Recursion in Networking

Joe Touch

Postel Center Director

USC/ISI

Research Assoc. Prof.

USC CS and EE/Systems Depts.

Information Sciences Institute

Outline

- Recursion is key
- Background on X-Bone VNs
- RNA
  - Intro.
  - Design
  - Related work
- Implications of Recursion

Information Sciences Institute

Recursion is key

Information Sciences Institute

What makes an architecture new?

- "Shaking the Hourglass"
  - All exchanges are 1 packet
  - Collosograms > RTT*delay
  - No LANs? (L2 is only pt-pt)
- What defines success?
  - Fixing what's 'broken'
  - Doing something new/different
  - The Internet / circuits as a degenerate case

Information Sciences Institute

Internet Architecture

Accused of ossification, but:

- Ossification = stability
- Flexibility is abundant:
  - Shim layers:
    - HIP, SHIM6, IPsec, TLS
  - Muxing layers:
    - SCTP, RDDP, BEEP
  - Connections:
    - MPLS, GRE, IKE, BEEP, SCTP
  - Virtualization:
    - L2VPN, L3VPN/X-Bone/RON/Detour, L7-DHTs

Information Sciences Institute

Motivation

- Desire to support new capabilities
  - Interlayer cooperation, dynamic layer selection, layering created by virtualization
- Desire to support emerging abstractions
  - Overlay layers don't map to 1-7
  - Support for recursive nodes (BARP, LISP, TRILL)
- Desire to coordinate services in diff. places
  - Security, soft-state, pacing, retransmission

Shannon Channel

- Two preselected parties
  - Homogenous endpoints

- Unidirectional channel
  - Preselected sender, preselected receiver

Information Sciences Institute

What is communication?

- Shannon: shared bits
  - Between fixed endpoints, known *a priori*

- Shared bits between two parties
  - How do we find the party to talk to?

Information Sciences Institute

What SCs Ignore

- What if you're not directly connected?
  - A) multihop
  - B) multilayer
- Why are multihop/multilayer interesting?
  - Scalable = multihop
  - Ubiquitous = multilayer
  - I.e., all scalable, ubiquitous comms!

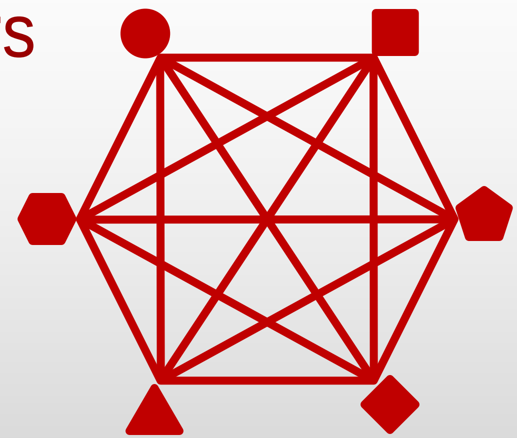
Information Sciences Institute

Observations

- Networking is *groups of interacting parties*
  - Groups are heterogeneous
  - All members want to interact
  - Groups can be dynamic (*i.e.*, virtual)
- Need an architecture that supports:
  - Heterogeneity
  - Interaction
  - Virtualization

