

HMS를 기반으로 한 웹 상의 병원정보 통합 및 검색

양정옥⁰ 홍동완 윤지희 주한규
한림대학교 컴퓨터공학과

{jerry, dwhong}@center.cie.hallym.ac.kr, {jhyoon, hkjoo}@sun.hallym.ac.kr

HMS-based Integration and Retrieval of Hospital Information on the Web

JeongUk Yang⁰ DongWan Hong Jeehee Yoon Hankyu Joo
Dept. of Computer Engineering, Hallym University

요약

HMS(Hallym Mediator System)는 XML을 기본 데이터 모델로 하여 인터넷에 산재하여 있는 분산 이질 정보에 대한 통합, 검색 기능을 제공하는 미디어이터 시스템이다. 분산이질 정보의 공통 스키마 구조로서 XML DTD를 사용하며, 각종 정보에 대한 가상의 통합 뷰(view) 생성 기능을 제공하여 웹 상의 통합된 가상 정보를 표현한다. 실용성 및 성능평가를 위하여, HMS를 기반으로 하는 병원정보 통합/검색 시스템을 구현하였다. 병원정보 통합/검색 시스템은 가상접근 기법(virtual approach)기반의 정보검색 시스템으로서, 일반 사용자는 웹 상의 각종 병원 정보를 정보의 위치에 상관없이 비주요 사용자 인터페이스를 통하여 제공 받게된다.

1. 서론

인터넷의 성장은 정보의 양을 증가시켰으며, 그 정보의 형태도 다양하고, 내용적으로도 같은 혹은 유사한 정보가 인터넷의 여러 곳에 중복, 산재되어 위치할 수 있다. 이들 정보들은 대부분 유용한 정보로서 공유되어야 하며, 이러한 공유에 의하여 새로운 데이터의 창출을 가져올 수 있다. 따라서 정보의 통합은 필수적인 요건으로 대두되고 있다. 최근 분산 이질 환경의 데이터를 통합하기 위한 방법으로서 미디어이터[1] 개념이 널리 활용되고 있다.

본 연구실에서는 XML(eXtensible Markup Language)을 기본 데이터 모델로 하여 인터넷에 산재하여 있는 분산 이질 정보에 대한 통합, 검색 기능을 제공하는 미디어이터 시스템 HMS(Hallym Mediator System)를 개발하였다[2,3]. HMS에서는 분산이질 정보의 공통 스키마 구조로서 XML DTD를 사용하며, 각종 정보에 대한 가상의 통합 뷰(view) 생성 기능을 제공하여 웹 상의 통합된 가상 정보 구조를 표현한다. 웹 상의 일반 사용자는 이와 같이 생성된 뷰 DTD를 근거로 분산이질 정보에 대한 구조적, 내용적 질의를 수행할 수 있다.

본 논문에서는 웹 상에 존재하는 병원정보의 통합 및 검색을 지원하는 병원정보 통합/검색 시스템 개발에 대하여 논한다. 병원정보 통합/검색 시스템은 HMS를 기반으로 구현 되었으며, 사용자는 웹 상의 각종 병원 정보를 정보의 위치에 상관없이 XML DTD형의 통합된 가상 정보구조를 통하여 제공받게 된다. 기존 연구에서 미디어이터 개념을 활용한 시스템 개발 사례[4,5,6]가 다수 제안된 바 있으나, 이와같이 실제적인 응용 시스템을 통한 시스템의 실용성 및 성능 평가 사례는 보고된 바 없다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구에 대하여 간단히 설명하고, 3장에서는 HMS의 시스템 구성 및 질의 언어에 대하여 설명한다. 4장에서는 병원정보 통합/검색 시스템의 구성, 정보 통합 및 질의 평가방식, 시스템 보안에 대하여 설명한다. 마지막으로 5장에서는 결론 및 차후 연구과제에 대하여 논한다.

