

Content De-duplication for CDNi

<http://tools.ietf.org/id/draft-jin-cdni-content-deduplication-optimization-01.txt>

WeiYi Jin (jin.weiya@zte.com.cn)

Mian Li (li.mian@zte.com.cn)

Bhumip Khasnabish (bhumip.khasnabish@zteusa.com)

Tuesday, July 31, 2012

CDNi WG, IETF84, Vancouver

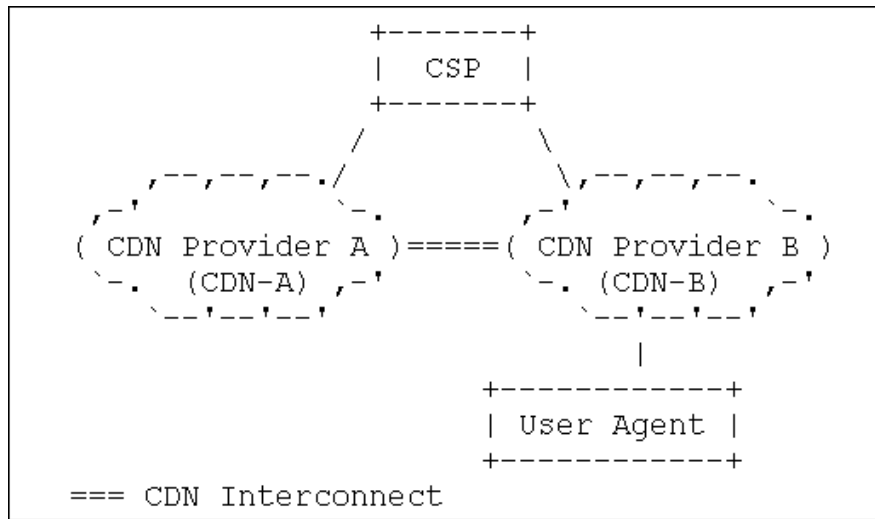
Outline

- Content de-duplication problem statement
- What needs-to-be and/or can-be done to overcome the problems of Content de-duplication
- Proposed next steps for IETF CDNI WG
- Q&A and open Discussion

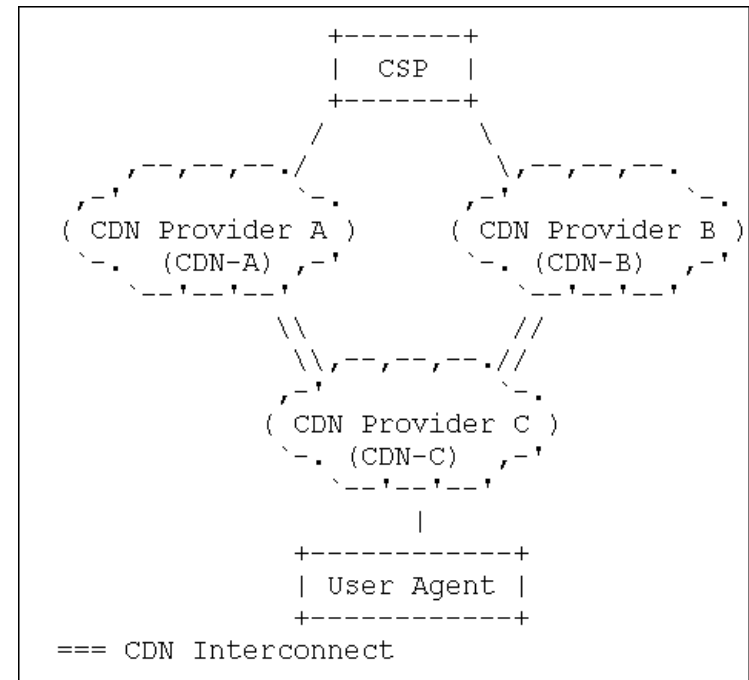
What's the Problem?

- Scenarios (for a single given CSP):
 - Interconnected CDN with the same CSP;
 - Interconnected CDN with the same CSP and with one dCDN;
 - Cascaded CDN
- Problem:
 - For a single given CSP, in some CDNi deployment, the dCDN may receive the content via 'push' from one of the uCDN via pre-position procedure, and then may also request for downloading the same content from another uCDN via another pre-position procedure or upon user's content request.
- Reason for the Problem:
 - Currently the content is identified by a specific URL. Different CDNs may identify the same content with different URLs;
 - The URL used to identify the content may be changed during the redirection processes between CDNs;
 - No unique identification is used to identify a content item so far

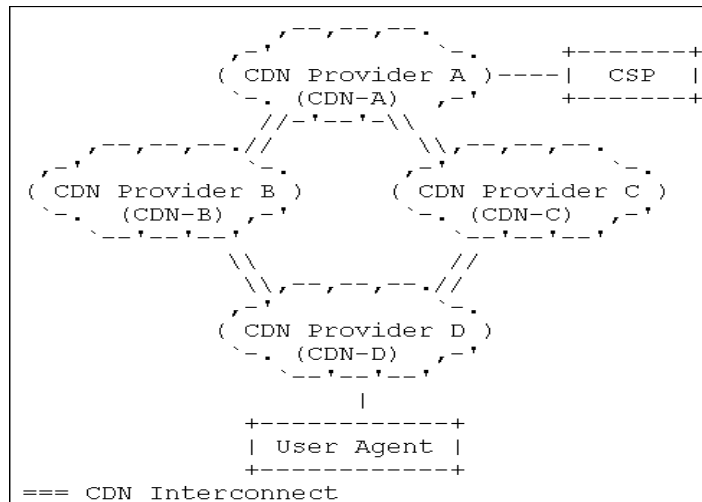
What's the Problem?



Scenario 1



Scenario 2



Scenario 3

Possible Approaches

- Specify a CDNI content naming mechanism to uniquely identify a content item for a single given CSP;
- define an optional Metadata object containing the Content Identifier;
- The content Identifier, as an optional content metadata, is distributed with/without the actual content item from uCDN to dCDN;
- When an action is needed (e.g. pre-positioning or content acquisition) for a resource URL, the dCDN would first acquire the metadata and then get the Content Identifier It can then make the determination on whether the content is already stored
- A simple approach with (as is described in section 4.3 of draft-jin-cdni-content-deduplication-optimization):
 - Extension of metadata object;
 - Extension of Content Identifier and Resource identifier in Request Routing and Metadata interfaces;
 - Some logical analysis in the internal dCDN

What Needs to be Done?

- Content duplication issue for a single given CSP should be made clear within CDNi scope;
- Content items should be uniquely identified within CDNi scope;
- CDNi framework/interface should provide mechanisms on content de-duplication optimization for a single given CSP

Proposed Next Step of CDNi WG

- Collect comments and feedback;
- Determine the necessity of content de-duplication in CDNi scope;
- Some technical discussion, maybe in mailing list;
- Further work or documents may include:
 - Include content de-duplication requirements in normative work;
 - Provide a CDNi content naming mechanism to uniquely identify a content item;
 - Enhance Metadata and Control models and interfaces to implement content de-duplication optimization;
 - Explore other possible solutions on content duplication issue

Q&A and Discussion