Abstract

This document defines an IGP based mechanism for discovering the MTU of a BIER sub-domain. This document defines extensions to OSPF and IS-IS, but other protocols could potentially be extended. MTU discovery is usually done for a given path, while this document defines it for a sub-domain. This allows the computed MTU to be independent of the set of receivers. Also, the MTU is independent of rerouting events within the sub-domain.
This document defines an IGP based mechanism for discovering the MTU of a BIER sub-domain. The discovered MTU indicates the largest possible BIER packet that can be sent across any link in a BIER sub-domain. This is different from [I-D.ietf-bier-path-mtu-discovery] which performs Path MTU Discovery (PMTUD) for a set of receivers. PMTUD is based on probing, and when there are routing changes, e.g., a link going down, the actual MTU for a path may become less than was previously discovered, and there will be some delay until the next probe is performed. Also, the set of receivers for a flow may change at any time, which may cause the MTU to change. This document instead discovers a BIER sub-domain MTU, which is independent of paths and receivers within the sub-domain.

Discovering the sub-domain MTU is much simpler than discovering the multicast path MTU, and is more robust with regards to path changes as discussed above. However, the sub-domain MTU may be a lot smaller than the path MTU would have been for a given flow. The discovery mechanisms may be combined, allowing the discovery of the path MTU for certain flows as needed.

The BIER sub-domain MTU defined here provides the maximum size of a BIER packet that can be forwarded through the sub-domain regardless of path. A BIER router that performs BIER encapsulation will need to subtract the encapsulation overhead to find the largest size packet that can be encapsulated. This would give the IP MTU, and may be
used for IP PMTUD by for instance sending an ICMP Packet too big
message if an IP packet will be too large after BIER encapsulation.

2. Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
"SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and
"OPTIONAL" in this document are to be interpreted as described in BCP
14 [RFC2119] [RFC8174] when, and only when, they appear in all
capitals, as shown here.

3. MTU discovery procedure

An interface on a router is said to be a BIER interface if the router
has a BIER neighbor on the interface. That is, there is a directly
connected router on that interface that is announcing a BIER prefix.
Further, the BIER interface is said to belong to a given sub-domain
if the router itself announces a prefix tagged with the sub-domain,
and there is BIER neighbor on the interface also announcing a prefix
tagged with the sub-domain.

The BIER MTU of an interface is the largest BIER packet that can be
sent out of the interface. Further, the local sub-domain MTU of a
router is the minimum of all the BIER MTUs of the BIER interfaces in
the sub-domain. Note that the local sub-domain MTU of a router is
only defined if it has at least one BIER interface in the sub-domain.

A BIER router announces a BIER prefix in either IS-IS or OSPF as
specified in [RFC8401] and [I-D.ietf-bier-ospf-bier-extensions].
They both define a BIER Sub-TLV to be included with the prefix.
There is one BIER Sub-TLV included for each sub-domain. This
document defines how a router includes its local sub-domain MTU in
each of the BIER Sub-TLVs it advertizes.

A router can discover the MTU of a BIER sub-domain by identifying all
the prefixes that have a BIER Sub-TLV for the sub-domain. It then
computes the minimum of the advertised MTU values for that sub-
domain. This includes its own local sub-domain MTU. This allows all
the routers in the sub-domain to discover the same sub-domain wide
MTU.

Note that a router should announce a new local MTU for a sub-domain
immediately if the value becomes smaller than what it currently
announces. This would happen if the MTU of an interface is
configured to a smaller value, or the first BIER neighbor for a sub-
domain is detected on an interface, and the MTU of the interface is
less than all the other local BIER interfaces in the sub-domain.
However, if BIER neighbors go away, or if an interface goes down, so
that the local MTU becomes larger, a router SHOULD NOT immediately announce the larger value. A router MAY after some delay announce the new larger MTU. The intention is that dynamic events such as a quick link flap should not cause the announced MTU to be increased.

