북한 서광사무처리 3.0 파일 구조 분석

최준형*, 강동수+
국방대학교 컴퓨터공학과
e-mail:freewannabe@naver.com*, greatkoko@kndu.ac.kr+

Analysis File Format of Seogwang Document Processor 3.0 in North Korea

Junhyeong Choi*, Dongsu Kang+
Dept of Computer Science & Engineering, Korea National Defense University

요 약

북한에서 운영하고 있는 오피스 프로그램인 서광사무처리 3.0은 ODF(Open Document Format) 파일 포맷을 입력으로 받아 문서를 처리한다. ODF는 여러 개의 XML(Extensible Markup Language) 파일로 구성되어 있고, 하위 노드들을 통해서 파일 구조를 정의한다. 이러한 서광사무처리 3.0의 ODF 파일 구조를 하위 프로그램별 입력받는 파일 확장자에 따라 공통 영역과 가변 영역으로 비교하고, CVE (Common Vulnerabilities and Exposures)를 통해 ODF와 XML 주요 취약점을 분석한다.

1. 서론

북한은 일찍부터 리눅스 기반의 운영체제인 붉은별(Red Star)을 자체적으로 개발하여 사용해오고 있으며, 붉은별에 내장된 응용 소프트웨어 중 대표적으로 오피스 프로그램인 서광사무처리가 있다.[1][2] 서광사무처리는 리브레오피스(LibreOffice)를 기반으로 개발되었고, 리브레오피스와 동일하게 ODF(Open Document Format) 파일 포맷을 입력으로 하여 작업을 수행한다. ODF는 2006년에 국제 표준으로 승인된 문서 파일 포맷으로서, 내부가 여러 개의 XML(Extensible Markup Language) 파일로 이루어져 있고, XML 파일의 하위 노드들을 통해서 파일 구조와 스타일을 정의하는 방식으로 구성되어 있다.

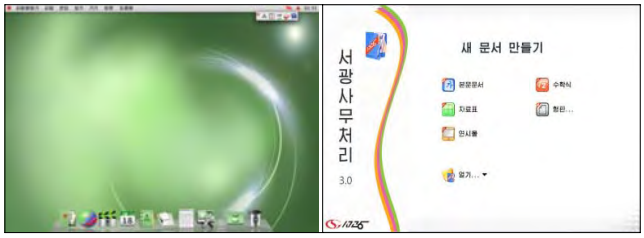
본 연구는 서광사무처리 3.0에서 사용하는 입력 파일 포맷인 ODF에 대한 분석과, 향후 분석된 내용을 바탕으로 소프트웨어 보안 취약점 테스팅 기법을 적용하여 서광사무처리 3.0의 보안 취약점을 찾아내기 위한 목적으로 진행되었다. 본 논문의 구성은 2장에서 서광사무처리 3.0과 ODF 파일 포맷에 대해 알아보고, 3장에서는 ODF 파일 포맷의 세부적인 분석 내용을 제시하며, 4장에서 결론과 향후 연구 방향을 기술한다.

2. 관련연구

2.1 북한 운영체제 붉은별과 서광사무처리

북한은 제3차 과학기술발전 5개년 계획(2008년-2012년)에서 우리식 컴퓨터 운영체제의 개발과 보급을 강조하였고, 핵심 소프트웨어 개발 기관인 조선컴퓨터센터(KCC : Korea Computer Center)를 필두로 하여 레드햇(Red Hat) 리눅스를 기반으로 한 운영체제인 붉은별을 2006년에 개발하였다. 이후 지속적인 개량을 통해 2008년, 2010년, 2011년에 각각 1.0, 2.0, 3.0 버전을 출시하였고, 2017년에는 선전 매체 홈페이지인 서광(sogwang.com/kp)에서 과학사업의 정보화 일환으로 붉은별 4.0이 개발되었다는 사실을 공표하였다.

붉은별 2.0 버전부터는 사용자의 편의를 위한 응용 소프트웨어가 다수 탑재되었는데, 오피스 프로그램인 서광사무처리도 그 중 하나이다. 서광사무처리는 평양인쇄공업대학에서 개발한 소프트웨어로 우리나라의 한글 오피스와 기능적으로 유사한 프로그램이다. 현재 붉은별 3.0 버전에 탑재되어 있는 서광사무처리 3.0에는 기능별로 여러 프로그램을 패키지화하여 제공하고 있다.[2] (그림 1)은 가상머신(Virtual Machine)을 이용하여 실행한 붉은별 3.0과 서광사무처리 3.0의 화면이다.



