

IPv6 기반 다국어 도메인 네임 시스템에 관한 연구

최지원*, 김기천*

*건국대학교 컴퓨터공학과

e-mail : jackeroo@konkuk.ac.kr ; kckim@konkuk.ac.kr

A Study on Internationalized Domain Name System based on IPv6

Ji-Won Choi*, Kee-Cheon Kim*

*Dept. of Computer Science & Engineering, Kon-Kuk University

요 약

인터넷의 급격한 발달로 인하여 인터넷 영역과 사용자 수가 전 세계적으로 급격히 확산 및 증가하고 있다. 국내의 사정만을 보더라도 광대역(초고속) 인터넷망과 PC 보급률의 증가로 2 천만 명을 넘는 인구가 인터넷을 사용하고 있는 것으로 집계되었다. 이와 같은 폭발적인 인터넷 사용자의 증가로 인하여 현재의 인터넷상에 많은 문제점들이 제시되기 시작하였으며, 이를 해결하기 위하여 다양한 연구가 활발하게 진행되고 있다. 그 중에서도 도메인 네임 시스템과 관련된 문제들은 인터넷의 핵심 기반 기술 중의 하나로서 많은 부분에서의 보완과 개선이 요구되고 있다. 본 논문에서는 현재 표준화로 될 것으로 예상되는 IDNA-punycode 기반으로 mBIND를 설계하였다.

1. 서론

기존의 인터넷 도메인 네임 시스템은 ASCII 문자 중 일부의 조합(알파벳(a-z, A-Z), 숫자(0-9), 특수문자(-))만으로[RFC 1035]이루어진 도메인 네임만을 지원함으로써 대한민국, 중국, 일본과 같은 비 영어권 국가들은 자국어로 도메인 네임을 표현할 수가 없었다. 즉, 모든 도메인 네임을 영어로 표시해야 하기 때문에 실제로 전 세계 인구의 68%에 해당하는 비영어권 지역의 사용자 영어를 어느 정도 알고 있어야만 하는 있으며, 이는 인터넷의 대중화, 도착화 및 글로벌 인터넷의 정착에 큰 장애 요인 중 하나로 작용하고 있다. 이러한 문제점을 해결하기 위하여 많은 국가들이 오래 전부터 다양한 연구를 수행하여 왔으며, 그 대안 중에 하나가 비영어권 국가들이 자국어로 된 도메인명을 사용하는 것이다. 비영어권 국가들이 자국어로 된 도메인을 사용

하고자 한다면 이들 도메인 이름을 현재의 인터넷 도메인 규정에 맞게끔 바꾸어 주는 시스템이 필요한 것이다. 이런 생각에 기초하여 CIR(Center for Internet Research)에서 추진하게 된 프로젝트가 바로 iDNS 이다. 즉, iDNS 은 비영어권 국가들이 알파벳이 아닌 코드를 사용하여 도메인 이름을 사용하고자 할 때 이를 현재 도메인 규약에 적합한 UTF-5 코드로 변환시켜 주는 시스템이다. 이의 결과로 많은 다국어 지원 도메인 네임 시스템(iDNS : Internationalized Domain Name System)이 개발되었다. iDNS 을 이용하면 중국어, 일본어, 한국어, 프랑스어, 독일어, 타밀어, 아랍어 등 어떤 언어로도 인터넷 도메인 이름을 등록할 수 있다. 또한 이 시스템은 현재 사용되고 있는 모든 인터넷 브라우저에서 어떤 충돌도 없이 사용될 수 있으며 현재 사용되고 있는 도메인 이름 시스템과의 호환성에도 큰 문제가 없다.

2. 관련 기술

2.1 인코딩 기술

현재 다국어 지원 도메인 네임 시스템 관련 기술분야는 다양한 연구 및 개발이 진행 중이다. 이러한 활동은 주로 각 국가의 인터넷 정보 센터 및 도메인 네임 서비스 관련 업체를 중심으로 이루어지고 있으며, 관련 조직으로는 IETF의 IDN(Internationalized Domain Name) 워킹 그룹, JET, MINC 등의 미팅/커미티 및 KRNIC, JPNIC, CNNIC, TWNIC 등 다양한 국가의 인터넷 정보 센터를 들 수 있다.

