
Web replication e Content Delivery Networks

Corso di ***Applicazioni Telematiche***

A.A. 2008-09 – Lezione n.3

Prof. Roberto Canonico

Università degli Studi di Napoli Federico II

Facoltà di Ingegneria

Approcci per aumentare la scalabilità dei siti web

- Caching
 - Web replication
 - Content delivery networks: realizzano una forma dinamica di web replication
-

Web replication

- Soluzione server-side per permettere la scalabilità Web:
 - il sito Web è replicato su più server, eventualmente dislocati su aree geografiche differenti. Questo consente generalmente di migliorare la QoS percepita dagli utenti.
 - Tre approcci possibili alla replicazione Web:
 - Mirroring esplicito
 - Mirroring statico trasparente
 - Replicazione dinamica trasparente
 - Ciascuna tipologia fornisce un diverso compromesso tra flessibilità, complessità e costo.
-

Problematiche legate alla replicazione

1. Come distribuire le richieste sui server?
 - ❑ Definizione delle architetture
 - ❑ Definizione delle tecniche di routing
 2. Quale server scegliere per soddisfare una richiesta?
 - ❑ Definizione di politiche di selezione dei server
 3. Su quanti server replicare lo stesso oggetto?
 - ❑ Definizione di politiche di replicazione
 4. Come garantire la consistenza delle repliche?
-

Replicazione mediante mirroring esplicito

- La più semplice forma di replication, largamente usata, è il mirroring esplicito
 - Il sito è interamente duplicato su altri server
 - Esempi: Tucows, GNU, Sourceforge, Linux, ...
 - Tuttavia questa soluzione presenta diversi svantaggi:
 - Gli utenti devono scegliere esplicitamente quale mirror usare
 - Il bilanciamento del carico non è controllabile
 - La decisione di rimuovere un mirror è difficile perchè può creare notevoli disagi agli utenti (es. link salvati dagli utenti possono diventare invalidi)
-

Esempio di mirroring esplicito

You are using mirror:
switch.dl.sourceforge.net

SWITCH

Location	Continent	Download
Prague, Czech Republic	Europe	 2683 kb
Brookfield, WI	North America	 2683 kb
Zurich, Switzerland	Europe	 2683 kb
Dublin, Ireland	Europe	 2683 kb
Keihanna, Japan	Asia	 2683 kb
Brussels, Belgium	Europe	 2683 kb
Minneapolis, MN	North America	 2683 kb
Phoenix, AZ	North America	 2683 kb

Cosa è una CDN

- Una Content Delivery Network è un'infrastruttura creata per distribuire efficacemente agli utenti di Internet i contenuti dei siti web più popolari
 - Una CDN si basa sulla distribuzione di repliche dei contenuti dal server principale del “Content Provider” ad una molteplicità di server disposti sulla rete da un “Content Delivery Operator”
 - Si presenta come un servizio a pagamento del quale usufruiscono i gestori dei siti web commerciali più popolari
 - Esempi: Akamai, Speedera, Inktomi
-

Obiettivi di una CDN

1. Alleviare il server web “master” dal carico degli utenti, in particolare proteggerlo da picchi di traffico improvvisi (flash crowds)
 2. Offrire i contenuti ai singoli utenti tramite server collocati in prossimità degli utenti (alla periferia della rete)
 3. Rendere il sistema di distribuzione dei contenuti più affidabile e robusto ai guasti
-

Interazioni che determinano il tempo di accesso ad una pagina web



- User enters `www.xyz.com`
- Browser requests IP address for `www.xyz.com`
- DNS returns IP address
- Browser requests HTML
- Content provider's web server returns HTML
- Browser obtains IP addresses for hostnames listed in URLs of objects embedded on page
- Browser requests embedded objects
- Content provider's web server returns embedded objects

Il modello *end-to-end* del Web

- La rete è un'entità passiva
- I servizi vengono forniti solo quando richiesti:
"just in time delivery model"