It is a concern that the sub-domain MTU will be based on the link with the smallest MTU. This means that if for instance a single link is accidentally configured with an extra small MTU, it will impact the sub-domain MTU and potentially all the flows through the sub-domain. As an example, an administrator might decide to use jumbo frames and has configured that on all the links. But accidentally forget to configure it on a new link before it is brought up. To provide some protection against this, an implementation SHOULD provide a configurable minimum BIER sub-domain MTU. When this is configured, the MTU discovery is still done according to the above procedure, but if the resulting MTU value is less than the configured minimum, the configured minimum MUST be used instead. If the discovery procedure later should provide an MTU larger than the minimum, then the discovered MTU MUST be used. An implementation SHOULD provide notification to the administrator when the discovered MTU is less than the minimum, as this is likely a configuration mistake that should be corrected.

4. IS-IS BIER Sub-Domain MTU Sub-sub-TLV

A router uses the BIER Sub-Domain MTU Sub-sub-TLV to announce the minimum BIER MTU of all its BIER enabled interfaces in a sub-domain. The BIER Sub-Domain MTU is the largest BIER packet that can be sent out of all the interfaces in a sub-domain. The Sub-sub-TLV MUST be ignored if it is included multiple times.

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Type: TBD

Length: 2

MTU: MTU in octets
5. OSPF BIER Sub-Domain MTU Sub-TLV

A router uses the BIER Sub-Domain MTU Sub-TLV to announce the minimum BIER MTU of all its BIER enabled interfaces in a sub-domain. The BIER Sub-Domain MTU is the largest BIER packet that can be sent out of all the interfaces in a sub-domain. The Sub-TLV MUST be ignored if it is included multiple times.

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             Type              |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             MTU               |           Reserved            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type: TBD2
Length: 4
MTU: MTU in octets

6. IANA considerations

An allocation from the "sub-sub-TLVs for BIER Info sub-TLV" registry as defined in [RFC8401] is requested for the IS-IS BIER Sub-Domain MTU Sub-sub-TLV. Please replace the string TBD in this document with the appropriate value.

An allocation from the "OSPF Extended Prefix sub-TLV" registry as defined in [RFC7684] is requested for the OSPF BIER Sub-Domain MTU Sub-TLV. Please replace the string TBD2 in this document with the appropriate value.

7. Acknowledgments

The authors would like to thank Greg Mirsky in particular for fruitful discussions and input. Valuable comments were also provided by Alia Atlas, Eric C Rosen, Toerless Eckert, Tony Przygienda and Xie Jingrong.

8. References
8.1. Normative References

[I-D.ietf-bier-ospf-bier-extensions]
Psenak, P., Kumar, N., Wijnands, I., Dolganow, A.,
Przygienda, T., Zhang, Z., and S. Aldrin, "OSPFv2
Extensions for BIER", draft-ietf-bier-ospf-bier-
extensions-18 (work in progress), June 2018.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
Requirement Levels", BCP 14, RFC 2119,
DOI 10.17487/RFC2119, March 1997,

[RFC7684] Psenak, P., Gredler, H., Shakir, R., Henderickx, W.,
Tantsura, J., and A. Lindem, "OSPFv2 Prefix/Link Attribute
Advertisement", RFC 7684, DOI 10.17487/RFC7684, November

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC
2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174,

Zhang, "Bit Index Explicit Replication (BIER) Support via
IS-IS", RFC 8401, DOI 10.17487/RFC8401, June 2018,

8.2. Informative References

[I-D.ietf-bier-path-mtu-discovery]
Mirsky, G., Przygienda, T., and A. Dolganow, "Path Maximum
Transmission Unit Discovery (PMTUD) for Bit Index Explicit
Replication (BIER) Layer", draft-ietf-bier-path-mtu-
discovery-04 (work in progress), June 2018.

Authors’ Addresses

Stig Venaas
Cisco Systems, Inc.
Tasman Drive
San Jose  CA 95134
USA

Email: stig@cisco.com
IJsbrand Wijnands  
Cisco Systems, Inc.  
De kleetlaan 6a  
Diegem 1831  
Belgium  

Email: ice@cisco.com  

Les Ginsberg  
Cisco Systems, Inc.  
Tasman Drive  
San Jose CA 95134  
USA  

Email: ginsberg@cisco.com  

Mahesh Sivakumar  
Juniper Networks  
1133 Innovation Way  
Sunnyvale CA 94089  
USA  

Email: sivakumar.mahesh@gmail.com