# SISTEMI P2P: INTRODUZIONE Lezione 1

Laura Ricci

24-2-2009



## INFORMAZIONI UTILI

L'esame può essere sostenuto dagli studenti delle Lauree Specialistiche in Informatica, Tecnologie Informatiche, Informatica per l'Economia e l'Azienda. Gli studenti della laurea triennale possono utilizzare i 9 crediti liberi per inserire l'esame nel proprio piano di studi, se non già utilizzati per altri esami/seminari.

#### Materiale Didattico:

- Libro di Testo:
  - P2P Networking and Applications, Jhon F.Budford, H.Yu, E.K.Lua, Morgan Kaufmann
- lucidi delle lezioni
- tutorial e materiale pubblicato sulla pagina delcorso

#### Modalità di esame:

Scritto o progetto (a scelta) + orale obbligatorio per tutti



## SVOLGIMENTO DEL PROGETTO

Durante il corso verranno proposti diversi progetti, che possono essere realizzati mediante i seguenti supporti:

- Peersim
  - simulatore orientato alla simulazione di reti P2P
  - altamente scalabile (fino a 10000 peer per semplici protocolli)
  - http://peersim.sourceforge.net/
- Overlay Weaver
  - Supporta diversi algoritmi di routing
  - Fornisce una semplice interfaccia verso diverse distributed hash tables
  - http://overlayweaver.sourceforge.net/

La simulazione di una rete P2P richiede l'implementazione di un insieme di interfacce/l'estensione di un insieme di classi JAVA



## SVOLGIMENTO DEL PROGETTO

- In alternativa è possibile progettare una rete P2P 'reale' e testarne il comportamento sulla rete locale del polo e/o su rete geografica.
- In questo caso è possibile utilizzare JXTA
  - Un supporto per lo sviluppo di applicazioni P2P
  - Estensione di JAVA: gruppi, advertisements, pipes
  - https://jxtadev.java.net/
- Esempi di proposte di progetto (anni precedenti)
  - realizzazione di
    - una chat in JXTA
    - overlay strutturati in JXTA
  - simulazione di
    - reti P2P non strutturate
    - Distributed Hash Tables



## SISTEMI PEER TO PEER: INTRODUZIONE

- Definizione: Un sistema Peer to Peer è un insieme di entità autonome (peers), capaci di auto-organizzarsi, che condividono un insieme di risorse distribuite presenti all'interno di una rete di computers. Il sistema utilizza tali risorse per fornire una determinata funzionalità in modo completamente o parzialmente decentralizzato.
- Peer= Pari, uguale (persona di pari grado, coetaneo)
- Risorse condivise:
  - Informazioni (Files)
  - Spazio di memorizzazione (Distributed File System)
  - Potenza di calcolo
  - Banda



## CONDIVISIONE DI RISORSE

- P2P: riguarda il dare e ricevere da una comunità. Ogni peer fornisce una risorsa ed ottiene in cambio altre risorse.
  - situazione più comune: si offre musica al resto della comunità, si ottiene altra musica in cambio (Napster, Gnutella,...)
  - un peer ha sia funzionalità di client, che di server (funzionalità simmetrica = SERVENT)
- Ma un peer può decidere di offrire gratuitamente risorse, ad esempio per partecipare ad una 'giusta causa'.
  - Ricerca di vita extra-terrestre
  - Ricerca sul cancro
- Le risorse condivise si trovano 'ai bordi' di Internet, cioè sono fornite direttamente dai peers, non esistono nodi 'special purpose' definiti solo per la loro gestione.



## CONDIVISIONE DI RISORSE

- La connessione dei peers al sistema è intermittente (transiente): le disconnessioni e le riconnessioni al sistema sono frequenti
- Le risorse offerte dai peers vengono aggiunte e tolte dinamicamente al sistema
- Ad un peer può essere associato un indirizzo IP diverso per ogni diversa connessione al sistema

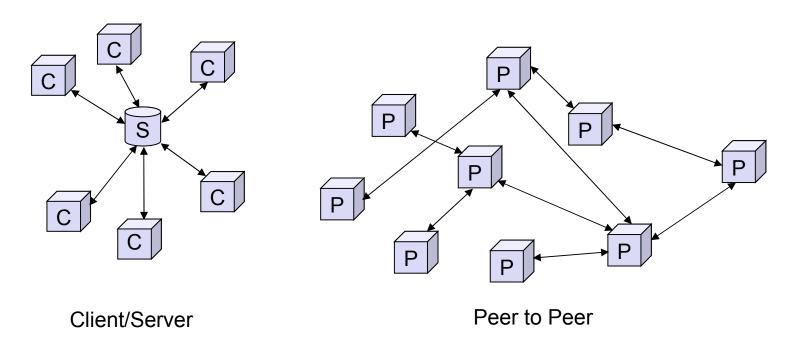
E' impossibile localizzare una risorsa mediante un indirizzo IP statico

Definizione di nuovi meccanismi di indirizzamento, definiti a livello superiore rispetto al livello IP

**Introduzione** 

## SISTEMI PEER TO PEER: ARCHITETTURA

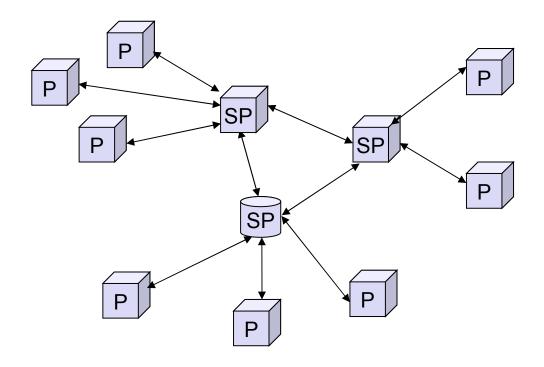
- I peer interagiscono direttamente tra di loro senza l'intervento di un server centralizzato
- Paradigma di interazione basato su una cooperazione decentralizzata, piuttosto che su coordinamento centralizzato





## ARCHITETTURE IBRIDE

- Per migliorare la performance del sistema, sono possibili soluzioni ibride
  - Definizione dinamica di Superpeers che indicizzano le risorse dei peer
  - Le risorse possono essere scambiate direttamente tra i peer