(그림 1) 붉은별과 서광사무처리 3.0 실행화면

2.2 리브레오피스(LibreOffice)

서광사무처리 3.0은 Document Foundation에서 개발한 무료 오픈소스 오피스 프로그램인 리브레오피스를 기반으

+ 교신저자: greatkoko@kndu.ac.kr

로 한다. 서광사무처리 3.0 내에 구성되어 있는 기능별 프로그램들은 리브레오피스에 구성되어 있는 프로그램과 역할 및 입력 파일의 구조적 측면에서 일치한다. 따라서 서광사무처리 3.0과 리브레오피스 간의 파일 호환이 가능하지만, 우리가 흔히 사용하는 한컴 오피스와의 호환은 불가하다. 두 프로그램 간의 비교는 <표 1>에서 명시하였다.

<표 1> 서광사무처리 3.0과 리브레오피스 비교

서광사무처리	리브레오피스	파일 포맷
본문문서	LibreOffice Writer	ODT
수학식	LibreOffice Math	ODF
자료표	LibreOffice Calc	ODS
형판	LibreOffice Draw	ODG
연시물	LibreOffice Impress	ODP

2.3 ODF(Open Document Format) 파일 포맷

서광사무처리 3.0에서 사용하는 파일은 개방형 문서표준 포맷인 ODF(Open Document Format) 형식을 취하고 있다. ODF는 상호 호환이 가능한 마크업 언어인 XML (Extensible Markup Language)을 사용하여 파일 포맷을 정의하며, 문서의 본문을 비롯한 스프레드시트, 프레젠테이션 등 오피스 기능을 ODT, ODS, ODP 등의 확장자 형태로 제공한다. ODF는 2005년에 유럽의 표준화 국제 컨소시엄인 OASIS(Organization for the Advancement of Structured Information Standards) 표준으로 승인되었고, 2006년에는 국제 표준화 기구인 ISO/IEC에서 국제 표준으로 승인되었다.[3][4][5]

3. 서광사무처리 3.0 분석

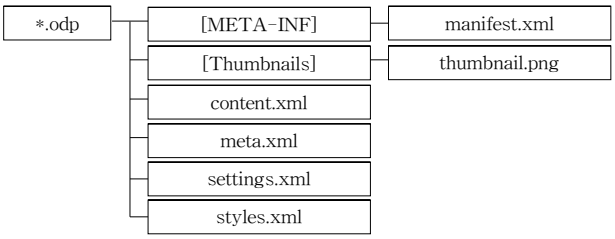
3.1 서광사무처리 3.0과 ODF 구성요소

서광사무처리 3.0을 구성하고 있는 프로그램은 본문문서, 수학식, 자료표, 형판 그리고 연시물 프로그램이며, 각 프로그램별 기능은 <표 2>와 같다.

<표 2> 서광사무처리 3.0 기능별 프로그램

명 칭	기 능	유사 프로그램
본문문서	Word Processor	한글, 워드
수학식	Mathematical Equation Editor	한글 수식편집기
자료표	Spreadsheet	한셀, 엑셀
형판	Image Editor	그림판
연시물	Presentation	한쇼, 파워포인트

다음으로 ODF 파일은 저장 용량의 축소를 위하여 압축 포맷인 ZIP으로 저장된다. 따라서 ZIP 압축 프로그램을 이용하여 압축을 해제하면 ODF 파일의 세부 구성요소를 확인할 수 있다.[6] ODF 파일의 구성요소는 각 프로그램마다 사용하는 기본 확장자에 따라 조금씩은 차이가 있으나, <표 3>과 같이 기본적으로 5개의 XML 파일(manifest, content, meta, settings, styles)과 1개의 PNG(Portable Network Graphics) 파일(thumbnail)로 이루어져 있다. 서광사무처리 3.0 중 연시물 프로그램을 이용하여 테스트 파일을 생성 후 압축을 해제한 결과는 (그림 2)와 같다.

