Delay-Tolerant Networking for Extreme Environments

Kevin Fall
Intel Research
Berkeley, CA
kfall@intel-research.net
http://www.intel-research.net

What is Extreme?

• Deep Space Communications
  – Beyond near-earth
  – Landers, Orbiters, Deep Space Probes

• Sensor Networks
  – Terrestrial: Ocean or Land Based
  – Extra-terrestrial objects (on planets, etc)

• High-Stress Physical Environments
  – Battlefield, Civil Emergency, Submarines
RFC1149

• “…encapsulation of IP datagrams in avian carriers” (i.e. birds, esp carrier pigeons)
• Delivery of datagram:
  – Printed on scroll of paper in hexadecimal
  – Paper affixed to AC by duct tape
  – On receipt, process is reversed, paper is scanned in via OCR

Implementation of RFC1149

• See http://www.blug.linux.no/rfc1149/
Ping Results

Script started on Sat Apr 28 11:24:09 2001
vegard@gyversalen:~$ /sbin/ifconfig tun0
tun0      Link encap:Point-to-Point Protocol
          inet addr:10.0.3.2  P-t-P:10.0.3.1  Mask:255.255.255.255
          UP POINTOPOINT RUNNING NOARP MULTICAST MTU:150 Metric:1
          RX packets:1 errors:0 dropped:0 overruns:0 frame:0
          TX packets:2 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0
          RX bytes:88  (88.0 b)  TX bytes:168  (168.0 b)
vegard@gyversalen:~$ ping -i 900 10.0.3.1
PING 10.0.3.1 (10.0.3.1) 56 data bytes
64 bytes from 10.0.3.1: icmp_seq=0  ttl=255 time=6165731.1 ms
64 bytes from 10.0.3.1: icmp_seq=4  ttl=255 time=3211900.8 ms
64 bytes from 10.0.3.1: icmp_seq=2  ttl=255 time=5124922.8 ms
64 bytes from 10.0.3.1: icmp_seq=1  ttl=255 time=6388671.9 ms
--- 10.0.3.1 ping statistics ---
9 packets transmitted, 4 packets received, 55% packet loss
round-trip min/avg/max = 3211900.8/5222806.6/6388671.9 ms
vegard@gyversalen:~$ exit

Script done on Sat Apr 28 14:14:28 2001

Comms Challenges

- Large Delays
- Intermittent and Scheduled Links
- Limited Power Nodes
- Bandwidth Asymmetry
- Limited Emission Requirements (LPI/LPD)
- Heterogeneous Network Architectures
- Link Security Needs
Extreme Delay Links

• The Problem:
  – Delivery may be prop or tx-time dominant
  – Both can be extreme:
    • Very long propagation → problems w/RTX
    • Very slow links → longer tx time, more storage

• Long propagation delay especially difficult
  – Can’t buy less latency
  – Adversely affects conventional reliable transports

TCP Dependence on RTT

• Slow-start ramp: time to window $W$:

\[ T = (RTT \cdot b) \log_2 \frac{W}{W_0} \text{ for initial window } W_0 \]

• Steady-State Throughput limited by

\[ B(p) \approx \min \left( \frac{W_{\text{max}}}{RTT}, \frac{1}{RTT \sqrt{\frac{2bp}{3} + T_0 \min(1,3,\sqrt{\frac{3bp}{8}}) p(1 + 32p^2)}} \right) \]

$p = \text{loss probability}, \ b = \text{pkts ACKd per ACK}, \ T_0 = \text{initial RTO}$
Asymmetric Links

• Significant differences in each direction
  – Latency (MAC behavior, path)
  – Bandwidth (cost/engineering, technology)
  – Loss characteristics (power, path, buffering)
• Problems in one direction affect the other
  – ACK congestion → lost ACKs
  – Burst ACK arrivals → burst sending
• Some cases have no reverse channel
  – Possibly applicable to erasure coding...