**Peer to Peer Ibrido** 

**Introduzione** 



## FILE SHARING: UNA 'KILLER APPLICATION'

- Condivisione di files: nasce con il rapido successo di Napster, alla fine degli anni 90, circa 10 anni dopo l'affermazione del World Wide Web
- Prima generazione: Napster
  - introduce un insieme di servers su cui gli utenti registrano i descrittori dei files che intendono condividere
  - la trasmissione dei contenuti (download/upload) avviene con modalità
    P2P
  - La presenza di directory centralizzate è stato il 'tallone di achille' di questa applicazione
  - Napster è stato giudicato responsabile per non aver rispettatole leggi sul copyright, proprio perchè avrebbe potuto individuare i contenuti scambiati illegalmente tra gli utenti mediante l'analisi delle directory centralizzate



## FILE SHARING: UNA 'KILLER APPLICATION'

- Seconda generazione
  - Non esiste alcun punto di centralizzazione
  - Sia la ricerca dei files che il trasferimento del contenuto avviene in modo completamente distribuito tra i peer
  - Gnutella, FastTrack/Kazaa, BitTorrent
  - Freenet: introduce diversi meccanismi per garantire l'anonimato dei peer
    - Sostituzione del mittente di un messaggio
    - Tecniche di encrypting
- Effetti laterali della diffusione delle applicazioni di file sharing: modifica radicale della modalità di fruizione di contenuti musicali
  - dai CD a musica online
  - iTunes



## FILE SHARING: UN'APPLICAZIONE KILLER

- Come funziona un'applicazione perla, condivisione di files musicali
- Tipico utilizzo del sistema P2P
  - Un utente U possiede un applicativo P2P sul suo notebook
  - Si interconnette ad Internet in modo intermittente: ottiene un nuovo indirizzo
    IP per ogni diversa connessione
  - Offre alla comunità alcune canzoni, registrandole in una directory condivisa, associando ad ogni canzone alcune chiavi per individuarla (titolo, autora, data di pubblicazione,...)
  - U è interessato a reperire una canzone ed invia una query al sistema
  - L'applicativo visualizza informazioni circa gli altri peers che posseggono la canzone richiesta
  - U sceglie un peer P (vedremo in seguito in base a qule criterio)
  - Il file viene copiato dal PC di P a quello di U
  - Mentre U effettua il dowload, altri utenti possono effettuare upload di qualche file messo a disposizione da U



## FILE SHARING: UN'APPLICAZIONE KILLER

#### Il software P2P consente:

- All'utente di definire una directory, nel proprio file system, dove memorizzare i files che vuole condividere con la comunità. Ogni altro peer può reperire files da quella directory
  - Il peer si comporta come un web server
- All'utente di copiare files dalle directory condivise dagli altri utenti
  - Il peer si comporta come un client
- Agli utenti di individuare il materiale richiesto, mediante queries sottoposte al sistema
  - Funzionalità analoga a Google
- L'applicativo sull'host dell'utente si comporta come un servlet



## FILE SHARING: PROBLEMI

- Molti applicativi P2P
  - sono scaricabili gratuitamente dalla rete, ma possono contenere meccanismi di spyware o di malware,
  - in realtà quindi il software non è gratuito, ma viene pagato attraverso un'invasione della privacy dell'utente, spesso inconsapevole.
- Spyware = un software che
  - raccoglie informazioni riguardanti l'attività online di un utente (siti visitati, acquisti eseguiti in rete etc) senza il suo consenso
  - trasmette tali informazioni tramite la rete ad un'organizzazione che utilizza tali informazioni per trarne profitto, solitamente attraverso l'invio di pubblicità mirata.
- Malware = un software che può causare danni più o meno gravi al computer su cui viene eseguito.



## FILE SHARING: PROBLEMI

#### Pollution (= Inquinamento)

- Viene immessa nella rete P2P una grossa mole di materiale corrotto
- I peer non riescono a distinguere il materiale integro da quello corrotto, scaricano e contribuiscono alla diffusione di materiale corrotto
- Il numero delle copie corrotte può superare quello di quelle integre
- Esistenza di vere e proprie pollution companies: Overpeer

#### Presenza di free riders

- Peer che utilizzano l'applicazione per scaricare contenuti dagli altri peer,
  ma non contribuiscono a fornire contenuti
- Soluzioni:
  - Meccanismo di incentivi (crediti in eMule)
  - Meccanismo per l'individuazione e la penalizzazione dei free riders
- Peer churn: presenza intermittente dei peer sulla rete



## SISTEMI PEER TO PEER: APPLICAZIONI

- P2P file sharing
  - Napster
  - Gnutella, KaZaa
  - eMule
  - BitTorrent
- Instant Messaging and Voice over P2P
  - IMP systems: Instant Message and Presence Applications
  - VoP2P: Skype
- P2P TV
  - Video Streaming applications
- P2P distributed Storage
  - OceanStore, Freenet
- P2P entartainement
  - Multiplayer games



#### IMP AND VoP2P

- IMP (Instant Messaging and Presence) applications sviluppate a partire dalla metà degli anni '90 sono spesso progettate secondo il modello client-server e
- Possono offrire servizi di voce/video desktop to desktop
  - Microsoft Messanger
  - Yahoo!Messanger
  - Jabber
- Scarsa diffusione del servizio voce/video
- Skype:
  - VoP2P client introdotto nel 2003
  - 10 milioni di utilizzatori negli ultimi anni
  - offre sia servizi di chiamata desktop-to-desktop che deskop-to-public switched-telephone-network
  - Buona qualità di voce è alla base del successo della applicazione
  - Include funzionalità tipiche di un'applicazione IMP



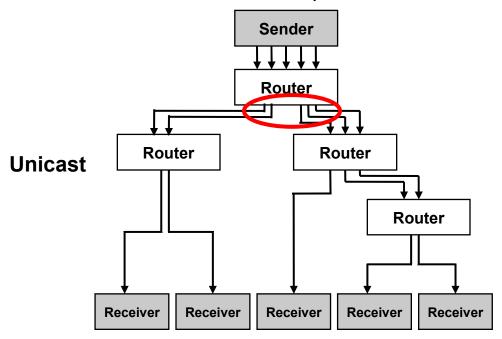
#### P2PTV: VIDEO STREAMING

- File sharing: consente si condividere anche file multimediali contenenti audio/video, tuttavia il video viene prima scaricato (almeno in parte) sul peer e visualizzato successivamente
- Video Streaming:
  - trasferimento e visualizzazione in real time dello stream video
  - cooperazione tra i peer per la distribuzione del contenuto (content distribution networks)
    - Bittorrent
  - bufferizzazione di un certo numero di video frames
- Applicazioni
  - Joost
  - Zatoo



## CONTENT DISTRIBUTION NETWORKS

 Un server pubblica un nuovo contenuto (esempio: nuova versione di un gioco, nuova release di un sistema operativo,...)



