XML을 이용한 디지털 증거의 수집 및 관리에 대한 연구*

임경수*, 이석희*, 이상진*
*고려대학교 정보보호대학원

Collection & Management of Digital Evidence using XML

KyungSoo Lim*, SeokHee Lee*, SangJin Lee*
*Graduated School of Information Security, Korea University.

요 약

디지털 포렌식에서 디지털 정보나 데이터는 범죄의 실마리를 풀 수 있는 증거로 사용되고 있어 사이버 범죄 현장에서 반드시 확보해야한다. 최근에는 사이버 범죄 뿐 아니라 민사, 형사 소송의 일반적인 범죄에서 중요한 역할을 담당하고 있지만, 범죄의 유형이 점차 다양해지고 복잡해짐에 따라 디지털 증거들을 수집, 관리하는데 어려움이 있으며, 디지털 범죄 수사 결과는 수사관, 분석관, 감정관 등 각 담당자마다 해석하고 필요로 하는 정보가 다르기 때문에 데이터의 가공과 다양한 형태의 보고서가 제공되어야 한다. 따라서 본고에서는 웹에서의 데이터 처리 표준으로 자리 잡은 XML을 이용하여 보다 편리하게 디지털 증거를 구조화하고 처리하는 방법과, 이를 효과적으로 데이터를 표현하는 방법을 제시하고자 한다.

I. 서론

세계적으로 IT 기술의 선진국으로 자리 잡은 우리나라는 이제는 유비쿼터스 시대의 실현을 앞두고 있다. 이러한 정보화 사회의 놀라운 발전의 이면에 사이버 범죄 또한 고도화되고 지능적으로 발전하고 있다. 경찰청이 올해 발표한 통계자료에 따르면 2005년 사이버 범죄의 증가는 5년 전에 비해 무려 36배 증가하였다.[9] 이러한 사이버 범죄를 효과적으로 차단하기 위해 더욱 디지털 포렌식 분야의 법적, 기술적인 노력이 대두되고 있는 현실이다.[10]

본론에 앞서 디지털 증거가 이용되는 디지털 범죄 수사(이하 디지털 포렌식)의 정의, 현황과 일반적인 포렌식 절차에 대해 설명한다. 그리고 이러한 포렌식 수사에 이용되는 디지털 증거를 XML을 이용하여 보다 효과적으로 사용될 수 있도록 구조화하고, 표현하는 방법에 대해 설명하고자 한다.

1. 디지털 포렌식 현황

디지털 범죄 수사(이하 디지털 포렌식)는 미국의 DFRWS (Digital Forensic Research Workshop)에 따르면, 범죄를 재현하거나 파괴적이고 비 인가된 행동들에 대한 예측을 손쉽게 하기 위해 디지털 증거물에 대한 보존, 수집확인, 식별, 분석, 해석, 문서화, 표현을 과학적으로 도출되고 증명된 방법을 사용하여 수행하는 것이라고 정의한다.[3] 정보화 사회가 고도화됨에 따라 디지털 포렌식은 과학수사와 수사과학 분야에서 필요성이 대두되었다. 이는 시스템, 네트워크, 정보처리 등 기술 분야를 연

* 본 연구는 정보통신부 대학 IT 연구센터 육성·지원 사업의 연구 결과로 수행되었습니다.

구하는 학계뿐만 아니라 사법계, 법조계에서 다양하게 연구되고 있다.

하지만 국내의 디지털 포렌식 수사는 아직 초기 단계라는 한계성으로, 법적인 제도나 수사에 착수할 가이드라인이 부족하며, 포렌식 기술에 대한 명확한 원천 기술도 아직 정립되지 않은 상태에 있다.[10] 이에 반해 미국은 1990년대 초부터 이미 디지털 포렌식이 도입되어 FBI(미연방수사국)를 중심으로 사이버 범죄 수사에 널리 이용되고 있다. 뿐만 아니라 법조계에서는 민사소송의 경우, 피고는 원고가 요구한 전자메일이나 전자문서 등의 디지털 증거에 대한 제출을 정당화 (e-discovery*)하는 등의 제도가 시행되어 올해 말에는 관련 법안이 통과될 예정이다. [10]

