

E-Learning in diretta live: una panoramica tecnica

Sergio Borghese, Francesco Lamonica, Stefano Lucetti, Fabio Mustacchio, Francesco Oppedisano, Federico Rossi

NetResults S.r.l., <http://www.netresults.it>

Pisa, Italy

s.borghese, f.lamonica, s.lucetti, f.mustacchio, f.oppedisano, f.rossi @netresults.it

Abstract— Dalla componente di codifica video, ai software di trasporto, alla connettività del trasmittente, alla connettività del ricevente fino al tipo di device utilizzato dal fruitore, il problema dello streaming live dei contenuti di una sessione formativa costituisce oggi un problema multidimensionale tutt'altro che semplice da trattare. L'articolo espone una panoramica delle problematiche tecniche connesse a questo tipo di sistemi di telecomunicazione.

Abstract (English version) — From video encoding, transport software, transmitter connectivity, receiver connectivity up to the type of device used by customers, the problem of streaming live content of a training session constitutes a multidimensional problem far that simple to treat. The article exposes an overview of technical problems related to this type of telecommunication systems .

Keywords— live content streaming; video encoding; unified communication and collaboration; e-learning; user experience

I. INTRODUZIONE

NetResults è un'azienda attiva nel settore del multimedia-over-IP e della "Unified Communication and Collaboration".

All'interno del progetto Città Educante collabora alla definizione della componente tecnologica ed in particolare delle tecnologie di streaming a distanza *in diretta* dei contenuti educativi.

Durante il primo anno di progetto, l'azienda ha effettuato uno studio approfondito delle tecnologie disponibili per la trasmissione a distanza in diretta di contenuti multimediali associati ad eventi live come una lezione formativa.

Dalla componente di codifica video, ai software di trasporto, alla connettività del trasmittente, alla connettività del ricevente fino al tipo di device utilizzato dal fruitore, il problema dello streaming live dei contenuti di una sessione formativa costituisce oggi un problema multidimensionale tutt'altro che semplice da trattare.

Per ogni dimensione, poi, si evidenzia una grande numerosità di possibili alternative, portando la scelta finale più verso una soluzione di "ottimo locale" che verso una ottimalità assoluta.

Inoltre si evidenzia una forte correlazione tra il valore oggettivo di alcuni parametri di rete e il livello di ergonomia che si riesce ad ottenere con questo tipo di sistemi.

Il presente articolo vuole sintetizzare i parametri tecnologici e qualitativi che possono influenzare le scelte tecnologiche nell'ottica di chi intende implementare una piattaforma di e-learning che supporti anche lo streaming in diretta dei contenuti. Quanto riportato costituisce parte dei risultati delle attività del primo anno di progetto.

II. PANORAMICA

A. Elementi del sistema in esame

Il sistema di telecomunicazione preso a riferimento per il presente studio prevede a livello più alto di interazione umana (Figura 1) l'esistenza di un'aula in cui un docente svolge la propria lezione in maniera frontale. In aula, inoltre, sono presenti alcuni discenti che, almeno in prima istanza, interagiscono col docente in maniera tradizionale ovvero non mediata da alcun device di comunicazione come smartphone o tablet.

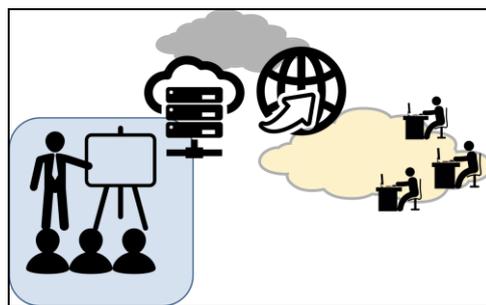


Figura 1: schema generale a livello di interazione umana

L'aula, inoltre, è attrezzata con opportuni sistemi di cattura video che consentono la trasmissione dei contenuti audio/video verso la rete Internet.

Attraverso la rete Internet medesima, poi, uno o più discenti remoti possono fruire i contenuti formativi in tempo reale rispetto alla loro erogazione.

Da un punto di vista funzionale, il sistema può essere osservato dalla prospettiva più tecnica fornita dalla Figura 2.