 Modello client server: collo di bottiglia rappresentato dall'unico server centralizzato



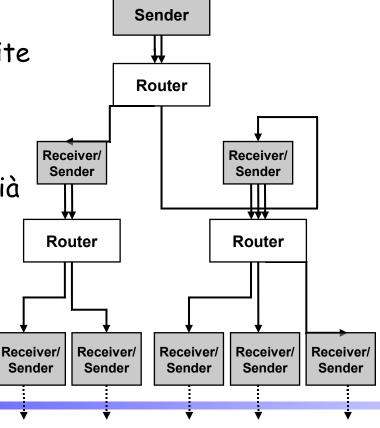
## Approcci P2P

 Ottenere un miglior bilanciamento nell'uso della banda di comunicazione, utilizzando canali di trasmissione meno utilizzati

#### Peer-to-Peer Content Distribution

 Le richieste iniziali di files sono servite da un server centralizzato

 Ulteriori richieste vengono inoltrate automaticamente ai peer che hanno già ricevuto e replicato, in precedenza, i files

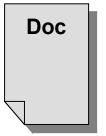




- Uso combinato del modello P2P e di quello client-server consente di ottimizzare gli accessi ad un server
- Approccio segmentato (esempio:BitTorrent)











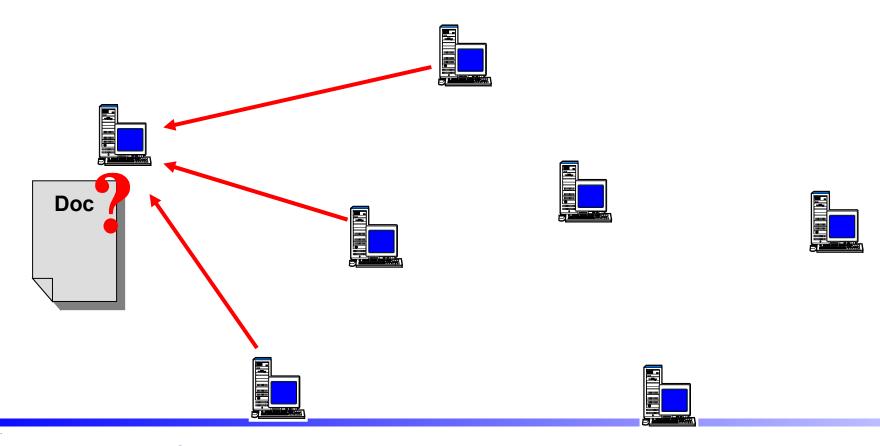








- Uso combinato del modello P2P e di quello client-server consente di ottimizzare gli accessi ad un server
- Approccio segmentato (esempio:BitTorrent)

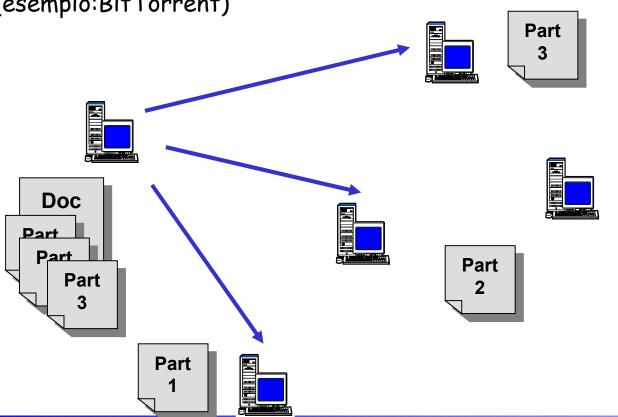


**Introduzione** 



 Uso combinato del modello P2P e di quello client-server consente di ottimizzare gli accessi ad un server

Approccio segmentato (esempio:BitTorrent)

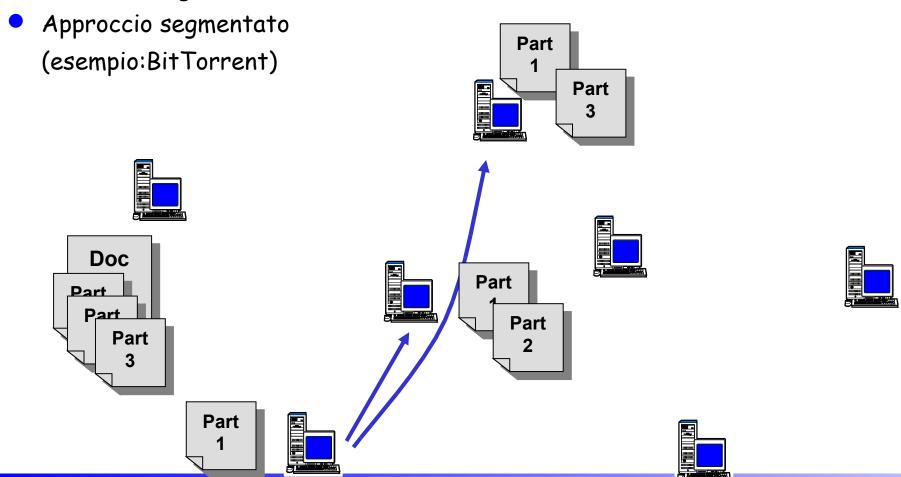






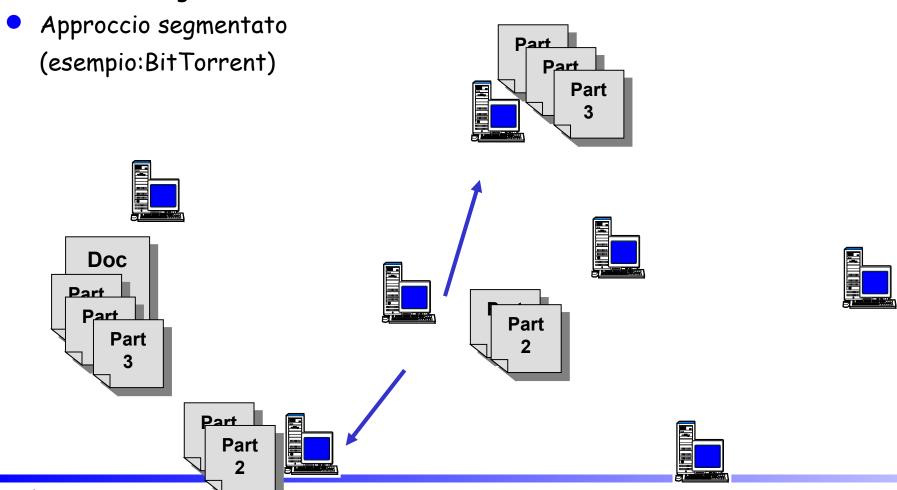


 Uso combinato del modello P2P e di quello client-server consente di ottimizzare gli accessi ad un server





