 사전 공격 방식을 가장 많이 사용하고 있는데 전수조사 공격방식은 패스워드의 길이와 복잡성에 따라 복구하는데 너무 많은 시간이 소요된다. 또한 사전파일을 이용한 패스워드 공격 방식은 사전파일에 너무 의존적이므로 전수조사 공격 방법보다 복구 성공률이 떨어 질 수 있다. 또한 우리나라의 실정에 맞는 패스워드 사전파일이 존재 하지 않아서 더욱 복구 성공률이 떨어지기 때문에 수사관이 효율적으로 패스워드 복구 할 수 있는 방안을 연구해야한다. 따라서 본 논문은 국내외의 효율적으로 패스워드 복구 하는 연구 사례 및 동향을 조사하고 더욱 효율적으로 패스워드를 복구를 위해 패스워드 생성 패턴을 조사하고 연구하였다. 그 결과 사람들이 패스워드를 생성할 때 자주 사용하는 일반화된 패턴이 존재하는 것을 확인하였다. 위의 결과물을 이용하여 용의자가 패스워드 기반 암호화를 한 증거파일에 대해 패스워드 복구가 좀 더 효율적으로 실행될 수 있는 사전파일 생성의 기초가 된 연구라 할 수 있다.

향후에는 심층적인 패스워드 패턴 분석과 함께 개인별 신상정보를 이용할 수 있는 방안 등을 연구하며, 개인에게 맞는 패스워드 사전파일을 자동으로 생성하는 알고리즘을 개발할 것이다. 더 나아가 각 개인별 정보를 활용하여 패스워드를 복구할 수 있는 도구를 개발할 것이다.

참 고 문 헌

[1] 사이버 범죄현황 ,http://www.ctrc.go. kr/, 사이버테러대응센터

[2] 문서 편집 프로그램, http://www.haansoft.com/, (주)한글과컴퓨터

[3] Office 프로그램, http://www.microsoft.com, Microsoft Corporation

[4] 문서 편집 프로그램, http://www.adobe.com/, Adobe System .Inc,

[5] 국내의 압축 프로그램, http://www.altools.co.kr/, (주)알툴즈

[6] 외국의 압축 프로그램, http://www.winzip.com/, Corel Corporation

[7] 외국의 압축 프로그램, http://www.rarsoft.com/, RARLAB

[8] 일반적으로 많이 활용하는 패스워드 복구 프로그램, http://www.elcomsoft.com/, Elcomsoft

[9] 국외의 패스워드 복구 프로그램, http://www.openwall.com/john/, John the Ripper

[10] 제 1, 2차 워크샵, 2007년 암호연구회, 한국정보보호학회