draft-narayanan-icnrg-bgp-uri-00

Ashok Narayanan
Stefano Previdi
Brian Field

ICNRG
Aug 1 2012

Agenda

- Problem statement
- Solution overview
- NDN use case overview
- CDN use case overview

Problem statement

- NDN data objects are identified by global content names
  - NDN names are similar to URIs
  - Arbitrary-octet components with defined separator
- NDN routers maintain an NDN-FIB to forward interest packets
  - NDN-FIB contains prefix names
  - Routers perform componentized longest prefix match

Problem statement

- CDNs would like to use BGP to carry information about content availability
  - "Integrating Routing with CDNs", Field et al, *IEEE Infocom NOMEN 2012*
  - Origin server advertises the URI prefixes of their content into routing
  - Cache servers can route client requests towards origin servers serving that content
  - All benefits of dynamic routing – no static cache tree configuration, rerouting around failure, etc.

Solution overview

- These two use cases (NDN and CDN) seem similar enough that the same solution may work for them
- We'll define a new BGP NLRI called "Content-URI", which carries a standard RFC3986 URI, representing content
- Content-URI NLRIs advertised by a BGP speaker have the same semantic as other NLRIs
  - "Any content matching this URI is reachable via this BGP speaker"
  - The BGP speaker must either store the content in question or know how to retrieve it

BGP details

- New BGP Capability identifying support of Content-URI [RFC3392]
  - Only BGP speakers identifying this capability will exchange Content-URI NLRIs
- Content-URI NLRIs will be carried in MP_REACH and MP_UNREACH attributes [RFC4760]
- Complete routes: Advertising a Content-URI NLRI indicates that *all known content* under the specified URI prefix can be retrieved via the advertising router

BGP details

- Origination: A BGP speaker can originate a Content-URI NLRI *iff* it :
  - Understands the scheme or namespace of the URI
  - Implements a well-known standard for transferring this type of content to a requestor
  - Stores this content locally or knows a path (external from BGP) to retrieve it
- Re-originate: A BGP speaker can re-originate a Content-URI NLRI (replace its address as the NEXT_HOP) *iff*:
  - Conditions a) and b) from above apply, AND
  - Has received a BGP advertisement for this Content-URI from a reachable NEXT_HOP

Content Retrieval for URLs

- If the URI is a URL, it includes a scheme which describes the mechanism of content retrieval
- BGP speakers who insert themselves as the NEXT_HOP of a Content-URI advertisement MUST support the scheme indicated by that URL
  - A http:// URL indicates that the object may be retrieved from the NEXT_HOP address using HTTP.
  - Other URL schemes that wish to use BGP MUST define a content retrieval mechanism and this MUST be supported by the BGP speaker

Content Retrieval for URNs

- The URI may be of URN form which does not include a scheme, but is prefixed by a namespace
- Content-URIs with a specific URN namespace MUST only be advertised by BGP speakers IF:
  - They support a well-defined mechanism & protocol to retrieve content with this namespace
  - E.g. “ccn:” indicates the use of NDN protocol to connect to the NEXT_HOP and retrieve the content
- **Namespace defines both the prefix matching technique and the content retrieval technique**

Flexibility

- Not bound to prefix matching
  - Hierarchical namespace does not necessarily imply longest prefix match (???)
  - New namespace can define more flexible match (e.g. Bloom filter)
- Not bound to hierarchical namespace
  - Any aggregatable namespace can be used (e.g Bloom filter)
- Bound to route distribution
  - BGP route distribution scalability limits apply

Treat this as a strawman

- Is this the right way to go long-term?
- Clear short-term benefits
- May be a good guinea pig

Backup slides

BGP details

- Adj-Rib-Out selection for Content-URI NLRIs is very similar to IPv4 NLRIs
- Standard NLRI selection rules apply same as IPv4
- Local-Rib construction for Content-URI NLRIs is more flexible than for IP
  - FIB can contain overlapping Content-URI prefixes
  - Request forwarding SHOULD prefer longer prefix matches to shorter prefix matches, but this MUST be controllable by policy and scheme

BGP details

- Two NLRIs with the same Content-URI string are identical for selection purposes
  - URL scheme or URN namespace MUST be included in comparison
- URI prefix length MUST NOT include scheme or namespace for Local-Fib purposes
  - "http://" functions like a default URL route for HTTP scheme
  - "ndn:" functions like a default URN route for NDN namespace
- Routers may aggregate Content-URIs via policy by removing trailing segments
  - MUST obey the completeness principle if they do,
  - MUST only aggregate on token boundary ('/'), and MUST NOT decode percent-encoded '/' characters as token boundaries