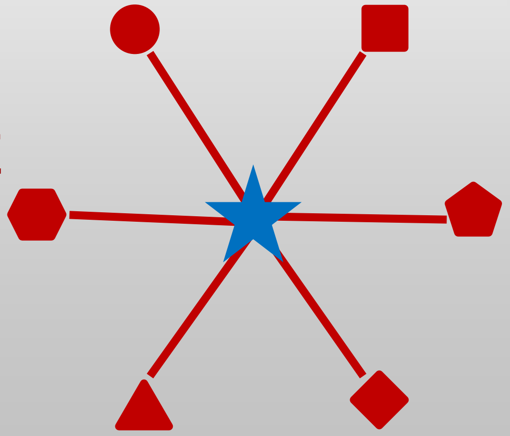
Heterogeneity leads to layering

- M different interacting parties need
  - M^2 translators



or

  - M translators + common format

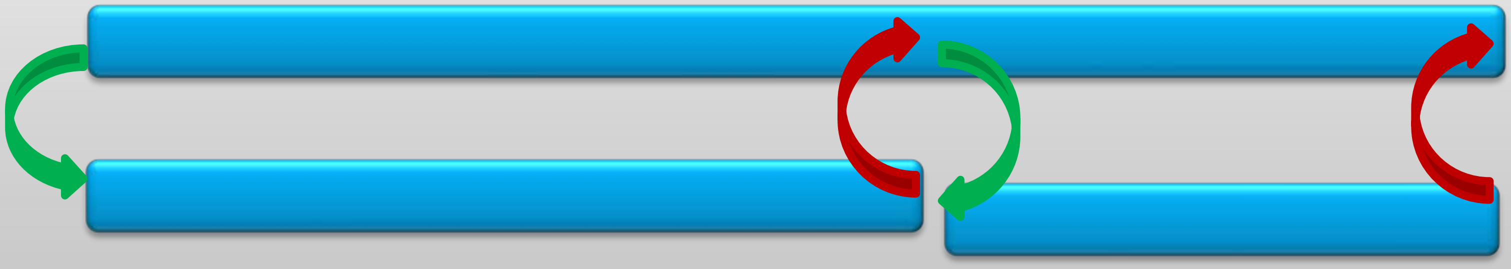


... *i.e.,* a layer

Information Sciences Institute

Layering leads to resolution

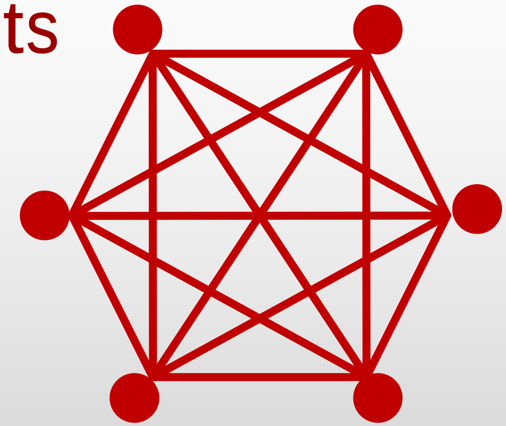
- IDs are local to a layer
  - Whether names, paths, locations
- Need to resolve IDs between layers
  - Google, DNS, ARP, LISP encap tables



Information Sciences Institute

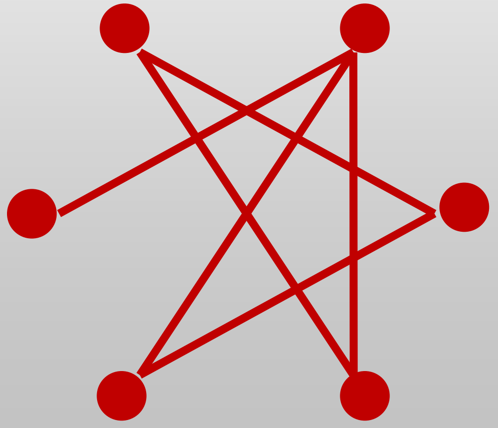
Interaction leads to forwarding

- N parties need
  - N^2 circuits



or

  - O(N) links + forwarding



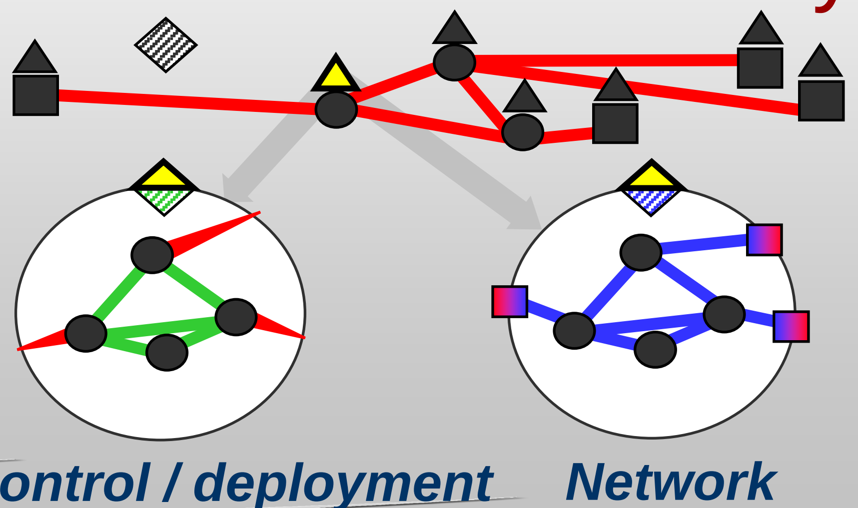
Information Sciences Institute

Virtualization leads to recursion

- N parties want to group in arbitrary, dynamic ways.

