Advanced Lecture on Internet Infrastructure

13. End to End Theorem in Distributed Information System Design

Masataka Ohta

mohta@necom830.hpcl.titech.ac.jp

ftp://chacha.hpcl.titech.ac.jp/infra13e.ppt

End to End Argument in the Original Paper (End to End Argument in System Design)

- The function in question can completely and correctly be implemented only with the knowledge and help of the application standing at the end points of the communication system. Therefore, providing that questioned function as a feature of the communication system itself is not possible. (Sometimes an incomplete version of the function provided by the communication system may be useful as a performance enhancement.)

The function?

- examples in the original paper
  - careful file transfer
  - encryption
  - duplicate message detection
  - message sequencing
  - guaranteed message delivery
  - detecting host crashes
  - delivery receipts
- information passing between sender and receiver
  - “function” is to know or change state of peer

the knowledge and help of the application standing at the end points of the communication system?

- what is “application”? how is layering?
  - In a system that includes communications, one usually draws a modular boundary around the communication subsystem and defines a firm interface between it and the rest of the system.

  only one interface (without layer) between sender/receiver and communication system and is firm
- layering treated skeptically
  - It is fashionable these days to talk about "layered" communication protocols
    - OSI layering is terrible

a firm interface between it and the rest of the system?

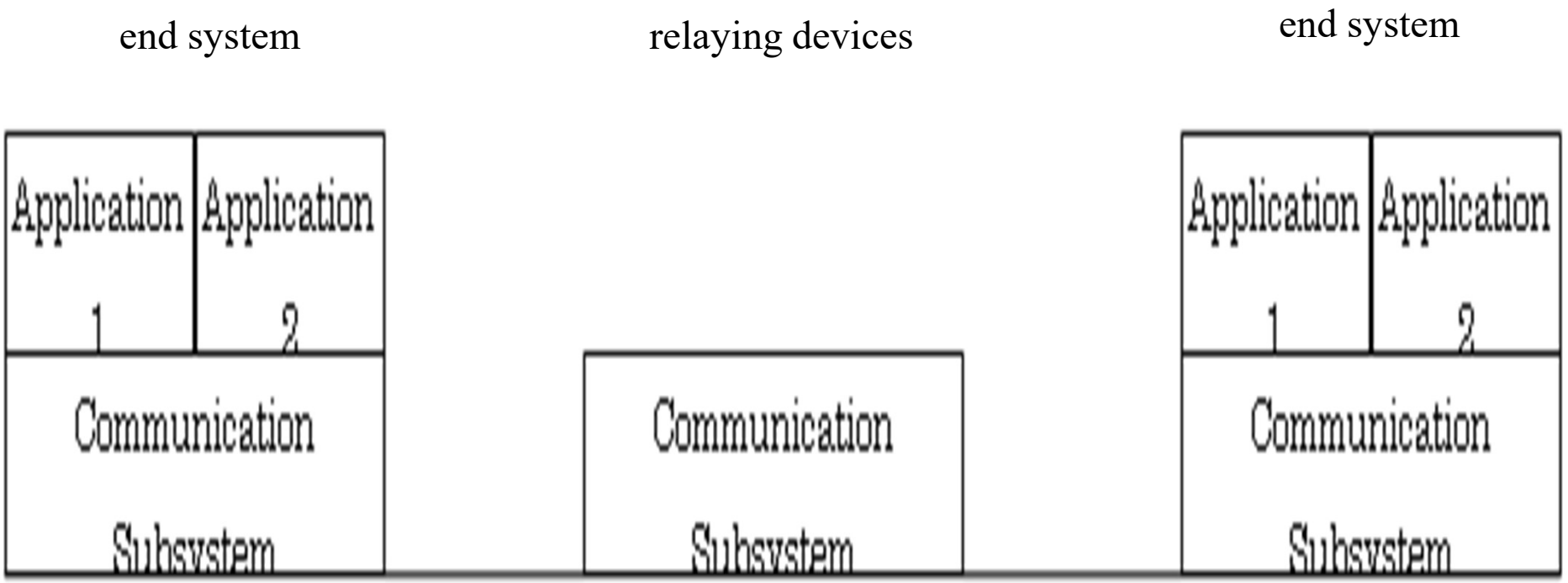
- interface is not flexible with little options
  - First, since the lower level subsystem is common to many applications, those applications that do not need the function will pay for it anyway. Second, the low-level subsystem may not have as much information as the higher levels, so it cannot do the job as efficiently

  that is, functions at lower level must be used by all the applications
- if "the low-level subsystem" "have as much information", the function may be provided completely and correctly
  - e. g. reliable byte stream by TCP

