SCALE and heterogeneity

Grid Computing (M)

Lecture 19

UNIVERSITY *of* GLASGOW

Case Study 3

Delay-Tolerant Networking

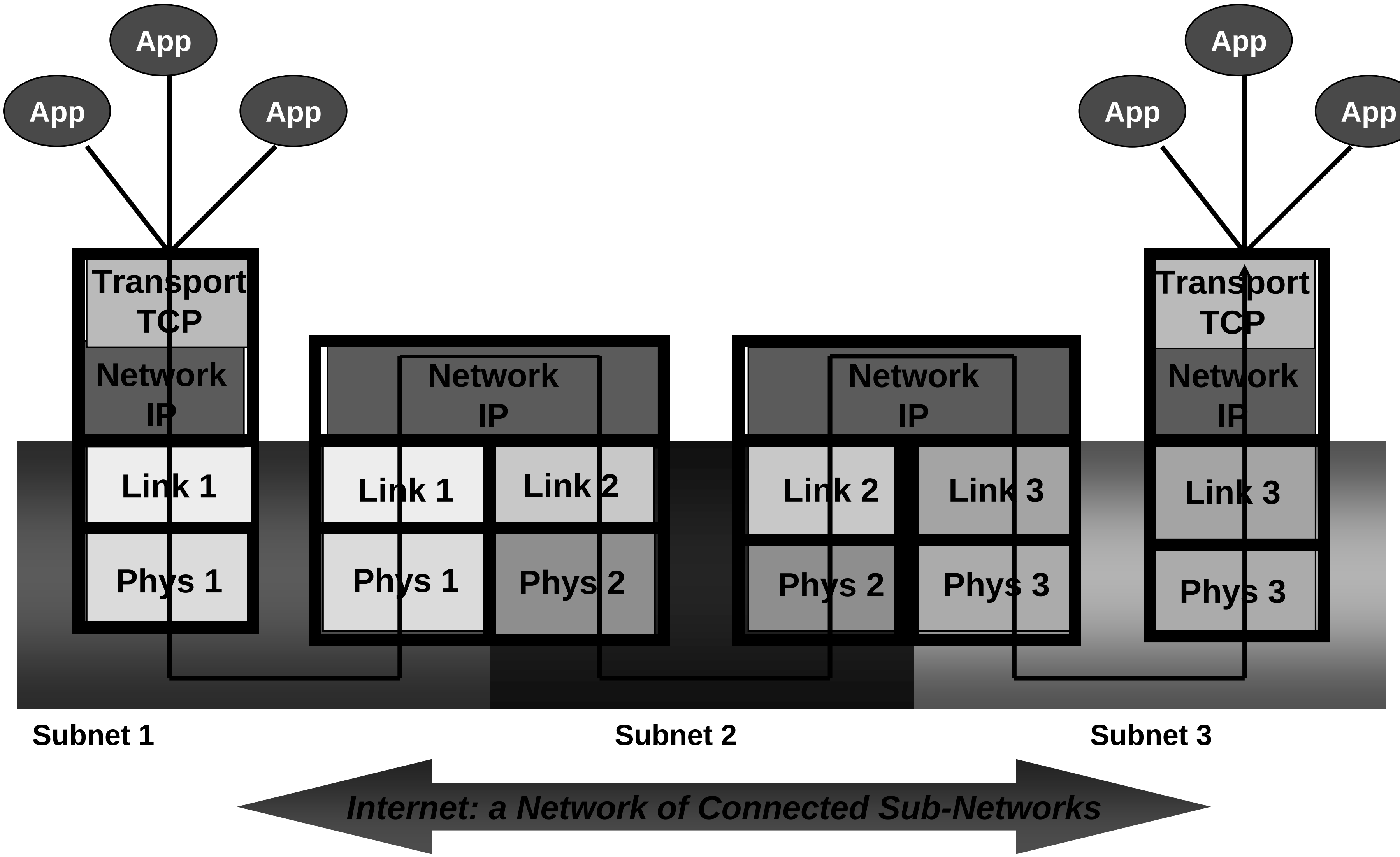
Material taken from “Delay-Tolerant Networks (DTNs), A Tutorial” by F. Warthman, Version 1.1, 5 March 2003 and a presentation entitled “The Interplanetary Internet” by Burleigh et al @ INET 2001.

