

글로벌 온라인 미디어 스트리밍 Best Practice

글.
이정애 라임라이트 네트워크 기술 부장

글로벌 미디어 스트리밍 서비스 시리즈

1. Global Online Media Streaming
2. HTTP Streaming, Current & Next
3. Codec, Current & Next
4. Building HTML5 Streaming Player

글로벌 온라인 미디어 스트리밍 트렌드

최근 글로벌, 특히 미주 지역의 미디어 스트리밍 소비 형태는 기존 미디어 시청 방식인 케이블에서 OTT 서비스 등의 인터넷을 통한 온라인 방식으로 빠르게 전환되고 있다. 이는 케이블 선을 끊는다는 의미의 코드 커팅(Cord Cutting)이라고도 불리며 최근 들어 국내에서도 가시화되고 있는 동향이다.

넷플릭스, 아마존 프라임 비디오, 그리고 훌루와 같은 OTT 서비스 공급 업체들은 2015년 기준으로 1.2억 명 유료 가입자와 \$100억 매출을 창출하였고, 2019년에는 3.3억 명의 유료 가입자와 \$310억 이상의 매출 규모를 예상하고 있다. 국내의 경우, 방송 3사의 콘텐츠 연합 플랫폼과 CJ E&M이 주요 OTT 서비스를 제공하고 있으며, 가입자 수는 매년 꾸준히 증가하고 있다. 한국의 방송 콘텐츠를 해외에서 서비스하고 있는 Viki.com 등의 OTT 서비스에서 한국 콘텐츠의 수요가 증가함에 따라서, 지상파 방송 3사는 한류 콘텐츠의 해외 배급을 위한 서비스를 직접 구축 및 운영하고 자 하는 움직임을 보이는 등 CP(Contents Provider)와 SP(Service Provider)의 경계가 점차 사라지는 추세이다.

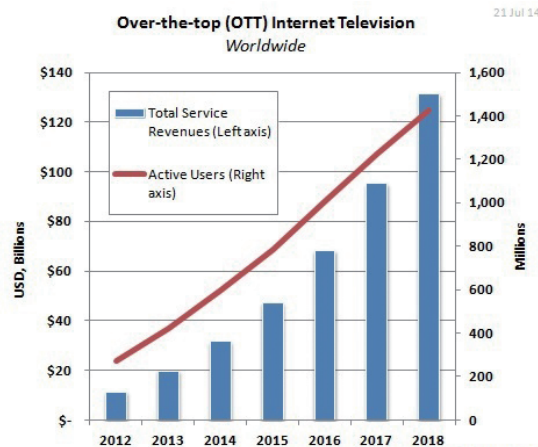


그림 1. OTT 서비스 추세 / 출처 : Generator Research

온라인 미디어 스트리밍을 운영하는데 있어서, 서비스의 규모가 커질수록 자체 솔루션이나 인프라를 구축하는 방법에서 클라우드 서비스에서 인프라와 미디어 스트리밍 솔루션을 제공 받는 방법으로 이동하게 된다. 전송하는 트래픽의 규모가 방대해질 경우 서비스 안정성과 빠르게 증가하는 트래픽 규모 수용을 위하여 단일 CDN보다는 멀티 CDN을 도입하는 경우도 늘어나고 있다. 2015년 10월 25일 야후가 전 세계에 첫 번째 무료 미식축구를 중계한 경우를 일례로 살펴보면 185개국에서 천오백만 명이 시청하여 7Tbps의 피크 트래픽을 서비스한 것으로 결과 집계가 되었는데, 이 정도 수준의 대용량 트래픽을 수용하기 위하여 자체 CDN 인프라를 포함하여 총 7개의 CDN 업체를 통하여 서비스를 제공했다고 한다. 또한, UHD/4K 콘텐츠와 VR 서비스에 대한 필

요와 관심이 증가함에 따라 해당 콘텐츠를 지원할 수 있는 H.264 이후의 차세대 코덱 및 좀 더 고화질의 콘텐츠 수급을 위한 방안 및 해당 콘텐츠의 재생을 위한 플레이어 지원에 대한 요구들 또한 늘어나고 있다.



