

데이터 변경에 유연한 이진 데이터 처리를 위한 XML 기반 객체 모델의 적용 방법

이용현*, 심준용*, 김세환*

*LIG넥스원 Maritime연구소 VM팀

e-mail: yhlee80@lignex1.com

Applications of XML based Object Model for flexible binary data processing

Yongheon Lee*, Jun-Yong Shim*, Sae-Hwan Kim*

*LIG Nex1 Co. LTD. Maritime R&D Center, VM Team

요 약

컴퓨터에서 사용하는 데이터의 기본 단위는 2진수 형식으로 표현되는 이진(binary) 데이터이며 이는 CPU 연산에서부터 이기종간 네트워크 통신에 이르기까지 폭넓은 영역에서 사용되고 있다. 이진데이터를 처리하는 소프트웨어를 개발하기 위하여 다양한 원시타입(Primitive Type)을 사용하는 프로그래밍 언어가 사용되는데, 일반적으로 이러한 프로그래밍 언어를 통한 개발은 소스코드에 구조체, 공용체 및 클래스로 이진 데이터를 정적으로 정의함으로써 추후 데이터 변경에 대한 유연성이 떨어지는 문제가 존재한다. XML 기반의 객체 모델은 프로그래밍 언어에서 사용되는 모든 타입 및 이진 데이터 구성 형식을 제공하고 이를 통해 이진 데이터를 동적으로 관리함으로써 이러한 유연성의 문제를 해결하고 있으며 실제 이진 데이터를 사용하는 분야에 적용할 수 있다.

1. 서론

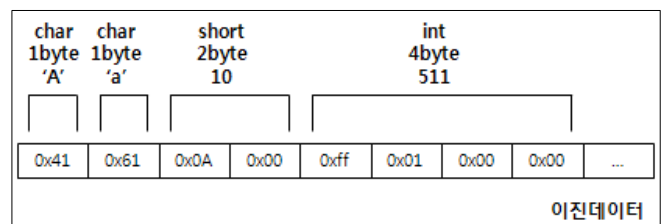
컴퓨터 시스템은 초창기부터 현재의 아키텍처까지 모두 2진수 형식의 데이터(이진 데이터)를 기본 단위로 사용하고 있다. 작게는 CPU의 레지스터 및 명령어에서부터 크게는 대용량 파일 저장 및 분산 시뮬레이션에 이르기까지 모든 컴퓨터 산업에 이러한 이진(binary) 데이터 형식이 적용된다. 이러한 데이터 처리를 위한 수많은 도구와 방법론이 존재하는데, 일반적으로 프로그래밍 언어와 개발 환경에서는 이를 용이하게 다룰 수 있는 API를 제공하고 있다. 소프트웨어 개발자는 이러한 API를 사용한 소스코드를 작성하여 다양한 이진 데이터를 처리할 수 있다.

한편 소프트웨어 개발은 시간의 흐름에 따라 그 규모가 점점 거대해지고 복잡해지고 있으며, 거대한 시스템간의 상호 연동이 매우 중요시되고 있는 추세이다. 이와 함께 이러한 소프트웨어의 개발 비용 및 개발 기간을 단축시킬 수 있는 환경과 능력이 요구되고 있다. 이를 위하여 소프트웨어 개발에 있어 재사용성(reusability)과 유연성(flexibility)이 필수 요소로 인식되고 있다. 이러한 추세에 비추어 볼 때 이진 데이터는 컴퓨터 및 소프트웨어 전반에 차지하는 비중이 매우 높으므로 이를 유연하게 처리할 수 있는 방법을 개발하고 적용할 수 있어야 한다. 본 논문에서는 XML 기반의 객체 모델[1][2][3]을 사용하여 이진 데이터를 데이터 변경에 대해 유연하게 처리할 수 있는 방법을 제시한다.

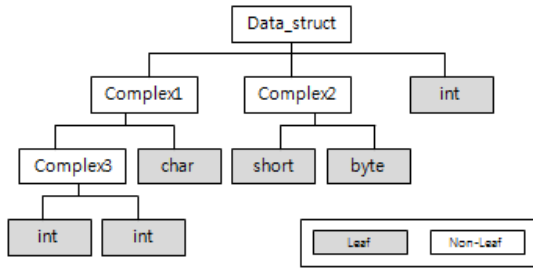
2. 관련 연구

C/C++, Java와 같은 프로그래밍 언어에서 이진 데이터를 다루는 기본 단위는 원시 타입(Primitive data type)이다. 일반적으로 1바이트를 표현하는 타입(e.g. char, bool, etc.)부터 8바이트를 표현하는 원시 타입(e.g. double, int64, etc.)을 제공한다. 이진 데이터는 이러한 원시 타입을 사용하여 개발자 및 사용자에게 친숙한 10진수 단위로 표현될 수 있으며 프로그래밍 언어가 제공하는 다양한 API에 의해 데이터 처리가 될 수 있다.