Assumptions underlying the TCP/IP stack

1. Continuous, bidirectional end-to-end path – a continuously available bidirectional connection between source and destination to support end-to-end interaction
2. Short round-trips – small and relatively consistent network delay in sending data packets and receiving the corresponding acknowledgement packets
3. Symmetric data rates – relatively consistent data rates in both directions between source and destination
4. Low error rates – relatively little loss or corruption of data on each link

If any one of these assumptions are violated in a particular environment, the efficacy of TCP/IP will be in question; if any two of them are violated, it is likely that TCP/IP will be useless in that environment

IP:
the “Thin Waist” of the Earth’s Internet



Characteristics of evolving networks

- Intermittent connectivity – if there is no end-to-end path between source and destination (network partitioning), the end-to-end communication used by the TCP/IP stack does not work.
- Long or variable delay – long propagation delays between nodes and variable queuing delays at nodes contribute to end-to-end path delays that can defeat Internet protocols and/or applications that rely upon quick return of acknowledgements or data
- Asymmetric data rates – if asymmetries are large, they defeat conversational protocols
- High error rates – bit errors on links require correction or retransmission of the entire packet; for a given link-error rate, fewer retransmissions are needed for hop-by-hop than for end-to-end retransmission (linear vs. exponential increase of load on each hop)

Store-and-forward message switching

- DTNs overcome these problems by using store-and-forward message switching; whole messages are moved (forwarded) from a storage place on one node (switch intersection) to a storage place on another node, along a path that *eventually* reaches the destination
- The storage places are persistent
- DTN routers need persistent storage for their message queues for one or more of the following reasons:
  - A communication link to the next hop may not be available for a long time
  - One node in a communicating pair may send or receive data much faster or more reliably than the other node
  - A message, once transmitted, may need to be retransmitted if an error occurs upstream, or if an upstream node declines acceptance of a forwarded message
- Acceptance of a message implies acceptance of responsibility for forwarding the message onward towards its destination