그림 2. 세계적인 OTT 서비스 제공 기업

온라인 미디어 서비스 중 글로벌 라이브 이벤트, 특히 올림픽이나 월드컵 같은 대형 라이브 이벤트의 경우에는 실시간 서비스에서 발생할 수 있는 이슈를 사전에 방지하기 위하여 인제스트 이중화부터 트래픽 전송 및 재생에 이르는 모든 포인트에 대하여 최적화하는 방안이 더욱 중요하게 대두된다.

성공적인 온라인 미디어 스트리밍이란?

위와 같이 빠르게 진화하는 미디어 서비스 시장에서, 실제 스트리밍 서비스를 성공적으로 운영하고 있는지 판단할 수 있는 기준은 어떤 것들이 있는지 살펴보자.

1. 최소한의 플레이 시작 시간 : 사용자가 재생 버튼을 누른 후에는 비디오 재생까지 최소한의 시간이 걸리기를 희망함
2. 최소한의 리버퍼링 : 스트리밍이 시작한 이후에는 끊김 없이 매끄럽게 시청하기를 원함
3. 최고의 품질(비트레이트) : 다양한 비트레이트의 미디어를 서비스하는 경우, 사용자 네트워크 환경에 최적화된 비트레이트로 최고의 품질을 보장받기를 원함
4. 오리진 서버 부하 방지 : 원본 미디어 파일을 공급하는 오리진 서버의 이슈로 인하여 스트리밍 서비스의 장애가 발생하는 것을 원하지 않음

온라인 미디어 스트리밍 서비스의 최적화 포인트

그럼, 성공적인 온라인 미디어 스트리밍 서비스를 위하여 최적화할 수 있는 포인트는 어떤 것들이 있는지 다섯 개의 카테고리로 나누어서 살펴보도록 하겠다.



플레이어 최적화

전 세계 주요 OTT 서비스 중 트래픽 규모가 큰 서비스에서는 멀티 CDN을 운용하기도 하는데, 이는 폭발적인 트래픽 수요, 하나의 CDN에 종속됨으로 생길 수 있는 장애 방지, 지역별로 좋은 성능을 보이는 CDN 선택 등을 위함이다. 좀 더 나은 서비스 품질을 보장하는 네트워크 환경을 찾기 위하여 RUM(Real User Metric) 데이터 즉 사용자 환경에서 측정된 성능 지표를 사용하는데, 스트리밍 서비스에서는 플레이어로부터 수집된 데이터를 기반으로 최적의 서비스를 운용할 수 있다. RUM 데이터에서 수집할 수 있는 사항들의 예로는 스트리밍이 정지되었을 때 멈춤 현상이 네트워크 이슈로 인한 멈춤인지 혹은 유저가 정지 버튼을 선택한 것인지 파악한다거나, 플레이어 상의 에러 코드 수집을 통하여 에러의 원인 파악을 하거나, 혹은 클라이언트 IP 및 기기 정보 등을 제공할 수 있는데, 이런 데이터들은 온라인 스트리밍 최적화를 위한 지표의 원천 데이터로 사용된다. HTTP 기반 스트리밍 서비스의 주요 기능인 가변 비트레이트 제공(ABR : Adaptive Bitrate)을 위하여, 플레이어는 네트워크 상황에 맞는 화질을 선택하여 최적화된 서비스를 제공할 수 있다. 이때, 비트레이트 조절이 너무 공격적이면 버퍼링 이슈가 생길 수 있고, 너무 방어적이면 고화질로의 전환이 빠르게 되지 않아서 서비스 품질 저하가 올 수 있다. 또한, 플레이어에서 HLS, HDS, MPEG-DASH 등의 출력 포맷 및 파일 명에 따라 유동적으로 재생할 수 있는 기능을 지원함으로써 멀티 디바이스 지원 요건을 만족 시킬 수 있다.

