

## Avalanche: peer-to-peer Microsoft progettato al sistema Network Coding.

### Prefazione:

Questo documento si basa sull'originale "Network Coding for Large Scale Content Distribution" firmato da alcuni ricercatori Microsoft. Il sistema di distribuzione di contenuti digitali in oggetto utilizza un network di nodi, peer-to-peer che codificano vari pacchetti del file originale condividendoli con i nodi vicini.

Solitamente la distribuzione di contenuti digitali in rete avviene attraverso l'implementazione di apparati appositi, su internet. Negli ultimi anni è emerso il paradigma in cui in maniera opposta, la distribuzione dei contenuti viene affidata a sistemi completamente distribuiti, dove gli utenti creano un network basato sulla cooperazione per condividere le proprie risorse.

Il punto di forza di questo sistema di distribuzione dei contenuti cooperativo risiede nella scalabilità:

- La banda per il download a disposizione di ogni utente aumenta all'aumentare degli utenti.
- La disponibilità di fonti e di contenuti cresce con il crescere del network.

Praticamente la capacità del sistema cresce insieme con la crescita della domanda. In questo modo i server che forniscono il contenuto originario non hanno bisogno di incrementare le risorse ogni volta che crescono gli utenti. Diversi tipi di contenuti come patch, video o altri software possono essere distribuiti abbassando i costi e senza intasare i server originari.

### Network Coding

Il Network Coding è un sistema in grado di ottimizzare il rendimento di una rete. Il sistema consente ai nodi intermedi, quelli posti tra fonte originale e destinatario, di codificare i pacchetti in maniera da rendere ottimizzato l'utilizzo delle risorse.

In pratica ogni volta che un utente deve inviare un pacchetto ad un altro utente, l'utente originario genera una combinazione lineare di tutte le informazioni disponibili. Il destinatario riceve una combinazione indipendente di pacchetti dai quali può ricostruire l'informazione originaria.

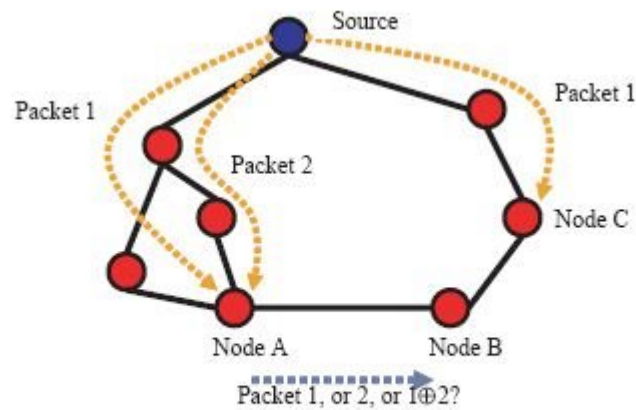


Fig. 1. Network Coding benefits when nodes only have local information.

In un sistema distribuito e cooperativo trovare un sistema che consenta la propagazione dei pacchetti in maniera ottimale è molto difficile. Questo perché viene a mancare il supporto offerto dall'eventuale presenza di un server centralizzato che gestisce le informazioni. Le decisioni riguardo l'instradamento dei pacchetti sono affidate a ciascun nodo anche se la difficoltà di instradamento aumenta con l'aumentare dei nodi.

## Tecnica

Il sistema in questione può essere utilizzato per distribuire contenuti in diverse maniere. Tramite blocchi originali del file non codificati, tramite blocchi di file codificati dalla fonte o tramite blocchi di file codificati sia dalla fonte che dai nodi della rete.

Supponiamo che un gruppo di utenti siano intenzionati alla ricezione di un file che in origine risiede su un solo server; la capacità del singolo server ovviamente è limitata ma gli utenti possono contribuire con la banda che hanno a disposizione per facilitare il download del file ad altri utenti.

Nel momento in cui il server non ha la capacità di servire tutti gli utenti in maniera simultanea, divide il file in un numero  $K$  di blocchi, effettuando upload di questi blocchi sulle macchine degli utenti. Gli utenti collaborano quindi tra loro per ricostruire il file iniziale.

Una delle condizioni è che ciascun utente non conosce l'identità degli altri utenti. Ogni utente conosce solo l'identità di una piccola porzione di tutti gli utenti, che denomineremo vicini.

La relazione tra vicini è simmetrica. Quindi se il nodo  $A$  è il vicino del nodo  $B$ , a sua volta il nodo  $B$  è il vicino del nodo  $A$ .

Ogni nodo può scambiare informazioni, parti di file o messaggi vari di protocollo, esclusivamente con i propri vicini. L'insieme di vicini è comunque un numero molto limitato.

Quando un utente si connette, un server centralizzato offre una lista casuale di utenti connessi, a questo punto l'utente può collegarsi a ciascuno di essi e creare il proprio rapporto di vicinanza.

Se un nodo perde uno dei vicini, può richiedere o consentire la connessione di altri vicini.

In questo scenario il principale collo di bottiglia è nella capacità di accesso e di collegamento ai vicini, proprio di ciascun utente.

La capacità di ricezione delle informazioni di ciascun utente è limitata alla sua capacità di download. Nella scenario descritto viene supposto che la capacità di upload e download sia simmetrica e indipendente. Test condotti su accessi di tipo asimmetrico hanno portato a risultati simili.