IETF의 IDN(Internationalized Domain Name) 워킹 그룹은 IETF의 워킹 그룹 중, Internet Area의 분류에 속하며, 다국어지원 도메인 네임을 사용하는데 필요한 사항들을 정의하고, 그 요구사항을 기반으로 프로토콜 및 인코딩 방식 등을 정의하는 목적을 가지고 있다. 지난 영국의 런던에서 열린 51차 IETF 회의 중 IDN 워킹 그룹 회의에서는 다음과 같은 사항이 논의되었다. IDN 워킹 그룹 회의에서는 다국어 지원 도메인 네임 시스템의 다양한 프로토콜 방식과 인코딩 방식 중에서 각각 하나씩을 적극 지원하기 시작함으로써, 표준화에 근접해 갔다고 할 수 있다. 프로토콜 분야에서는 IDNA 방식에 다수가 동의하였으며, 인코딩 분야에서는 AMC-ACE-Z를 다수가 지지하였다.

우선, 다국어 도메인(IDN : Internationalized Domain Name)을 다국어 지원 도메인 네임 시스템(IDNS : Internationalized Domain Name System)에 보내기 위하여 표현 방식을 변경하는 인코딩 방식의 경우 기존의 DNS에 영향을 최소화하기 위하여 8bit 기반의 UTF-8 대신, 기존의 DNS를 그대로 사용할 수 있는 7bit 기반의 ACE(ASCII Compatible Encoding)방식을 다수가 지지하였으며, 그 중에서도 DUDE (Differential Unicode Domain Encoding)과 AMC-ACE-Z가 경합을 벌였다. 이 중 퍼포먼스 및 수행 결과 등에서 우세한 성능을 보인 AMC-ACE-Z가 다수의 지지를 받았다.

다음으로, 프로토콜에 대한 표준화를 들 수가 있다. 이번 회의에서는 인코딩 등의 처리를 클라이언트 어플리케이션에서 주로 처리하는 IDNA와 DNS 서버에서 주로 처리하는 uDNS(IDNS), 그리고 Unique Identifiers/Names를 이용하여 서비스를 제공하자는 uName이 논의되었는데 이 중에서 클라이언트 어플리케이션에서 인코딩처리를 수행하는 IDNA가 다수의 지지를 받았다. 그러나 기존의 (legacy) OS나 어플리케이션을 IDNA 방식으로 모두 전환하기에는 비용 및 시간적인 측면에서 큰 어려움이 따르므로, 기존 OS/어플리케이션의 패치를 제공하거나 인코딩 되지 않은 도메인명이 들어올 경우 서버에서 인코딩을 처리해 주어야 한다는 논의가 있었다. 이는 앞으로 개발될 인터넷 어플리케이션뿐만이 아닌 기존의 어플리케이션을 사용하는 사용자도 고려하기 위한 것이다.

이와 함께 다국어 지원 도메인 네임 시스템 상에서 효율적인 도메인명 해석을 위하여 도메인명에 대한 준비사항과 지양해야 할 캐릭터 등에 대한 가이드라인을 제시하여 주는 NAMEPREP에 대하여도 많은 논의가 이루어 졌다. 이는 한국어, 중국어, 일본어 등 다양한 언어가 존재하고 있고, 도메인 이름은 자연어에 대한 애기가 아니므로 실제 도메인명에 사용되는 캐릭터는 한정되어 있기 때문이다. 또한 각 언어는 IETF IDN 워킹 그룹 Charter에 없기도 하므로(IDN 워킹 그룹 표준화 범위 외라는 의견) 최종적으로 현재까지 나와 있는 각 언어의 nameprep draft를 통합하는 작업이 힘든 과정이기는 하지만 꼭 필요하다는 결론을 도출하였다.

IETF IDN 워킹 그룹 회의에서는 AMC-ACE-Z와 IDNA가 많은 지지를 받았으며, NAMEPREP 관련 문서(nameprep, tsconf, hangeul, jpchar)의 관계가 명확히 이루어져야 한다는 의견과 reordering은 향후 메일링 리스트를 통한 토론을 진행하자는 의견이 제시되었다. 제 52차 IETF 회의(2001.12.10)에서는 최종적으로 IDNA, Nameprep, Punycode에 대한 draft가 last call로 가도록 합의되었고 2002년 10월 25일 IESG에서는 IDN proposed standards로 표준을 확정하였다. (IDNA-14, Nameprep-11, Punycode-03, 아래 버전 및 I-D 명칭 참조)

다국어도메인뿐 아니라, 프로토콜 식별자, 회사명/성명, 다국어도메인네임 등 문자열의 사용과 범위를 정하는 Stringprep 과정은 매핑(Mapping) → 유니코드 정규화(Normalization) → 금지문자 체크(Prohibition) → 양방향 문자체크(Checking Bidi)의 과정을 거치게 되는데 다국어 도메인의 Nameprep 과정은 Stringprep의 과정 및 테이블을 참조하게 됩니다. 최근에 변동된 사항은 참조할 유니코드의 버전이 3.2로 전면 개정된 점과 아랍어/라틴어와 같이 오른쪽에서 왼쪽으로 읽히게 되는 스크립트에 대한 처리과정(Bidi)이 추가된 점이다.