일반적으로 프로그래밍 언어에서는 여러 원시 타입을 묶어 하나로 표현할 수 있는 방법을 제공한다. 이는 통상 구조체(struct), 공용체(union), 클래스(class)와 같은 호칭으로 통용되며 이들은 다시 다른 구조체/공용체/클래스와 묶어 새로운 사용자 정의 타입을 만들어낼 수 있다.[2] 이는 원시 타입을 Leaf 노드로 갖는 트리(Tree) 구조와 비교할 수 있다.



(그림 1) 타입과 이진 데이터의 관계



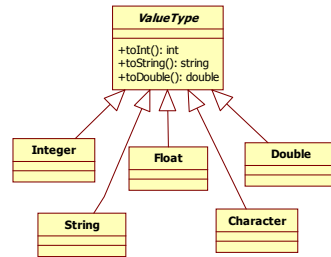
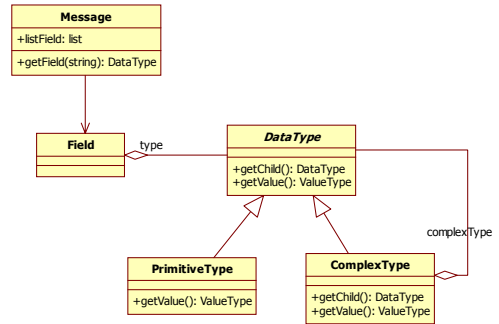
(그림 2) 타입의 트리 구조

이렇게 정의된 사용자 정의 타입과 프로그래밍 언어에서 제공하는 API들을 통하여 다양한 이진 데이터를 처리할 수 있다. 하지만 이는 소스코드에 기술되어 결정된 타입 및 API를 컴파일러에 의해 해석하는 정적인 방식으로써, Run-time에 데이터 타입을 변경하지 못하며 타입 변경을 위해서는 소스코드를 수정하고 다시 컴파일을 수행해야 하는 문제점을 가지고 있다. 이는 이진 데이터 처리의 변경에 대한 유연성이 떨어지는 것으로써 응집도(Cohesion)와 재사용성이 높은 소프트웨어를 개발할 수 없도록 만든다.

3. XML 기반 객체 모델 적용

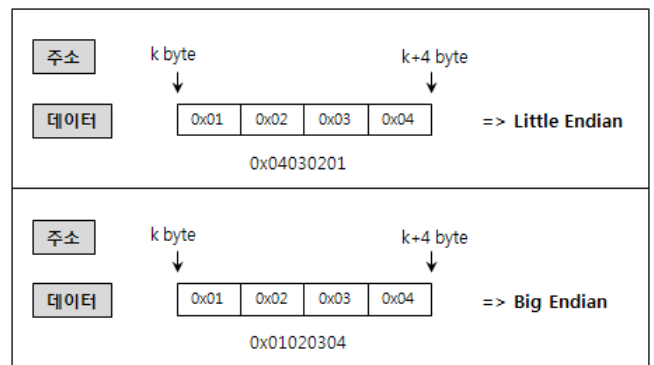
XML 기반 객체 모델은 당사가 보유한 플러그인 기반 프레임워크[3]를 통하여 구현되는 플러그인 간의 메시지 통신을 위한 인터페이스를 제공한다. 이는 플러그인 간에 송수신되는 메시지의 데이터 구조를 XML 스키마[4]에 의해 규정하고 XML 문서에 메시지 데이터를 기록하는 방식으로써, 메시지 파서(Parser)를 통해 데이터 구조를 해석하고 실제 이진 데이터를 동적으로 다룰 수 있는 구조를 취하고 있다. XML 기반 객체 모델에서는 이진 데이터 처리를 위한 프로그래밍 언어의 원시 타입 및 이를 조합한 사용자 정의 복합 타입(Complex Type)을 정의할 수 있으며 실제 이진 데이터를 처리하기 위한 기능(Encoding, Decoding)을 제공한다.