  ... such groups are inherently virtual

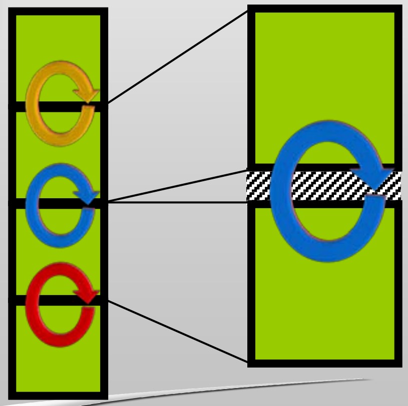
... and virtualization is inherently recursive

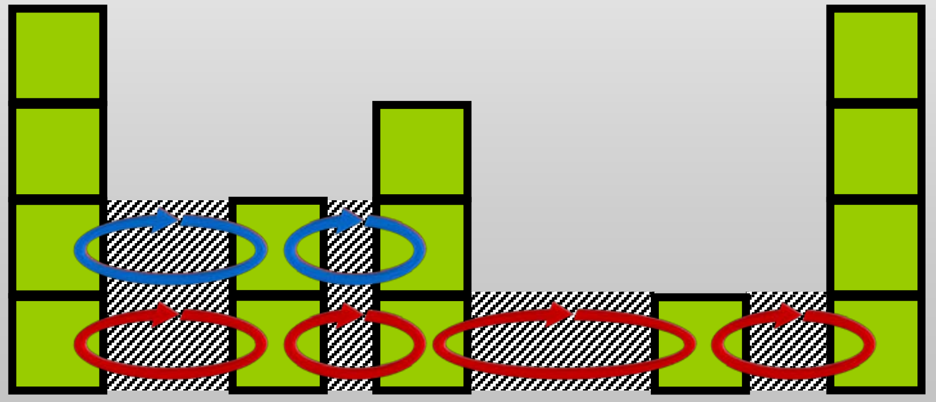


Information Sciences Institute

Recursion unifies layering, forwarding, & resolution

- Layering (left)
  - Heterogeneity via O(N) translators
  - *Supported by successive recursive <u>resolution</u>*
- Forwarding (right)
  - N^2 connectivity via O(N) links
  - *Supported by successive iterative <u>resolution</u> (tail recursion)*





Information Sciences Institute

What makes this an architecture?

- Abstraction for virtualization
  - Tunnel as link
  - Partitioned router as virtual router
  - Partitioned host + internal router as virtual host
- Abstractions for recursion
  - Recursive router implemented as a network of vrouters with vhosts at the router interfaces
  - Recursion within the protocol stack
- General templates (metaprotocol, ID tree)
  - Instantiates as different layers or forwarding

Information Sciences Institute

X-Bone Virtual Nets

Information Sciences Institute

Virtual Net Req'ts

- **Internet-Compliant Architecture**
  - Hosts add/delete headers
  - Routers transit (constant # headers)
- **Supports New Capabilities**
  - Concurrence (multiprocessing)
  - Revisitation (multiple roles in one net)
  - Recursion (to hide topology and/or mgt.)

Information Sciences Institute

VN Principles

- *TENET 1. Internet-like*
  - VIs = VRs + VHs + tunnels
  - Emulating the Internet
- *TENET 2. All-Virtual*
  - Decoupled from their base network
- *TENET 3. Recursion-as-router*
  - Some of VRs are VI networks

Information Sciences Institute

VN Corollaries

- Behavior:
  - VH adds/deletes headers
  - VRs transit (constant # headers)
- Structure:
  - VIs support concurrence
  - VIs support revisitation
- Each VI has its own names, addresses
  - Address indicates overlay context