2. 관련 연구

정보통합 : 최근 웹 정보 통합을 위한 시스템 구성 방식으로, 데이터 웨어하우스(Data Warehousing)방식과 가상 접근(Virtual Approach)[7] 기법에 관한 연구가 활발히 이루어지고 있다. 데이터 웨어하우스 기법은 웹 소스로부터의 데이터를 시스템으로 가져와서 통합, 저장 한 후에, 웨어하우스에 대하여 모든 질의를 적용하는 시스템 구현 방식으로서, Lore(Lightweight Object Repository) 시스템[4] 등을 예로 들 수 있다. 가상 접근기법은 모든 정보가 분산 이질 형태로 웹 상에 그대로 존재한 상황에서, 가상 통합 뷰에 대한 질의가 실행시 각 소스에 대한 질의로 분해되어 평가되는 시스템 구성 방식으로 MIX(Mediation of Information using XML) 시스템

[5] 등을 예로 들 수 있다. 일반적으로 데이터 소스의 수가 많고 변화가 빈번하며, 소스에 대한 제약이 어려운 데이터 통합의 경우에는, 가상 접근 기법이 웨어하우스 기법보다 적합한 시스템 구현 형태로 알려져 있다. 그러나 가상접근 기법은 웨어하우스 기법에 비하여 효율면에 문제가 있을 수 있으므로 질의 평가를 위한 최적화 작업[6]을 위한 연구가 수반되어야 하며, 실용화를 위한 다각적인 연구가 필요하다.

의료 정보 시스템 : 최근 국내 병원에서는 내부적으로 산출된 모든 의료 데이터처리를 전산 자동화하는 추세이다. 각 부서에서 산출된 의료 데이터는 병원내의 타부서간 공유되어야 하며 또한 환자가 타 병원으로 이송되거나 수탁 진료가 이루어지는 경우 외부 병원과도 공유되어야 한다. 또한 시스템간의 데이터 공유가 가능하면 일반 사용자도 웹을 통하여 진료 예약이나 진단 결과 검색 등이 가능하여진다.

일반적으로 의료 데이터는 텍스트 데이터, 영상데이터, 일관적인 형태의 이진화 데이터(custom-formatted binary data) 등 다양한 데이터 포맷을 가지며 개개의 구성 시스템도 이질의 플랫폼으로 구성되어 있어, 각각의 시스템을 간단히 연계하여 자료를 공유하는 방법이 존재하지 않는다. 그 해결방안으로 시스템간의 스키마 공유, 표준 포맷의 데이터 전송을 들 수 있다.

이와 같은 의료 데이터의 전자적교환 표준으로 HL7(Health Level 7)[8]이 알려져 있다. HL7 표준은 데이터 표준, 메시징 표준, 문서 표준으로 이루어져 있으며, 데이터의 저장 및 교환을 위한 문서 표준으로서는 SGML(Standard Generalized Markup Language), HTML(HyperText Markup Language), XML 문서를 응용 범위로 참고 있다. 특히 2002년 상반기에 발표될 버전 3.0에서는 XML문서를 외부 전송 문서 표준으로 결정하고 있다.

3. HMS(Hallym Mediator System)

3.1 시스템 개요

HMS[2,3]는 가상 접근기법을 기반으로 하는 웹 정보통합/검색 시스템이다. 시스템의 기본 기능은 통합 뷰 생성 및 관리 기능과 정보검색 기능으로 나눌 수 있다.

통합 뷰 생성/관리 기능 : 미디어이터 관리자는 웹 상의 각종 소스에 산재되어 있는 정보를 통합, 가공하여 응용 목적에 적합한 가상의 통합 뷰를 정의, 구축하여야 한다. HMS에서는 웹 상의 정보 소스에 대한 각종 정보(소스 내용, 속성 정보, 제약 조건, 소스 신뢰도, 질의처리 능력)를 메타 데이터 형태로 수집, 저장, 관리한다. 관리자는 이들 메타 데이터를 기반으로 사용자 인터페이스에서 제공되는 가상 DTD 브라우징 기능, 뷰 DTD 편집 기능 등을 이용하여 통합 뷰를 정의한다. 통합 뷰를 생성하는 과정에서 웹 상의 각종 소스와 통합 뷰와의 소스 매핑 정보가 자동 추출, 관리된다.

정보 검색 기능 : 사용자는 사용자 인터페이스를 이용하여

적합한 뷰를 구성시킨 후, 분산이질 정보에 대한 구조 정보 및 내용 정보에 대한 질의를 수행할 수 있다. 뷰 DTD를 근거로 작성된 사용자의 질의는 뷰 정의시에 생성된 소스 매핑 정보를 참조하여 소스정보에 대한 실제 질의로 변환되어, 실행된다.