 Uso combinato del modello P2P e di quello client-server consente di ottimizzare gli accessi ad un server

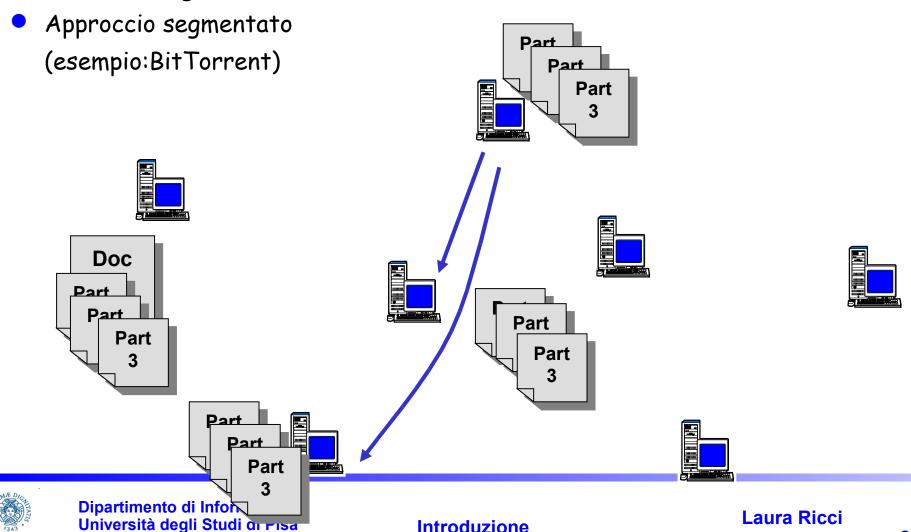




Dipartimento di Informatica

Università degli Studi di Pisa

Uso combinato del modello P2P e di quello client-server consente di ottimizzare gli accessi ad un server



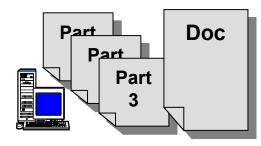
**Introduzione** 

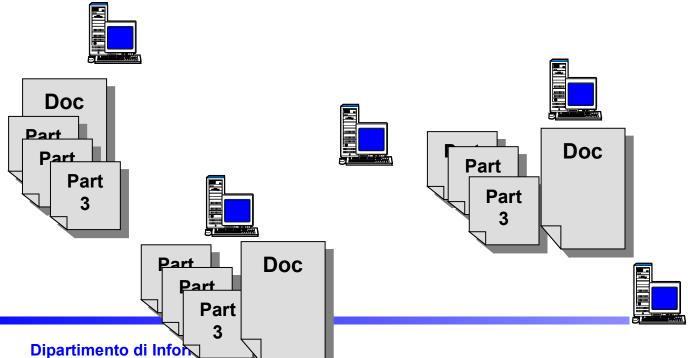
 Uso combinato del modello P2P e di quello client-server consente di ottimizzare gli accessi ad un server

Approccio segmentato

Università degli Studi di Pisa

(esempio:BitTorrent)





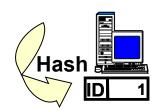


**Laura Ricci** 

- Una P2P Storage Network è un cluster di computers connessi in rete che utilizza tutta la memoria resa disponibile dai peer per definire un servizio distribuito di memorizzazione
  - Esempi: PAST, Freenet, OceanStore, XtreemOS Directory Service
- Organizzazione:
  - Associazione peer-identificatori mediante funzioni hash
  - Ogni peer offre una parte del proprio spazio di memorizzazione, o paga una quota di denaro
  - A seconda del suo contributo, ad ogni peer viene assegnato un volume massimo di dati che possono essere aggiunti alla storage network
  - Assegnazione documenti-identificatori mediante una funzione hash calcolata sul nome o su parte del contenuto del file
  - La memorizzazione e la ricerca di files nella rete è guidata dagli identificatori associati dalla funzione hash a peers e files.