Information Sciences Institute

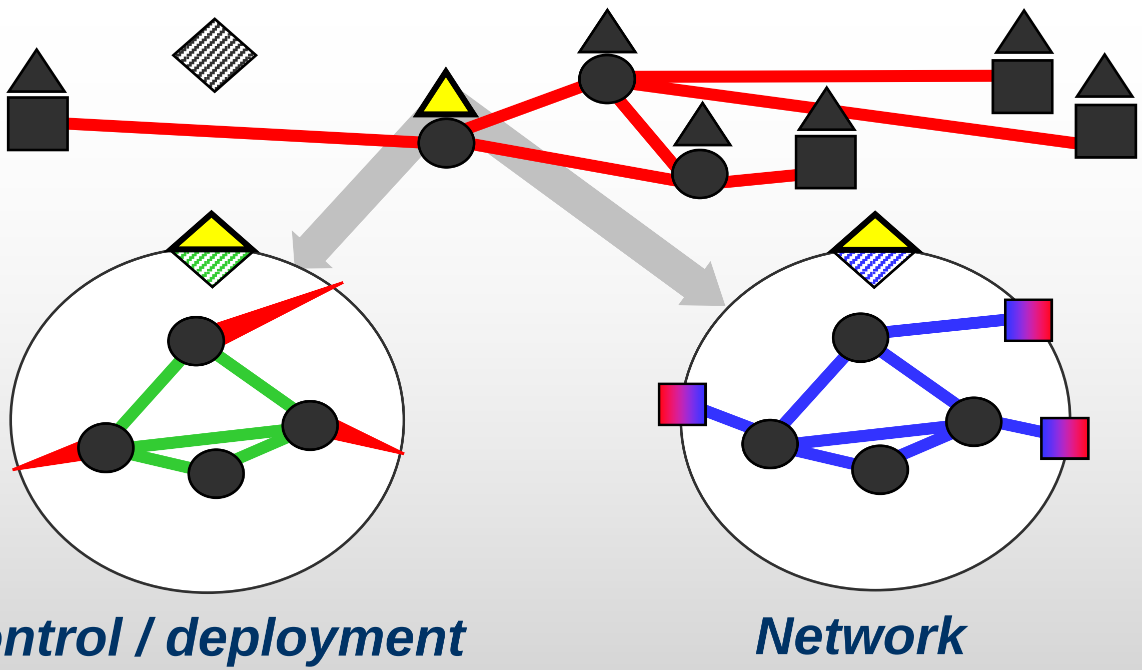
VN Architecture

- Components:
  - VH -> hosts include a hidden router
  - VL -> 2 layers of encaps. (strong link, weak net)
  - VR -> partitioned forwarding
- Capabilities:
  - Revisitation -> multihoming for VNs
  - Recursion -> router as network, i.e., Rbridges, LISP

>> RUNNING CODE (FreeBSD, Linux, Cisco)