3.2 HML(Hallm Mediator Language)

HMS에서는 뷰 정의와 정보검색을 위한 미디어터 언어 HML[3]를 제공한다. HML은 웹 질의어로서 XML-QL[9], Yal, UnQL등의 XML질의 언어와 등등한 기능을 가지는 고수준 선언적 질의언어로서, HMS시스템에서 최적화되어 동작을 할 수 있도록 개발, 구현되었다.

4. 병원정보 통합/검색 시스템

4.1 시스템 개요

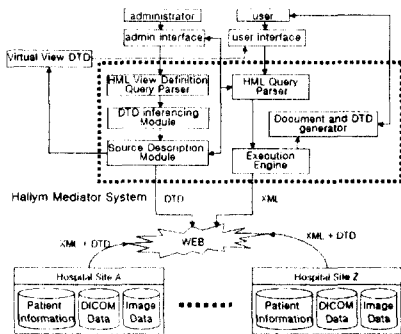
병원정보 통합/검색 시스템은 HMS를 기반으로 하며, [그림1]과 같이 사용자 인터페이스, 미디어터, 래퍼를 기본 구조로 한다. 사용자 인터페이스는 통합 뷰 생성을 위한 관리자용 인터페이스와 정보검색을 위한 일반 사용자 인터페이스로 이루어져 있다. 미디어터는 뷰 정의/질의 파싱모듈, DTD 추론 모듈, 질의 실행 모듈, 결과(XML문서와 DTD) 생성 모듈, 소스 매핑정보 관리 모듈 등으로 이루어져 있으며, 래퍼는 인터넷상의 정보 매핑을 위하여 소스 정보별로 작성된다. 각 모듈의 기능과 구현 방식은 다음과 같다.

(1) DTD 추론 모듈 : DTD 추론 모듈은 정확한(tight) 최소의 뷰 DTD를 생성한다. 임의의 소스 DTD $d_1 \sim d_n$ 과 이에 대한 뷰 정의가 주어진 경우, 소스 DTD $d_1 \sim d_n$ 로부터 뷰 정의를 만족하는 뷰 DTD는 여러 개가 존재할 수 있다. DTD $d_1 \sim d_n$ 과 뷰 정의를 만족하는 뷰 DTD v_1, v_2 에 대하여 $v_1 \subset v_2$ 의 관계가 성립할 경우, v_1 이 v_2 보다 더 정확하다고 말한다. 본 시스템의 DTD 추론 모듈은 표준적인 DTD에서 정확한 뷰 DTD를 생성하도록 최적화되어 있다[3].

(2) 소스 매핑 정보 관리 모듈 : 사용자에게 질의 데이터에 대한 위치 투명성을 제공하기 위하여 메타 정보인 소스매핑 정보를 사용한다. 소스 매핑 정보는 웹 상의 정보 소스에 대한 각종 정보(소스 내용, 속성 정보, 제약 조건, 소스 신뢰도, 질의 처리 능력), 뷰 DTD와 소스 DTD와의 내용 관계를 나타내는 매핑정보, 뷰 DTD의 엘리먼트들에 대한 값 질의 정보 등으로 이루어진다. 소스 매핑 정보 관리 모듈은 질의 처리시 뷰정의문으로부터 소스 매핑 정보를 자동 추출하여, 일반 데이터와 같이 XML 문서 형태로 저장, 관리한다.

(3) HML 질의 파서 : 사용자는 직관적인 비주얼 사용자 인터페이스를 통하여 질의를 생성할 수 있다. 즉, 사용자는 HML문법을 알지 못해도 질의 구성이 가능하다. HML 질의 파서는 사용자가 인터페이스를 통해 생성한 질의 정보와 소스 매핑 정보 관리 모듈에서 받은 메타 정보를 이용하여 완전한 HML질의를 구성한다. 완성된 HML 질의는 파싱되어 각각의 문서별로 실행 전략이 수립되어 실행엔진에 전달된다.

(4) 실행 엔진 : 실행 전략에 따라 실행 엔진은 분할된 서브 질의에 대한 결과를 각 래퍼에 요청한다. 래퍼는 서브 질의에 해당하는 정보를 검색하여 이를 XML문서 형태로 실행엔진에 전송하며, 실행 엔진은 이들 결과를 취합한다.