2. 디지털 포렌식의 일반적인 절차

디지털 포렌식 절차는 시스템이나 네트워크 침해 사고 대응과 관련되어 활발히 연구가 진행 중이며, 아직 국내의 표준 절차는 정립되지 않았다. 일반적으로 정의된 수사 절차는 사건 준비 단계, 사건 대응 단계, 물리적 범죄 현장 조사 단계, 디지털 범죄 조사 단계, 분석 및 보고 단계[3]로 나누어지며, 본고에서 주로 다룰 디지털 범죄 조사 단계는 세부적으로 디지털 증거의 수집 및 분석으로 나누어진다.

하지만 디지털 증거는 수집해야 할 데이터의 종류가 다양하고 복잡하기 때문에 이를 효과적으로 분석하기 위해서는 디지털 데이터들을 구조화하여 효과적으로 처리해야 할 필요성이 있다. 또 이렇게 수집된 증거들은 마지막 검토 및 보고 단계에서 수사 결과를 전달받는 대상에 따라 처리하기도 쉬워진다.

3. XML에 대한 소개

XML(eXtendable Markup Language)은 웹의 표준을 정하는 W3C에서 정의한 언어로, HTML과 같이 태그를 이용하는 마크업 언어이다. 하지만 가장 큰 차이점은 HTML 1.0이 십여 가지의 정의된 태그만을 포함하는 것과 달리, XML은 태그를 사용자가 정의할 수 있다. 예를 들어, 기업의 해외 지점들 사이의 정보를 XML로 교환한다면, 다음과 같이 <Office>, <Department>등과 같이 정의하여 사용할 수 있지만, HTML은 단지 <BODY>, <FORM>등의 태그만 사용해야 하는 한계성이 있다.

```
<Office>
    <Name>Nanonull Europe, AG</Name>
    <Desc>
        <Para>In May 2000, Nanonull<italic>Europe</italic> was set up in Vienna.
    </Desc>
    <Location>EUK</Location>
    <Address_EU>
    <Phone>+ 43 1 555 5000</Phone>
    <Fax>+ 43 1 555 5009</Fax>
    <EMail>euoffice@nanonull.com</EMail>
    <Department>
    <Department>
        <Name>IT &amp; Technical Support</Name>
        <Person>
            <First>Gemil</First>
            <Last>Djervo</Last>
            <Ti
            <PhoneExt      rvo@nanonull.com</EMail>
            <Title      </Shares>
            <LeaveTotal>22</LeaveTotal>
            <LeaveUsed>5</LeaveUsed>
            <LeaveLeft>17</LeaveLeft>
        </Person>
        <Person>
        <Person>
    </Department>
```

[그림1] XML 사용 예

따라서 XML의 장점은 사용자 정의 마크업을 이용하여 데이터 처리와 교환을 용이하게 하고, 구성요소들의 이름만으로도 데이터를 직관적으로 알 수 있는 점이다.

여기에 XML 문서의 구조를 정의하고 유효성을 검증할 수 있는 DTD, XML schema 기술이 있으며, XML 데이터 부분 중 필요한 부분만 추출하는 DOM, SAX Parser기능을 이용하여 효과적으로 데이터를 처리할 수 있다. 또 CSS, XSLT와 같이 XML 데이터를 다양한 관점에서 표현되어 질수 있도록 스타일시트를 사용할 수도 있다.

II. 디지털 증거에 XML 적용

디지털 증거는 휘발성 데이터와 비휘발성 증거로 나누어지는데, 휘발성 데이터(혹은 Live Data, 이하 라이브 데이터)는 수사 현장의 개인 컴퓨터가 전원이 켜져 있는 상태에서 획득해야 의미가 있는 데이터로, 시스템의 날짜나 시간 정보, 접속된 네트워크의 상태, 내부 라우팅 테이

*

블, 실행 중인 프로세스, 실행 중인 서비스, 시스템이 사용 중인 메모리 등이다. 반대로 비휘발성 데이터는 현장에서 획득하지 않아도 컴퓨터의 하드디스크 이미지에서 취득할 수 있는 데이터를 말한다. 예를 들어, 파일시스템의 날짜/시간 정보와 레지스트리 데이터, 파일 속성, 사용자 계정, 사용자 로그인 등이다.[1][3]