Cascaded Intermittent Links

• Prob success (iid fail prob $p_f$) over $k$ links:
  \[ p_s = 1 - p_f; p_s(k) = (1 - p_f)^k \]
• For E2E delivery must have all links up
• But, expected # of failed links is
  \[ kp_f \]
Routing Issues

- End-to-end path may not exist
  - Lack of many redundant links [there are exceptions]
  - Traditional routing assumes at least one path exists, fails otherwise

- Routing Algorithms More Complex
  - Scheduled links and contact opportunities
  - Need to match between pending messages, send opportunities and message priority
  - Available power may affect link selection decision
  - Typical routing algorithms optimize 1 metric. Those that don’t [e.g. BGP] can become unstable.
Limited Power

• Extreme devices tend to use batteries

<table>
<thead>
<tr>
<th>Battery Type</th>
<th>Cell Voltage (V)</th>
<th>Storage Density (Wh/kg)</th>
<th>Temp Range (°C)</th>
<th>Rapid Charge (hrs)</th>
<th>Memory Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>NiCD</td>
<td>1.2</td>
<td>40-60</td>
<td>-10 to +50</td>
<td>.5 to 1</td>
<td>Yes</td>
</tr>
<tr>
<td>NiMH</td>
<td>1.25</td>
<td>60-80</td>
<td>-10 to +50</td>
<td>2 to 3</td>
<td>No</td>
</tr>
<tr>
<td>Li-Ion</td>
<td>3.6</td>
<td>100</td>
<td>-20 to +60</td>
<td>3 to 6</td>
<td>No</td>
</tr>
<tr>
<td>Li-Poly</td>
<td>3</td>
<td>140</td>
<td>-30 to +55</td>
<td>8 to 15</td>
<td>No</td>
</tr>
</tbody>
</table>

• Power Requirements

<table>
<thead>
<tr>
<th>Intel 2011</th>
<th>RFMTR1000</th>
<th>LQUWM1k</th>
<th>LQUWM7k</th>
<th>Sojourner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rx Cur</td>
<td>300mA</td>
<td>4.6mA</td>
<td>45mA</td>
<td>43mA</td>
</tr>
<tr>
<td>Tx Cur</td>
<td>500mA</td>
<td>12mA</td>
<td>60mA</td>
<td>1.09A</td>
</tr>
<tr>
<td>Sleep Cur</td>
<td>25mA</td>
<td>5uA</td>
<td>480uA</td>
<td>348uA</td>
</tr>
<tr>
<td>Tx Power</td>
<td>63mW</td>
<td>0.75 mW</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

Expected Lifetime

• An example with the RFM Radio:
  – 2 AA cells (2900 mAh each), 3v
  – 100% Duty Cycle Xmit: 242 hrs (10 days)
  – 100% Duty Cycle Rcv: 630 hrs (26 days)
  – 100% Duty Cycle Sleep: 580k hrs (66 yrs)

• Clearly:
  
  *Power management is fundamental*
Heterogeneous Architectures

• Most extreme systems do not (won’t) run IP
  – Oceans: just beginning to investigate routing
  – Space: very limited routing [e.g. rover to lander]
  – Sensors: novel or simple routing, low power
  – Too much overhead, no need for global routing
• Most have domain-specific naming scheme
  – Typically, a flat node ID or name
• But we don’t want to scrap existing (Internet) software

Delay-Tolerant Architecture

• Goals
  – Interoperability across network architectures
  – Reliability robust to link and node failure
• Components
  – Flexible Naming Scheme
  – Reliable Message Overlay with Routing
  – Per-hop Authentication with CoS
  – Interoperability Gateways
Naming

- Naming *region*:
  - Local region naming
  - Inter-region prefix
- Tuples of the form (A, T)
  - A: administrative [valid inside a region]
  - T: topological [valid globally]
- Inter-region next hop determinable by local $f(T)$ directly
- A is externally opaque and region-specific
  - May encompass esoteric routing [e.g. diffusion]
  - Could be object names, addresses, etc.

Reliable Message Overlay

- End-to-End Message Service: “Bundles”
  - “postal-like” delivery over regional transports
  - Optional class of service/notification
- Key Idea: *Custody Transfer*
  - *Custodian* owns reliable-delivery guarantee
  - Bundles transferred between custodians toward destination
  - Sender may free resources upon successful custody transfer
Bundles

- Bundles
  - Arbitrarily long messages delivered end-to-end between DTN capable nodes over distinct (but possibly identical) transport layers.
  - May have associated delivery characteristics. Thus, CoS is always at bundle granularity.
  - Bundles may be fragmentary and require reassembly to be complete.
  - Authenticated/verified during delivery.

