

# MPEG-2 표준을 위한 오류 데이터 분류

곽태희\*      최병주  
이화여자대학교      컴퓨터학과  
{thgwak, bjchoi}@mm.ewha.ac.kr

## The Taxonomy of Dirty Data for MPEG-2 TS

Taehee Gwak\*      Byoungju Choi  
Dept. of Computer Science & Engineering, Ewha Womans University

### 요 약

DASE(Digital TV Application Software Environment)는 데이터 방송을 위한 국제 표준으로 MPEG-2 TS(Moving Picture Experts Group-2 Transport Stream) 형식의 데이터를 처리한다. 소스코드 대신 입력 데이터 명세 정보만을 공개하는 특성상 DASE 시스템의 오류를 테스트하기 위해서는 테스트 데이터에 오류를 삽입하는 방법이 적합하고 이를 위해 MPEG-2 표준에 적용 가능한 오류 항목의 개발이 요구된다. 본 논문에서는 관계형 데이터 베이스를 위한 오류 데이터 분류인 Kim's et al 분류를 근거로 하여 MPEG-2 표준을 위한 오류 항목을 개발하였다. 이는 DASE 시스템의 오류 삽입 테스트 기법에 유용하게 사용될 수 있을 것이다.

### 1. 서론

DASE(Digital TV Application Software Environment)[1]는 데이터 방송을 위한 국제 표준으로 지상파 디지털 TV방송을 위한 MPEG-2 TS(Moving Picture Expert Group-2 Transport Stream)[2] 형식의 데이터를 처리하는 시스템이다. 이러한 DASE 시스템은 복잡한 구조의 데이터를 대량으로 처리하기 때문에 다양한 에러가 존재할 가능성이 높으므로 상용화되기 전에 철저한 테스트가 필요하다. 그러나 소스 코드를 공개하지 않는 DASE의 특성상 소스 코드를 이용한 기존의 테스트 방법을 적용하기에는 무리가 있다.

오류 삽입기법[3]은 소프트웨어 코드의 어떤 위치에 임의로 오류를 삽입한 후 소프트웨어가 어떻게 동작하는지를 관찰하는 것으로서 프로그램에 의해 수행되는 경로가 매우 복잡하거나 수행 경로를 찾기 어려운 경우에 매우 유용하다. 따라서 소스 코드 대신 입력 데이터의 명세 정보만을 공개하는 DASE 시스템의 경우 이러한 오류 삽입 기법이 적합함을 알 수 있다. 즉, DASE 시스템의 오류를 테스트하기 위해서는 테스트 데이터 자체에 오류를 삽입하는 방법이 적합하다. DASE 시스템의 테스트 데이터로 이용하기 위해서는 MPEG-2 표준의 신택스가 반드시 지켜져야 하므로 신택스는 위반하지 않으면서 단순히 필드 값 자체에 오류를 심는 방법이 채택되어야 한다. 그리고 이를 위해 MPEG-2 표준에 적용할 수 있는 오류 항목의 개발이 우선적으로 요구된다. 본 논문에서는 MPEG-2 표준을 위한 오류 항목을 개발함으로써 DASE 시스템의 테스트를 위한 오류 삽입 테스트 기법의 기반을 마련하였다.

본 논문의 구성은 다음과 같다. 2장에서는 MPEG-2 표준의 특징을 살펴보고 3장에서는 MPEG-2 표준에 적용될 오류 항목의 추출 방법에 대해 기술한다. 4장에서는 추출된 오류 항목을 분석하고 5장에서는 결론 및 향후 연구 과제를 기술한다.

### 2. MPEG-2 표준의 특징

MPEG은 국제 표준 기구(ISO) SC29의 W.G. 11에서 만들어진 영상과 음향의 압축 표준 규격이다. MPEG 표준 중 MPEG-2 표준은 상호 운용성(interoperability), 스케일러빌리티(scalability), 확장성(extensibility)의 세 가지 개념을 기본으로 마련되었기 때문에 넓은 적용 영역과 고품질 영상을 제공한다. MPEG-2는 여러 산업 분야에서 적용될 수 있으며 그 중 응용이 가장 적합한 분야가 방송 분야이다. MPEG-2 표준을 이용한 디지털 방송은 압축 효과로 인하여 다채널 방송이 가능해지고, 다양한 내용을 전송할 수 있다.

