RFC 4838 Delay-Tolerant Networking, Architecture

Carlo Caini

carlo.caini@unibo.it

DEI/ARCES University of Bologna, Italy

General remarks

- RFC 4838 is the most important document on DTN as it defines the DTN Bundle Protocol (BP) architecture
- It is an “informational” RFC
  - “This memo provides information for the Internet community. It does not specify an Internet standard of any kind.”
- A glance to the authors:
  - V. Cerf, Google (vice president) /Jet Propulsion Laboratory
  - S. Burleigh, A. Hooke, L. Torgerson, NASA/Jet Propulsion Laboratory
  - R. Durst, K. Scott, The MITRE Corporation
  - K. Fall Intel Corporation
  - H. Weiss, SPARTA, Inc
- Note: in the following slides titles and text have been almost completely extracted by the RFC itself to allow the reader to immediately reference to the original text.
  - This also preserves original clarity and language style.
  - Author’s personal comments or additions are indicated with CC

1. Introduction

- This document describes an architecture for delay and disruption- tolerant interoperable networking (DTN).
- The architecture embraces the concepts of occasionally-connected networks that may suffer from frequent partitions and that may be comprised of more than one divergent set of protocols or protocol families.
- The basis for this architecture lies with that of the Interplanetary Internet, which focused primarily on the issue of deep space communication in high- delay environments. We expect the DTN architecture described here to be utilized in various operational environments, including those subject to disruption and disconnection and those with high-delay; the case of deep space is one specialized example of these, and is being pursued as a specialization of this architecture.
- Other networks to which we believe this architecture applies include sensor-based networks using scheduled intermittent connectivity, terrestrial wireless networks that cannot ordinarily maintain end-to-end connectivity, satellite networks with moderate delays and periodic connectivity, and underwater acoustic networks with moderate delays and frequent interruptions due to environmental factors.
- We define an end-to-end message-oriented overlay called the "bundle layer" that exists at a layer above the transport (or other) layers of the networks on which it is hosted and below applications. Devices implementing the bundle layer are called DTN nodes.

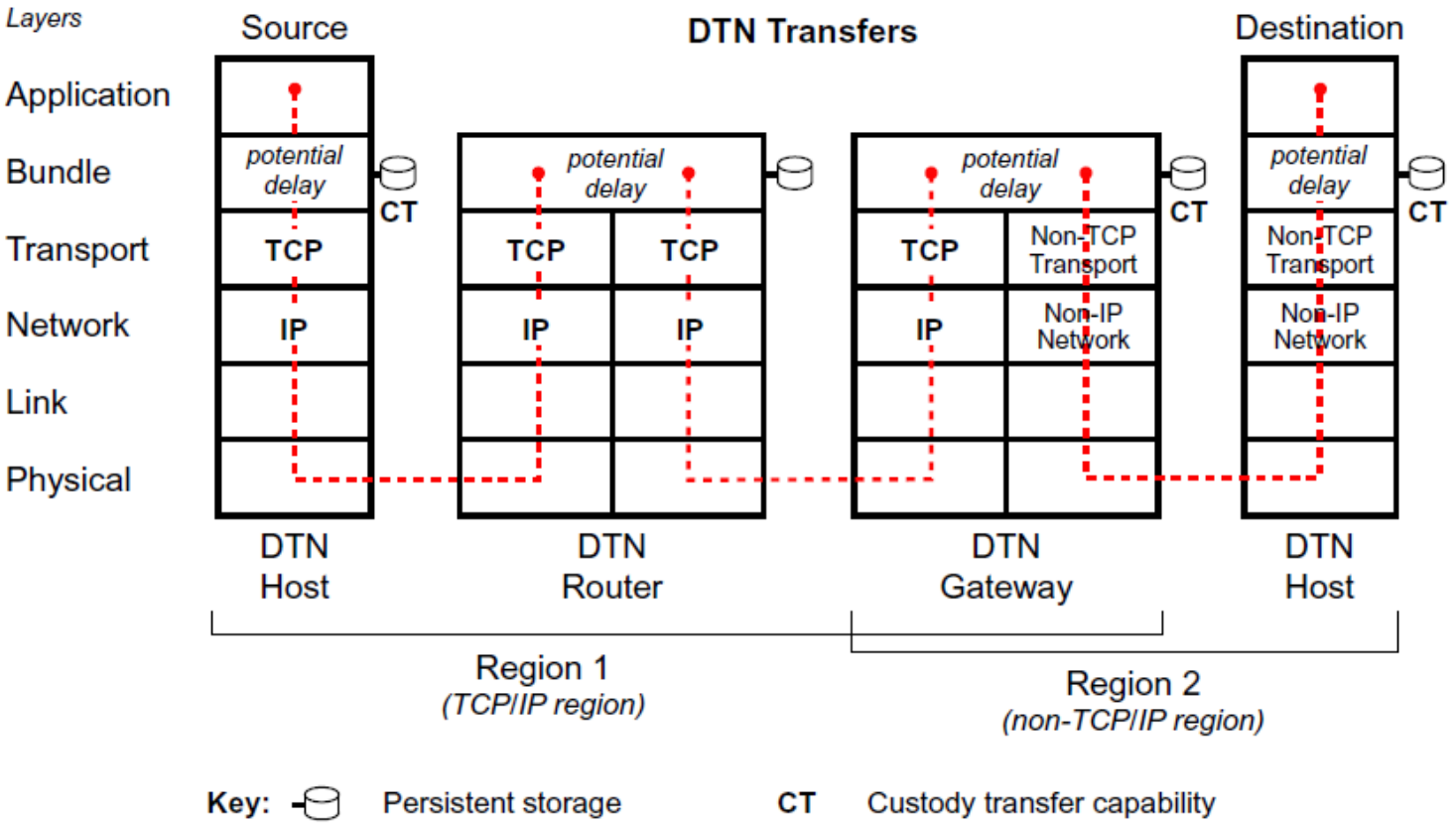
2. Why an Architecture for Delay-Tolerant Networking?