XML기반 객체 모델은 XML 문서에 정의된 원시 타입과 열거 타입(Enumeration Type)을 조합하여 사용자 정의 복합 타입을 정의하고 이를 반복 수행할 수 있는 구조를 위해 GoF[5]의 컴포지트 패턴(Composite Pattern)을 적용하여 구현되었다. 이는 복합 타입 클래스가 각 원시 타입, 열거 타입 및 복합 타입을 추상화한 데이터 타입 클래스를 자식(Child) 노드로 가지는 구조이며 이러한 재귀적(Recursive) 구조를 통해 전체 데이터 타입의 트리 구조가 생성되고, 트리 탐색(Tree Traversal)을 통해 타입 하위의 모든 원시/열거 타입을 접근할 수 있다. 이러한 구조를 통해 이진 데이터를 구성하는 모든 데이터의 구성을 트리 형태로 동적으로 관리할 수 있으며 플랫폼 및 프로토콜에 맞는 형태로의 변환이 용이해진다.

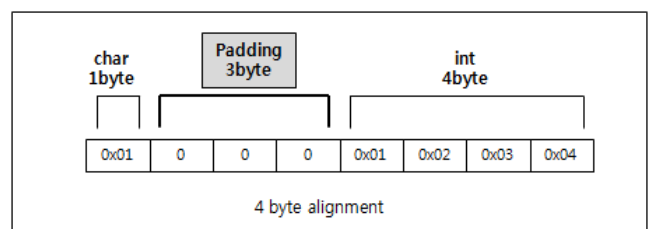


(그림 3) XML기반 객체 모델의 클래스 다이어그램

일반적으로 TCP/IP와 같은 네트워크 프로토콜, 데이터 로깅과 같이 데이터 스트림을 통해 이진 데이터를 처리하는 경우 연관된 하드웨어의 CPU 및 플랫폼에 따라 효율성을 높이기 위해 몇 가지 규칙을 적용하고 있다. 이는 이진 데이터의 바이트 순서를 나타내는 엔디안(endian)[6]과 CPU의 데이터 처리 속도를 높이기 위한 구조체 바이트 정렬(byte alignment)[7]이며 이러한 이진 데이터의 규칙을 XML 기반 객체 모델에서 처리가 가능하다.



(그림 4) Endian(Little/Big) 개념

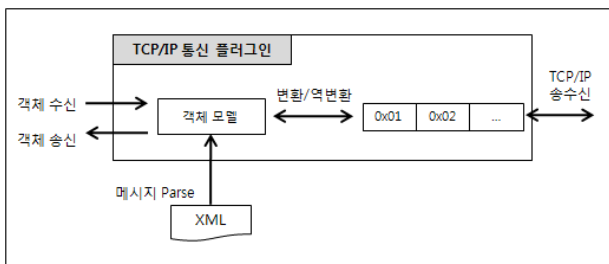


(그림 5) 바이트 정렬 개념

XML 기반 객체 모델은 플러그인 기반 프레임워크의 메시지 인터페이스이므로 이진 데이터 처리 기능을 XML 기반 객체 모델로 처리하기 위해서는 이를 플러그인 형태로 구현하게 된다. 이 플러그인은 플러그인 프레임워크를 통하여 외부 플러그인과 객체(메시지) 통신을 수행하며 송수신되는 객체를 XML 기반 객체 모델을 통하여 이진 데이터 형태로 변환(Encoding)/역변환(Decoding)한다. 객체 모델은 XML 문서에 기술된 원시/열거/복합 타입의 구조를 적용한 변환/역변환을 수행하며, 데이터 구조 변경시 단지 XML 문서의 내용만을 수정하면 되므로 소스코드의 변경은 필요하지 않다. 이러한 과정을 통하여 데이터 변경에 유연한 이진 데이터 처리를 수행할 수 있다.

3.1. TCP/IP 이진 데이터 통신

TCP/IP 통신을 위한 데이터 프레임(패킷)은 일반적인 이진 데이터가 사용되며 통상 big endian 형태를 가진다. 이를 위해 데이터 정의를 위한 메시지 명세 XML 문서를 작성하고 이를 TCP/IP 통신 플러그인의 메시지 파서를 통해 동적으로 관리한다. 플러그인을 통해 송수신되는 객체 데이터를 이진 데이터로 변환/역변환 함으로써 TCP/IP 통신을 수행할 수 있다. 통신하는 이진 데이터의 변경이 필요한 경우 데이터가 정의된 XML 문서만 수정하면 된다.