2.2 AMC-ACE-Z

AMC-ACE-* 인코딩 방식은, M, O, R, W, V 버전이 드래프트로 공개되어 있으며, 가장 최신의 버전은 AMC-ACE-Z 버전이다. AMC-ACE-* 인코딩 방식은 인코딩 효율과 정확성 면에서 가장 많은 점수를 받음으로써 차후 표준안이 될 가능성이 가장 많은 인코딩 방식이다. AMC-ACE-* 시리즈는 인코딩/디코딩 알고리즘의 구현이 용이하며, 최적화된 알고리즘을 통하여 레이블 길이제한(63 캐릭터)에 적용하기에 매우 효과적이다.

2.3 IPv6 DNS transition issues

IPv6 주소는 AAAA 레코드를 사용하여 표현하고, 리버스 존은 PTR 레코드를 사용하여 nibble 포맷의 경우 ip6.arpa를 사용한다. 모든 recursive DNS는 IPv4와 듀얼 스택을 지원하여야 하고, single DNS의 경우 최소한 IPv4로 접속이 가능하여야 한다. 이것은

IPv6 만 가능한 DNS 는 안 된다는 뜻이다. 주소는 [IPv6ADDRARCH]에서 link local, site local, global 주소로 세가지로 정의 되어 있다.

리버스 패스는 IPv6 의 경우 그 길이가 매우 길기 때문에 다음과 같이 와일드 카드를 써서 나타낼 수 있다.

```
*.0.1.2.3.4.5.6.7.8.9.a.b.c.ip6.arpa. IN PTR
customer-42.local-ISP.net
```

IPv6 순방향 및 역방향 맵핑

A 레코드 대신 AAAA 를 128 비트 IPv6 주소를 저장하기 위해 사용하고 새로운 IPv6 역방향 맵핑 도메인인 ip6.int 도 있다.

AAAA 레코드와 ip6.int 역방향 맵핑의 단점은 네트워크 리넘버링(번호 재부여)을 어렵게 만든다.

예) NLA 를 바꾼다면 영역 데이터 파일에 있는 AAAA 레코드를 모두 변경해야 한다. 24 비트가 NLA 를 나타내기 때문이다.

또, NLA 가 TLA 를 바꾸는 경우도 마찬가지다.

A6 레코드와 순방향 맵핑

리넘버링을 쉽게 하기 위해 A6 레코드는 IPv6 주소 중 호스트의 네트워크 인터페이스에 할당된 마지막 64 비트 (인터페이스 ID)만 기입한다. 그리고 나머지 부분은 도메인 네임을 이용해 언급한다.

DNAME 레코드와 역방향 맵핑

IPv6 주소에 대한 역방향 맵핑은 DNAME 레코드와 비트 문자열 레이블을 포함한다.

비트문자열 레이블은 어떤 도메인 네임에서 연속된 긴 2 진수를 간결하게 표현하는 방법이다. 비트문자열 레이블은 연속된 레이블의 비트를 짧게 16 진수, 8 진수, 2 진수 또는 점구분 문자열(dotted-octet)로 연결한다. 문자열은 “[“ 와 ”] ” 로 둘러싸여 있어 전통적인 레이블로 구분된다.

DNAME 과 비트문자열 레이블은 IPv6 주소를 나타내는 긴 도메인 네임의 일부분과 매치하기 위해 사용된다. 또 그 IPv6 주소를 갖는 호스트를 관리하는 기관이 제어하는 도메인 네임을 반복적으로 탐색하는 데도 사용된다.