- The existing Internet protocols do not work well for some environments, due to some fundamental assumptions built into the Internet architecture:
  - that an end-to-end path between source and destination exists for the duration of a communication session
  - (for reliable communication) that retransmissions based on timely and stable feedback from data receivers is an effective means for repairing errors
  - that end-to-end loss is relatively small
  - that all routers and end stations support the TCP/IP protocols
  - that applications need not worry about communication performance
  - others…
- The DTN architecture is conceived to relax most of these assumptions

2. Why an Architecture for Delay-Tolerant Networking? (continued)

- Design principles
  - Use variable-length (possibly long) messages (not streams or limited-sized packets)
  - Use a naming syntax that supports a wide range of naming and addressing conventions to enhance interoperability.
  - Use storage within the network to support store-and-forward over multiple paths.
  - Provide security mechanisms that protect the infrastructure from unauthorized use.
  - Provide coarse-grained classes of service, delivery options, and a way to express the useful lifetime of data to allow the network to better deliver data in serving the needs of applications.
- Application design principles:
  - Applications should minimize the number of round-trip exchanges.
  - Applications should cope with restarts after failure while network transactions remain pending.
  - Applications should inform the network of the useful life and relative importance of data to be delivered.

3. Architectural description



3.1 Virtual Message Switching Using Store-and-Forward Operation

- A DTN-enabled application sends messages of arbitrary length, also called Application Data Units or ADUs. The relative order of ADUs might not be preserved (CC:!).
- ADUs are transformed by the bundle layer into one or more protocol data units called "bundles", which are forwarded by DTN nodes.
- Bundles have a defined format containing two or more "blocks" of data.
  - One is the “primary block”, which serves as a sort of header, one is the payload, others (optional) serve different purposes
  - The term "block" is used instead of "header" because blocks may not appear at the beginning of a bundle due to particular processing requirements (e.g., digital signatures).
- Bundles may be split up ("fragmented") into multiple constituent bundles (also called "fragments" or "bundle fragments") during transmission.
  - Fragments are themselves bundles, and may be further fragmented. Two or more fragments may be reassembled anywhere in the network, forming a new bundle.