Ogni volta che si presenta un trasferimento di un blocco dal server o dall'utente ad un altro utente viene presa una decisione su come sarà scaricato il blocco. Il server o l'utente non conosce in dettaglio i blocchi che ciascun nodo possiede. Ogni utente ha informazioni su blocchi che ha scaricato o sui blocchi dei vicini, queste informazioni sono la base dell'algoritmo che gestisce il trasferimento.

Il sistema si basa su sistemi euristici, già adottati da altri sistemi:

- Random block: il blocco da trasferire è scelto in maniera casuale tra i blocchi presenti nella fonte.
- Local rarest: il blocco da trasferire è quello meno diffuso tra i vicini. Se si verifica che più di un blocco è il meno diffuso, viene scelto un blocco casuale tra questi.
- Global rarest: il blocco da trasferire è scelto tra un campione di meno diffusi tra i vicini. In questa maniera viene trasferito il blocco probabilmente più raro.

Il diffuso sistema BitTorrent usa la combinazione delle tecniche Random block e Local rarest.

Quando il server codifica i dati da trasferire la tecnica è simile a quella di BitTorrent. Il server offre pacchetti di file codificati e non il singolo file originale.

Il server genera un nuovo blocco codificato ogni volta che ha l'esigenza di caricare sulla macchina di un utente.

Qui gli utenti devono scaricare un numero  $K$  di blocchi sia dal server che dai nodi, che andranno a ricostruire il file originario.

Nel caso della codifica effettuata dai nodi, sia il server che l'utente effettua la codifica. Quando un nodo o il server deve inviare i dati ad un altro nodo, produce una combinazione lineare dei blocchi che possiede

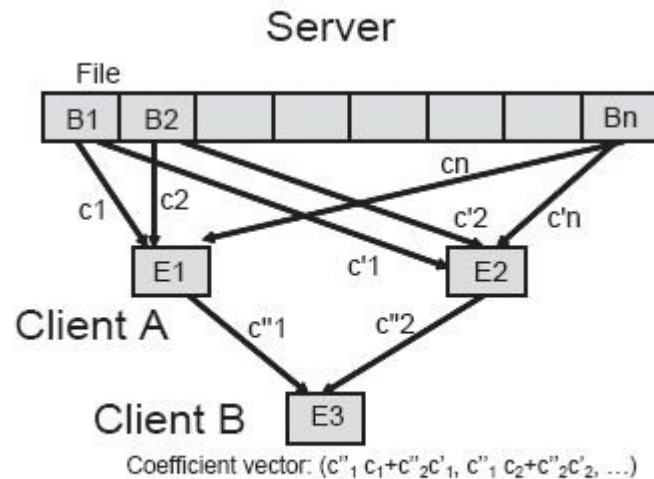


Fig. 2. Sample description of our network coding system.

La ricostruzione del file a partire dai blocchi è un processo simile alla risoluzione di un'equazione lineare.

Il beneficio nell'utilizzo della codifica da parte dei nodi è dovuto alla casualità introdotta ogni volta che si genera un blocco codificato. Va sottolineato che senza questo tipo di codifica ogni utente avrebbe bisogno di decidere quale file scaricare, in base alle proprie esigenze del momento. Con la codifica da parte dei nodi si ha una combinazione lineare dei blocchi, che combina ogni blocco senza attribuire minore o maggiore popolarità ad ognuno. È importante ricordare che in questa maniera non ci interessa quale blocco trasmettere agli altri nodi.

Un difetto degli attuali sistemi collaborativi di distribuzione dei contenuti risiede in atteggiamenti definiti free-riding. Qui un utente beneficia delle risorse degli altri utenti, senza però contribuire con risorse proprie. Questo atteggiamento degrada la qualità del sistema, per questo motivo alcuni circuiti di distribuzione scoraggiano atteggiamenti del genere. Nel sistema in oggetto le risorse per lo scambio dei dati vengono assegnate maggiormente a chi effettua upload e quindi offre contenuti. Inoltre ciascun utente, come nel sistema BitTorrent, non può consentire l'upload di file verso un altro utente finché quello non ha consentito il download verso l'utente originario.

## Conclusioni

Il sistema di distribuzione utilizza in maniera innovativa la tecnologia del Network Coding. Lo scopo è quello di offrire la distribuzione di file di grandi dimensioni in un contesto dinamico, dove ciascun nodo collabora con gli altri nodi. Il sistema non prevede la presenza di un server o comunque di un'intelligenza centrale che coordina lo scambio di dati. Lo scambio di dati e le decisioni sulla propagazione dei blocchi è affidato ai singoli nodi. Questo sistema è ottimale per quelle situazioni in cui i nodi sono eterogenei e la connessione/disconnessione dei nodi non è sincronizzata, dove ci sono colli di bottiglia naturali e esiste un meccanismo che scoraggi la pratica del free-riding.

I miglioramenti sono del 20-30 % riguardo alla distribuzione da un server centralizzato. Il sistema **Avalanche** è un sistema sperimentale che consente di testare la distribuzione di grandi quantità di file attraverso la tecnica del Network Coding.

### Chiarimenti

Questo documento non garantisce nessuna attinenza con la ricerca "**Network Coding for Large Scale Content Distribution**". L'autore non è responsabile per eventuali omissioni, rimaneggiamenti o errori.

"**Network Coding for Large Scale Content Distribution**" disponibile qui:  
[http://www.research.microsoft.com/~pablo/papers/nc\\_contentdist.pdf](http://www.research.microsoft.com/~pablo/papers/nc_contentdist.pdf)