Compressed Bundle Header Encoding (CBHE)

Abstract

   This document describes a convention by which Delay-Tolerant
   Networking (DTN) Bundle Protocol (BP) "convergence-layer" adapters
   may represent endpoint identifiers in a compressed form within the
   primary blocks of bundles, provided those endpoint identifiers
   conform to the structure prescribed by this convention.

   Compressed Bundle Header Encoding (CBHE) compression is a
   convergence-layer adaptation. It is opaque to bundle processing.
   Therefore, it has no impact on the interoperability of different
   Bundle Protocol implementations, but instead affects only the
   interoperability of different convergence-layer adaptation
   implementations.

   This document is a product of the Delay-Tolerant Networking Research
   Group and has been reviewed by that group. No objections to its
   publication as an RFC were raised.

Status of This Memo

Copyright Notice

   Copyright (c) 2011 IETF Trust and the persons identified as the
   document authors. All rights reserved.

   This document is subject to BCP 78 and the IETF Trust's Legal
   Provisions Relating to IETF Documents
   (http://trustee.ietf.org/license-info) in effect on the date of
   publication of this document. Please review these documents
   carefully, as they describe your rights and restrictions with respect
   to this document.

Table of Contents

1. Introduction

   This document describes a convention by which Delay-Tolerant
   Networking (DTN) Bundle Protocol (BP) [RFC5050] "convergence-layer"
   adapters may represent endpoint identifiers (EIDs) in a compressed
   form within the primary blocks of bundles, provided those endpoint
   identifiers conform to the structure prescribed by this convention.

   Each DTN bundle's primary block contains the following four BP
   endpoint identifiers, of which any two, any three, or even all four
   may be lexically identical: the endpoint identifiers of the bundle's
   source, destination, report-to endpoint, and current custodian. Each
   EID is a Uniform Record Identifier (URI) as defined by [RFC3986].
   More specifically, each BP EID is a URI consisting of a "scheme name"
   followed by ":", followed by a sequence of characters --
   historically, termed the "scheme-specific part" (SSP) in DTN
   specifications -- conforming to URI syntax as defined by RFC 3986.

A degree of block compression is provided by the design of the
primary block: the scheme names and scheme-specific parts of the four
endpoints' IDs -- up to eight NULL-terminated strings -- are
concatenated at the end of the block in a variable-length character
array called a "dictionary", enabling each EID to be represented by a
pair of integers indicating the offsets (within the dictionary) of
the EID's scheme name and scheme-specific part. Duplicate strings
may be omitted from the dictionary, so the actual number of
concatenated NULL-terminated strings in the dictionary may be less
than eight, and two or more of the scheme name or scheme-specific
part offsets in the block may have the same value. Moreover, the
eight offsets in the primary block are encoded as Self-Delimiting
Numeric Values (SDNVs), which shrink to fit the encoded values; when
the total length of the dictionary is less than 127 bytes, all eight
offsets can be encoded into just eight bytes.

However, these strategems do not prevent the scheme names and
especially the scheme-specific parts themselves from being lengthy
strings of ASCII text. It is therefore still possible for the length
of a bundle's primary header to be a very large fraction of the total
length of the bundle when the bundle's payload is relatively small,
as is anticipated for a number of DTN applications such as space
flight operations (and as is in any case true of bundles carrying BP
administrative records).

The Compressed Bundle Header Encoding (CBHE) convention was developed
to improve DTN transmission efficiency for such applications by
further reducing the number of bytes used by convergence-layer
adapters to represent EIDs in the primary blocks of bundles.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
"SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this
document are to be interpreted as described in RFC 2119 [RFC2119].

2. Compression Convention

2.1. Constraints

The only valid scheme name for BP EIDs identified to date is "dtn".
Although no specification of valid SSP syntax for URIs composed
within the "dtn" scheme has yet been formally defined, the syntax on
which rough agreement has been reached in practice is unsuitable for
CBHE's compression procedures. For the purposes of CBHE, then, this
document defines an additional URI scheme named "ipn". As noted in
Section 4, IANA has registered this new URI scheme.

Compressed Bundle Header Encoding (CBHE) is possible only when all
endpoint IDs in the primary block of a given bundle are "CBHE
conformant". The following two forms of endpoint ID are CBHE
conformant: (a) the null endpoint ID "dtn:none" and (b) any endpoint
ID formed within the "ipn" scheme.

The SSP of every URI formed within the "ipn" scheme must comprise:

1. A sequence of ASCII numeric digits representing an integer in the
    range 1 to (2^64 - 1), termed the "node number" of the URI.