디지털 방송을 하기 위해서는 방송 내용인 데이터 서비스와 프로그램 정보를 MPEG-2 표준 형식으로 구성하고 이를 동기를 포함하는 통합된 하나의 비트열로 만들어야 한다. MPEG-2 시스템에는 두 종류의 방식이 있는데 하나는 프로그램 스트림(Program Stream)이고 다른 하나는 트랜스포트 스트림(TS: Transport Stream)이다. 이 중 TS는 복수의 프로그램을 구성할 수 있기 때문에 디지털 방송을 위한 스트림은 TS로 생성한다. 디지털 방송을 하기 위한 MPEG-2 TS의 데이터

구조는 그림 1[4]과 같이 크게 PSI(Program Specific Information), PSIP(Program and System Information Protocol), Data Service의 세 부분으로 구성된다.

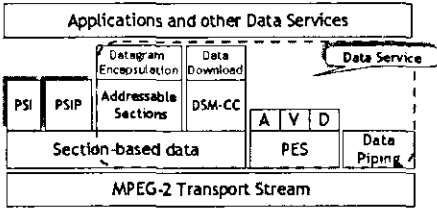


그림 1 MPEG-2 TS의 구조

Data Service부분은 디지털 방송 시 기존의 오디오, 비디오 데이터가 아닌 다양한 응용 프로그램이나 데이터 서비스를 기술하기 위한 부분으로 DSM-CC Data download protocol, Addressable section, PES, Data piping의 4가지 프로토콜로 구성되며, 각 프로토콜은 여러 개의 테이블 형식으로 기술된다 [5,7]. PSI와 PSIP는 프로그램 정보인 가상 채널들의 시스템 정보와 서비스의 내용, 서비스 시간 등을 기술하기 위한 부분이다[1,6,7]. PSI는 기본적으로 PAT, PMT, CAT의 3개의 테이블로 구성되며, PSIP는 STT, RRT, VCT, EIT, DET, MGT, ETT의 7개의 테이블로 구성된다. 이러한 기본적인 테이블 이외에 세부적인 정보 기술을 위한 다수의 디스크립터(descriptor) 등이 있으며 방송 데이터를 이러한 형식으로 기술하기 위해서 위와 같은 수많은 테이블들이 유기적으로 연결되어 사용된다.

이와 같이 MPEG-2 표준은 모든 데이터를 테이블과 같은 형태로 나타내어 저장하고 각 테이블 사이에 계층적인 관계가 존재하는 관계형 데이터 베이스와 매우 유사하다. 따라서 이와 같은 유사성에 착안하여 successive hierarchical refinement 방식을 이용한 기존 관계형 데이터 베이스의 오류 데이터 분류(이하 Kim's et al 분류라 칭함)[8]로부터 MPEG-2 표준의 필드 값에 적용 가능한 오류 항목을 추출할 수 있다.

3. MPEG-2 표준을 위한 오류 항목 분류 방법

MPEG-2 표준을 위한 오류 항목의 추출은 다음과 같이 3단계로 나누어져 진행된다. 1단계는 33가지의 Kim's et al의 오류 데이터 분류로부터 MPEG-2 표준에 적용 가능한 오류 항목을 추출하는 단계이다. 2단계는 1단계에서 추출한 오류 항목들이 MPEG-2 표준이 가지는 속성을 모두 포함할 수 있는지를 검토하는 단계이고 이러한 작업을 통해 3단계에서는 부족한 항목이 추가되어 오류 항목 분류가 보완된다.

3.1 Kim's et al 분류로부터 오류 항목 추출

Kim's et al의 오류 데이터 분류에서 MPEG-2 표준에 적용할 수 없는 항목들을 제거하는 방식으로 필요한 오류 항목을 추출한다. 즉, MPEG-2 표준의 명세 정보는 단일 데이터 베이스로 표현되므로 Kim's et al 분류에서 두 개 이상의 데이터 베이스가 통합되거나 입력 데이터의 표시 체계가 일정하지 않을 때 발생하는 오류 항목들을 우선적으로 제거한다. 또한 본 연구에서는 테스트를 위해 의도적으로 필요한 오류만을 삽입하는 것이므로 의도하지 않은 오류 항목과 트랜잭션 과정 중에 우연히 발생할 수 있는 오류 항목, 그리고 MPEG-2 표준의 요구 사항과 관련이

없어 적용할 수 없는 오류 항목, 예를 들어 다각형 표기 오류, 들을 제거하였다.

이와 같은 방식에 의해 추출된 오류 항목은 다음 6가지이다.