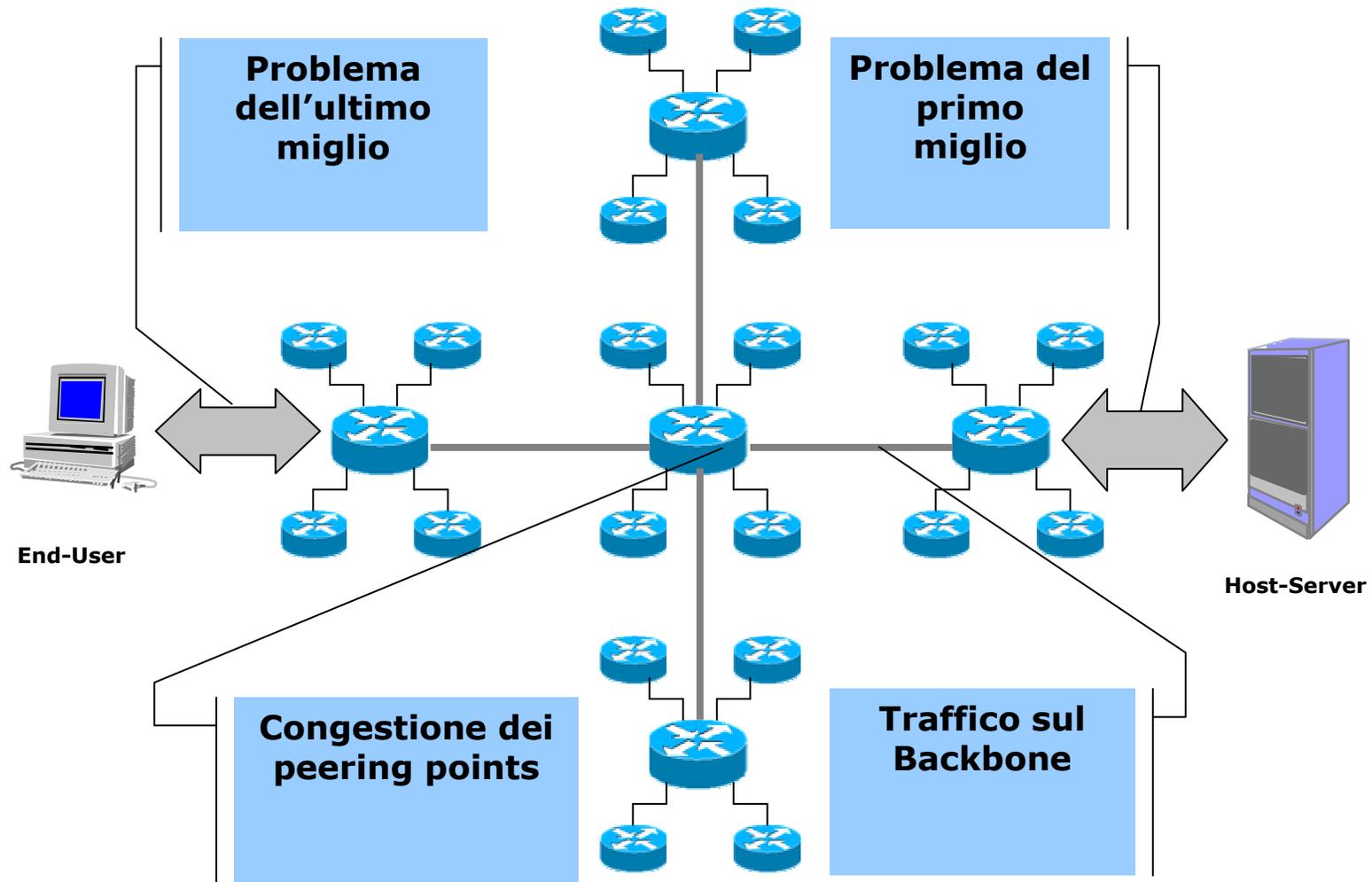
Vantaggi

- Le modifiche apportate ai dati dal Content Server si riflettono immediatamente nei documenti consegnati
- Il Content Server può *tracciare* le richieste dei client e calcolare la popolarità dei servizi forniti
- Possibilità di differenziare l'accesso ai servizi