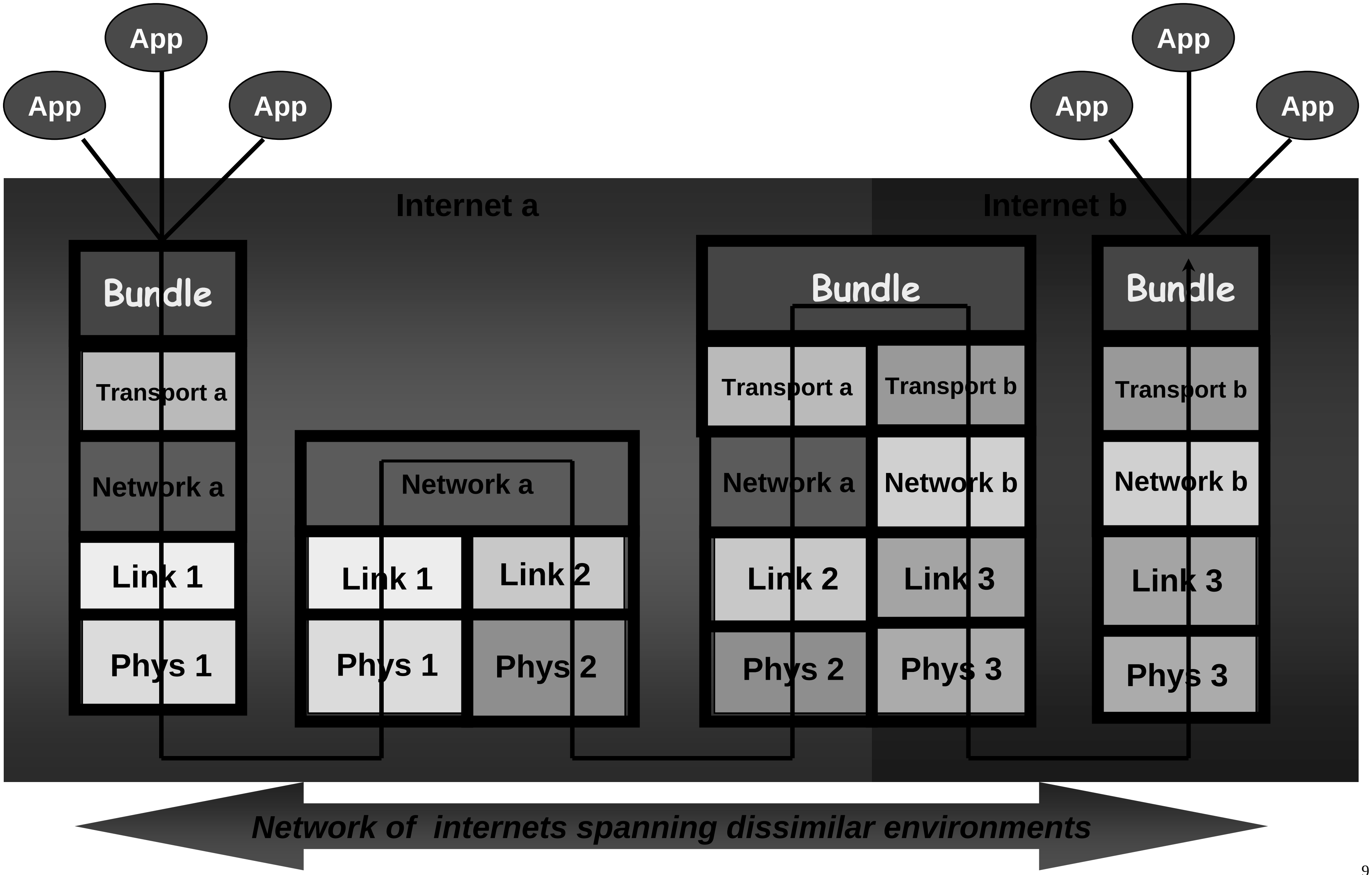
Intermittent connectivity

- An ever increasing number of communicating devices are in motion and/or operate on limited power; e.g. interplanetary probes, mobile *ad hoc* networks
- When nodes are in motion, links can be obstructed by intervening bodies; when nodes must conserve power, links are shut down; these events cause intermittent connectivity, leading to network partitions
- Types of contacts:
  - Opportunistic – contacts at unscheduled times; moving people, vehicles, aircraft, or satellites may make contact and exchange information when the are within line-of-sight and close enough to communicate using their available power
  - Scheduled – in space, almost everything is in motion and speed-of-light delays are significant; if potentially communicating nodes move along predictable paths, they can predict or receive time schedules of their future positions and arrange their future communication sessions

The Bundle Layer

- The DTN architecture implements store-and-forward message switching by overlaying a new protocol layer – called the bundle layer – on top of heterogeneous, region-specific lower layers
- The bundle layer stores and forwards entire bundles between nodes; a single bundle-layer protocol is used across all networks (regions) that make up a TDN
- The layers below the bundle layer (transport and below) are chosen for their appropriateness to the communication environment of each region
- Bundles consist of three things:
  - A source-application's user data
  - Control information, provided by the source appl for the dest appl, describing how to process, store, dispose of, and otherwise handle the user data
  - A bundle header, inserted by the bundle layer, for use by the bundle protocol
- The bundle protocol is a non-conversational protocol – bundle layers communicate between themselves using simple sessions with minimal or no round-trips

Bundles: A Store and Forward Overlay
The "Thin Waist" of a Delay-Tolerant Internet



DTN Nodes

- A node is an entity with a bundle layer. Three types of nodes:
  - Host – sends or receives bundles, but does not forward them. The bundle layers of hosts that operate over long-delay links require persistent storage in which to queue bundles until outbound links are available. May support custody transfers.
  - Router – forwards bundles within a single DTN region, and may optionally be a host. Routers that operate over long-delay links require persistent storage. May support custody transfers.
  - Gateway – forwards bundles between two or more DTN regions, and may optionally be a host. Must have persistent storage and support custody transfers. Provide conversions between the lower-layer protocols of the regions they span.

Custody transfers

- The bundle layer supports node-to-node retransmissions by means of custody transfers.
- These are arranged between bundle layers of successive nodes along the path from source to destination at the request of the source application.
- When the current custodian of a bundle sends it to the next node, it requests a custody transfer and starts a time-to-acknowledge timer.
- If the next-hop bundle layer accepts custody, it returns an acknowledgement to the sender.
- If no ACK is received before the timer expires, the sender retransmits the bundle.
- The value assigned to the timer can either be distributed to nodes with routine information or computed locally.