이러한 비휘발성 데이터는 획득한 증거로부터 추후에 분석을 해도 되는 특성을 지니지만, 라이브 데이터는 사건 현장에서 즉시 획득해야 의미를 지닌다. 따라서 이러한 라이브 데이터를 보다 효과적으로 수집, 배포하기 위해 XML로 데이터를 구조화하고 이를 웹으로 전송하면 효율적이다. 이러한 라이브 데이터를 XML 스키마로 정의하면 [그림4]과 같고, 이를 이용한 XML 데이터는 [그림3]와 같다.

그림[3]은 위에서 정의된 스키마를 이용하여, 라이브 데이터를 XML로 표현한 것으로, 단순히 포렌식 툴을 이용한 텍스트 결과보다 훨씬 이해하기 쉽다. 또 이렇게 구조화된 XML 문서는 일반적인 데이터베이스를 다루듯이 데이터를 추출하거나, 또한 합치거나 수정할 수 있어 편리하다. 그림[4]는 XML을 개발할 때 가장 널리 사용하는 XMLSpy[13]라는 프로그램에서 스키마를 정의한 후 도식화한 그림이다. 위와 같이 라이브 데이터를 XML로 표현하면, 디지털 증거들에 대해 직관적으로 이해할 수 있고, 또 수사관이 현장에서 수집할 데이터들을 점검해 볼 수 있는 체크리스트 역할도 될 수 있다.