[그림1] HMS 구조도

(5) 래퍼 : 본 시스템에서 사용되는 의료 데이터는 환자 기초 정보와 의사의 진단 정보, 의료 장비로부터 산출된 영상 데이터 등으로 구분된다. 래퍼는 소스별로 작성되며 예를 들어 관계형 데이터베이스에 저장된 환자 기초정보 등은 다음과 같이 검색된다. 사용자 인터페이스를 통하여 검색하기를 원하는 항목과 조건에 대해서 작성된 질의는 HML 질의 파서에 의하여 분석되어, XML 문서 형태로 변환된다. 이 XML 문서에는 질의의 의미적, 구조적 정보가 포함된다. XML 문서를 전해 받은 래퍼는 전달된 구분 정보를 분석하여 이를 관계형 데이터베이스의 일반 질의문(SQL) 형태로 변환한다. 관계형 데이터베이스 시스템에 의하여 검색된 결과 데이터는 다시 래퍼에 의하여 XML 문서형태로 작성되어 실행엔진에 전달된다.

(6) XML 문서/DTD 생성기: 실행 엔진에 의하여 검색된 결과는 결과 생성 모듈에 의하여 사용자 질의의 생성절에 명시한 엘리먼트 구조에 맞게 XML문서로 변환된다. 이때 결과 XML 문서와 함께 DTD정보가 자동 생성된다. 따라서 이를 이용하여 생성된 결과 XML문서의 유효성 검사 및 결과 XML문서에 대한 재 질의가 가능해진다.

4.2 정보통합과정

본 시스템에 의한 웹 상의 병원정보 통합/검색 과정을 예를 들어 단계별로 설명하면 다음과 같다. 시스템 관리자는 관리자 인터페이스를 통하여 일반 사용자에게 제공할 가상 통합 뷰를 작성한다. 각 병원의 의료정보 구조를 나타내는 원본 DTD들이 관리자 인터페이스에서 트리 형태로 보이게 되며, 시스템 관리자는 이 원본 DTD들을 브라우저 하에서 뷰 정의의 질의를 구성한다. [그림2]는 병원의 영상정보와 환자의 신상정보를 나타내는 DTD의 예이며, [그림3]은 [그림2]의 DTD를 근거로 관리자가 구성한 HML 뷰 정의문의 예이다. 뷰 정의문은 각 해당 DTD로부터 1999년 1월 1일부터 2000년 12월 31일까지의 해당 기간에 보관된 모든 영상정보와 환자 정보, 환자 이름, 환자 주민번호, 담당의사이름, 진로일, 진단정보를 추출하라는 의미를 가진다.

dicom.dtd : DICOM정보	patient.dtd : 환자정보
<ELEMENT dcms (dcms)*>	<ELEMENT patients (patient)*>
<ELEMENT dcms (name, providers)*>	<ELEMENT patient (chnum, name, jumin, icode, name, doctor, date, study)*>
<ELEMENT name (#PCDATA)>	<ELEMENT chnum (#PCDATA)>
<ELEMENT providers (provider*, img_data)*>	<ELEMENT name (#PCDATA)>
<ELEMENT provider (group, groupname)*>	<ELEMENT icode (#PCDATA)>
<ELEMENT group (service)*>	<ELEMENT icode (#PCDATA)>
<ELEMENT groupname (#PCDATA)>	<ELEMENT name (#PCDATA)>
<ELEMENT service (id, element_name, value)>	<ELEMENT doctor (#PCDATA)>
<ELEMENT id (#PCDATA)>	<ELEMENT date (#PCDATA)>
<ELEMENT element_name (#PCDATA)>	<ELEMENT study (#PCDATA)>
<ELEMENT value (#PCDATA)>	<ELEMENT study (#PCDATA)>
<ELEMENT img_data (#PCDATA)>	

[그림2] 예제 DTD

이와 같이 구성된 질의는 뷰 정의 파서에 의하여 파싱이 되고 DTD 추론계획이 수립이 된다. 추론 계획에 의하여 정확한 뷰 DTD가 생성되며, 동시에 뷰 정의문의 값반 질의문에 관한 정보를 나타내는 값질의 정보와 소스 매핑정보가 자동 추출되어 소스 매핑 정보로 관리된다. 다음의 [그림4]는 DTD추론 모듈에 의해 생성된 통합 뷰 DTD를 나타낸다.

```

<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<createview name="patients" >
  <view>
    <dcms data>
      <dcms.elements> $dcms />
      <dcms.value> $v />
      <dcms.image> $id />
    </dcms data>
    <patient data>
      <chnum> $chn />
      <name> $nm />
      <jumin> $jm />
      <icode> $ic />
      <date> $dt />
      <study> $st />
    </patient data>
  </view>
  <where>
    <dcms>
      <providers>
        <provider>
          <group>
            <service>
              <element_name> $en />
              <value> $v />
            </service>
          </group>
        </provider>
      </providers>
    </dcms>
  </where>
  <select>
    <dcms data> $id />
  </select>
  <in http://hyacinth.ce.hallym.ac.kr/dicom.dtd
  </in>
</createview>
  