3.1 Virtual Message Switching Using Store-and-Forward Operation (continued)

- Bundle sources and destinations are identified by (variable-length) Endpoint Identifiers (EIDs)
  - Bundles also contain a “report-to” and a “custodian” EID.
  - An EID may refer to one or more DTN nodes (i.e., for multicast or "report-to" destinations).
- “store-and-forward” in IP vs. DTN
  - In IP there is an assumption that links are continuous
    - the "storing" will not persist for more than a modest amount of time, on the order of the queuing and transmission delay.
  - In contrast, the DTN architecture does not expect that network links are always available or reliable
    - instead expects that nodes may choose to store bundles for some time. In most cases, stored bundles will survive system restarts.

3.1 Virtual Message Switching Using Store-and-Forward Operation (continued)

- An essential element… is that bundles have a place to wait in a queue until a communication opportunity ("contact") is available. This highlights the following assumptions:
  - that storage is available and well-distributed throughout the network
  - that storage is sufficiently persistent and robust to store bundles until forwarding can occur
- Problem:
  - Congestion management and denial of service mitigation
    - Node storage in essence represents a new resource that must be managed and protected

3.2 & 3.3 Nodes, Endpoints, EID and registrations

- A DTN node (or simply "node" in this document) is an engine for sending and receiving bundles -- an implementation of the bundle layer.
  - Applications utilize DTN nodes to send or receive ADUs carried in bundles.
- Nodes may be members of groups called "DTN endpoints". A DTN endpoint is therefore a set of DTN nodes.
- An Endpoint Identifier (EID) is a name, expressed using the general syntax of URIs, that identifies a DTN endpoint.
  - Each node is also required to have at least one EID that uniquely identifies it.
- An application's desire to receive ADUs destined for a particular EID is called a "registration“.

3.2 & 3.3 Nodes, Endpoints, EID and registrations

- Each EID is expressed as a URI [rfc3986]; it consists of a “scheme” and a “scheme-specific part (SSP)”
- In practice two schemes are in use
  - “dtn”, general use
    - dtn://susy.dtn/ping (where dtn://susy.dtn is the node identifier and ping is the demux token, equivalent to TCP or UDP ports)
  - “ipn” , used in deep space networks (NASA)
    - ipn:10.3 (where 10 is the node number and 3 is the “service” number)

Does not tell you where in the network the address actually is

- CC: As names, EIDs are not required to have a topological meaning.
- CC: When registering, an application tells the BP that incoming bundles with a specific “demux token” (or service number) are destined to her (like ports).

3.3.2 Late binding

- “Early binding” in IP
  - the name-to-address translation is performed prior to data being sent into the network by means of a DNS lookup.
- “Late binding” in DTN
  - Binding means interpreting the SSP of an EID for the purpose of carrying an associated message towards a destination.
  - For example, binding might require mapping an EID to a next-hop EID or to a lower- layer address for transmission.
  - "Late binding" means that the binding of a bundle's destination to a particular set of destination identifiers or addresses does not necessarily happen at the bundle source.
- CC: Late binding is a concept difficult to understand, but extremely powerful.
  - E.g. I can send a bundle to a device that does not physically exist and thus has no IP address yet; the bundle can be routed on the basis of the DTN name to a gateway where is stored; when the device “materializes”, it first establishes a TCP/IP connection to the gateway and then tells the gateway its DTN name; as its IP address can be derived by the source field of incoming IP packets, the BP eventually binds the dtn name of the waiting bundle to the IP address of the device.

3.5 Priority Classes

- Three (“cardinal”) priority classes.
  - Bulk - Bulk bundles are shipped on a "least effort" basis. No bundles of this class will be shipped until all bundles of other classes bound for the same destination and originating from the same source have been shipped.
  - Normal - Normal-class bundles are shipped prior to any bulk-class bundles and are otherwise the same as bulk bundles.
  - Expedited - Expedited bundles, in general, are shipped prior to bundles of other classes and are otherwise the same.
- The priority class of a bundle is only required to relate to other bundles from the same source.
  - CC: It means that a high priority bundle from one source will be handled preferentially to a lower priority bundle sent from the same source.
    - If the sources are different, there is not any obligation.