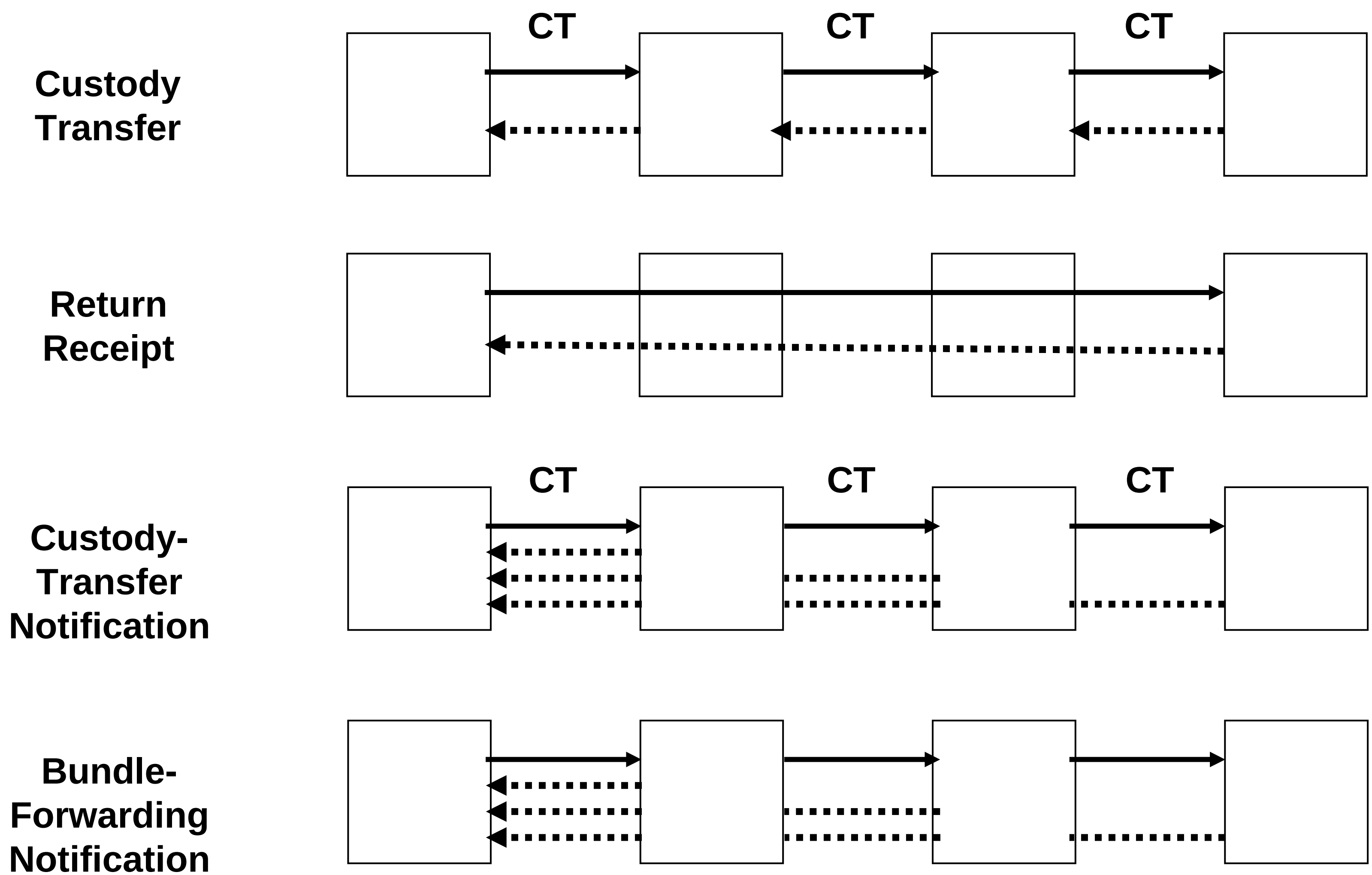
Custody transfers (2)

- A bundle custodian must store a bundle until
  - Another node accepts custody
  - The bundle's time-to-live expires; this is expected to be MUCH LONGER than a custodian's time-to-acknowledge
- Custody transfers do not provide guaranteed end-to-end reliability; this is only achieved if a source requests both custody transfer and return receipt.
- In such a case, the source must retain a copy of a bundle until receiving a return receipt from the destination; if it does not receive a return receipt, it initiates retransmission

Classes of Service

- The bundle layer provides six classes of service for a bundle:
  - Custody transfer – delegation of retransmission responsibility to an accepting node; the accepting node returns a custodial acceptance ACK to the previous custodian
  - Return receipt – confirmation to the source that the bundle has been received by the destination application
  - Custody-transfer notification – notification to the source when a node accepts a custody transfer of the bundle
  - Bundle-forwarding notification – notification to the source whenever the bundle is forwarded to another node
  - Priority of delivery – bulk, normal, or expedited
  - Authentication – the method, if any, used to verify the sender's identity and the integrity of the message

Classes of Bundle Service (2)