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- edited with XMLSpy v2005 U (http://www.xmlspy.com) by any (Ru-Board) -->
<Incident Date="2006-06-20" Title="Test Incident" ID="06062001" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="C:\Documents and Settings\Luke\My Documents\LiveData2.xsd">
  <LiveData>
    <System>
      <OS>WindowsXp</OS>
      <Ver>SP2</Ver>
      <Date>2006-6-20</Date>
      <Time>10:25:33</Time>
    </System>
    <Network>
      <Connection>
        <Protocol>TCP</Protocol>
        <Local_Address>163.125.165.117:4453</Local_Address>
        <Foreign_Address>163.125.165.114:4453</Foreign_Address>
        <State>ESTABLISHED</State>
      </Connection>
      <Connection>
        <Protocol>TCP</Protocol>
        <Local_Address>163.125.165.117:1180</Local_Address>
        <Foreign_Address>163.125.165.118:1180</Foreign_Address>
        <State>CLOSE_WAIT</State>
      </Connection>
      </OpenPorts>
      <OpenPorts>
        <Protocol>TCP</Protocol>
        <Local_Address>0.0.0.0:3372</Local_Address>
        <Foreign_Address>0.0.0.0:0</Foreign_Address>
        <State>LISTENING</State>
      </OpenPorts>
      <ExecPorts>
        <PID>1111</PID>
        <Process>iroffer</Process>
        <Port>1174</Port>
        <Protocol>TCP</Protocol>
        <Path>C:\windows\system32\os2\dll\iroffer.exe</Path>
      </ExecPorts>
      <ExecPorts>
        <PID>1044</PID>
        <Process>inetinfo</Process>
        <Port>1091</Port>
        <Protocol>TCP</Protocol>
        <Path>C:\windows\system32\inetsrc\inetinfo.exe</Path>
      </ExecPorts>
      <RoutTables>
        <Network_Destination>0.0.0.0</Network_Destination>
        <Netmask>0.0.0.0</Netmask>
        <Gateway>103.98.91.1</Gateway>
        <Interface>103.98.91.41</Interface>
        <Metric>1</Metric>
      </RoutTables>
      <RoutTables>
        <Network_Destination>103.98.91.0</Network_Destination>
        <Netmask>255.255.255.0</Netmask>
        <Gateway>103.98.91.41</Gateway>
        <Interface>103.98.91.41</Interface>
        <Metric>1</Metric>
      </RoutTables>
    </Network>
    <RunningProcess>
      <Name>cmd</Name>
      <PID>1272</PID>
      <Priority>8</Priority>
      <Thread>1</Thread>
      <Handle>25</Handle>
      <Memory>666</Memory>
      <User_Time>0:00:00</User_Time>
      <Kernel_Time>0:00:20</Kernel_Time>
      <Elapsed_Time>2:41:47</Elapsed_Time>