1. 없어진 데이터 (missing data): null 이 허용되지 않은 필드에 필드 값으로 null 이 들어간 경우
2. 잘못된 데이터 타입의 사용 (use of wrong data type): 필드 값의 데이터 타입이 잘못된 경우
3. 잘못된 범주의 데이터 (wrong categorical data): 필드 값이 정상 값의 범위를 벗어나는 경우
4. 중복된 데이터 (duplicate data): 중복이 금지된 두 개 이상의 필드 값이 중복된 값을 갖는 경우
5. 시간이 맞지 않는 데이터(outdated temporal data): 필드 값의 시간이 맞지 않는 경우
6. 다중 테이블에 걸쳐진 불일치성 (inconsistency across multiple tables/files): 여러 개의 테이블에 속하는 필드 값이 서로 동일해야 하는데 불일치 하는 경우

3.2 MPEG-2 표준의 필드 속성을 이용한 검토

본 연구에서 개발한 오류 항목은 MPEG-2 표준 전체에 다양하게 적용될 수 있어야 하고 이를 위해서는 MPEG-2 표준 필드 값들의 모든 속성을 포함할 수 있어야 한다. MPEG-2 필드 속성은 방송 내용과의 관련성과 의존성 등을 기준으로 다음 4가지 타입으로 분류되며 이러 필드 속성과 3.1에서 추출한 6가지 오류 항목을 표1과 같이 연관 지을 수 있다.

MPEG-2 표준의 필드속성		관련 오류 항목
필드 속성 1	사람이 최소한으로 계획해야 하는 필드	1, 2
필드 속성 2	필드 속성 1 의 내용을 기반으로 값이 정해지는 필드	5
필드 속성 3	필드 속성 1 에 독립적, 표준에 따라 유효 값을 취하는 필드	3, 4, 6
필드 속성 4	필드 속성 1 에 독립적, 표준에 따라 고정 값을 취하는 필드	

표 1 MPEG-2 표준의 필드 속성과 관련 오류 항목

3.3 오류 항목의 보완

MPEG-2 표준의 필드 속성과 오류 항목과의 사상 결과 필드 속성 4와 관련된 오류 항목이 빠져있음을 알 수 있다. 따라서 이에 대한 오류 항목을 추가할 필요가 있다. 즉, 유효 값이기는 하나 필드 값은 선정할 때 적용되는 MPEG-2 표준의 규칙을 어긴 경우의 오류 항목이 추가되어야 한다. MPEG-2 표준의 필드 속성을 이용한 검토를 통해 보완된 최종 오류 항목은 다음과 같은 7가지이다.

1. 없어진 데이터 (missing data)
2. 잘못된 데이터 타입의 사용 (use of wrong data type)
3. 잘못된 범주의 데이터 (wrong categorical data)
4. 중복된 데이터 (duplicate data)
5. 시간이 맞지 않는 데이터(outdated temporal data)
6. 다중 테이블에 걸쳐진 불일치성 (inconsistency across multiple tables/files)
7. 생성 규칙을 어긴 데이터 (rule violation): 필드 값을 선정할 때 적용되는 MPEG-2 표준의 규칙이나 제약 사항을 어긴 경우

4. 오류 항목의 분석

본 절에서는 본 논문에서 개발한 MPEG-2 표준을 위한 오류 항목의 적용 범위와 우선 순위를 기술함으로써 오류 항목을 분석한다.

4.1 오류 항목의 적용 범위

오류 항목은 MPEG-2 TS 표준의 모든 범위에서 발생할 수 있는 오류를 커버할 수 있어야 한다. 표2는 MPEG-2 테이블들의 내용과 요구 사항을 기반으로 적용 가능 오류 항목을 나타낸 것이다. 즉, 오류 항목 1,2,3은 MPEG-2 표준 전체에, 오류 항목 4,6은 다른 테이블과 연관관계가 있는 테이블에, 오류 항목 5는 시간정보를 나타내는 테이블에 마지막으로 오류 항목 7은 고정 값을 둔 테이블에 적용이 가능하다. 따라서 본 논문에서 개발한 오류 항목은 MPEG-2 표준 전체에 다양하게 적용될 수 있다.

MPEG-2 TS 표준		내용	적용 오류 항목 번호
PSI	PAT	방송될 프로그램의 번호와 그에 해당하는 stream의 PID를 가지고 있는 정보	1,2,3,4,6,7
	CAT	PAT는 program number와 PMT의 PID를 연관 지어 줌	1,2,3,7
	PMT		1,2,3,4,6,7
PSIP	STT	시간과 날짜에 대한 정보 제공	1,2,3,5
	RRT	지리/지역별로 이벤트에 대한 등급 정보 제공	1,2,3
	MGT	STT를 제외한 모든 PSIP 테이블의 버전, 길이 및 PID 정보를 제공	1,2,3,4,6,7
	VCT	Transport Stream으로 전송되는 가상 채널들의 속성에 대한 정보를 제공	1,2,3,4,6,7
	EIT	가상 채널에서 정의한 이벤트들에 대한 세부 정보(제목, 시작시간 등)를 제공	1,2,3,4,5,6,7
	DET	가상 채널에서 정의한 데이터 이벤트들에 대한 세부 정보(제목, 시작시간, 데이터 서비스에 필요한 버퍼 모델 등)를 제공	1,2,3,4,5,6,7
	ETT	가상 채널과 이벤트들의 세부 정보를 제공하는 확장 텍스트 메시지를 제공	1,2,3,7
	DSM_CC		1,2,3
PES	데이터 방송 프로토콜	1,2,3	
Addressable_section		1,2,3	