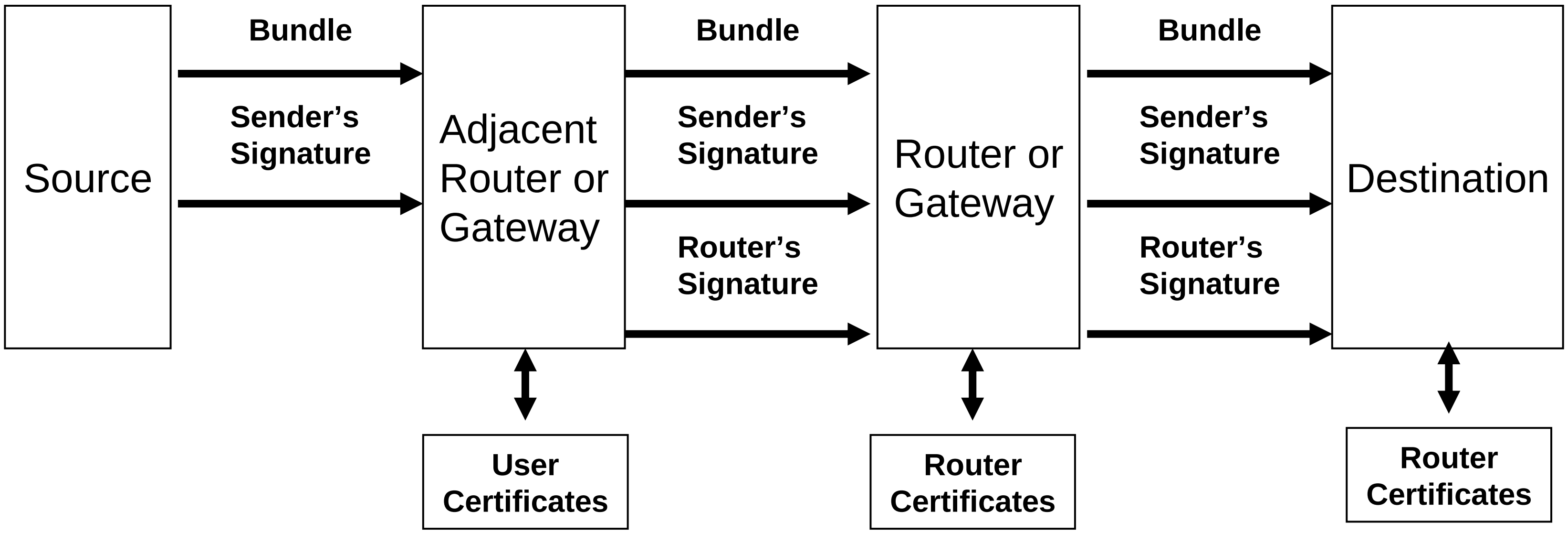
DTN Regions, Naming and Addressing

- A DTN is a network of networks
- Each network is a region in which communication characteristics are homogeneous
- Each region has a unique region ID which is knowable among all regions of the DTN, and is part of each node's name
- DTN gateways are members of two or more regions, and are the only means of moving messages between regions
- Each DTN node has a 2-part name, consisting of a region ID and an entity ID.
- Routing between regions is based only on region IDs; routing with regions is based only on entity IDs
- E.g. {earth.sol.int, src.dcs.gla.ac.uk:1234}

Security

- Forwarding nodes (routers and gateways) are authenticated, and sender information is authenticated by forwarding nodes – in this way, network resources can be conserved by preventing the carriage of prohibited traffic at the earliest opportunity
- Assuming public-key crypto, each user has a private and public key pair and a number of certificates. Each forwarding node has a key-pair and certificates. The steps for securing the transmission of a bundle is as follows:
  - The source sends its bundle, signed using its private key; if the receiving node does not already have a copy of the sender's certificate, it obtains one from the sender or a certificate authority
  - The first forwarding node that receives the sender's bundle verifies the sender's identity and CoS rights, using its stored copies of adjacent-user certificates and CA public keys. The node than adds its own signature and forwards onwards
  - Each subsequent forwarding node verifies only the identity of the previous forwarding node, using its stored copies of adjacent-router certificates; then it replaces the prior node's signature with its own.

Security (2)


Source
Bundle
Sender's Signature
Adjacent Router or Gateway
Bundle
Sender's Signature
Router's Signature
Router or Gateway
Bundle
Sender's Signature
Router's Signature
Destination
User Certificates
Router Certificates
Router Certificates

Summary

- Violating the assumptions behind the TCP/IP stack, especially with regards to intermittent or long-delay connectivity, requires a paradigm shift – i.e. store-and-forward message switching instead of end-to-end transport of application data
- Thin waist shifts from IP to Bundle Layer
- Compound naming structure required for bundle layer to work.
- Support for various levels of responsibility delegation required to enable different styles of interaction with different levels of guarantee
- Security now involves vetting the infrastructure in addition to vetting the source and destination entities