In essa possiamo notare come l'aula sia attrezzata, generalmente, con sistemi di cattura e compressione audio/video in grado di comunicare con una centrale di servizi che raccoglie i flussi ed implementa la *business logic*. Una caratteristica di fondamentale importanza in relazione alla componente di cattura e compressione audio/video riguarda la risoluzione del video prodotto ed il rapporto di compressione ottenibile oltre, naturalmente, a fattori di grande importanza come il ritardo di processing e la qualità ottenibile.

Dagli studi effettuati in seno al progetto è emerso che non è possibile definire una "best choice" per qualunque situazione. In particolare, si sono evidenziate due grosse categorie di flussi video: flussi video derivanti da **scene naturali** come una lezione tenuta da un docente o una partita di calcio, ecc. e flussi

video che catturano **lo schermo di un computer o una lavagna interattiva** (LIM).

Queste due macro-casistiche determinano un vero e proprio spartiacque tecnologico.

Per quanto riguarda la cattura di scene di tipo naturale, sono consuete **risoluzioni di cattura** che vanno dal 720p al fullHD (1920x1080) con un **frame rate** che va dai 15 ai 30 fps. Il video prodotto, poi, viene compresso con codec standard tra cui i più famosi sono:

1. **H.264/AVC** ([1]), sicuramente il più diffuso ed interoperabile
2. **H.264/SVC** ([1]), molto promettente dal punto di vista tecnologico perché “multi-layer” ma molto poco diffuso allo stato attuale
3. **H.265** ([2]), si tratta del successore di H.264 ma ancora è poco diffuso
4. **VP8** ([3]) si tratta del codec open source usato per lo standard WebRTC ([4])
5. **VP9** ([5]) si tratta del successore di VP8

Il processing associato alla compressione video è probabilmente una delle attività più pesanti nel mondo delle telecomunicazione. Oltre determinate risoluzione, infatti, è inusuale trovare sistemi di compressione software e si trovano più spesso soluzioni accelerate in hardware con sistemi FPGA. Dallo studio è emerso che mediante un personal computer di comune utilizzo non è pratico effettuare compressioni video a risoluzioni superiori a 720p@20fps soprattutto quando la scena che viene catturata è molto movimentata (ad es. eventi sportivi).

Per sopperire a queste difficoltà esistono alcuni sistemi che consentono la cattura e la compressione in real-time tramite il codec H.264/AVC.

In seno al progetto ne sono stati testati alcuni con risultati che non possono essere qui esposti per motivi di spazio:

1. Epiphan ([6]) **DVI2USB**, con ingresso HDMI e interfacciamento USB3
2. AVerMedia ([7]) **CaptureHD H727**, con ingresso HDMI e interfacciamento PCI
3. Auvideo ([8]) **E110** con ingresso HDMI e interfacciamento ethernet

In relazione al bitrate, poi, i nostri test hanno evidenziato come per avere un flusso video di qualità accettabile in risoluzione 720p@25fps sono necessari almeno 2Mbps e che, comunque, la variabilità del bitrate al cambiare della quantità di moto presente nella scena non è così sensibile. In pratica, anche scene quasi statiche creano flussi con bitrate significativi.

Nel mondo della cattura e compressione di schermi desktop o LIM le cose cambiano in modo sostanziale.

Le risoluzioni FullHD, infatti, sono ormai all’ordine del giorno mentre stanno diventando sempre più diffusi schermi *hiDPI* che raggiungono risoluzioni ben superiori. Questo vale per il mondo desktop: il mondo delle LIM, in realtà, è molto più

limitato dalla tecnologia dei proiettori *short throw* spesso associati a questi apparecchi e che difficilmente raggiungono risoluzioni proibitive per i sistemi di codifica.