오리진 최적화

오리진은 VOD 서비스에서는 원본을 제공해주는 클라우드 스토리지 등의 저장 장소이고, 라이브 스트리밍에서는 스트림을 인제스트 해주는 인코더이다. 오리진은 단순히 미디어 서비스의 원본을 제공하는 역할로만 생각하기 쉽지만, 스트리밍 서비스를 사용자에게 전달하는 CDN에 캐시 컨트롤(Cache-Control) 정보를 제공하는 원천 소스이기도 하다. 뿐만 아니라, 오리진은 VOD의 프리 패키징 혹은 라이브의 전송 방식 변환 등 동적 변환을 지원해 주는 기능을 제공하는 경우도 있다. 오리진의 캐시 컨트롤 정보는 해당 미디어 소스를 CDN에 캐시하는 주기 혹은 유효성 재확인을 위하여 리프레쉬 하는 주기로 사용되는데, 라이브 스트리밍의 Best Practice 예를 하나 설명하자면 Manifest(Playlist) 파일은 실제 Segment 파일 대비 1/2의 TTL을 유지하는 것을 권장한다. 일반적으로 HLS의 .ts Segment 파일은 10초 단위로 생성되는데, 예를 들어 Manifest 파일을 5초 주기로 업데이트하면 CDN에 캐시되는 1~2초를 감안하더라도 플레이어리스트의 업데이트가 Segment 파일의 생성 주기보다 뒤쳐지지 않게 된다. 실제, 다수의 대용량 라이브 스트리밍을 운영해 본 경험상, 3~6초 사이의 Manifest 파일 업데이트 주기를 적용하면 라이브가 원활하게 진행되는 것을 볼 수 있었다.

오리진 최적화의 또 다른 중요한 요소는 사용자의 동시 접속 요청을 받은 CDN이 오리진으로 과도한 부하를 유발하지 않도록 오리진 부하 방지를 위한 방안을 마련하는 것이다. 특히 멀티 CDN을 사용하는 서비스의 경우, 하나의 CDN에서 발생한 오리진 부하로 인하여 다른 CDN이 오리진에서 에러를 응답받아 전체 서비스에 영향을 미칠 수 있다. 오리진 부하 이슈를 방지할 수 있는 가장 손쉬운 방법은 오리진에 충분한 대역폭을 미리 확보해두는 것이겠지만, 오리진에서도 너무 많은 대역폭 확보가 필요하다면 트래픽 비용을 절감하기 위하여 CDN을 도입하여 서비스를 구성하는 목적에 위배된다. 그러므로 CDN에서 오리진 부하 발생을 방지할 수 있는 방안을 마련할 필요가 있다. 오리진 부하 방지 방안의 몇 가지 예를 들어보면 1) 동일한 URL의 HTTP 요청이 동시에 들어왔을 때 CDN과 오리진 사이의 커넥션을 제어할 수 있는 옵션 제공, 2) 오리진에서 에러 응답 시 CDN에서 에러 코드 자체를 캐시하여 에러 요청을 오리진에 지속적으로 보내지 않는 방법. 예를 들어, 어떤 미디어의 원본이 오리진에 존재하지 않아서 '404 Not Found' 응답이 리턴될 경우, CDN에서 404 응답 코드를 10초간 캐시하여 해당 기간은 존재하지 않는 파일에 대하여 오리진으로 추가 요청을 보내지 않는 방안, 3) 비인기 콘텐츠의 경우 TTL을 길게 부여하거나 클라우드 스토리지를 활용하여 글로벌 여러 곳에 사전에 원본을 배치하여 CDN과 직접 연동하는 방법 등이 있다.

클라우드 스토리지의 경우, 하나의 Active 한 주 저장소와 Inactive 한 백업 저장소의 구조로 운영하는 것보다는 다수의 지역에

Active 한 스토리지를 위치할 수 있다면, 비인기 콘텐츠가 CDN에 캐시 되어있지 않은 경우에도 사용자와 보다 가까운 위치에서 원본 파일에 접근할 수 있는 구조를 제공할 수 있다.

또한, 미디어 이외의 자막, 메타데이터, Thumbnail 이미지 등에 대한 캐시 정책을 어떻게 설정하면 좋을지도 온라인 미디어 서비스에서 고민해봐야 할 부분이다.

미디어 서비스 보안

성공적인 미디어 스트리밍에서 고려해야 할 사항 중에 보안이 빠질 수 없다. 특히 콘텐츠 제공 업체에게 TV 프로그램이나 영화 파일을 공급받아서 서비스를 제공하는 OTT 서비스나 유료 결제 혹은 회원 가입이 필요한 콘텐츠의 경우 보안의 중요성은 더욱 강조될 것이다.