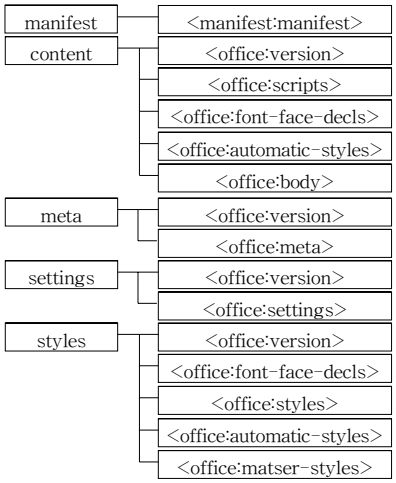


(그림 2) ODP 파일 압축 해제 결과

<표 3> ODF 파일 세부 구성요소

구 분	포 함 내 용
manifest.xml	· XML 파일 구조 명시 · ODF 파일 내부에 삽입된 파일 종류 및 위치 표시
content.xml	· 실제 본문 내용을 포함하는 파일
meta.xml	· 속성에서 입력할 수 있는 문서에 관한 메타데이터 · 문서 작성 일자 등
settings.xml	· 문서에 대한 추가 속성 · 인쇄 설정, View 설정 등
styles.xml	· 문서에 적용된 스타일 저장
thumbnail.png	· 문서의 기본 여백 이미지 저장

서광사무처리 3.0에서 사용하는 ODF 중 수학식 프로그램을 제외한 ODT, ODS, ODG, ODP 파일 포맷 모두 <표 3>의 XML 파일 5개를 동일하게 가지고 있으며, 각 XML 파일의 하위 노드 또한 (그림 3)과 같이 프로그램별로 구조가 동일하다.



(그림 3) XML 파일 하위 노드

3.2 ODF 확장자별 구조 분석

앞서 언급한 바와 같이 서광사무처리 3.0의 하위 프로그램에서 입력받는 파일 확장자(ODT, ODF, ODS, ODG, ODP)에 따라 파일 구조는 대부분 동일하지만, 일부 가변 영역이 존재한다. 수학식 프로그램은 다른 프로그램들과 다르게 폰트, 여백, 색상 등 스타일 설정 기능이 없기 때문에 수학식 프로그램에서 사용하는 .odf 파일 내부에는 styles.xml 파일이 존재하지 않는다. 또한 본문문서와 자료표 프로그램에서 사용하는 .odt, .ods 포맷은 다른 포맷들과 다르게 manifest.rdf 파일이 별도로 존재하고, 그 내용은 (그림 4)와 같다.