Svantaggi

- Elevato tasso di duplicazione del traffico Internet
 - Carico elevato per i server che distribuiscono contenuti molto popolari
-

Limiti dell'approccio centralizzato



Limiti dell'approccio centralizzato (2)



L'approccio CDN



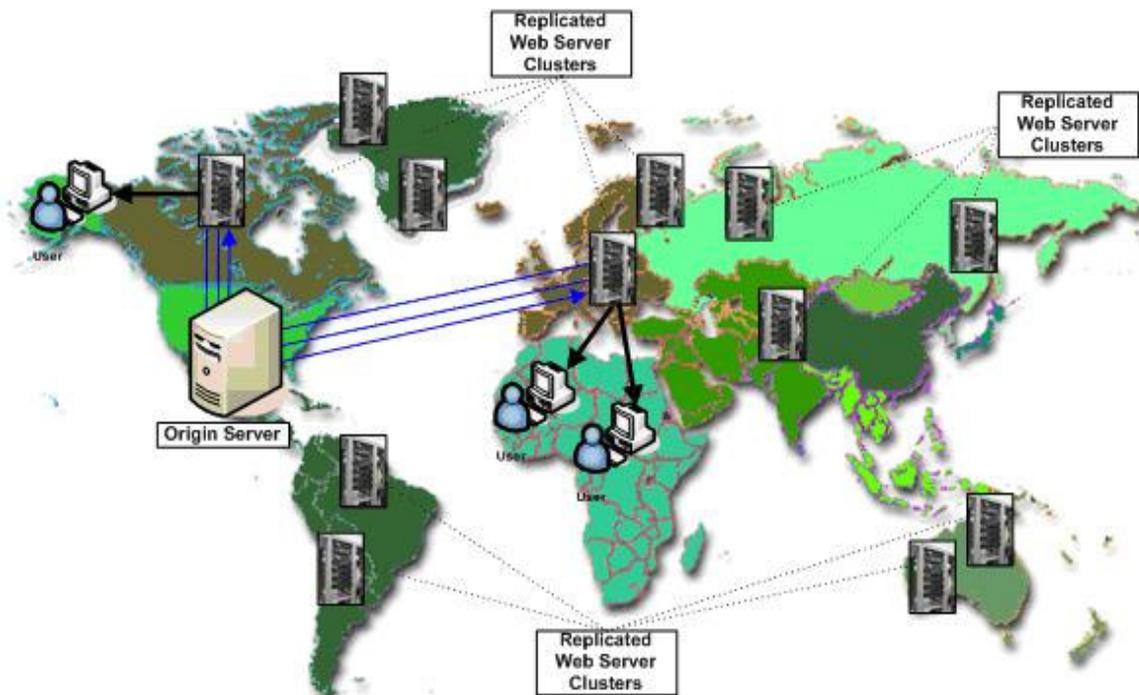
Content Delivery Networks

- Tramite una infrastruttura, spesso privata, distribuiscono, in maniera capillare i contenuti di uno specifico Content Server
- Utilizzano forme proprietarie di caching basate su una complessa gestione del DNS, caratterizzata, tra l'altro, dalla conoscenza dell'indirizzo IP del Client
- Gestione centralizzata dei contenuti

Servizi offerti

- Le Content Delivery Network offrono ai Content Server la possibilità di raggiungere, con una certa QoS, una vasta utenza
 - Le CDN, d'altra parte, propongono a ISP di medie e grandi dimensioni, di collaborare, spesso gratuitamente, alla loro struttura
-