#### Costruzione







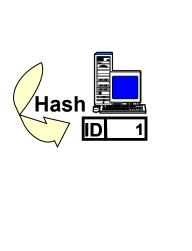


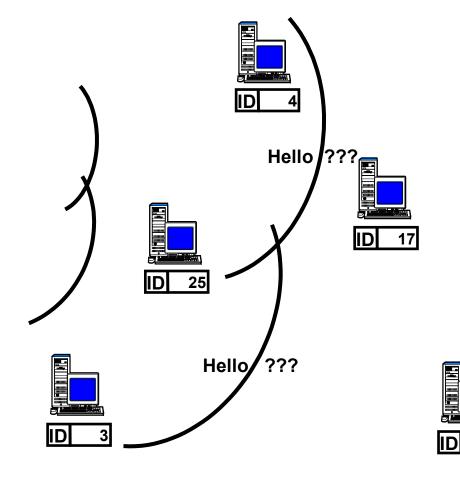






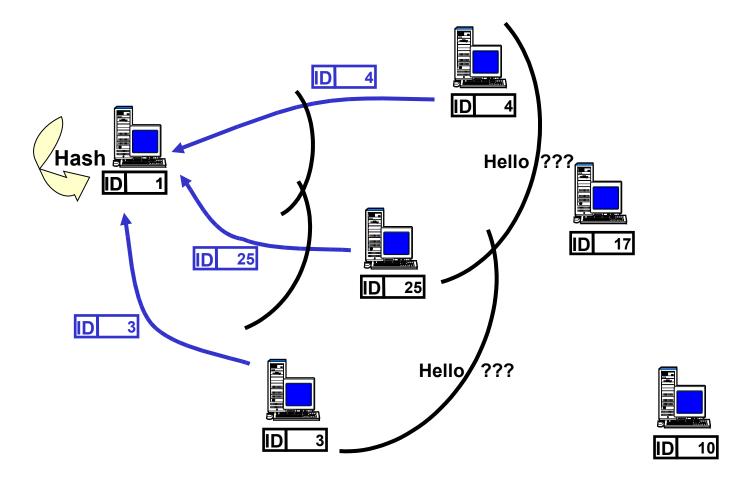
## Costruzione





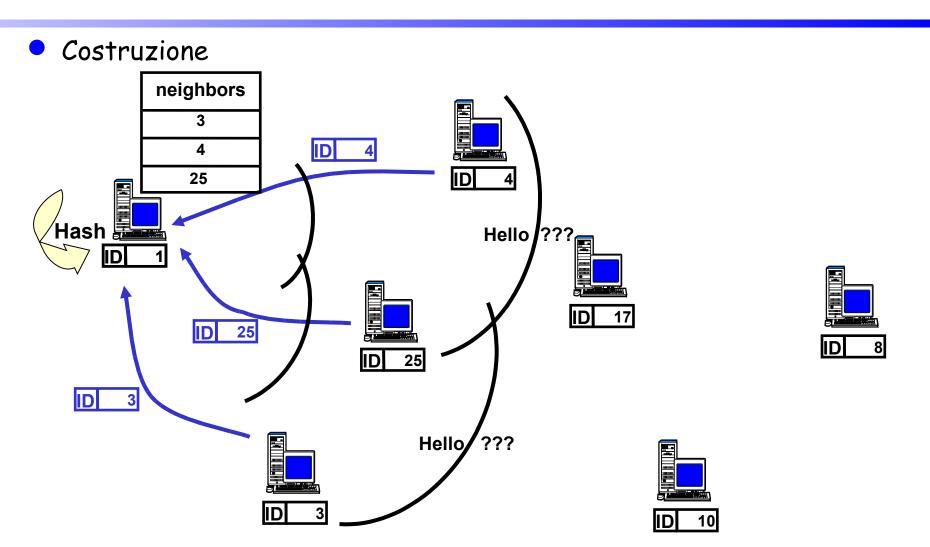


#### Costruzione

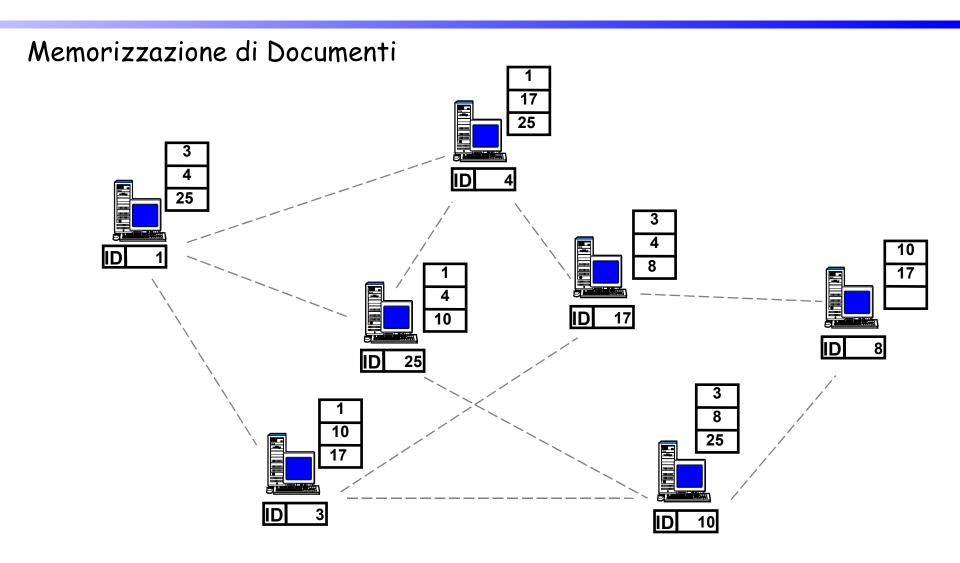




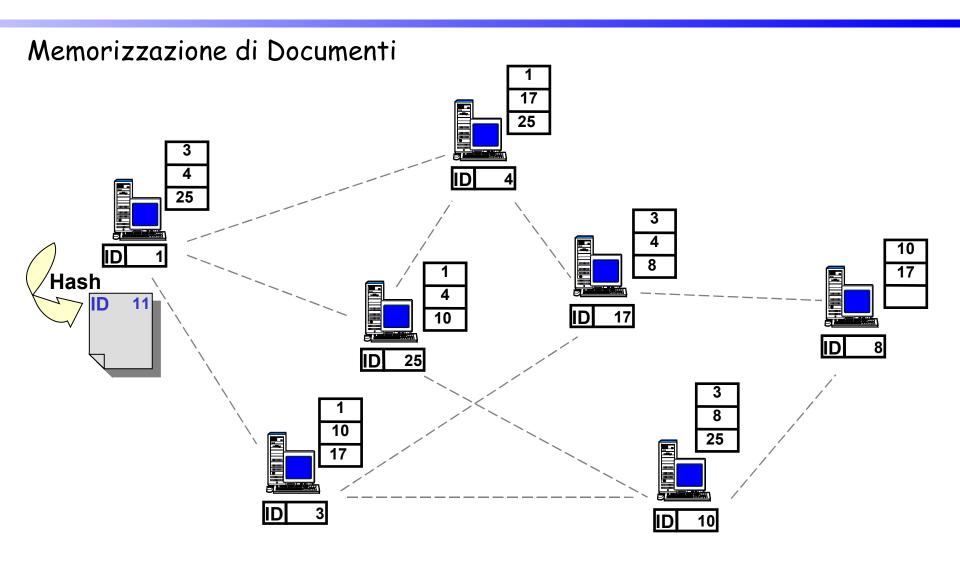




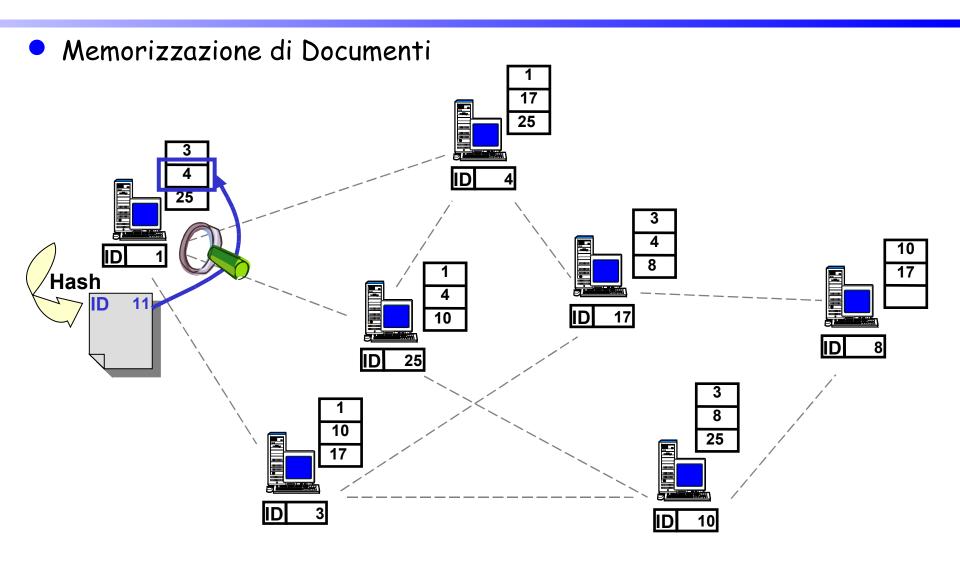




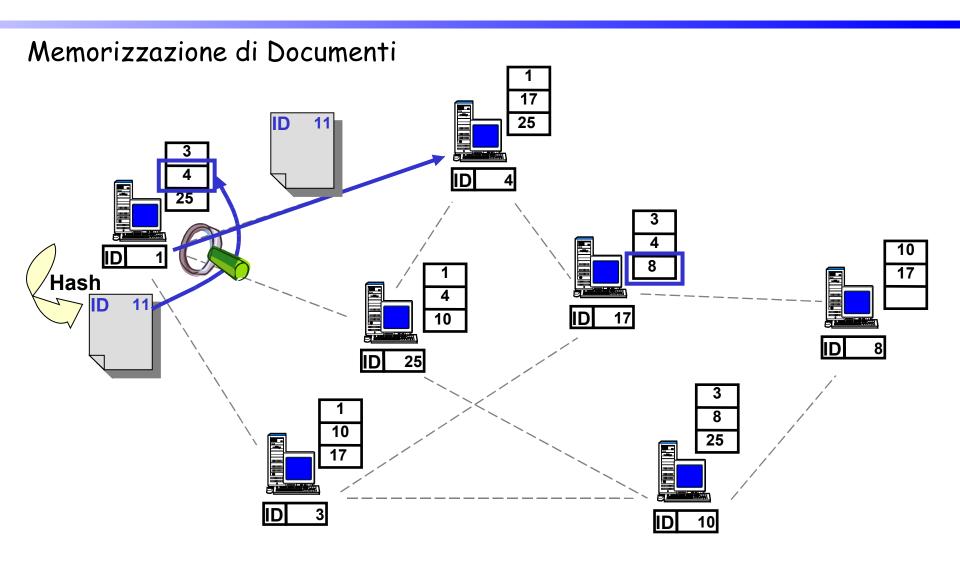




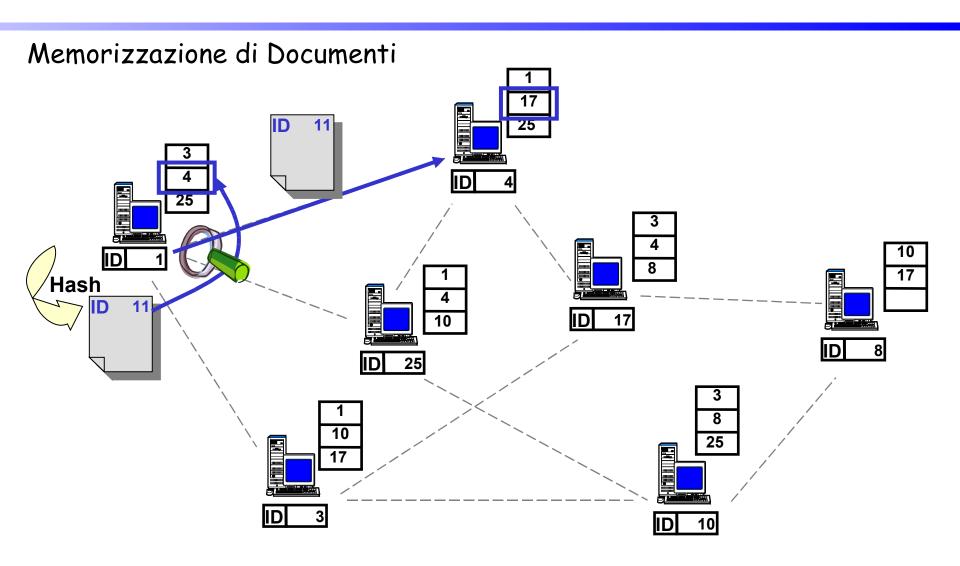




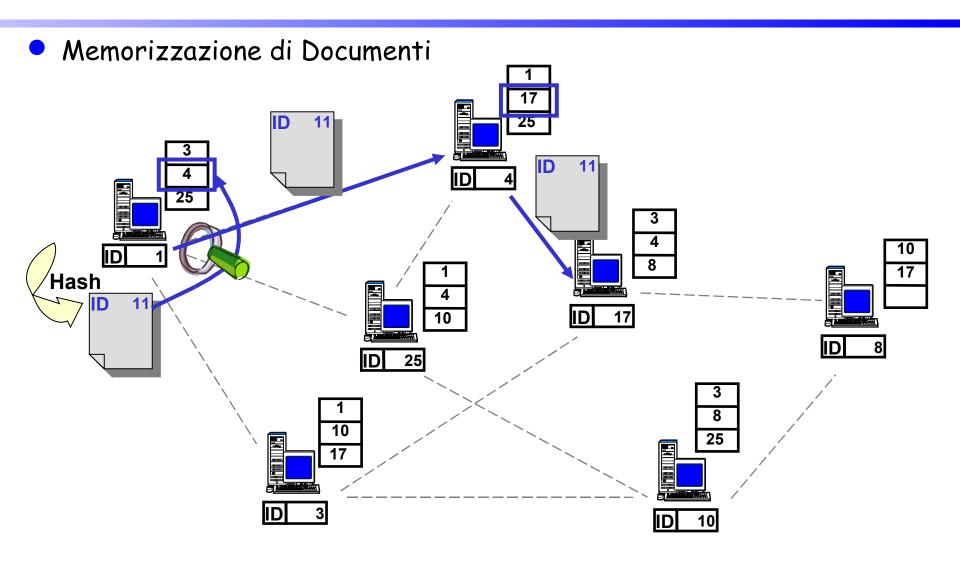




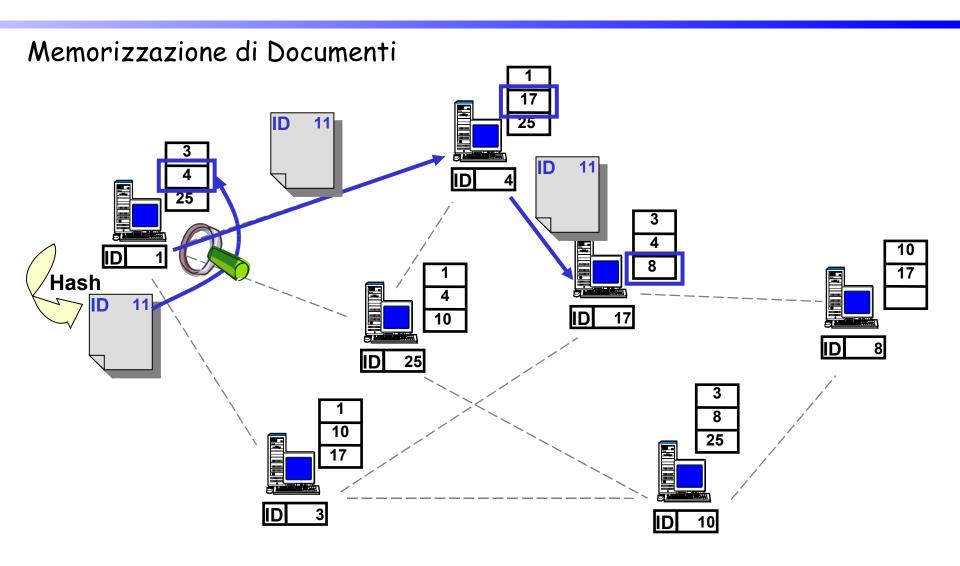




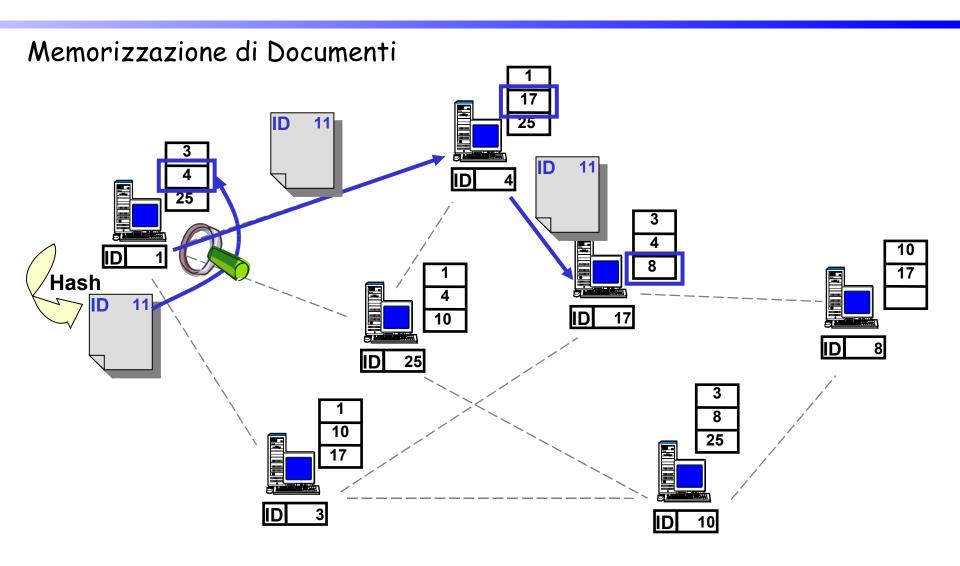




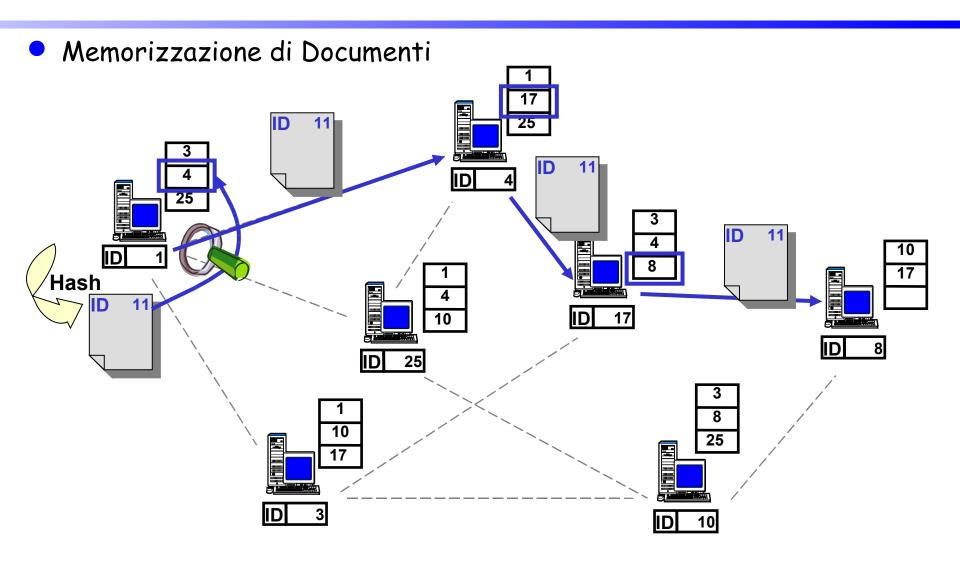