BGP details

- Standard BGP attributes may apply to Content-URI NLRIs as follows:
  - NEXT_HOP: Specifies the address which provides reachability to the identified content prefix
  - ORIGIN, AS_PATH, MED, LOCAL_EXIT_PREF: Same as RFC4271
  - AGGREGATOR and ATOMIC_AGGREGATE: Not that useful any more

URI details

- BGP will only carry URIs expressed in octets
  - Character limitations as described in Section 2 of RFC3986 apply to URIs in BGP
  - Arbitrary octets MUST be encoded using Percent-Encoding
- URL scheme is significant
  - A well-known scheme (RFC4395) signifies a mechanism for retrieving the content from the speaker
  - Unknown or unsupported schemes MUST NOT be readvertised in Adj-Rib-Out with a local NEXT_HOP. They MAY be readvertised without changing NEXT_HOP, although this might cause reachability issues.

Multiple schemes?

- Content-URI with http:// scheme only allows retrieval via HTTP, not HTTPS or SPDY
- Proposal 1: Require multiple advertisements for the same content retrievable by multiple protocols
  - Triples the number of routes
  - Perhaps a compact scheme? E.g. “http+s://”?
- Proposal 2: define a node capability advertised by the BGP speaker
  - IPv4 NLRI with capability attached as community string
  - Nodes advertise the capability to do “HTTP-to-HTTPS translation
  - Client can combine a http:// Content-URI with “HTTP-to-HTTPS” community to request the content via HTTPS

NDN use case

- First NDN use case assumes the presence of TCP/IP in order for BGP to function
  - Future: come up with a BGP model that operates completely in the NDN model
- NDN will reserve the “ccn:” URN namespace [RFC3406]
- All NDN prefixes will be advertised as Content-URI NLRIs of the form “ccn:/….”
- A BGP speaker advertising NDN Content-URIs can receive NDN interest packets matching these URI prefixes at the address specified in the NEXT_HOP
  - EBGP speakers need NDN faces that are congruent with BGP peering relationships
  - IBGP speakers rely on IGP to identify the interim face that reaches the IBGP peer advertising a specific Content-URI

Future NDN use cases

- More complex schemes for NDN routes (such as Bloom filter, prefix+Bloom filter) have been proposed
- These can also be carried in BGP within the Content-URI object
- For each such scheme, a new URN sub-namespace should be reserved:
  - “ccn:bloom:<bloom-filter>”
  - “ccn:prefbloom:<prefix>_<bloom-filter>”
- Note that in this case, the URN namespace defines both the prefix matching technique and content retrieval scheme
- Same rules apply (determining scheme to retrieve content, forwarding routes, etc)

CDN use case

- Origin servers advertise URI prefixes for the content that they originate, using Content-URI NLRIs into IBGP
  - All cache nodes are IBGP speakers and can build a <URI,NEXT_HOP> top-level reachability database
  - Cache nodes then refer to IGP to compute a dynamic path for requests to pass through caches towards a nearby origin for a request.
  - Definition of this scheme deferred to separate document.

EBGP use of Content-URI objects

- Content-URI can be carried between ASBRs in EBGP
- Scalability of large-scale Content-URI deployment in the Internet has not been evaluated
- Hence, deployments SHOULD restrict Content-URIs within their administrative boundaries
  - Content-URI advertisements SHOULD carry the “no-export” community
  - If not, the network operator MUST deploy appropriate filtering such that Content-URI advertisements are not advertised to global Internet
- Note that without agreement between providers, the CONTENT_URI_CAP capability will not be advertised across AS boundaries, which acts as a natural filter for Content-URI NLRIs.

EBGP use of Content-URI

- A CDN provider with multiple AS attachment points may use MED & LOCAL_EXIT attributes with Content-URI NLRIs, for the same purpose as they are used in IP
- ASBRs may supply the address of a nearby CDN node as the NEXT_HOP of a Content-URI NLRI (if the ASBR does not itself support a HTTP proxy).
- To achieve reachability, the ASBR SHOULD also advertise IP routes in EBGP to reach the address of the cache node.
  - CDNs may also use BGP-LS [draft] to obtain interior network information for their own or neighboring ASes, in order to compute paths through the CDN tree towards an origin