Information Sciences Institute

Recursive Internet

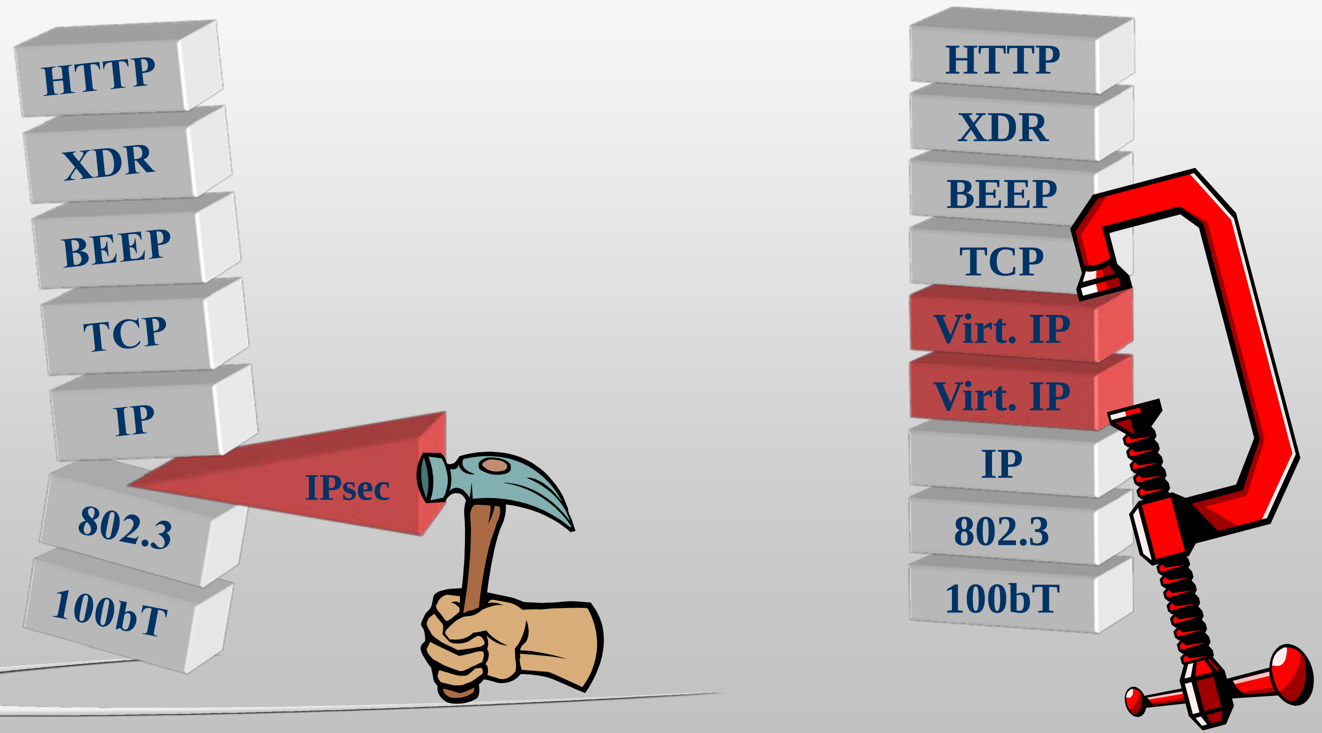


- Recursion as a router (vs. ASes)
- Network recursion examples
  - L3 = BARP (X-Bone), LISP (IRTF)
  - L2 = Rbridges/TRILL

Information Sciences Institute

Recursion requires new layers – where? Why?

- Wedge between (IPsec, left) or replicate (virtualization, right)



Information Sciences Institute

Challenges of Layering

- Which to add...
  - IPv4/IPv6, TCP/DCCP/SCTP
- When to add...
  - Security, muxing, cong. control
- Real vs. virtual
  - What’s the difference?

Information Sciences Institute

RNA Intro.

Information Sciences Institute

Motivation for RNA

- Layers of a stack becoming more similar
  - Security, soft-state, pacing, retransmission
- Desire to support new capabilities
  - Interlayer cooperation, dynamic layer selection
- Desire to support emerging abstractions
  - Overlay layers don't map to 1-7
  - Support for recursive nodes (BARP, LISP, TRILL)

Is layering more than a coding artifact?

Information Sciences Institute

One module to reuse

- "Resolve" unifies:
  - Layer address translate/resolution
    - ARP, IP forwarding lookup
    - BARP/LISP/TRILL lookup
  - Layer alternates selection
    - IPv4/IPv6, TCP/SCTP/DCCP/UDP
  - Iterative forwarding
    - IP hop-by-hop, DNS recursive queries
- "Process data" unifies:
  - Shared state, security, management
  - Flow control, error control

```
LAYER(DATA, SRC, DST)
   Process DATA, SRC, DST into MSG
   WHILE (Here <> DST)
      IF (exists(lower layer))
         Select a lower layer
         Resolve SRC/DST to next layer S',D'
         LAYER(MSG, S', D')
      ELSE
         FAIL /* can't find destination */
      ENDIF
   ENDWHILE
   /* message arrives here */
   RETURN {up the current stack}
```