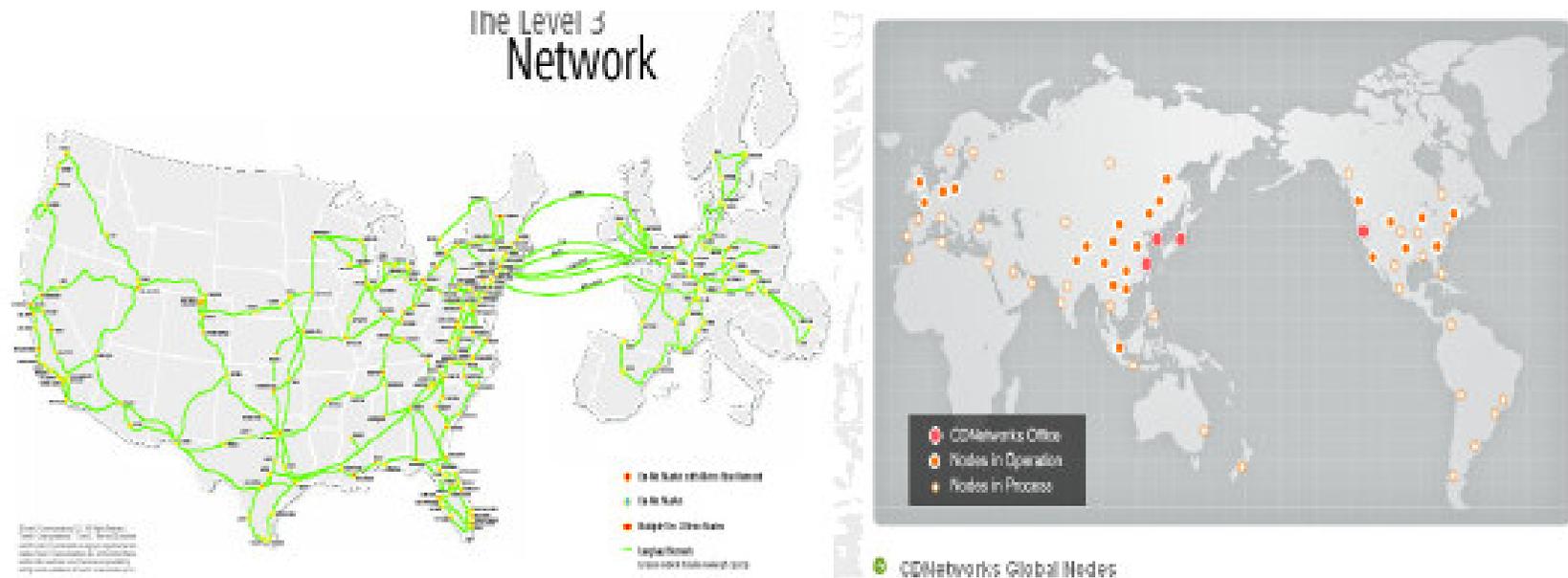
Content Delivery Networks



- ottimizzano l'uso delle risorse di Internet avvicinando il contenuto all'utente
- benefici
 - per gli utenti
 - per gli ISP
 - per i Content Provider
- obiettivi
 - bassa latenza
 - bassi costi
 - raggiungibilità
 - protezione da flash events