3.6 Postal-Style Delivery Options and Administrative Records

- Administrative records are used to report status information or error conditions related to the bundle layer.
  - Bundle Status Reports
    - Informative bundles sent by BP to the “report-to” EID
      - Bundle Reception, Custody Acceptance, Bundle Forwarded, Bundle Deletion, Bundle Delivery, Acked by application
  - Custody signal
    - Used for signaling custody acceptance or refusal (see below); sent by BP to the “current-custodian” EID
- Administrative records must reference a received bundle.
  - A bundle is uniquely identified by a tuple of 3 elements (5 if is a fragment), as shown later (see 3.12)

Status reports (CC's comments)

- Bundle Status Reports (BSR) are an essential feature of DTN.
  - They allow controllers/researchers to track a bundle like a courier parcel (see the example below).
  - Node that all BSRs are always sent to the "report-to:" EID, which may, or may not, coincide with the source.
    - The distinction between the report-to and the source is powerful. Bundles can be tracked by an external node, acting as a controller.
  - When the report-to EID coincides with the source, the «delivered» (dlv) BSR works as a return-receipt which can be used by the DTN application to enforce end-to-end reliability
    - True end-to-end reliability is not provided by the BP alone; only custodian-by-custodian.

TABLE I: EXCERPT OF A.CSV LOG FILE: STATUS REPORTS TRIGGERED BY BUNDLE "3970.44" SENT FROM VM1 TO VM2.

RX TIME	Report_SRC	Report TST	Rep. SQN	Rep. Type	Bndl_SRC	Bndl TST	Bndl SQN	Bndl FO	Bndl FL	Dlv	Ct	Rcv	Fwd	Del
0	dtn://vm1.dtn	3970	45	S_R	dtn://vm1.dtn/dtnperf:/src_10860	3970	44	0	0		3970			
3.85	dtn://vm2.dtn	3973	42	S_R	dtn://vm1.dtn/dtnperf:/src_10860	3970	44	0	0			3973		
3.93	dtn://vm1.dtn	3974	49	S_R	dtn://vm1.dtn/dtnperf:/src_10860	3970	44	0	0				3974	
4.46	dtn://vm2.dtn	3973	44	S_R	dtn://vm1.dtn/dtnperf:/src_10860	3970	44	0	0		3973			
4.47	dtn://vm2.dtn	3973	45	S_R	dtn://vm1.dtn/dtnperf:/src_10860	3970	44	0	0	3973				

3.7 Primary Block Fields

- Bundle structure (fully defined in RFC 5050)
  - Primary block (like a header: it must be replicated in all fragments)
    - Creation Timestamp - a concatenation of the bundle's creation time and a monotonically increasing sequence number such that the creation timestamp is guaranteed to be unique for each ADU originating from the same source.
    - Lifespan. The lifespan of a bundle is expressed as an offset relative to its creation time. If a bundle is stored in the network (including the source's DTN node) when its lifespan is reached, it may be discarded.
    - Class of Service Flags - indicates the delivery options and priority class for the bundle.
    - Source EID - EID of the source (the first sender).
    - Destination EID - EID of the destination (the final intended recipient(s)).
    - Report-to EID- the EID identifying where reports (return- receipt, route-tracing functions) should be sent.
    - Custodian EID - EID of the current custodian of a bundle (if any).
  - Payload block
  - Optional (Extension) blocks

3.8 Routing and Forwarding

- Connections between DTN nodes can be intermittent, with variable delay, with variable Tx speed (capacity in this RFC), and asymmetrical (included unidirectional links).
  - The interval during which two nodes are connected is called a "contact".
  - In addition, the product of the capacity and the interval is known as a contact's "volume".
- If contacts and their volumes are known ahead of time, intelligent routing and forwarding decisions can be made.
  - CC: Space networks (deterministic contacts)
  - CC: One algorithm: Contact Graph Routing (CGR), designed by NASA-JPL
- When delivery paths through a DTN graph are lossy or contact intervals and volumes are not known precisely ahead of time, routing computations become especially challenging.
  - CC: Opportunistic networks (random contacts).
  - CC: A plethora of routing proposals… most of which of scarce significance. However, a few exceptions exist.