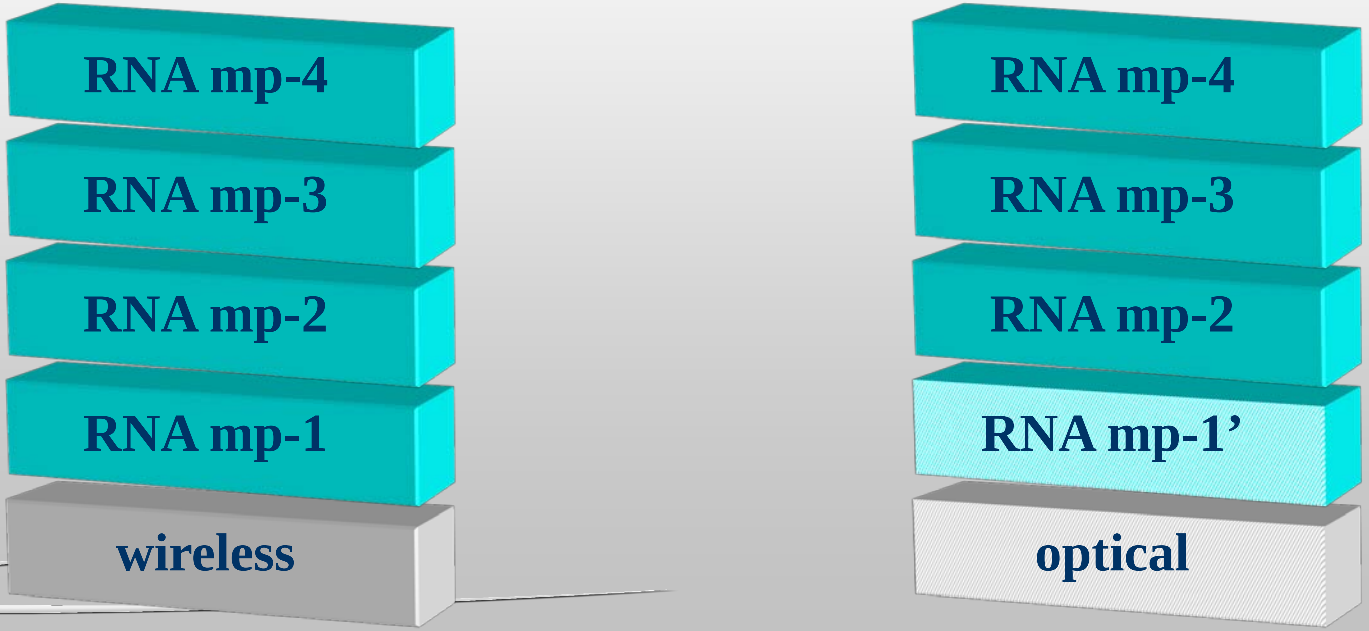
Next-hop Resolution

Next Layer Resolution

Information Sciences Institute

RNA Stack

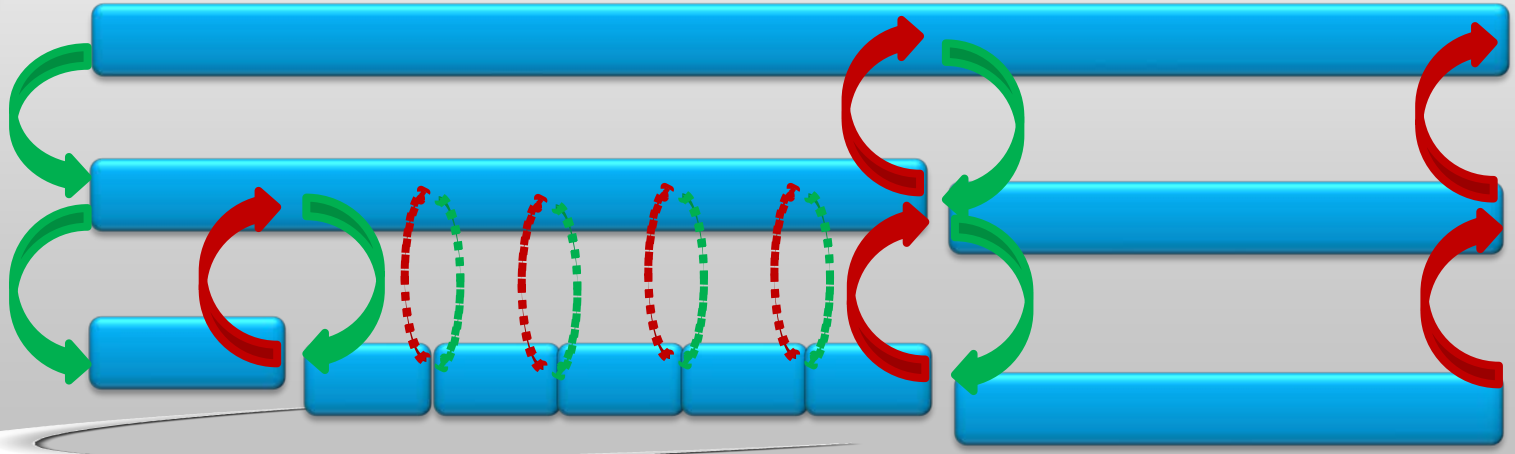
- One MP, many instances
  - Needed layers, with needed services
  - Layers limit scope, enable context sensitivity
  - Scope defined by reach, layer above, layer below