2.4 Punycode

Punycode 는 일반적으로 특별한 인자 값을 가진 Bootstring 의 한 단계이다. 비 ASCII 문자열이나 ASCII 문자로 이루어진 도메인 이름은 특별한 ACE 프리픽스로 시작하는 ACE label 에서 표현이 가능하다. 프리픽스 뒤에 오는 나머지 이름은 일정한 규칙에 만족을 시키는 유니코드 문자가 Punycode 인코딩 방식이다. Punycode 는 bootstring 으로 불리는 일반적인 알고리즘의 한 방식으로, 부트스트링은 일반적인 코드의 작은 세트로 구성된 문자열을 유일무이한 표현을 하기 위해서 더욱 큰 문자세트를 끌어 들이는 것이다. Punycode 는 IDNA 를 지원하기 위해서 특별한 인자 값을 가진 부트스트링이다. Punycode 는 완전성, 유일성, 가역성, 유용한 인코딩, 단순성을 만족시킨다.

완전성 : 확장된 모든 문자열은 기본문자열로 표현이 가능하다.

유일성 : 확장된 문자열은 많아도 한가지의 기본 문자열이다.

가역성 : 어떠한 변환된 확장된 문자열도 이전 상태로 변환이 가능하다.

유용한 인코딩 : 확장된 문자열의 길이가 제한 길이에 만족한다.

단순성 : 인코딩과 디코딩 방법은 구현하기 쉬운 이치에 맞아야 한다.

3. IPv6 기반 IDNS 스펙 설계

지난 IETF 회의에서는 다국어(8bit)를 영어, 숫자 및 하이픈(7bit)으로 인코딩하는 방안으로 제시된 ACE 방안들(DUDE, Punycode, RACE, MACE 등)에 대한 평가결과, 압축률과 Namespace 의 길이 등을 고려할 때, Punycode 가 가장 우수한 것으로 나타났다. 또한 ACE 를 수행하는 곳도 네임서버측과 어플리케이션측에서 하는 방안이 논의되었는데 이는 기존 DNS 체계의 안정성을 위해 Infrastructure(네임서버, 메일서버 등)는 변경하지 않고 도메인 이름을 이용하는 어플리케이션을 모두 수정하여 다국어 도메인을 서비스하는 방안인 IDNA(Internationalizing Domain Names in Applications)가 많은 지지를 받았다.

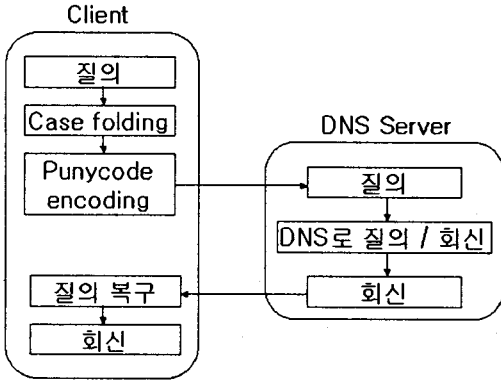
IETF IDN WG 메일링 리스트에서는 IDNA 이외에도 네임서버 인프라에서 사용될 다국어도메인용 별도 클래스(Class) 및 RR(Resource Record)를 생성하는 제안 및 동적 CNAME 을 이용한 네임서버 존 운영방식과 DNS 에서 UTF-8/ACE 존 데이터의 설정에 따른 장단점 비교 및 문제점 등에 대해 활발한 토론이 진행되었고, 현재는 IDNA 가 현 DNS 프로토콜에 변화를 유발하는지 여부에 관해 토론을 하고 있다.

본 논문에서는 현재 표준화로 될 것으로 예상되는 IDNA - punycode 기반으로 mBIND 를 설계하였다. 따라서 질의는 클라이언트 측에서 punycode 로 encoding 하여 질의하게 되고 DNS 에서는 encoding 된 zone file 에서 질의와 비교를 한 후에 응답을 하면 클라이언트 측에서 응답에 대하여 그 결과를 복구하는 구조이다. 이때 IPv6 의 경우 새로운 타입 AAAA 와 reverse zone file 에서 IPv6 의 주소가 처리 가능하여야 한다.

3.1 mBIND 구조

1) mBIND 클라이언트

mBIND 프로토콜 스펙을 따르는 클라이언트 측의 resolver 는 canonicalization 을 위해 case folding 을 하여야 한다.



case folding 을 client 측에서 하는 이유는, 단순한 ASCII 문자의 case folding 에 비해 유니코드의 case folding 은 복잡하고, 많은 부하가 될 수 있기 때문에, 모든 case folding 을 서버에서 하는 것은 무리가 있기 때문이다. 유니코드의 case folding 은 어떤 알고리즘에 의해 계산되는 것이 아니라, mapping 테이블을 사용하여 구해진다.