3.8.1 Types of Contacts

- Persistent Contacts (or 'always-on‘). Persistent contacts are always available (i.e., no connection- initiation action is required to instantiate a persistent contact).
- On-Demand Contacts. On-Demand contacts require some action in order to instantiate, but then function as persistent contacts until terminated.
- Intermittent
  - Scheduled Contacts. A scheduled contact is an agreement to establish a contact at a particular time, for a particular duration.
    - An example of a scheduled contact is a link with a low-earth orbiting satellite.
  - Opportunistic Contacts. Opportunistic contacts are not scheduled, but rather present themselves unexpectedly.
    - For example, an unscheduled aircraft flying overhead and beaconing, advertising its availability for communication, would present an opportunistic contact.
  - Predicted Contacts. Predicted contacts are based on no fixed schedule, but rather are predictions of likely contact times and durations based on a history of previously observed contacts or some other information.

3.9 Fragmentation and Reassembly

- DTN fragmentation and reassembly are designed to improve the efficiency of bundle transfers by ensuring that
  - contact volumes are fully utilized (proactive fragmentation)
  - retransmission of partially-forwarded bundles is avoided (reactive fragmentation)
- Proactive Fragmentation
  - A DTN node may divide a block of application data into multiple smaller blocks and transmit each such block as an independent bundle. This approach is called proactive fragmentation because it is used primarily when contact volumes are known (or predicted) in advance.
- Reactive Fragmentation
  - DTN nodes sharing an edge in the DTN graph may fragment a bundle cooperatively when a bundle is only partially transferred… This is called reactive fragmentation because the fragmentation process occurs after an attempted transmission has taken place.

3.9 Fragmentation and Reassembly (continued)

- Fragments are bundles and as that they can be further fragmented
- The primary block must be replicated in each segment
- Fragments must be re-assembled before final delivery
- CC: The mechanism is very similar to IPv4 fragmentation
  - Issues
  - Problems with security (bundle integrity/authentication in particular)
  - Proactive Fragmentation is not implemented in DTN2/DTNME.
  - Reactive Fragmentation is not implemented in ION.
    - Even if reactive fragmentation is not present in an implementation, the ability to reassemble fragments at a destination is required
    - On reactive fragmentation see IETF RFC 7242 on TCPCL https://tools.ietf.org/html/rfc7242

3.10 Reliability and Custody Transfer

- The most basic service provided by the bundle layer is unacknowledged, prioritized (but not guaranteed) unicast message delivery.
- It also provides two options for enhancing delivery reliability:
  - end-to-end acknowledgments
    - Applications wishing to implement their own end-to-end message reliability mechanisms are free to utilize the acknowledgment.
    - CC's comment: BSR are sent to the report-to EID which does not necessarily coincide with the source…
  - custody transfer.
    - The custody transfer feature of the DTN architecture only specifies a coarse-grained retransmission capability, described next.
    - CC: eliminated in BPv7 draft

3.10 Custody Transfer (description)

- Transmission of bundles with the Custody Transfer Requested option specified generally involves moving the responsibility for reliable delivery of an ADU's bundles among different DTN nodes in the network.
- For unicast delivery, this will typically involve moving bundles "closer" (in terms of some routing metric) to their ultimate destination(s), and retransmitting when necessary.
- The nodes receiving these bundles along the way (and agreeing to accept the reliable delivery responsibility) are called "custodians".
- The movement of a bundle (and its delivery responsibility) from one node to another is called a "custody transfer".

3.10 Custody Transfer (description, cont.)

- Custody transfer allows the source to delegate retransmission responsibility and recover its retransmission-related resources relatively soon after sending a bundle.
  - CC: Very good for sensors with limited storage and… for spies, likely unwilling to be caught with secret documents in their hands!
- Not all nodes are required by the DTN architecture to accept custody transfers, so it is not a true 'hop-by-hop' mechanism. For example, some nodes may have sufficient storage resources to sometimes act as custodians, but may elect to not offer such services when congested or running low on power.
- Custody transfer provides a (weak) mechanism for enhancing the reliability of message delivery (with respect to the sole use of reliable convergence layers).
  - when custody transfer is requested, the bundle layer provides an additional coarse-grained timeout and retransmission mechanism and an accompanying (bundle-layer) custodian-to-custodian acknowledgment signaling mechanism.
  - When an application does not request custody transfer, this bundle layer timeout and retransmission mechanism is typically not employed, and successful bundle layer delivery depends solely on the reliability mechanisms of the underlying protocols.