    </RunningProcess>
    <RunningService>
      <Name>HTTP Filter </Name>
      <Display_Name>HTTP SSL<Display_Name/>
    </RunningService>
    <Users>
      <Name>admin</Name>
      <LogOnDate>2006-06-20</LogOnDate>
      <LogOnTime>08:50:33</LogOnTime>
    </Users>
  </LiveData>
</Incident>
```

[그림 2] XML로 구조화한 라이브 데이터

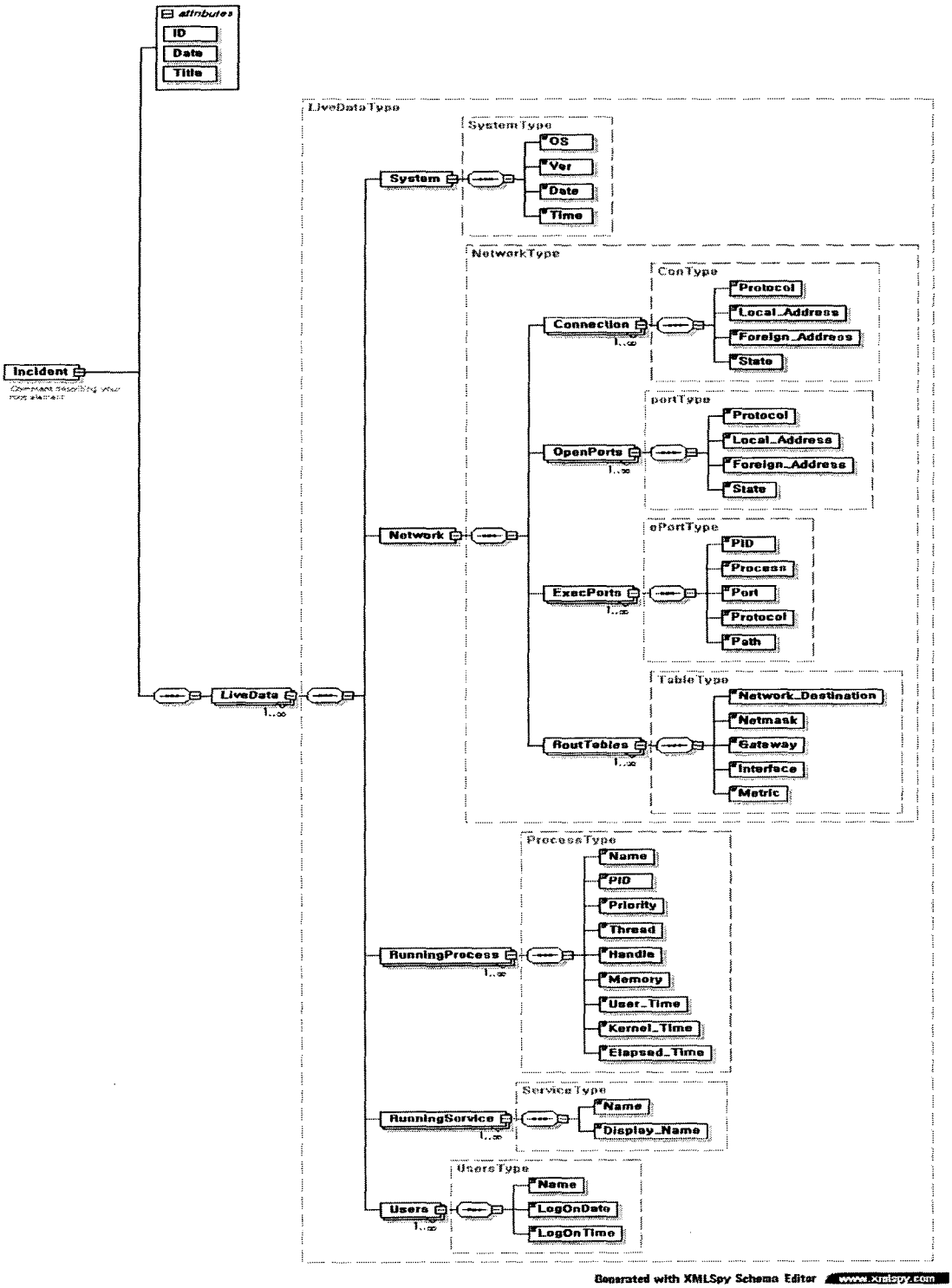

그림 3 .라이브 데이터를 XML 스키마로 정의

III. 결론

본 논문은 디지털 증거를 XML을 이용하여 구조화하고 이를 실제로 표현하는 방법을 제안해 보았다. 이와 같이 수집한 디지털 정보를 저장하기위한 데이터 구조를 XML로 정의하는 이유는 다음과 같다.

첫째, 디지털 증거를 단순히 텍스트로 표현하는 것이 아니라, XML로 구조화하면 직관적으로 이해하기 쉽고, 데이터를 처리하기에도 편리하다.

둘째, 디지털 범죄의 수사 결과에 대한 보고 양식 표준이 아직 정립되지 않았으므로, XML을 이용하면 웹에서의 데이터 교환이 효율적이다. 또한 향후 개발될 포렌식 툴에서 이렇게 구조화된 증거 수집양식을 지원한다면, 데이터 교환이나 이동 및 가공이 편리해진다.

셋째, XML 문서를 작성하는 사용자는 XML Parser를 이용, 필요한 데이터만 추출하여 다양한 관점에서 수사 결과를 표현, 보고할 수 있다.

넷째, 네트워크로 연결된 시스템에 XML 웹서비스를 이용한 포렌식 시스템을 구축하면, 사용자의 PC로부터 디지털 증거들을 주기적으로 수집, 모니터링 하여 사이버 범죄를 미연에 방지할 수 있다.

따라서 디지털 증거를 XML 기술을 이용하면, 다양한 부분에서 디지털 범죄 수사를 효율적으로 진행해 나갈 수 있으며, 향후 포렌식 툴들의 수사결과를 XML 문서로 표준화할 경우, 디지털 포렌식 기술 분야의 시너지 효과는 상당할 것으로 기대된다.

[참고문헌]

[1] Chris Prosise, Kevin Mandia, Incident Response & Computer Forensics (2nd Edition), McGraw-Hill, 2003-07-17

[2] Harlan Carvey, Windows Forensics and Incident Recovery, Addison-Wesley, 2004-07-15

[3] Brian Carrier and Eugene H. Spafford, "Getting Physical with the Digital Investigation Process", International Journal of Digital Evience, Fall 2003, Volume 2, Issue 2.

[4] A. Chris Bogen and David A.Dampier, "Preparing for Large-Scale Investigations with case Domain Modeling", presented at Digital Forensics Research Workshop, New Orleans, LA, 2005.

[5] V. Baryamureeba and F. Tushabe, "The Enhanced Digital Investigation Process Model",http://www.dfrws.org/(current 2005 January 11), 2004.

[6] David Hunter, Andrew Watt, Jeff Rafter, Beginning XML (3rd Edition), Wrox Press, 2004-08-01

[7] Steven Holzner, Real World XML, Pearson Education, 2003-01-15

[8] 신민철, XML 웹서비스, 프리렉, 2003-10-02

[9] 경찰청 사이버테러대응센터 홍보자료, http://www.ctrc.go.kr, 2006

[10] 임종인(고려대학교 정보보호대학원 원장, 디지털포렌식학회 회장), "외국의 디지털 포렌식 정책", 제1회 디지털 포렌식 세미나, 한국디지털포렌식학회, 2006-6-15

[11] 변정수(경찰청 디지털포렌식센터/사이버테러대응센터 연구관), "디지털 포렌식 도구와 기법", 제1회 디지털 포렌식 세미나, 한국디지털포렌식학회, 2006-6-15

[12] 충남 지방경찰청 사이버수사대, "디지털 포렌식 실제", 제1회 디지털 포렌식 세미나, 한국디지털포렌식학회, 2006-6-15

[13] Altova XMLSpy 2005 Standard Edition, http://www.altova.com,