In questo contesto, comunque, l’uso di codec come quelli descritti in precedenza non sortisce i risultati sperati sotto molti punti di vista:

1. I codec utilizzati per scene naturali sono inerentemente *lossy* ovvero la compressione si ottiene attraverso una *sgranatura* dell’immagine che in contesti desktop o LIM può pregiudicare la leggibilità dello schermo
2. I codec utilizzati per scene naturali, inoltre, non riescono ad ottenere fattori di compressione paragonabili a quelli ottenuti dai codec specifici per questo tipo di scene
3. I codec utilizzati per scene naturali sono resistenti alla perdita di pacchetti ma in presenza di tali perdite l’immagine subisce distorsioni non accettabili nei contesti desktop/LIM

Per i motivi appena esposti, si ritiene che l’uso di sistemi che si basano su codec specifici per questo tipo di scene sia da preferirsi rispetto al caso di sistemi che “forzano” l’uso dei codec standard per questi contesti.

Uno dei codec individuati per questo tipo di lavoro è lo ScreenPressor (<http://www.infognition.com/ScreenPressor/>). Si tratta di un codec più orientato al salvataggio dei file che alla comunicazione real-time ma costituisce comunque un ottimo esempio di codec specifico. E’ un codec lossless (a meno di decimare, facoltativamente, il numero di bit di colore) e riesce ad ottenere fattori di compressione e frame rate molto importanti. La Figura 3 (reperibile sul sito del produttore) delinea un confronto in termini di rapporto di compressione tra ScreenPressor ed altri comuni codec utilizzati per la registrazione di *screencast*. Pur non avendo replicato tutti i test esposti dal produttore, i test di laboratorio effettuati confermano che si tratta di un codec molto efficiente sia in termini di compressione che in termini di qualità dell’immagine ottenuta. Esso, inoltre, si è rivelato efficiente anche con risoluzioni superiori al fullHD.

E’ bene, infine, notare come la scelta dei codec (di qualunque tipologia) debba tenere in considerazione l’aspetto del modello di licensing e/o di *patent indemnification*. Spesso, infatti, i codec sono sviluppati a valle di processi di brevettazione che poi si ripercuotono sul costo finale della soluzione prescelta. La disponibilità di implementazioni hardware o software dei codec, inoltre, non implica necessariamente che le relative obbligazioni di indennizzo siano state ottemperate. Nelle scelte tecniche, pertanto, è di fondamentale importanza la rilevazione tempestiva di questa potenziale fonte di extra-costi.

Tornando allo schema funzionale di Figura 2, il flusso video compresso viene inviato in tempo reale alla centrale di servizio attraverso la connessione ad internet presente in aula. Tale collegamento, pertanto, avrà un ruolo cruciale nella qualità del servizio fornito end-to-end.

La centrale di servizio implementa, innanzi tutto, la funzione di MCU (Multipoint Control Unit) che ha in carico il

processing e l'eventuale ritrasmissione verso i ricevitori dei flussi video. Fino a questo punto, il sistema descritto assomiglia ad un sistema di videoconferenza reperibile sul mercato presso svariati produttori; infatti molte piattaforme in produzione implementano questa componente del sistema adattando sistemi di videoconferenze commerciali.

Allo stato attuale degli studi effettuati, tuttavia, nessun sistema utilizza codec specifici per la componente desktop/LIM per cui spesso si possono riscontrare inefficienze in questo specifico contesto.

Oltre a svolgere la funzionalità di MCU, la centrale di servizio dovrà ospitare la componente di back-end, attraverso cui, in generale, i discenti potranno accedere ai contenuti registrati in modalità on-demand. I contenuti inviati dal sistema d'aula, pertanto, dovranno essere processati per essere adattati alla piattaforma di streaming on-demand e, successivamente immagazzinati sui sistemi di storage e opportunamente "taggati" per facilitarne la successiva ricerca da parte dei discenti. I contenuti, inoltre, dovranno essere associati ad altri contenuti come ad esempio appunti o slide o altro materiale associato alla medesima lezione/evento.

A seconda, quindi, delle funzionalità ricercate nel "Service Data Center", si potranno trovare quanto meno i seguenti sistemi integrati tra loro:

1. Sistema **MCU**
2. Piattaforma di **streaming on-demand**
3. Piattaforma di **E-learning**
4. Sistema di **data storage**

Come già accennato, comunque, il focus del presente lavoro è la componente real-time ovvero la componente MCU.