CDN ed ISP

- Single ISP CDN
 - Level3
- Multi-ISP CDN
 - CDNNetwor, EdgeCast



CDN e Content Provider

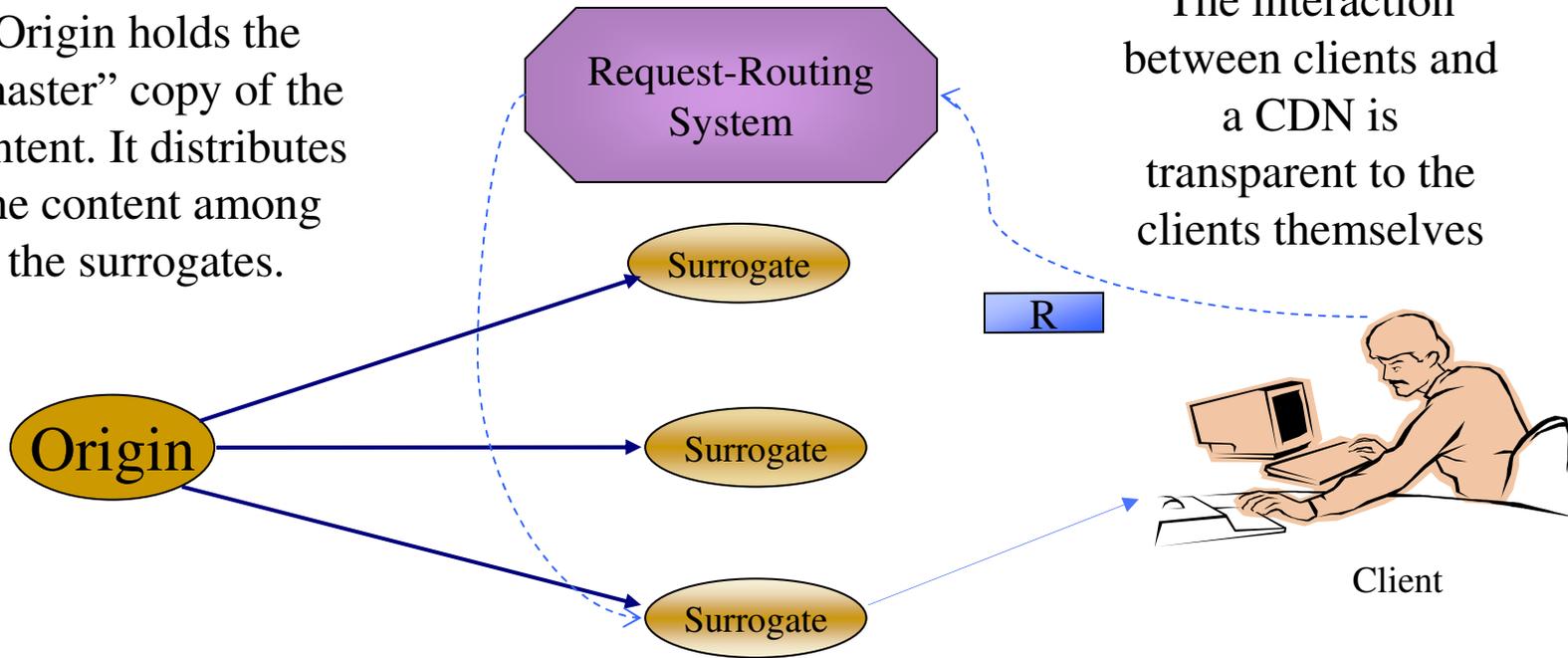
- Hosting CDN
 - L'Origin Server è nella gestione del CDN operator
 - Relaying CDN
 - First Hit at Origin Server
 - First Hit at CDN
-