클라이언트 측의 resolver 의 작동 순서는 다음과 같다.

- (1) 유니코드로 된 다국어 질의 또는 ASCII 로 된 질의를 받는다. resolver 는 이를 case-folding 하여 변형하기 전에 원래의 질의 형태를 저장한다. Canonicalization 을 위해 질의를 case-folding 하지만, 요청을 되돌려 줄 때는, case-folding 이 되지 않은 원형의 상태로 되돌려 주기 위해 저장을 한다.
- (2) Case-folding 을 실시한다.
- (3) punycode 인코딩을 한다.
- (4) Name server 로 질의를 하여, 회신을 받는다.
- (5) 저장해 놓았던 원형의 질의로 복구한다.
- (6) 질의를 한 응용프로그램에게 질의 결과를 회신한다.

2) mBIND 서버 resolver

mBIND 프로토콜 스펙을 따르는 서버의 작동 방식은 다음과 같다.

- (1) 쿼리의 "puny--" prefix 를 확인하여 인코딩 여부를 판별한 후 인코딩이 되어 있는 경우 다음 단계로 넘어가며, 인코딩이 되어 있지 않을 경우 현재 인코딩을 하지 않고 서버로 질의하는 클라이언트에 대하여 올바른 응답을 보내주기 위하여 인코딩을 수행한다.
- (2) mBIND 서버에서는 punycode 로 인코딩 되어 있는 자신의 존 파일에서 검색하거나 다른

DNS 로 전달한다.

- (3) 요청 받은 클라이언트 또는 서버에게 회신을 한다.

4. 결론

국내에서는 2000~1 년도에 JET 의 멤버인 CNNIC, JPNIC, TWNIC, KRNIC 이 시범 서비스 등을 시행하였으나 국제 표준 미비 등의 이유로 활성화가 되지 않았다. IETF 에서는 2002 년 10 월 25 일 52 차 IETF 회의를 통해 다국어 도메인에 대해 일반 영문도메인과 같은 규칙으로 운영하기 위하여 IDNA(Internationalized Domain Name Application) 아키텍처와 퓨니코드(Punycode) 인코딩 방식을 IDN 의 국제 표준으로 확정하였다. IDN 을 처리하기 위해 DNS 서버에서 처리하는 방식과 클라이언트에서 처리하는 방식 중 클라이언트에서 다국어 도메인을 영문과 숫자로 인코딩(ACE 인코딩:ASCII Compatible Encoding)하여 DNS 에 질의를 하여 돌아온 값을 디코딩하여 사용하는 IDNA 방식을 표준으로 제정하였다. 앞의 내용과 관련해서 IDNA 에 관련된 문서[RFC3490] 및 Punycode[RFC3492], Nameprep [RFC3491] 등에 관한 문서가 RFC 로 확정되어 있다.

KRNIC 의 주도로 국내 ENUM 에 대한 정보 확산 및 사용성이 꾸준히 제기되고 있으며 정보통신부에서는 정보통신표준화 중점과제로 URI/ENUM 을 선정하여 지원하고 있다. 또한, ENUM 도입을 위해 다양한 회의를 개최하고 있으며, 한국통신, 데이콤, 온세통신, SK 텔레콤, 한국통신프리텔, 하이텔 등 여러 IT 업체에서 ENUM 에 관련한 서비스의 시험망을 구축하여 실험중이다. 현재 ENUM 의 스펙이 이미 RFC 로 정의되어 있기 때문에 ENUM 자체의 기술 부분보다 ENUM 의 활용방안 및 운용방안에 대한 연구가 활발히 진행되고 있다. 또한 국내 참여업체들이 모두 기존의 PSTN 및 H.323, SIP 와의 연동에 중점을 두고 Testbed 를 구축하고 있다.

그리고 향후 서비스의 신규 출현에 따른 지원 규격 및 차세대 All-IP 환경에 대한 활용 가능성에서 연구의 필요성을 제기하고 있다.

참고문헌

- [1] P. Hoffman 외, "RFC 3490 Internationalizing Domain Names in Applications (IDNA)", March 2003
- [2] P. Hoffman 외, "RFC 3491 Nameprep: A Stringprep Profile for Internationalized Domain Names (IDN)", March 2003.
- [3] A. Costello, "RFC 3492 Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA)", March 2003.