2. An ASCII period ('.') character.

3. A sequence of ASCII numeric digits representing an integer in the
    range 0 to (2^64 - 1), termed the "service number" of the URI.

The node number notionally identifies a BP node. However, since CBHE
is not used universally in Delay-Tolerant Networking, it must not be
assumed that all BP nodes are identified by node numbers.

Negative integers and integers larger than (2^64 - 1) cannot be used
as node numbers because they cannot be encoded into the SDNVs that
are used for representation of scheme name and SSP offsets in the
primary blocks of bundles and therefore could not be compressed as
described later in this specification. Node number zero is reserved
for representation of the null endpoint ID in the compressed form
described later; decompressing a compressed null EID must always
yield the standard null endpoint ID URI "dtn:none".

The service number notionally functions as a de-multiplexing token.
When the bundle payload is a protocol data unit of some protocol that
has its own de-multiplexing identifiers, the service number may
function in a manner similar to that of the protocol number in an IP
packet, characterizing the encapsulated protocol; alternatively, the
service number may function in a manner similar to that of the port
number in a UDP datagram. Service numbers enable inbound bundles'
application data units to be de-multiplexed to instances of
application functionality that are designed to process them, so that
effective communication relationships can be developed between bundle
producers and consumers.

A service number must not be negative or exceed (2^64 - 1) for the
same reason that a node number must not do so.

For example, "ipn:9.37" would be a CBHE-conformant endpoint ID.

Conversion of a CBHE-conformant EID to and from a tuple of two
integers is therefore straightforward: all characters in the EID
other than the node number and service number are constant (as
defined by the "ipn" scheme definition), and the node number and
service number are taken as the two integers of the tuple. This ease
of conversion enables an array of pairs of integers to serve the same
function as a dictionary of ASCII string EIDs.

Note, however, that CBHE decompression cannot faithfully recreate the
dictionary of a compressed primary block from an array of integer
pairs unless the order of the scheme names and scheme-specific part
strings in the dictionary of the original, uncompressed block is
known. (The Bundle Protocol Specification does not require that the
strings in the dictionary appear in any particular order and does not
require that redundant strings be omitted from the dictionary.)
Therefore, a further precondition to CBHE compression is that the
strings in the dictionary of the bundle to be compressed must be
exactly as follows, in this order and without addition:

1. The scheme name of the destination endpoint ID.

2. The scheme-specific part of the destination endpoint ID.

3. The scheme name of the source endpoint ID, if and only if
    different from any prior string in the dictionary.

4. The scheme-specific part of the source endpoint ID, if and only
    if different from any prior string in the dictionary.

5. The scheme name of the report-to endpoint ID, if and only if
    different from any prior string in the dictionary.

6. The scheme-specific part of the report-to endpoint ID, if and
    only if different from any prior string in the dictionary.

7. The scheme name of the current custodian endpoint ID, if and only
    if different from any prior string in the dictionary.

8. The scheme-specific part of the current custodian endpoint ID, if
    and only if different from any prior string in the dictionary.

Note: this constraint implies that a bundle that includes any
extension blocks containing EID-references to endpoint IDs other than
the block's destination, source, report-to, and current custodian
cannot be CBHE compressed since such compression would result in a
dictionary that would deviate from this structure.

2.2. Method

   When the constraints enumerated above are met, the CBHE block
   compression method can be applied by the convergence-layer adapter
   (CLA) at the time the bundle is transmitted via a convergence-layer
   protocol. In a CBHE-compressed primary block, the eight SDNVs that
   normally contain EIDs' scheme name and SSP offsets within the
   dictionary are instead used to contain the eight integer values
   listed below, in the order shown:

   1.   The node number of the destination endpoint ID, or zero if the
        destination endpoint is the null endpoint.

   2.   The service number of the destination endpoint ID, or zero if the
        destination endpoint is the null endpoint.

   3.   The node number of the source endpoint ID, or zero if the source
        endpoint is the null endpoint.

   4.   The service number of the source endpoint ID, or zero if the
        source endpoint is the null endpoint.

   5.   The node number of the report-to endpoint ID, or zero if the
        report-to endpoint is the null endpoint.

   6.   The service number of the report-to endpoint ID, or zero if the
        report-to endpoint is the null endpoint.

   7.   The node number of the current custodian endpoint ID, or zero if
        the current custodian endpoint is the null endpoint.

   8.   The service number of the current custodian endpoint ID, or zero
        if the current custodian endpoint is the null endpoint.