CDN: redirection

- DNS outsourcing
 - URL rewriting
 - HTTP redirection
-

CDN infrastructure

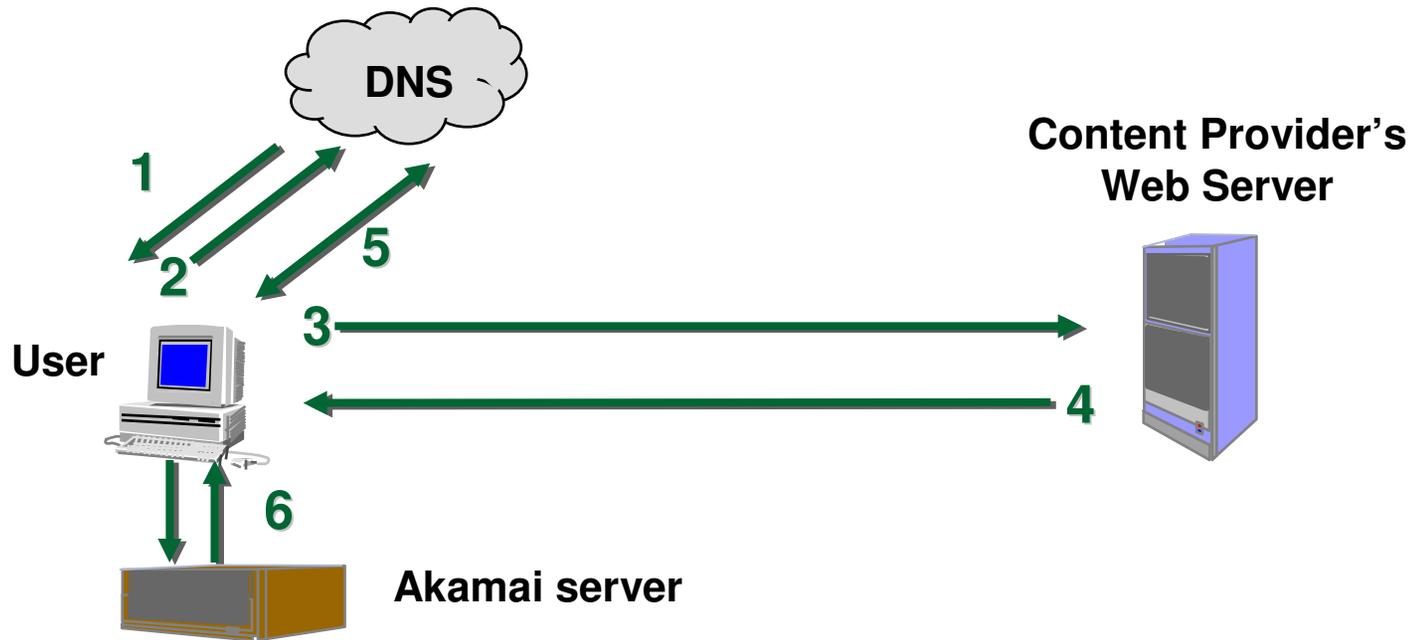
Origin holds the “master” copy of the content. It distributes the content among the surrogates.



The interaction between clients and a CDN is transparent to the clients themselves

Surrogates are geographically distributed servers that maintain copies of the content

Come funziona Akamai



- User enters `www.xyz.com` and browser requests IP address for `www.xyz.com`
- DNS returns IP address
- Browser requests HTML
- Content provider's web server returns page with *Akamaized* URLs
- Browser obtains IP address of optimal Akamai server for embedded objects
- Browser obtains objects from optimal Akamai server

Cosa significa “Akamaizzare” i riferimenti

```
<html>
<head>
<title>Welcome to xyz.com!</title>
</head>
<body>
  
  
  <h1>Welcome to our Web site!</h1>
  <a href="page2.html">Click here to enter</a>
</body>
</html>
```