A tale unità, infatti, dovranno connettersi i dispositivi utilizzati dai discenti per la fruizione degli eventi formativi.

La situazione ideale prevede la massima flessibilità nell'accesso ai contenuti sia da postazione desktop che in mobilità. Sarebbe quindi auspicabile che il sistema MCU funzionasse secondo standard aperti ed accessibili anche mediante software di terze parti e/o gratuiti e/o open source anche a patto di rinunciare ad una o più funzionalità avanzate. In altri termini sarebbe auspicabile una "openness" non intesa come gratuità ma come apertura degli standard di comunicazione. Un esempio potrebbe essere la fruizione attraverso lo standard **WebRTC**. In questo modo sarebbe possibile la fruizione da PC, tablet e smartphone dei più diffusi sistemi operativi.

Proseguendo nella panoramica end-to-end del sistema, altri due elementi di cruciale importanza per la qualità finale della fruizione è la disponibilità di potenza di calcolo nel data-center e la capacità del collegamento verso Internet del data-center stesso. Questi due elementi, infatti, determineranno la scalabilità del sistema ed il degrado qualitativo percepito quando la numerosità dei discenti inizia a crescere.

Infine, la disponibilità di processing sul device di ricezione del discente e la capacità del suo collegamento ad Internet determinato la qualità finale percepita. In relazione a

quest'ultima componente, comunque, il sistema può solo aiutare la fruizione attraverso un adattamento dinamico del contenuto alle possibilità contingenti del discente. E' comunque evidente che si tratta di una componente su cui, in ultima analisi, il sistema non può intervenire.

B. Parametri del sistema

Il punto di vista, in questo momento, è focalizzato sui discenti e sulla capacità del sistema di trasmissione dati di fornire loro un contenuto il più possibile fedele a quello catturato dai sistemi d'aula. E' argomento di altre componenti di studio la disamina delle tecnologie che possono migliorare la qualità di quanto catturato già in partenza; per motivi di spazio, tuttavia, tale argomento non verrà trattato nel presente lavoro.

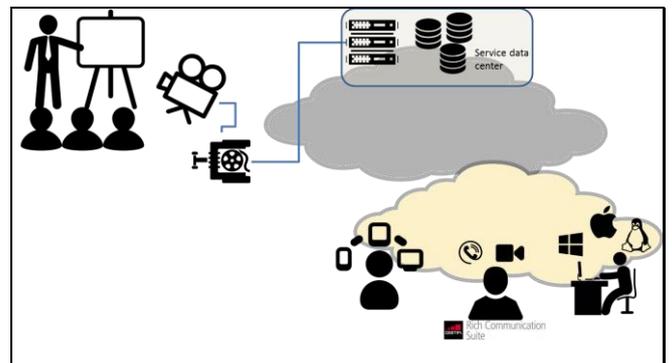


Figura 2: schematizzazione funzionale del sistema

Quella che segue è una lista, necessariamente non esaustiva, dei parametri che influenzano la qualità della fruizione dei contenuti da parte dei discenti. Come suggerito dalla Figura 2 gli elementi possono essere suddivisi nelle seguenti macro-categorie:

1. Parametri che riguardano l'**attrezzatura d'aula**
2. Parametri che riguardano gli **allestimenti di "data center"**
3. Parametri che riguardano l'**attrezzatura dei discenti**
4. Parametri che riguardano la qualità della **connessione di rete end-to-end**

I paragrafi che seguono, elencheranno molti dei parametri appartenenti ad ognuna di queste categorie.

Naturalmente, tutto quanto descritto va poi inserito in impianto concettuale che tenga conto anche di aspetti di tipo economico finanziario come gli investimenti iniziali, i costi ricorrenti ed il *time to market*.

