Delay-Tolerant Networking

An Approach to Interplanetary Internet

Scott Burleigh, Adrian Hooke, Leigh Torgerson, Kevin Fall, Vint Cerf, Bob Durst, Keith Scott and Howard Weiss

Presented by: James Kazmierczak
Introduction: The Problem

• A scientist wishes to upgrade information on a station on Mars
  – Information must travel from the scientist's workstation to a deep space antenna
  – From Antenna to a relay satellite in low Mars orbit
  – From Mars orbit to the station.
The Journey

• First step typically done with TCP/IP
  – Small propagation latencies
  – high data rates
  – end-to-end connectivity

• Second step however…
  – Very large propagation latencies (measured in hours or even days)
  – low data rates (8-256 kb/s)
  – intermittent connectivity
  – Due to orbits, possible time-disjoint periods for reception and transmission

• Third step likely to be on a Wireless lan, so TCP/IP might again be the best choice
  – Data rates will likely be low
Solutions?

• Some sort of retransmission protocol capable of handling these factors is needed
  – Forward Error Correction (FEC) can be used,
    • Will consume limited bandwidth whether or not it is needed
    • Provides no protection against sustained outages
  – Some sort of Automated Repeat Request (ARQ) is needed
CCSDS File Delivery Protocol

- Abbreviated as CFDP
  - Developed by Consultative Committee for Space Data Systems
- Designed for reliable file transfer across interplanetary distances
  - CFDP-RP is a subset of CFDP
    - Would implement just the retransmission aspects of CFDP
    - Currently Hypothetical
CFDP Overview

• No connection protocol
  – Time required to establish a connection may exceed duration of communication opportunity

• Does not wait for acknowledgments
  – Acks may take hours/days to arrive
  – Each message must contain a common transaction identifier

• Retransmission buffers retained in nonvolatile storage
  – May send retransmissions out of order
Existing Protocols- TCP

• TCP/IP
  – Connection must be negotiated
  – Establishing takes at least one round trip
  – If latency exceeds duration of communication opportunity, no data will flow
  – Delivers data in transmission order
    • If retransmission is required, all other data will be slowed down as well
TCP’s retransmission protocol

• Suppose we have three-hop connection ABCD
  – 500 ms A-B
  – 8 minutes B-C (480,000 ms)
  – 100 ms C-D
• TCP retransmission is only possible if original sender retains a copy until it knows it will not be needed
  – A single packet of data must be retained on A for 961,200 ms (AB, BC, CD, DC, CB, BA)
  – In turn, this means memory is needed to store all the packets transmitted in that time...
    • Suppose A is transmitting 256 kb images every 250 ms (1 Mb/s)
    • A’s retransmission buffers will consume over 961 Mb of memory (Can not assume memory is cheap)
Existing Protocols- Routing

• Hierarchy System

• Various routing protocols
  – Select paths in changing topologies
  – Assume network is not partitioned
    • losses of connectivity are assume to be structural rather than operation or temporary
    • If there is currently no direct or indirect route to a network element it will not be included in any computed path
Routing Issues

• Highest level is BGP (Border Gateway protocol)
  – Built on TCP
    • Performance in high-delay environments is limited by the TCP operational issues
    • If TCP can’t keep a connection, BGP performs poorly
    • Premature timeouts lead to false negative conclusions about connectivity
Delay Tolerant Network Architecture

• New protocols needed
  – Use exactly those protocols at all layers that are best suited to operation within each environment
  – Create a general purpose application level gateway infrastructure
  – Can not be based on any end-to-end expectation of:
    • Continuous connectivity, low or constant transmission latency, low error rate, low congestion, high transmission rate, symmetrical data rate, common name or address expression syntax or semantics, data arrival in transmission order
Three Main Principles of DTN architecture

1) Postal Model of Communications
   - Messages should be self-contained Units
   - Applications should issue messages asynchronously, and should not wait for responses before sending the next message
   - Example:
     - Requesting a transmission of the file would not initiate a dialog (such as in FTP)
       - Instead, it would bundle together into a single message the name of the file, and any other meta data that may be required, such as username/password, encoding instructions etc.
Three Main Principles of DTN architecture continue

2) Tiered Functionality

– Protocols designed for use within various environments already exploit whatever favorable conditions the environments offer

• DTN architecture relies on those capabilities to the greatest extent possible
• The Bundling protocol (one layer higher on the stack), performs any required additional functions that local protocols can’t
Three Main Principles of DTN architecture continue

3) Terseness
   – Bandwidth cannot be assumed to be cheap
   – Protocols should use processing power over bandwidth whenever possible/necessary
Main Structural Elements of DTN: Tiered Forwarding