미디어 서비스에서 보안이라는 주제를 떠올리면, 제일 먼저 DRM(Digital Rights Management)이 생각날 것이다. 온라인 미디어 서비스에서는 DRM과 같이 미디어 콘텐츠 자체를 보호하는 장치 이외에 서비스 접근 제한, 스트리밍 보안 및 인프라 보안을 적용하는 방법을 고려해 볼 수 있다.

먼저 Access Controls Lists(ACLs)라고 부르는 서비스 접근 제한을 적용할 수 있다. 특정 IP 혹은 IP 대역을 차단 혹은 허용하는 IP Whitelisting 혹은 Blacklisting 기능을 CDN에 적용하거나 대륙이나 국가별로 접근을 허용 혹은 차단하는 지역별 접속 권한 설정을 적용할 수 있으며, HTTP 헤더의 Referrer URL값을 이용하여 특정 사이트에서만 미디어가 재생될 수 있도록 제어하는 기능 혹은 User-Agent 값을 이용하여 특정 디바이스에서만 재생을 허용하는 기능 등을 적용할 수 있다. 예를 들어, 콘텐츠의 저작권이 미국에만 있고 '방송과기술.com'이라는 사이트에서만 재생할 수 있는 권한을 부여해야 하는 요건이 있다면, 위에 기술된 접근 제한 기술을 복합적으로 적용할 수 있을 것이다. 또한, 스트리밍 보안 방식인 HTTPS 혹은 HLS Encryption을 사용하여 스트리밍 자체를 Secured 한 레이어를 통하여 보다 안전하게 전송할 수 있다. 다음으로, 미디어에 접속하는 URL에 인증값 및 유효기간을 적용하여 CDN 엣지 서버에서 인증되지 않거나 유효기간이 지난 콘텐츠에 대하여 에러 응답을 내려주는 보안 URL(혹은 인증 URL이라고 불리는) 기능을 사용하여 URL에 대한 비정상적인 접근을 차단할 수 있다. 위에 언급하였던 콘텐츠 자체를 보호하는 DRM의 경우, 실시간으로 DRM 패키징을 수행하여 스토리지 저장 공간을 절약하고 (프리 패키징된 DRM 파일의 저장이 불필요하므로) 미디어 파일의 배포 시간을 단축할 수 있다는 장점이 있어서, VOD와 라이브에 on-the-fly DRM을 적용하는 방안도 최근에 많이 적용 중이다.

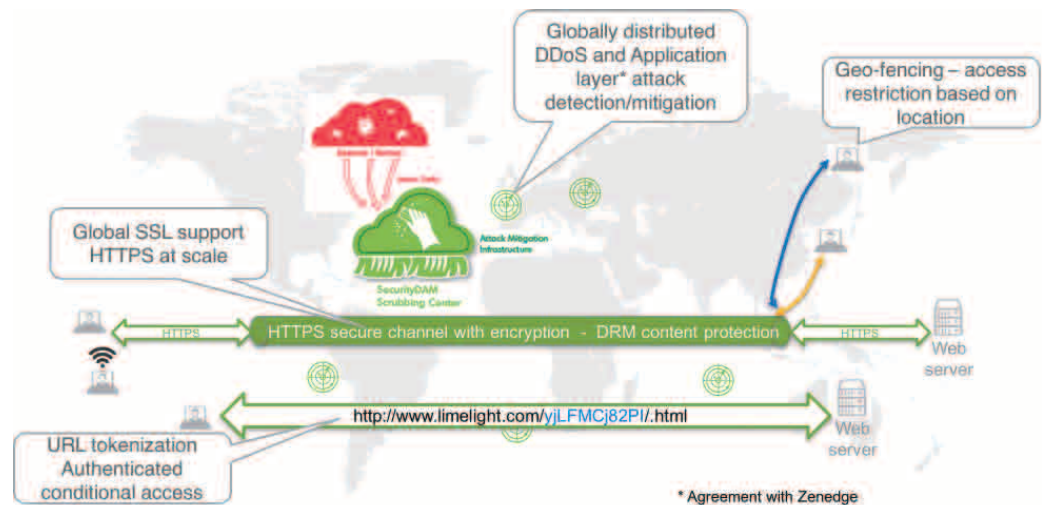


그림 3. 온라인 미디어 서비스 보안

DRM과 같은 콘텐츠 보안과 스트리밍 보안 이외에, 모든 서비스가 클라우드로 이동함에 따라 인프라 보안에도 많은 관심이 쏠리고 있다. DDoS 공격에 의하여 미디어 서비스 사이트 자체가 다운되거나 OWASP(Open Web Application Security Project) Top 10 공격 같이 사용자의 정보를 탈취하고자 하는 시도에 대비하여 WAF(Web Application Firewall)를 오리진 앞 단에 위치하여 인프라를 구성하는 것도 하나의 주요한 보안 방안이 될 것이다.