ak



Efficacia: un esempio

Total page 87,550 bytes
Total Akamai Served 68,756 bytes

Logos 3,395 bytes

Navigation Bar 9,674 bytes

Banner Ads 16,174 bytes

Gif links 22,395 bytes

Fresh Content 17,118 bytes

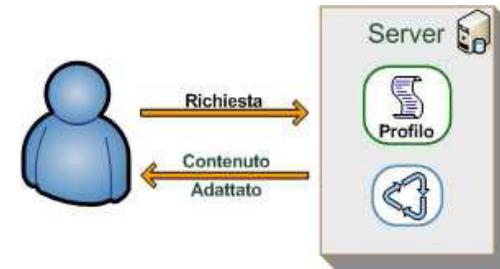


78% Page Served by Akamai

Content Adaptation

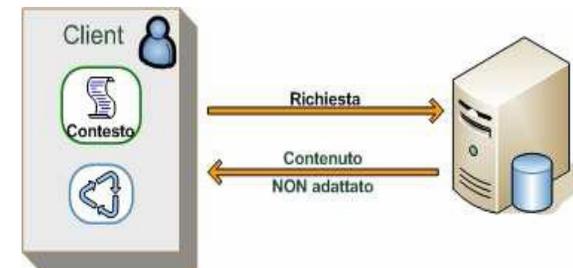
■ Server Based

- ❑ Adattamento offline, diverse versioni del contenuto pre-adattate
- ❑ Poca flessibilità rispetto alla varietà di dispositivi richiedenti



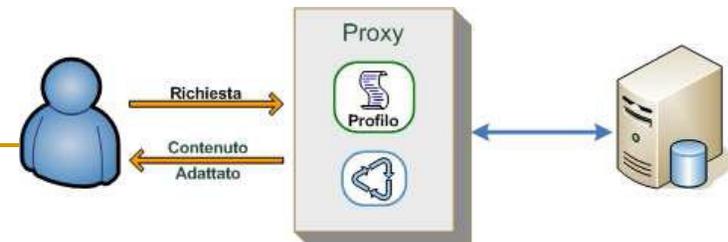
■ Client Based

- Ricezione del contenuto non adattato
- Inutilizzabile per client con ridotte capacità elaborative



■ Intermediary Based

- Adattamento avviene sul proxy
- Possibilità di caching dei contenuti sul Proxy e diminuzione del carico sul Server



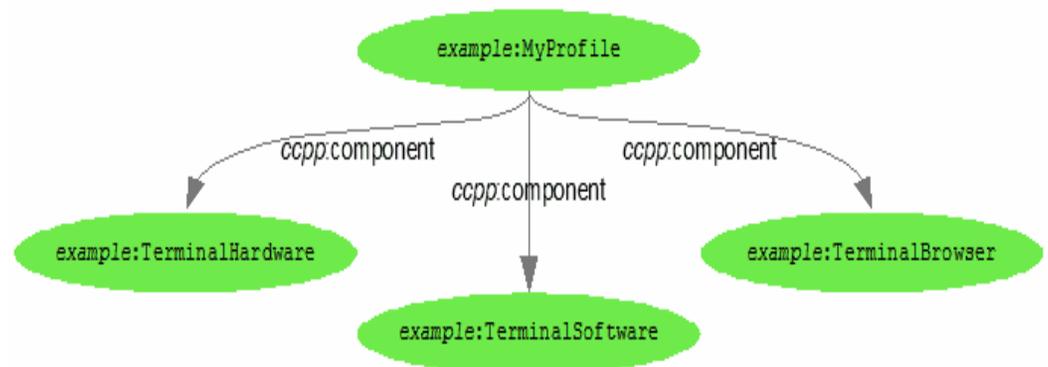
Profilazione terminali utente: CC/PP

■ Profilo CC/PP

- Insieme delle caratteristiche HW e SW di un dispositivo mobile presentate in formato XML o RDF

□ Organizzazione gerarchica

- Terminal Hardware
- Terminal Software
- Terminal Browser



Trasmissione profili CC/PP

- Logica di trasmissione dei profili CC/PP non standardizzata
 - Utilizzo di particolari header nella richiesta HTTP
 - Esistono tre tecniche di trasporto dei profili
 - WHTTP
 - header x-wap-profile per il link al profilo
 - HTTPex
 - header OPT e MAN per il namespace
 - "ns"-profile per trasportare il link al profilo
 - WSP
 - Usa gli stessi header di HTTPex, ma eventuali differenze vengono inoltrate in formato compresso proprietario WBXML
-