• Terminology
  – DTN nodes - analogous to hosts and routers
  – DTN region – informally defined as a set of DTN nodes that can communicate among themselves using a single common protocol
Tiered Forwarding

• Relies on existing protocols (such as the internet) for forwarding bundles between DTN nodes

• Gateway nodes straddling boundaries between regions convert data
  – has an interface that can communicate with both regions protocols
Tiered Forwarding

- Gateway nodes also perform bundling’s store-and-forward operations
  - Data may need to be stored for hours or days in nonvolatile memory
  - Deferred transmission may be unavoidable - Continuous connectivity can not be assumed
    - Power management, orbital dynamics, etc.
Tiered Naming and Addressing

• Need a DTN Regional Identifier in addition to a DTN node identifier
• Destination address contains a concatenated Regional ID and a Regional Destination ID called a tuple
  – Regional Destination ID’s are ‘late bound’
  • Allows new regions with new naming and addressing systems to be added
Tiered Routing

• Route computation must be sensitive to future link establishments, or Contacts.
• Contacts can be anticipated by:
  – May be scheduled network management
  – May be discovered in real time within regions
  – May be predicted based on region specific awareness (mobility patterns or orbital dynamics)
  – May be computed stochastically based on prior contact history
Tiered ARQ

- DTN relies on regional protocols (such as TCP or CFDP-RP) for assured transmission of bundles
  - These mechanisms may not always be sufficient in regions with large round-trip latencies
- Certain nodes may become custodians of certain bundles
  - Guarantee memory space for that bundle
Tiered Security

• Nodes practice mutual suspicion
  – Can not prevent introduction of unauthorized data, but can prevent its spread

• Applications (above bundling) may require additional security
  – Can not query key servers, negotiate shared keys, etc.
  – One possibility is to send a certificate containing cryptographic key material
    • This technique might violate the terseness principle
Tiered Congestion Control

- TCP has carefully engineered congestion avoidance
- In admission controlled environments, congestion is controlled by management of resources rather than protocols
  - Access is scheduled and controlled, usually planned in advance rather than real time
- DTN architecture relies on the effectiveness of existing protocols within regions
  - Yet to be seen if this will be effective congestion control
  - More mechanisms may be needed at the bundling layer
Resilient Delivery

- Due to long transmission latencies, destination services may not be running when a bundle finally arrives
  - Or a source service may no longer be active when a reply arrives
- DTN nodes must be equipped with deferred delivery
  - Nodes may have to store bundles until destination services start (or restart) and are able to receive it
  - May even be necessary for Bundling to reanimate destination services
Postal Service Levels

• Postal Communications work well, so let's take a page from their book
  – Three levels of delivery priority
    • Low, Standard, High
      – Can optionally be sent to a ‘reply-to’ instead of original sender
  – Notice of initial transmission (notice of mailing)
  – Notice of delivery of the ultimate destination (return receipt)
  – Report of route taken (delivery record)
Results

- Simple, yet highly adaptable and extensible structure
DTN without Bundling

• DTN departs from the Internet model of communication… but what if it didn’t?
  – Build a reliable link system using TCP/IP tunnels
  – Build an IP virtual interface to the TCP/IP RLT (reliable link tunnel) system
    • Also build it for CFDP-RP regional protocols
DTN without Bundling continued

• Solving Tiered naming and addressing…
  – Carry regional destination IDs as URLs in HTTP 1.1 layered on top of UDP
  – URL resolution at the HTTP layer of the destination regions gateway can determine the final IP address
DTN without Bundling continued

• Other new virtual interfaces could handle:
  – Management of nonvolatile storage (for tiered forwarding)
  – Custodial retransmission (for tiered ARQ)
  – Mutual suspicion (for tiered security)
  – Deferred delivery and service reanimation (for resilient delivery)
  – Postal service notifications, together with differentiated service capabilities (for postal service levels)
DTN without Bundling continued
Conclusion

• Non-bundling DTN would be familiar to application developers
• Still need to be developed with DTN architecture in mind, but the interface would be a known one
• However, proposed virtual layers are a poor idea
  – A single bundling layer would be vastly simpler, and easier to manage and port
Conclusion

• Work began in early 1998
• Internet Research Task Force organized in October of 2002
• Over the past year (2003) created a prototype demonstrated to be tolerant of system reboots and simulated connectivity lapses of up to 60 minutes
• Publicly available at http://www.dtnrg.org/
  – Latest code version is Dec. 2005 (v. 2.1.99)
Questions?