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF xmlns:rdf=
"http://www.w3.org/1999/02/22-rdf-syntax-ns#">
<rdf:Description rdf:about="styles.xml">
<rdf:type rdf:resource=
"http://docs.oasis-open.org/ns/office/1.2/meta/odf#StylesFile"/>
</rdf:Description>
<rdf:Description rdf:about="">
<ns0:hasPart xmlns:ns0=
"http://docs.oasis-open.org/ns/office/1.2/meta/pkg#"
rdf:resource="styles.xml"/>
</rdf:Description>
<rdf:Description rdf:about="content.xml">
<rdf:type rdf:resource=
"http://docs.oasis-open.org/ns/office/1.2/meta/odf#ContentFile"/>
</rdf:Description>
<rdf:Description rdf:about="">
<ns0:hasPart xmlns:ns0=
"http://docs.oasis-open.org/ns/office/1.2/meta/pkg#"
rdf:resource="content.xml"/>
</rdf:Description>
<rdf:Description rdf:about="">
<rdf:type rdf:resource=
"http://docs.oasis-open.org/ns/office/1.2/meta/pkg#Document"/>
</rdf:Description></rdf:RDF>
```

(그림 4) manifest.rdf

RDF(Resource Description Framework)는 웹에 있는 객체의 메타 데이터를 표현하기 위한 언어 규격으로, manifest.rdf 파일을 읽어보면 문자 인코딩 방식(UTF-8)과 styles.xml, content.xml에 들어갈 문서의 표준을 웹을 이용해 정의하고 있다. 형판과 연시물 프로그램은 그래픽 기반의 프로그램이므로 manifest.rdf가 존재하지 않는다.

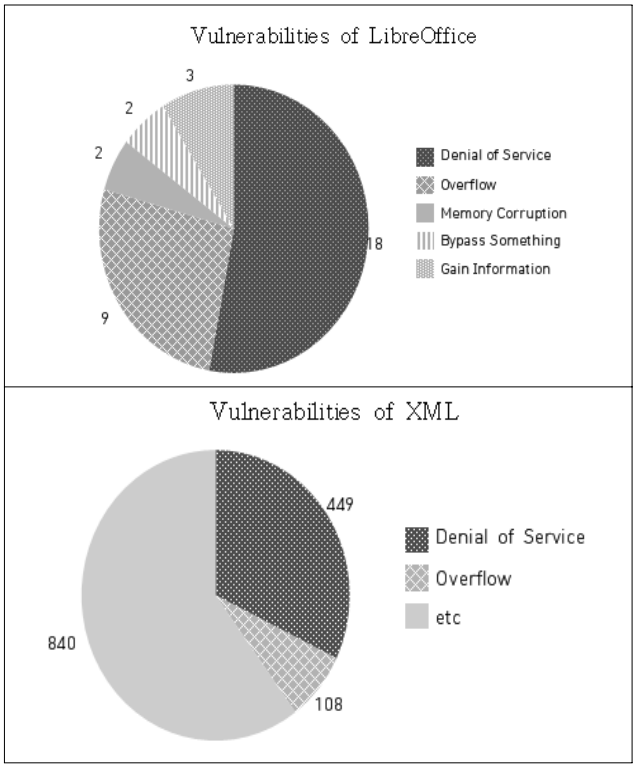
<표 4> 서광사무처리 3.0 하위 프로그램별 파일 구조

구 분	XML					mani-fest.rdf
	mani-fest	con-tent	meta	sett-ings	styles	
본문문서	○	○	○	○	○	○
수학식	○	○	○	○	×	×
자료표	○	○	○	○	○	○
형판	○	○	○	○	○	×
연시물	○	○	○	○	○	×

<표 4>와 같이 서광사무처리 3.0의 하위 프로그램별 공통 영역(manifest, content, meta, settings)과 가변 영역으로 나누어 볼 수 있는데, 만약 단일 하위 프로그램에서 공통 영역에 대한 보안 취약점을 식별하게 되면, 서광사무처리 3.0 패키지 프로그램 전체의 보안 취약점이라고 판단해 볼 수 있을 것이다.

3.3 CVE(Common Vulnerabilities and Exposures) 분석

CVE Detail(cve.mitre.org)에 의하면 현재까지 발견된 단순 리브레오피스나 ODF 자체의 취약점은 서비스 거부(DOS: Denial of Service) 18건, 오버플로우 9건, 기타 7건 등 총 34건 밖에 없지만, 위에서 언급한 XML 기반의 파일 구조 관점에서 보게 되면 취약점은 서비스 거부 449건, 오버플로우 108건 등 총 1,397건으로 늘어난다. 이는 곧 북한에서 사용하는 서광사무처리 3.0의 보안 취약점 발견 가능성이 높아진다는 것을 의미한다.



(그림 5) LibreOffice 및 XML 취약점

4. 결론 및 향후 연구

본 연구에서 서광사무처리 3.0의 보안 취약점 식별을 위해 입력 파일 포맷인 ODF의 파일 구조를 살펴보고, CVE를 통해 XML 기반의 파일 구조가 많은 보안 취약점을 가지고 있다는 것을 확인하였다. 향후 서광사무처리 3.0 하위 프로그램의 XML 기반 파일 구조 중 공통 영역에 대한 보안 취약점을 CVE, CWE를 통해 더욱 세부적으로 분석하고, 소프트웨어 테스팅 기법인 퍼즈 테스팅(Fuzz Testing)을 적용하여 취약점을 식별할 예정이다.

참고문헌

[1] Gu-Yeon Jeong, Gi-Tae Lee “Innovation of Science and Technology and North Korea's Asymmetric Threat : Rise of Cyber Warfare and Unmanned Aerial Vehicle” KINU Research Series 16-04, 2016.

[2] Choon-Geun Lee, Jong-Seon Kim, Dal-Li Nam “Policies to Promote South-North ICT Cooperation” Science and Technology Policy Institute 14-28, 2014.

[3] Yun-Yong Jang, So-Yang Kim, Won-Sung Sohn, Dong-Sun Nam, Soon-Bum Lim “Document Structure Analysis for Support ODF in Haansoft Hangul” 한국멀티미디어학회 국제학술대회 논문지, pp.258-261, 2008.

[4] 정제호, 손원성, 임순범 “ODF와 OOXML을 중심으로 한 사무용 전자문서 국제표준화 동향” 한국정보과학회지 26(6), pp.20-28, 2008.

[5] Meenu Pandey “Version Aware LibreOffice Documents” University of Wisconsin Milwaukee, 2014.

[6] 박찬주, 강동수 “보안 취약점 분석을 위한 붉은별(Red Star) 서광문서처리체계 파일 구조 분석” 한국정보처리학회지 25(1), pp.110-112, 2018.