CDN 최적화

CDN(Contents Delivery Network)은 온라인 미디어 스트리밍 서비스에서 가장 핵심을 담당하는 영역이라고 할 수 있다. 온라인 미디어 서비스에 CDN 인프라를 도입할 경우 고려해야 할 사항은 크게 두 가지인데, 폭발적인 동시 접속에도 유연하게 대응할 수 있는 규모를 확보하고 있는 지와 대용량 라이브러리 보유 시 비인기 콘텐츠의 비율이 높은 경우에도 높은 캐시 히트율로 서비스가 가능한 지이다.

VOD 서비스의 경우 인기 드라마 혹은 예능 방영 직후 캐시 되지 않은 콘텐츠에 대한 동시 접속 요청에 대한 처리 능력도 CDN 인프라 선택 및 최적화 설정 시 고려해야 할 부분이다. 대용량 라이브러리를 보유하고, 지난 시즌 드라마같이 재생 요청이 빈번히 들어오지 않는 콘텐츠들도 효과적으로 CDN 인프라에서 서비스할 수 있는지도 주요 고려 포인트가 될 것이다.

반면에 올림픽, 월드컵, 미식축구 혹은 메이저리그 야구 결승전과 같이 전 세계 시청자들의 주목을 받는 주요 스포츠 경기 혹은 미국 대선 토론이나 대통령 스캔들 이슈 같은 주요 정치 사안에 대한 뉴스들은 순간 동시 접속자 수를 예측하기 어렵고 그 규모 또한 일반적인 VOD 영상 시청 패턴과 다를 것이다.

라임라이트 네트워크와 같은 글로벌 CDN 업체들은 위와 같은 상황에 대비하여 여러 가지 최적화 방안을 제시하고 있으며, 아래에 적용 가능한 몇 가지 최적화 기술들의 예를 나열해 보았다.

- TCP 스택 튜닝, 버퍼 사이즈 조절 등의 네트워크 최적화 옵션 적용
- 계층적인 POP (Point of Presence) 혹은 캐시 서버 인프라 구성
- HTTP 기반 스트리밍의 경우 Segment 파일 (예 : .ts)의 캐시 정보를 한 POP 내에서도 중앙 집중식으로 관리하여, 하나의 미디어 콘텐츠를 재생할 때 하나의 Playlist에 있는 ts들의 정보를 마스터 캐시 서버에서 관리
- 오리진에서 리턴된 에러 코드 자체를 캐시 하거나 오리진 부하를 방지할 수 있는 CDN 옵션을 적용하여 최적화
- 콘텐츠의 캐시 주기가 만료되었을 경우 무조건 리프레쉬 해주는 것이 아니라, 먼저 콘텐츠의 변경 여부를 확인하여 재 캐시 결정

온라인 미디어 서비스 인프라 구성 방안

마지막으로 온라인 미디어 서비스 인프라 구성 시 자체 구축할 수 있는 영역과 OVP(Online Video Platform), CDN 등 미디어 클라우드 서비스 제공 업체들을 이용하여 다양하게 플랫폼을 구성할 수 있는 방안에 관하여 알아본다.

첫 번째로, 클라우드 기반의 Online Video Platform을 활용하여 엔드 투 엔드 비디오 워크플로우를 구현하는 방법이 있다. 온라인 비디오 플랫폼은 일반적으로 콘텐츠의 인제스트, 트랜스코딩 및 DRM 패키징, 메타데이터 및 썸네일 이미지 등의 콘텐츠 관리, 광고 연동, 디바이스에 맞는 미디어 프로토콜 변환, 플레이어 제공, 통계 및 실제 스트리밍을 담당하는 CDN까지 하나의 플랫폼 안에서 통합적으로 온라인 미디어 서비스를 구현할 수 있는 기능을 제공한다. 온라인 비디오 플랫폼으로 서비스를 구축할 경우에는 해당 플랫폼이 실제 스트리밍 서비스를 제공하는 CDN과 얼마나 긴밀하게 연결되어 있는가를 검토하는 것이 중요한 포인트이다. 실제 서비스 운영 시 버퍼링 혹은 네트워크 이슈 등이 발생했을 때, 비디오 플랫폼과 CDN 사이의 장애 영역 파악 등의 대응에 중요한 변수가 되기 때문이다.