표 2 MPEG-2 표준과 적용 가능 오류 항목

4.2 오류 항목의 우선 순위

오류 항목 별로 적용될 수 있는 MPEG-2 표준 테이블이 다르므로 여러 개의 오류 항목 중 하나의 항목을 임의로 선택하는 것은 문제가 있다. 모든 필드에 적용 가능한 오류

항목과 하나의 필드에만 적용 가능한 오류 항목이 적용될 수 있는 경우 임의로 전자의 오류 항목을 우선적으로 적용하게 된다면 후자의 오류 항목은 적용할 수 없게 되기 때문이다. 따라서 오류 항목 적용 범위의 희소성을 기준으로 오류 항목의 우선 순위를 정할 필요가 있다. 각 오류 항목의 적용 범위를 기반으로 한 오류 항목의 우선 순위는 표3과 같다.

오류 항목	MPEG-2 표준 적용 대상	오류 항목 우선 순위
없어진 데이터	MPEG-2 표준 전체	4
잘못된 데이터 타임의 사용	MPEG-2 표준 전체	4
잘못된 범주의 데이터	MPEG-2 표준 전체	4
중복된 데이터	PSI의 PAT, PMT PSIP의 VCT, EIT, DET, MGT	2
시간이 맞지 않는 데이터	PSIP의 STT, EIT, DET	1
다중 테이블에 걸쳐진 불일치성	PSI의 PAT, PMT PSIP의 VCT, EIT, DET, MGT	2
생성 규칙을 어긴 데이터	PSI의 PAT, CAT, PMT PSIP의 VCT, EIT, DET, MGT, ETT	3

표 3 오류 항목 별 우선 순위

5. 결론 및 향후 연구 과제

본 논문에서는 MPEG-2 표준을 위한 7가지 오류 항목을 개발하였다. 이를 위해 MPEG-2 TS와 유사한 구조의 관계형 데이터 베이스를 위한 오류 데이터 분류인 Kim's et al 분류를 근거로 삼았고 MPEG-2 표준 자체의 4가지 필드 속성을 이용하여, 추출된 오류 항목을 검토, 보완하였다.

본 연구를 통해 개발된 MPEG-2 표준을 위한 오류 항목은 MPEG-2 TS를 처리하는 DASE 시스템의 테스트에 유용하게 이용될 수 있을 것이다. 향후 오류 삽입 기법을 기반으로 하는 테스트 방안을 완성하고 오류 항목을 실제 환경에 적용한 실험을 통해 타당성을 검증할 필요가 있다.

참고문헌

- [1]ATSC Draft, ATSC T3/S17-Doc. 075, Digital Television Receiver Application Software Environment for Terrestrial and Cable Broadcast of Data and Interactive Services.
- [2]ISO/IEC 13818-1|ITU-T Rec. H.222.0, Information Technology - Generic coding of moving pictures and associated audio - Systems, 1994.
- [3]Sudipto Ghosh, Aditya P. Mathur, Joseph R. Horgan, J. Jenny Li, W. Eric Wong, "Fault Injection Testing of Distributed Systems - A case Study," Proceedings of Quality Week Europe, Nov. 1997
- [4]ATSC Standard T3/S13 - ATSC Data Broadcasting Implementation Guidelines Draft Description, 1999. (Doc.011)
- [5]ATSC Draft T3-504 - ATSC Data Broadcast Standard, 2000.
- [6]ATSC Standard A/65 - Program and System Information Protocol for Terrestrial Broadcast and Cable, 1997.
- [7]ISO/IEC 13818-6 - MPEG-2 Digital Storage Media command & Control, Chapter 2, 4, 6, 7, 9 and 11, 1998.
- [8]Won Kim, Byoungju Choi, Eui-Kyeong Hong, Soo-Kyung Kim, Doheon Lee, "A Taxonomy of Dirty Data," Data Mining and knowledge Discovery, 2001, Accepted