Routing, Forwarding and Custody Transfer

- “Classic” Concepts (Internet):
  - **Routing**: selecting best next hop for every possible destination
  - **Forwarding**: sending packet to best next hop
    - Typically, “on demand” [statistical multiplexing]
    - Forwarders know *a-priori* next hop for every destination

- DTN Concepts:
  - **Routing**: selecting best DTN next hop for destination
  - **Forwarding**: sending a bundle p2p when possible
  - **Custody Transfer**: reliable intra-DTN delivery *(with storage)*
DTN Node Types

- Non Persistent Node [NP node]
  - no stable storage
  - Build/consume bundles, forwards bundles, participates in time synchronization
  - May forward or cache bundle or bundle parts
  - Never assumes custody
- Persistent Node [P node]
  - stable storage
  - Does everything an NP node does
  - Always accepts custody of a bundle on success
  - Notifies prior custodian of custody transfer
- Exception: SRC/DST always accept custody

Bundle Routing Example
Comparison

• Trials until success [end-to-end]:

\[ E(X) = \frac{1 - (1 - p_f)^k}{(1 - p_f)^k} \]

• Trials until success [link-by-link]:

\[ E(X) = \sum_k \frac{q}{p} = k \frac{p_f}{(1 - p_f)} \]

(assuming \( p_f < 1 \))

Expected Retries

<table>
<thead>
<tr>
<th>Pf</th>
<th>LINK-LINK</th>
<th>E2E</th>
<th>LINK-LINK</th>
<th>E2E</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.11 0.43 1.00 2.33 9.00</td>
<td>0.11 0.43 1.00 2.33 9.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.22 0.86 2.00 4.67 18.00</td>
<td>0.22 0.86 2.00 4.67 18.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.33 1.29 3.00 7.00 27.00</td>
<td>0.33 1.29 3.00 7.00 27.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.44 1.71 4.00 9.33 36.00</td>
<td>0.44 1.71 4.00 9.33 36.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.56 2.14 5.00 11.67 45.00</td>
<td>0.56 2.14 5.00 11.67 45.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pf</th>
<th>LINK-LINK</th>
<th>E2E</th>
<th>LINK-LINK</th>
<th>E2E</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.00 0.36 9.00 21.00 81.00</td>
<td>0.00 0.36 9.00 21.00 81.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>1.11 4.26 10.00 23.33 90.00</td>
<td>1.11 4.26 10.00 23.33 90.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>1.22 4.71 11.00 25.67 99.00</td>
<td>1.22 4.71 11.00 25.67 99.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>1.33 5.14 12.00 28.00 108.00</td>
<td>1.33 5.14 12.00 28.00 108.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>1.44 5.57 13.00 30.33 117.00</td>
<td>1.44 5.57 13.00 30.33 117.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

LINK-LINK formula: \( kp_f/(1-p_f) \)

E2E formula: \( \lfloor \frac{1 - (1 - p_f)^k}{(1 - p_f)^k} \rfloor \)
Internetwork Operation

Source

Internet

TCP
IP
Ethernet
UWB-specific

Bundle Protocols
Sensor Transport
Diffusion Routing

Sensor Region

Destination

Application

Bundle Protocols
Sensor Transport
Diffusion Routing

Network Option

UWB-specific

Interoperability Gateways

Internet

Extreme Environment

FTP
TCP
IP

Application Proxy

Internet

TCP
IP
Ethernet
UWB-specific

Bundle File Transfer
Bundle Transport
Bundle Routing

Bundle Protocols
Sensor Transport
Diffusion Routing

Extreme Network
CoS the USPS Way…

DTN CoS

- Classes of Service for a Bundle:
  - Types: Expedited, Regular, Bulk
  - Options: send notification, keep delivery record, inform on delivery

- Stamps encode CoS, are not forgeable, and are obtained by sender from trusted service

- DT routers can verify CoS in stamp using network “forwarding service” key
Postage Stamps

- Each bundle contains a cryptographically-signed “postage stamp”
  - Similar to Kerberos tickets
- Provides authorization to use the network at a particular class of service for a particular message
- Postage stamps are verified at each P node
  - NP nodes may not store any complete bundle
  - Endpoint P nodes are special (later)

Related Work

- Protocol Architecture
  - ARPANET design, NewArch, IPN
- Naming, Addressing and Routing
  - Intentional naming, CHORD, CAM
  - IPNL, TRIAD, RON
  - Diffusion Routing, MPLS (sort of)
- Extreme Links
  - Many (WHOI, JPL, UCB, UW, MIT, …)
Futures

• DTN work based on earlier IPN Architecture
  – Interplanetary Internet www.ipnsig.org
  – Mitre, JPL, MCI and others
  – DTN generalizes to non-space environments
• Investigations
  – Army TI and Special Forces Ops
  – Heterogeneous UCB/Intel/JPL Sensor Nets
  – UWB Developments
    • BWRC, Intel, UCSD, Rutgers, USC

Thank you for listening…