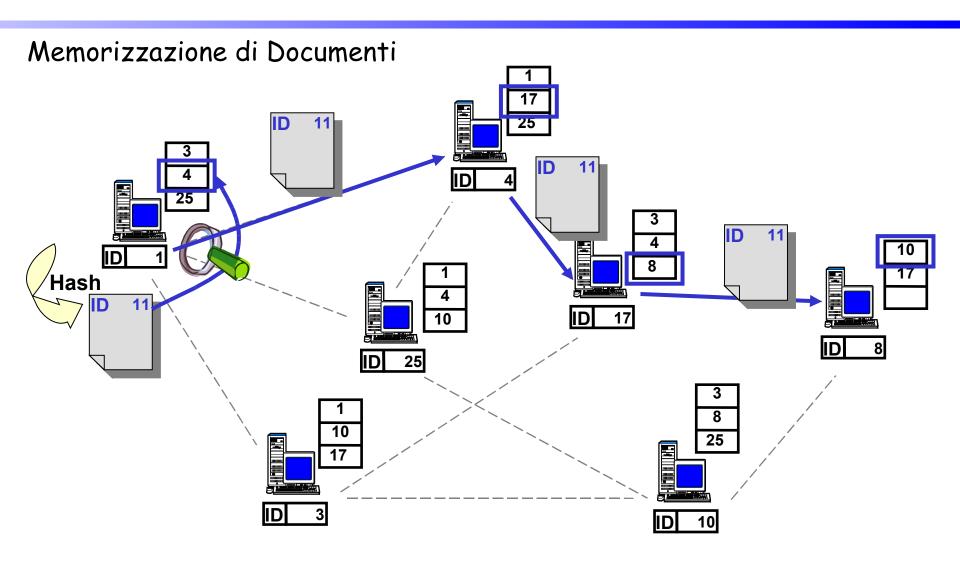




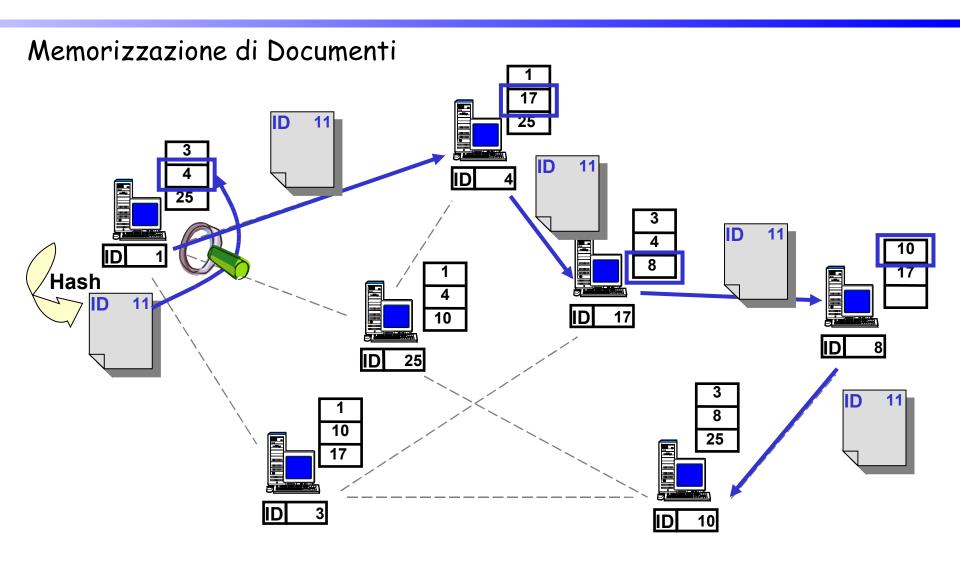




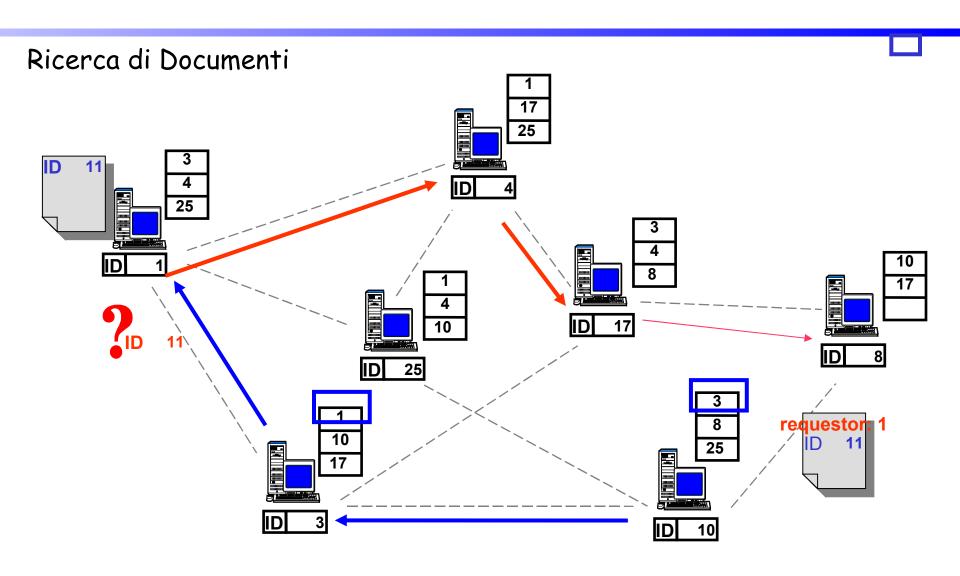














#### **OVERLAY P2P**

- Un protocollo P2P definisce l'insieme dei messaggi che i peer si scambiano, il loro formato, la loro semantica
- Caratteristica comune a tutti i protocolli P2P:
  - identificazione dei peer mediante un identificatore unico, generalmente creato mediante l'applicazione di una funzione hash
  - i protocolli sono definiti a livello applicazione dello stack TCP/IP e definiscono una strategia di routing
- Overlay Network: rete logica definita tra i peer a livello applicazione