두 번째로, 콘텐츠 관리나 광고 연동 같은 CMS 기능은 내재화하여 구축하고, 실시간 비트레이트 변환 혹은 포맷 변환을 CDN을 이용하여 서비스하는 방법이 있다. 복잡한 CMS 요건이나 빈번히 발생하는 상품 혹은 외부 연동 변경 요건이 예측되는 경우 해당 요건의 적용은 내재화된 시스템에서 적용하지만, 미디어 인코더 등의 장비 구입에 대한 부분은 클라우드 서비스를 이용함으로써 인프라 도입에 대한 부담은 적게 가져갈 수 있는 장점이 있다.

마지막으로, CMS 및 트랜스코딩 기능은 자체로 구축한 후 CDN의 기본 기능인 미디어 스트리밍의 전송만을 연동하는 방안이 있을 것이다. 미디어 스트리밍을 제공하는 CDN이 얼마나 미디어 서비스에 특화되었는지, 글로벌 미디어 스트리밍에 경험이 있는지 등이 고려해야 할 포인트가 될 것이다.

위의 세 가지 방안 중 어떤 방안으로 서비스를 구성할지 여부는 현재 보유하고 있는 플랫폼 현황, 미디어 서비스 도입을 위한 인력 구조, 복잡한 상품 요건 혹은 다언어/다자막 지원 및 글로벌 인프라 필요 여부 등 서비스의 복잡도에 따라서 결정하면 된다.

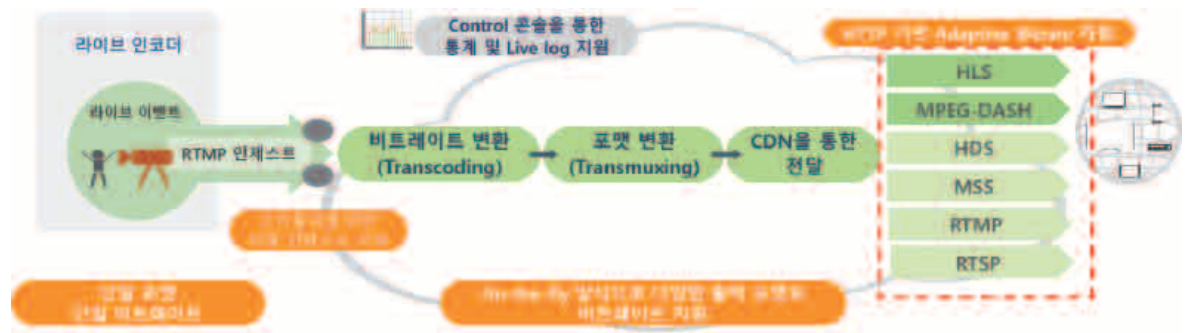


그림 4. 효과적인 라이브 스트리밍 서비스 구성 예 - 실시간 비트레이트 및 프로토콜 변환 적용 시



그림 5. 효과적인 VOD 스트리밍 서비스 구성 예 - 실시간 프로토콜 변환 적용 시

지금까지 글로벌 미디어 스트리밍 서비스의 최적화 방안에 대하여 플레이어 및 오리진 최적화, 보안, CDN 및 인프라 구성 방안에 대하여 짚어 보았다.

다음 연재에서는 빠르게 진화하는 미디어 스트리밍 기술에 발맞추어 시장 점유율을 점점 높여가고 있는 HTTP 기반 스트리밍 방식 중 MPEG-DASH의 보편화 전망 및 CMAF의 등장, UHD/4K 화질의 미디어도 원활하게 서비스할 수 있는 H.264 이후의 차세대 코덱 기술 및 HTTP 기반 스트리밍 서비스에 최적화된 HTML5 플레이어 개발 방안에 관하여 다룰 예정이다. 📺