C. Parametri che riguardano l'attrezzatura d'aula

1. **Risoluzione del video** catturato
2. **Qualità dell'audio** catturato
3. Qualità della **compressione audio/video** utilizzata
4. **Capacità del collegamento Internet dell'aula**

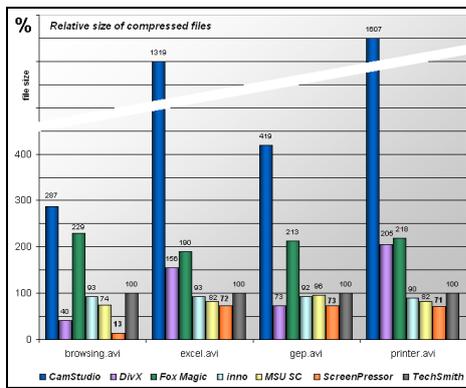


Figura 3: confronto nel rapporto di compressione di vari codec (Fonte: Infognition)

D. Parametri che riguardano gli allestimenti di "data center"

1. Architettura di **content delivery** utilizzata
2. Possibilità di **registrare il contenuto su storage** per futura fruizione on-demand e relativa integrazione con sistemi di back-end per la gestione dei discenti
3. **Capacità computazionale** dei sistemi server utilizzati in termini di processing per eventuali transcodifiche
4. **Capacità del collegamento Internet dei server**
5. Numero di **contemporaneità** di discenti remoti
6. Capacità dei server di **adattare il media** al device ed al tipo di accesso utilizzato da ogni singolo discente
7. Capacità dei server di **adattare il tipo di codifica** alla qualità del link di accesso del singolo discente
8. Possibilità per i discenti di **intervenire in aula con domande** o osservazioni
9. Necessità di supportare **software generici** per la fruizione da parte dei discenti.
10. Possibilità di consentire la fruizione degli eventi attraverso **servizio telefonico standard oppure "aumentato" attraverso i nuovi standard RCS** (Rich Communication Suite) oppure attraverso altri **servizi telefonici OTT** (Over the Top) come **Skype**.

E. Parametri che riguardano l'attrezzatura dei discenti

11. **Capacità del collegamento Internet utilizzato dai discenti**
12. **Device utilizzato dai discenti** per la fruizione della formazione e possibilità per i discenti di fruire delle lezioni **in mobilità**

13. Alternative di **sistema operativo desktop** utilizzato dai discenti

F. Parametri che riguardano la qualità della fruizione dei contenuti end-to-end

A seconda del livello di fruibilità che il sistema vuole garantire, dovranno essere definiti i parametri tecnici del sistema in termini di parametri qualitativi end-to-end.

Tra questi parametri possiamo annoverare:

1. La qualità dell'audio in termini di **Listening MOS** (Mean Opinion Score), misurata oggettivamente attraverso sistemi come il PESQ o il POLQA
2. La qualità dell'audio in termini di **Conversational MOS**
3. La qualità del video in termini di MOS, misurata oggettivamente attraverso sistemi come il PEVQ
4. Lo **streaming MOS** definito in base ai buffer underrun dei sistemi riceventi
5. La perdita di pacchetti (**packet loss**)
6. Le distorsioni audio da **eco e riverbero**
7. Il ritardo bidirezionale (**RTT**)
8. La varianza del ritardo (delay **jitter**)
9. Il **bitrate** necessario per ottenere determinati livelli di qualità
10. Il **sincronismo** audio/video

III. CONCLUSIONI

L'articolo descrive le principali problematiche associate alla definizione di un sistema di e-learning con supporto della comunicazione in diretta dei contenuti. Quanto riportato costituisce un breve estratto degli studi effettuati durante il primo anno di attività in seno al progetto Città Educante dall'azienda NetResults S.r.l.

RINGRAZIAMENTI

Si ringrazia tutto lo staff del progetto Città Educante.

CREDITS:

Icons made by <http://www.freepik.com> from <http://www.flaticon.com> and licensed under "Creative Commons BY 3.0"

BIBLIOGRAFIA

- [1] <http://www.itu.int/rec/T-REC-H.264>
- [2] <http://www.itu.int/rec/T-REC-H.265/en>
- [3] <http://www.webmproject.org/docs/vp8-sdk/>
- [4] <http://www.webrtc.org/>
- [5] <http://www.webmproject.org/vp9/>
- [6] <http://www.epiphany.com/>
- [7] <http://www.avermedia-usa.com/AVerTV/>
- [8] <http://www.auvidea.com/>