  - La rete definisce dei links logici tra i peer, che non corrispondono a collegamenti fisici
  - ad ogni link logico possono corrispondere
    - più hops fisici
    - attraversamento di un insieme di routers



## OVERLEY P2P: CLASSIFICAZIONE

#### Sistemi P2P non strutturati (Gnutella, Kazaa,...)

- Un nuovo peer si connette in modo casuale ad un certo numero di peer già attivi all'interno del sistema
- La rete costruita dinamicamente dai peers (overlay network) risulta non strutturata
- Algoritmi di Ricerca di informazioni sulla rete: basati su
  - directory centralizzato (Napster).
  - flooding (Gnutella),

Costo della ricerca= lineare in N, dove N è il numero di nodi della rete

Problema: scalabilità.



## OVERLAY P2P: CLASSIFICAZIONE

#### Sistemi P2P strutturati

- La scelta dei vicini a cui un nuovo peer si deve collegare è effettuata in base ad un determinato criterio
- La rete dei peer (overlay network) risulta strutturata
- obiettivo:garantire la scalabilità
  - La struttura della rete garantisce che la ricerca di una informazione abbia complessità limitata (ad esempio O(log N)).
  - Complessità limitata anche nel caso di aggiunta di un nuovo peer, o eliminazione di un peer
- Utilizzo di distributed hash tables