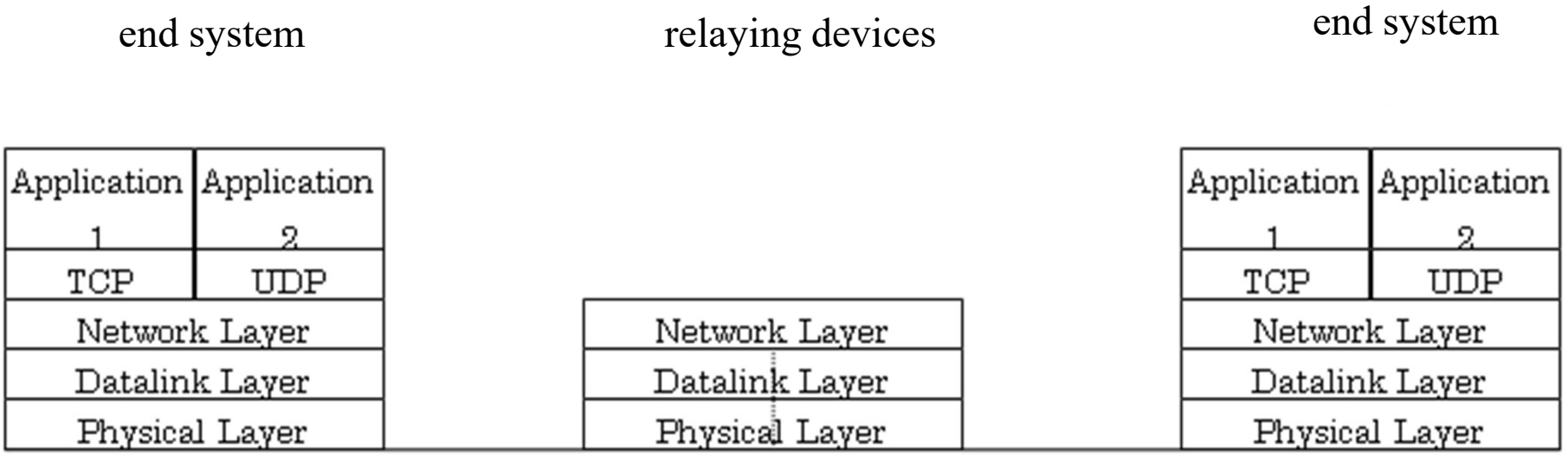
the knowledge and help of the application standing at the end points of the communication system?

- argument, as is, not applicable protocols used inside of “communication system”
  - extension to distributed system, in general, necessary
  - in the paper
    - We begin by considering the communication network version of the argument.
    - In a broader context the argument seems to apply to many other functions of a computer operating system, including its file system. Examination of this broader context will be easier if we first consider the more specific data communication context, however.

  general argument is avoided because it may be difficult, but, actually, easy



layering structure assumed by the original paper



modern layering structure of the Internet

End to End Theorem

- in distributed information system, a subsystem can know or change state of other subsystems only with knowledge and help of other subsystems
- function of a subsystem can not be provided completely or correctly by other subsystems without knowledge of the subsystem
- (Sometimes an incomplete version of the function provided by other subsystem may be useful as a performance enhancement)

End to End Theorem with Layering

- in distributed information system, a subsystem can know or change state of certain layer of other subsystems only with knowledge and help of the layer of other subsystems。

Proof of the End to End Theorem?

- in distributed information system, a subsystem can know or change state of other subsystems only with knowledge and help of other subsystems

- should be obvious
  - function of a subsystem can not be provided completely or correctly by other subsystems without knowledge of the subsystem

should also be obvious, if communication involves errors

Proof (or, definition of “distributed information system”? (1)

- distributed information system is a set of subsystems connected by communication links
  - just define elements of distributed information system
- each subsystem can know and change its state but can not directly know or change state of other subsystems
  - because they are “distributed”, it is obvious not “directly”
  - first half of the theorem is directly deduced

Proof (or, definition of “distributed information system”? (2)

- each subsystem can exchange information with other subsystems through communication links adjacent to it and can know local state of the communication links
  - define what are “communication links”
- each subsystem and each communication links suffer from noise and failures not to be able to send information or send false information
  - last half of the theorem is deduced

Warp-up

- by intensive reading on the original paper of the end to end argument
  - relation ships with layering are clarified
  - is proved as a therorem extended for general distributed information systems
- the end to end theorem is directly applicable to
  - arguments on layered protocols
  - arguments on protocols within networks

Quantization of “completely and correctly”

- infinite completeness and correctness needs infinite amount of communication
- the original paper (qualitative)
  - a checksum that has sufficient redundancy to reduce the chance of an undetected error in the file to an acceptably negligible value.
- quantization of error rate: quality improves with repeated retries as
  - (end to end error rate of communication)^(time/RTT)
  - recent most information (w.r.t. RTT) should be obtained or retry should be attempted

Application of the End to End Theorem to PMTUD

- difficulties of PMTUD
  - how can we get complete and correct MTU?
    - how frequently should we check MTU increased
- RTT!
  - or, if transmission rate is less frequent, maybe less
  - should be implemented at transport layer
  - very frequent
    - high load with IPv6 style design to generate ICMPv6
    - should have an IP header field to record PMTU

Application of the End to End Theorem to Unicast IGP RP

- “function” is computation of to destinations
- function of a subsystem can not be provided completely or correctly by other subsystems without knowledge of the subsystem
  - DV style computation by network is no good, with LS style, route will converge within several RTT
    - minimum HELO interval of OSPF (1s) too large
      - should be as short as link RTT
      - unless bandwidth consumed by HELO is too much (<0.1%?)