Information Sciences Institute

Retain layering

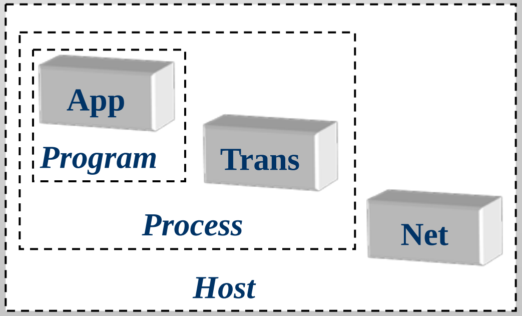
- One metaprotocol, many instances
  - Needed layers, with needed services
  - Layers limit scope, enable context sensitivity
  - Scope defined by reach, layer above, layer below
  - Resolution connects the layers (red/green)



Information Sciences Institute

Scope defines a layer

- Its endpoints
  - A “hop” @layer N = E2E extent of layer N-1
- The layer above
  - What services this layer provides
- The layer below
  - What services this layer requires
- E.g.: Shared state at diff. layers for diff. services
  - Application binding
  - Transport delivery
  - Net security

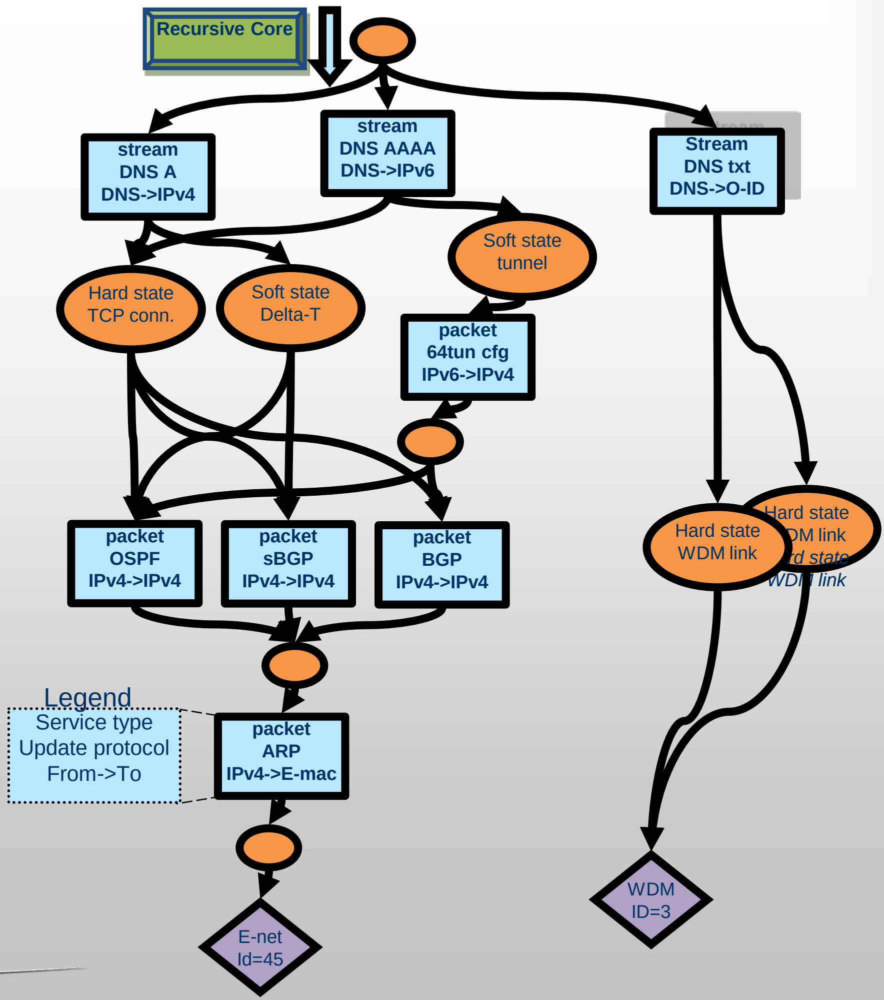


The difference is scope

Information Sciences Institute

IDs constrain structure

- Tree of ID spaces
  - Link at resolvers
- State inbetween
  - Connections, provisioning
- Table management
  - ID use coordination
  - Routing
  - Resolution



Information Sciences Institute

What makes this an architecture?