```

[그림3]뷰 정의 예제

이와 같이 생성된 뷰 DTD는 사용자 인터페이스를 통해서 사용자에게 보여지고, 일반 사용자는 뷰 DTD를 보면서 직관적

인 질의를 구성한다. 사용자 인터페이스는 [그림5]와 같이 '환자들이 Magna Karta, 진료일이 1999년 7월 25일인 환자의 영상정보와 환자이름, 환자번호, 차트번호, 진료일, 진단정보, 담당의사를 출력하라'는 질의를 나타내고 있다. 일반 사용자는 뷰 DTD의 엘리먼트들이 웹상의 어느 정보인지를 알 필요가 없다. 비주얼 사용자 인터페이스를 통하여 구성된 사용자 질의는 HML 질의 파서에 의하여 소스 매핑정보를 참조하여 각 소스별로 작성된 완전한 HML 질의문으로 변환된다. 생성된 질의는 다시 HML 질의 파서에 의해 파싱이 되고, 실행 계획이 수립되어 실행엔진으로 수평계획을 보낸다. 실행 엔진은 실행 전략에 따라 분할된 서브 질의에 대한 결과물을 각 래퍼에 요청한다. 각 래퍼는 XML 형태로 요청받은 서브 질의를 각 시스템 형태에 맞추어 변환, 처리한 후 평가 검색결과 다시 XML 형태로 변환하여 실행엔진에 전달한다. 단, 병원 시스템이 XML문서를 외부 전송 문서 표준(HL7)으로 사용하고 있는 경우, 별도의 래퍼를 필요로 하지 않는다. 실행 엔진은 래퍼들로부터 전달받은 결과 데이터들을 XML 문서/DTD 생성기로 전달한다. 여기에서 사용자 질의문의 생성 질의 명시된 형태로 결과 문서와 DTD가 생성되며, 스타일시트를 구성하여 결과 문서로 사용자에게 최종적으로 전달한다. [그림6]은 이와 같은 과정에 의하여 생성된 결과 문서의 예제를 보인다.

```

<ELEMENT vldid (#text)>
<ELEMENT view (dicom_data patient_data)
<ELEMENT dicom_data (dicom_value dicom_elements dicom_image)
<ELEMENT patient_data (pchnum, pdoctor, pumin, pname, pstudy, pdtale)
<ELEMENT pchnum (#chnum)
<ELEMENT pdoctor (doctor)
<ELEMENT pumin (#umin)
<ELEMENT pname (#name)
<ELEMENT pstudy (#study)
<ELEMENT pdtale (#date)
<ELEMENT dicom_value (#value)
<ELEMENT dicom_elements (element_name)
<ELEMENT dicom_image (#img_data)
<ELEMENT element_name (#PCDATA)
<ELEMENT value (#PCDATA)
<ELEMENT img_data (#PCDATA)
<ELEMENT chnum (#PCDATA)
<ELEMENT name (#PCDATA)
<ELEMENT umin (#PCDATA)
<ELEMENT doctor (#PCDATA)
<ELEMENT date (#PCDATA)
<ELEMENT study (#PCDATA)
    
```

[그림4] 뷰 DTD

4.3 보안

의료 정보는 개인정보, 진료 기록 등으로 구성될 수 있으며, 개인 사생활 보호, 인권침해의 문제로 보안을 요구하게 되고 인터넷과 같은 공개망을 사용하면서도 필요한 사용자나 컴퓨터만이 의료 정보를 사용하기 위해서 보안이 필요하다.