   Further, the dictionary is omitted from the primary block and the
   primary block's dictionary length is set to zero.

   Upon reception, the receiving convergence-layer adaptation de-
   compresses the block by simply reversing the process so that the
   bundle presented to the bundle protocol agent has the standard form
   (i.e., the dictionary is reconstituted).

3. Specification

   CBHE compression is a convergence-layer adaptation. It is opaque to
   bundle processing. Therefore, it has no impact on the
   interoperability of different Bundle Protocol implementations, but
   instead affects only the interoperability of different convergence-
   layer adaptation implementations.

   Bundle Protocol convergence-layer adapters that conform to the CBHE
   specification must implement the following procedures.

3.1. Transmission

   When and only when required by the bundle protocol agent to transmit
   a bundle whose primary block's endpoint IDs satisfy the constraints
   identified in Section 2.1, the CLA MAY encode the primary block of
   the bundle in accordance with the CBHE compression convention
   described in Section 2.2 UNLESS the CLA to which the bundle is to be
   transmitted is known not to be CBHE conformant. Note that CBHE
   compression may be applied only if the receiving CLA is known or
   presumed to be CBHE conformant, i.e., able to decode the encoded
   primary block. Knowledge as to whether or not a receiving CLA is (or
   might be) CBHE conformant may be asserted by node administration
   and/or may be inferred from reception of a CBHE-compressed bundle as
   noted in Section 3.2.

3.2. Reception

   Upon receiving a bundle whose dictionary length is zero (and only in
   this circumstance), a CBHE-conformant convergence-layer adapter:

   1. MAY infer that the CLA from which the bundle was received is CBHE
       conformant.

   2. MUST decode the primary block of the bundle in accordance with
       the CBHE compression convention described in Section 2.2 before
       delivering it to the bundle protocol agent.

   Note that when a CLA that is not CBHE conformant receives a bundle
   whose dictionary length is zero, it has no choice but to pass it to
   the bundle agent without modification. In this case, the bundle
   protocol agent will be unable to dispatch the received bundle,
   because it will be unable to determine the destination endpoint; the
   bundle will be judged to be malformed. The behavior of the bundle
   protocol agent in this circumstance is an implementation matter.

4. IANA Considerations

   IANA has registered a provisional registration (per [RFC4395]) for a
   URI scheme for CBHE, with the string "ipn" as the scheme name, as
   follows:

   URI scheme name: "ipn"

   Status: provisional

   URI scheme syntax:

   This specification uses the Augmented Backus-Naur Form (ABNF)
   notation of [RFC5234], including the core ABNF syntax rule for DIGIT
   defined by that specification.

   ipn-uri = "ipn:" ipn-hier-part
   ipn-hier-part = node-nbr nbr-delim service-nbr ; a path-rootless
   node-nbr = 1*DIGIT
   nbr-delim = "."
   service-nbr = 1*DIGIT

   None of the reserved characters defined in the generic URI syntax are
   used as delimiters within URIs of the IPN scheme.

   URI scheme semantics: URIs of the IPN scheme are used as endpoint
   identifiers in the Delay-Tolerant Networking (DTN) Bundle Protocol
   (BP) [RFC5050] as described in Section 2.1.

   Encoding considerations: URIs of the IPN scheme are encoded
   exclusively in US-ASCII characters.

   Applications and/or protocols that use this URI scheme name: the
   Delay-Tolerant Networking (DTN) Bundle Protocol (BP) [RFC5050].

   Interoperability considerations: as noted above, URIs of the IPN
   scheme are encoded exclusively in US-ASCII characters.

   Security considerations:

   o Reliability and consistency: none of the BP endpoints identified
      by the URIs of the IPN scheme are guaranteed to be reachable at
      any time, and the identity of the processing entities operating on
      those endpoints is never guaranteed by the Bundle Protocol itself.
      Bundle authentication as defined by the Bundle Security Protocol
      is required for this purpose.

o Malicious construction: malicious construction of a conformant
   IPN-scheme URI is limited to the malicious selection of node
   numbers and the malicious selection of service numbers. That is,
   a maliciously constructed IPN-scheme URI could be used to direct a
   bundle to an endpoint that might be damaged by the arrival of that
   bundle or, alternatively, to declare a false source for a bundle
   and thereby cause incorrect processing at a node that receives the
   bundle. In both cases (and indeed in all bundle processing), the
   node that receives a bundle should verify its authenticity and
   validity before operating on it in any way.