3.10 Custody Transfer (advanced, cont.)

- When a node accepts custody for a bundle that contains the Custody Transfer Requested option, a Custody Transfer Accepted Signal is sent by the bundle layer to the Current Custodian EID contained in the primary bundle block.
- In addition, the Current Custodian EID is updated to contain one of the forwarding node's (unicast) EIDs before the bundle is forwarded.
- When an application requests an ADU to be delivered with custody transfer, the request is advisory.
  - In some circumstances, a source of a bundle for which custody transfer has been requested may not be able to provide this service. In such circumstances, the subject bundle may traverse multiple DTN nodes before it obtains a custodian. Bundles in this condition are specially marked with their Current Custodian EID field set to a null endpoint.
- In cases where applications wish to require the source to take custody of the bundle, they may supply the Source Node Custody Acceptance Required delivery option.
  - This may be useful to applications that desire a continuous "chain" of custody or that wish to exit after being ensured their data is safely held in a custodian.
  - CC': In DTN the decision of sending a bundle is decoupled from the actual Tx. It is delegated to BP, even when the custody option is not requested. From the point of view of the application sending a bundle is a "Fire and forget" mechanism. Very powerful service offered by the BP to the application.

3.10 Custody Transfer (advanced)

- Problems with unidirectional links
  - In a DTN network where one or more custodian-to-custodian hops are strictly one directional (and cannot be reversed), the DTN custody transfer mechanism will be affected over such hops due to the lack of any way to receive a custody signal (or any other information) back across the path, resulting in the expiration of the bundle at the ingress to the one-way hop.
  - This situation does not necessarily mean the bundle has been lost; nodes on the other side of the hop may continue to transfer custody, and the bundle may be delivered successfully to its destination(s). However, in this circumstance a source will receive an expiration report for the bundle, and possibly conclude (incorrectly) that the bundle has been discarded and not delivered.
  - Although this problem cannot be fully solved in this situation, a mechanism is provided to help ameliorate the seemingly incorrect information that may be reported when the bundle expires after having been transferred over a one-way hop. This is accomplished by the node at the ingress to the one-way hop reporting the existence of a known one-way path using a variant of a bundle status report. These types of reports are provided if the subject bundle requests the report using the 'Report When Bundle Forwarded' delivery option.
  - CC: however, the old custodian could receive the custody acceptance via a third node in a triangular layout (e.g. via C. in the layout A->B-C-A).

3.11 DTN Support for Proxies and Application Layer Gateways

- In cases where existing Internet applications can be made to tolerate delays, local proxies can be constructed to benefit from the existing communication capabilities provided by DTN. Making such proxies compatible with DTN… allows existing TCP/IP-based applications to operate unmodified over a DTN-based network.
  - Comment: the difficulty of making TCP/IP compatible with DTN long delays is largely underestimated in the statement above.
- When DTN is used to provide a form of tunnel encapsulation for other protocols, it can be used in constructing overlay networks comprised of application layer gateways.
  - The application acknowledgment capability is designed for such circumstances. This provides a common way for remote application layer gateways to signal the success or failure of non-DTN protocol operations initiated as a result of receiving DTN ADUs. Without this capability, such indicators would have to be implemented by applications themselves in non-standard ways.

3.12 Timestamps and Time Synchronization

- The DTN architecture depends on time synchronization among DTN nodes for
  - bundle and fragment identification
  - routing with scheduled or predicted contacts
  - bundle expiration time computations
  - application registration expiration.
- Unique bundle identification is used for a number of purposes, including custody transfer and reassembly of bundle fragments.
  - The 3-tuple (source EID, timestamp, timestamp seq. number) identify an original non fragmented bundle
  - The additional 2-ple (fragment offset and fragment length) is necessary to distinguish two fragments of the same bundle (e.g. in administrative records).
  - Note that fragment offset and administrative data unit length (not fragment length!) are used to reassemble fragments into the original bundle

Following sections

- CC’s summary:
  - 3.13 Congestion control (only general considerations)
  - 3.14 Security, see RFC 9172 Bundle Protocol Security (BPSec), 2022
  - 4, 5, 6, 7, 8 sections seem not fully developed and are neither recalled nor commented here.