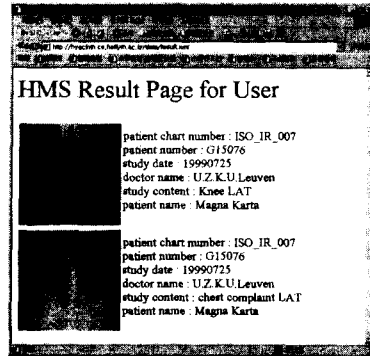
사용자 인증 : 일반적인 인증방법은 아이디와 패스워드를 이용하여 인증하는 방법이다. 이러한 인증방법은 패스워드에 대한 추측이나 패스워드 전수 조사에 취약한 단점을 보인다. 사용자 인증은 전자서명에 기반한 challenge-response 기법[10]을 사용한다. 사용자는 공개키 암호화 기법에 사용되는 자신의 공개키, 비밀키 쌍을 가지고 있으며 비밀키를 안전하게 보호하고 있다고 가정한다. 공개키 알고리즘은 RSA[11]를 사용하며 해쉬 함수는 SHA-1[12]을 사용한다.



[그림5] 사용자 질의 예제

정보전송 보안 : HMS시스템에서는 웹 기반으로 XML문서를 전송하는 방식으로 사용자와 시스템, 시스템과 시스템이 통신한다. 따라서, 본 시스템은 문서단위의 암호화가 가능한

SSL(Secure Socket Layer)/TLS[13]를 사용하여 문서를 통한 통신의 보안을 해결한다. SSL은 netscape사에 의하여 웹 보안을 위하여 개발되었으나 일반적인 프로토콜의 보안에도 이용될 수 있다.



[그림6] 결과 문서

5. 결론

본 시스템은 윈도우 NT기반에서 JDK1.2를 이용하여 구현하였으며, XML파서로는 IBM Parser for Java를 이용하였다. 본 시스템은 의료 정보에 대한 직관적이고 쉬운 질의 환경을 제공하는 것을 그 목적으로 두고 있으며, 이를 위해 XML DTD를 기반으로 하는 가상 뷰 기능을 제공하고 있다. 본 시스템은 현재 프로토타입이 가동 중이며, 보안 모듈과 인증 과정 등 시스템 기능 강화에 관한 연구를 진행 중이다.

참고문헌

- [1] Wiederhold, G., "Mediators in the Architecture of Future Information Systems," IEEE Computer Vol.25, No.3, 1992, pp. 38-49.
- [2] 홍석일, 양정욱, 홍동완, 윤지희, "XML의 View DTD를 이용한 웹 상의 정보통합 및 검색," 한국정보과학회 학술 발표 논문집, Vol 27, No 2, 2000, pp. 231-233
- [3] 양정욱, 홍동완, 이덕형, 윤지희, "웹 정보 통합 및 검색을 위한 XML기반 미디어터 시스템의 개발," 한국데이터베이스 학회 2001년 춘계 Conference, 2001, pp. 281-294
- [4] McHugh, J., Abiteboul, S., Goldman, R., Quass, D., Widom, J., "Lore: A Database Management System for Semistructured Data," SIGMOD Record, Vol.26, No.3, 1997, pp. 54-66.
- [5] Baru, C. K., Gupta, A., Ludascher, B., Marciano, R., Papakonstantinou, Y., Velikhov, P., Chu, V., "XML-Based Information Mediation with MIX," SIGMOD Conference, 1999, pp. 597-599.
- [6] Florescu, D, Levy, A., Mendelzon, A., "Database Techniques for the World-Wide Web: A Survey," SIGMOD Record, Vol.27, No.3, 1998, pp. 59-74.
- [7] Ludascher, B., Papakonstantinou, Y., Velikhov, P., "Navigation-Driven Evaluation of Virtual Mediated Views," EDBT, 2000, pp. 150-165.
- [8] "Health Level 7," <http://www.hl7.org>
- [9] Deutsch, A., Fernandez, M., Florescu, D., Levy, A., Suci, D., "XML-QL - A query language for XML," In Proc. Of the Int. World Wide Web Conf., Canada, 1999
- [10] ISO/IEC 9798-3, Information Technology Security techniques Entity authentication mechanism Part 3: Entity authentication using a public-key algorithm, International Organization for Standardization, 1993.
- [11] R. Rivest, A. Shamir, and L. M. Adleman, A method for digital signature and public-key cryptosystems, *Communications Of the ACM*, vol. 21, pp. 120-126, 1978.
- [12] FIPS 180-1, Secure hash standard, Federal Information Processing Standards Publication 180-1, U.S. Dept. of Commerce / NIST, 1995.
- [13] A. Freier, P. Karlton, and P. Kocher, The SSL Protocol Version 3.0, Internet draft, 1996.