o Back-end transcoding: the limited expressiveness of URIs of the
   IPN scheme effectively eliminates the possibility of threat due to
   errors in back-end transcoding.

o Rare IP address formats: not relevant, as IP addresses do not
   appear anywhere in conformant IPN-scheme URIs.

o Sensitive information: because IPN-scheme URIs are used only to
   represent the identities of Bundle Protocol endpoints, the risk of
   disclosure of sensitive information due to interception of these
   URIs is minimal. Examination of IPN-scheme URIs could be used to
   support traffic analysis; where traffic analysis is a plausible
   danger, bundles should be conveyed by secure convergence-layer
   protocols that do not expose endpoint IDs.

o Semantic attacks: the simplicity of IPN-scheme URI syntax
   minimizes the possibility of misinterpretation of a URI by a human
   user.

Contact:
   Scott Burleigh
   Jet Propulsion Laboratory,
   California Institute of Technology
   scott.c.burleigh@jpl.nasa.gov
   +1 (800) 393-3353

Author/Change controller:
   Scott Burleigh
   Jet Propulsion Laboratory,
   California Institute of Technology
   scott.c.burleigh@jpl.nasa.gov
   +1 (800) 393-3353

References: S. Burleigh, "Compressed Bundle Header Encoding (CBHE)",
RFC 6260, May 2011.

5. Security Considerations

   The Bundle Security Protocol (BSP) may, under some conditions, insert
   additional endpoint ID strings into the dictionary of a bundle and
   reference those strings in BSP extension blocks. Because a bundle
   that includes any extension blocks containing EID-references to
   endpoint IDs other than the block's destination, source, report-to,
   and current custodian cannot be CBHE compressed, bundles cannot be
   compressed under those conditions.

   Specifically, the specification detailed above implies that a bundle
   cannot be CBHE compressed when the security-source endpoint for the
   Bundle Authentication Block (BAB) is noted in the dictionary
   (typically, because there is no other way for the receiving bundle
   protocol agent to determine the security-source endpoint); when the
   security-destination endpoint for the BAB is noted in the dictionary
   (in the rare case that the receiving endpoint is not the security-
   destination endpoint); when the security-source endpoint for the
   Payload Integrity Block (PIB), Payload Confidentiality Block (PCB),
   or Extension Security Block (ESB) is not the source endpoint; or when
   the security-destination endpoint for the PIB, PCB, or ESB is not the
   destination endpoint.

   Also, the CBHE-conformance inference mechanism identified in
   Section 3.2 introduces a possible denial-of-service attack.
   Malicious code could issue a CHBE-compressed bundle whose source EID
   falsely identifies the bundle origin as some node whose CLA is not
   CBHE conformant; a CBHE-conformant CLA receiving this bundle might
   incorrectly infer that the CLA at the purported source node was CBHE
   conformant and might then begin CBHE compressing all bundles that it
   sends to that node, thus preventing those bundles from being
   dispatched by the node's bundle protocol agent. Nodes can defend
   against such an attack by requiring Bundle Authentication Blocks and
   discarding any inference of CBHE conformance for the CLAs at nodes
   from which inauthentic bundles are received.

   These caveats aside, CBHE introduces no new security considerations
   beyond those discussed in the DTN Bundle Protocol RFC 5050 [RFC5050]
   and Bundle Security Protocol RFC 6257 [RFC6257] Specifications.

6. References

6.1. Normative References

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

   [RFC3986] Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform
              Resource Identifier (URI): Generic Syntax", STD 66,
              RFC 3986, January 2005.

   [RFC5050] Scott, K. and S. Burleigh, "Bundle Protocol
              Specification", RFC 5050, November 2007.

   [RFC5234] Crocker, D. and P. Overell, "Augmented BNF for Syntax
              Specifications: ABNF", STD 68, RFC 5234, January 2008.

   [RFC6257] Symington, S., Farrell, S., Weiss, H., and P. Lovell,
              "Bundle Security Protocol Specification", RFC 6257,
              May 2011.

6.2. Informative References

   [RFC4395] Hansen, T., Hardie, T., and L. Masinter, "Guidelines and
              Registration Procedures for New URI Schemes", BCP 35,
              RFC 4395, February 2006.

Appendix A. Acknowledgments

Author's Address

   Scott Burleigh
   Jet Propulsion Laboratory, California Institute of Technology
   4800 Oak Grove Drive, m/s 301-490
   Pasadena, CA 91109
   USA

   Phone: +1 818 393 3353
   EMail: Scott.C.Burleigh@jpl.nasa.gov