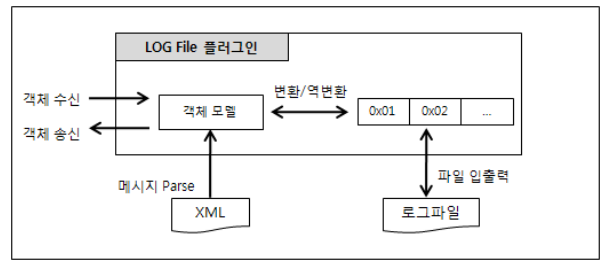


(그림 6) TCP/IP 플러그인을 통한 이진데이터 처리

3.2. 파일 로깅 및 로그 파일 분석

하드 디스크와 같은 저장장치에 저장되는 파일은 이진 형식이다. 일부 파일은 ASCII code와 같이 쉽게 해석할 수 있는 형태로 인코딩한 경우도 있지만, 저장된 데이터는 본질적으로 이진 형태를 가진다. 따라서 파일에 저장된 데이터는 이진 데이터의 엔디안 및 바이트 정렬의 특징을 그대로 가지고 있으며 이는 XML 기반 객체 모델을 통하여 처리할 수 있다.

TCP/IP 통신 플러그인과 마찬가지로 파일 로깅 및 로그 파일 분석 기능 역시 플러그인으로 구현할 수 있으며 파일에 저장되거나 분석할 이진 데이터는 XML 기반 객체 모델을 사용하여 처리할 수 있다. 또한 파일에 저장되는 이진 데이터의 구조가 변경되는 경우 XML 파일의 수정만으로 처리가 가능하므로 데이터 변경에 매우 유연하게 대처할 수 있다.



(그림 7) 로그파일 플러그인을 통한 이진데이터 처리

위와 같이 XML기반 객체 모델을 적용하여 데이터를 처리한 네트워크 통신과 파일 스트림은 대표적인 이진 데이터 적용 분야이다. 이외에도 프로그래밍 언어에서 제공하는 원시타입을 조합하여 이진 데이터를 다루는 모든 분야에 XML 기반 객체 모델을 적용 가능하다.

4. 결론

이진 데이터는 컴퓨터 시스템에 있어 가장 기본적인 요소이다. 소프트웨어를 구성하는 사용자 정의 형식의 데이터 역시 이진 형식으로 표현할 수 있으며 이는 파일 및 네트워크 통신 등에서 이러한 이진 형식의 데이터를 사용하고 있다. 본 연구에서 제시한 XML 기반 객체 모델은 이진 데이터를 표현하는 방법을 기존의 정적인 방식 대신 XML을 통한 동적인 방법을 사용함으로써 데이터 변경에 유연하게 대처할 수 있다. 그리고 XML 기반 객체 모델과 플러그인 기반 프레임워크를 통하여 개발 기간 및 비용을 단축시킬 수 있는 개발 환경을 제공하고 있다. 차후 이를 다양한 분야에서 유연성을 확보할 수 있는 도구로 발전시켜 나갈 것이다.

참고문헌

- [1] 이용현, 심준용, 이승영, 김세환 "Publish-Subscribe 메시지 프로토콜을 위한 XML 기반의 Object Model 설계" 한국시물레이션학회 춘계학술대회, 2010.05
- [2] 이용현, 심준용, 김세환 "RPR-FOM 기반 HLA 시물레이터 개발을 위한 XML 기반 Object Model의 적용" 한국시물레이션학회 추계학술대회, 2010.10
- [3] 이용현, 심준용, 김세환 "사용자 정의 형식을 지원하는 XML 기반 객체 모델의 구조 설계" 한국정보처리학회 추계학술발표대회, 2010.11
- [4] <http://www.w3.org/XML/Schema>
- [5] Erich Gamma, Richard Helm, Ralph Johnson, John Vlissides, "Design Patterns: Elements of Reusable Object-Oriented Software", Addison-Wesley, 1994
- [6] <http://en.wikipedia.org/wiki/Endianness>
- [7] P. Ranjan Panda, H. Nakamura, N.D. Dutt, A. Nicolau "A data alignment technique for improving cache performance" 1997 IEEE International Conference on Computer Design(ICCD'97)