- Basic components
  - Metaprotocol + MDCM, ID space tree, etc.
  - Instantiates as different layers or forwarding
- Abstraction for virtualization
  - Tunnel as link
  - Partitioned router as virtual router
  - Partitioned host + internal router as virtual host
- Abstraction for recursion
  - Recursive router implemented as a network of vrouters with vhosts at the router interfaces

Information Sciences Institute

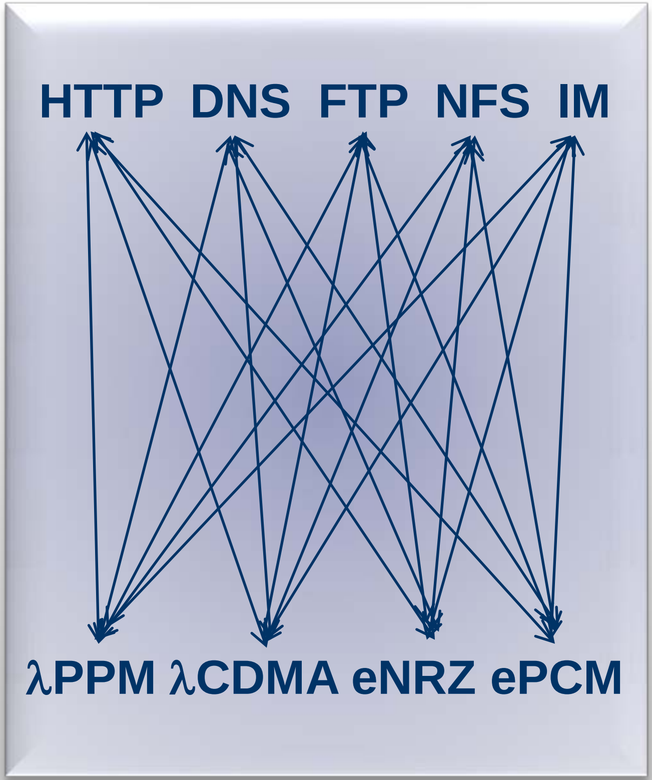
What does RNA enable?

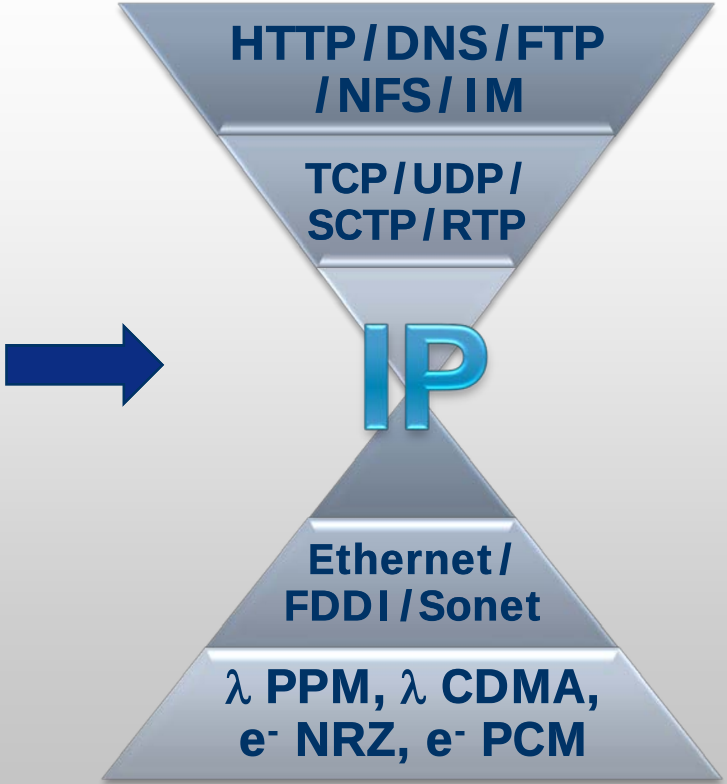
- Integrate current architecture
  - ‘stack’ (IP, TCP) *vs.* ‘glue’ (ARP, DNS)
- Support needed improvements
  - Recursion (AS-level LISP, L3 BARP, L2 TRILL)
  - Revisitation
- Supports “old horses” natively
  - Dynamic ‘dual-stack’ (or more)

Information Sciences Institute

The Hourglass Principle

- Common interchange format between layers





Information Sciences Institute

Multiple hourglasses

- "Waist" is relative
  - The common interchange = the waist

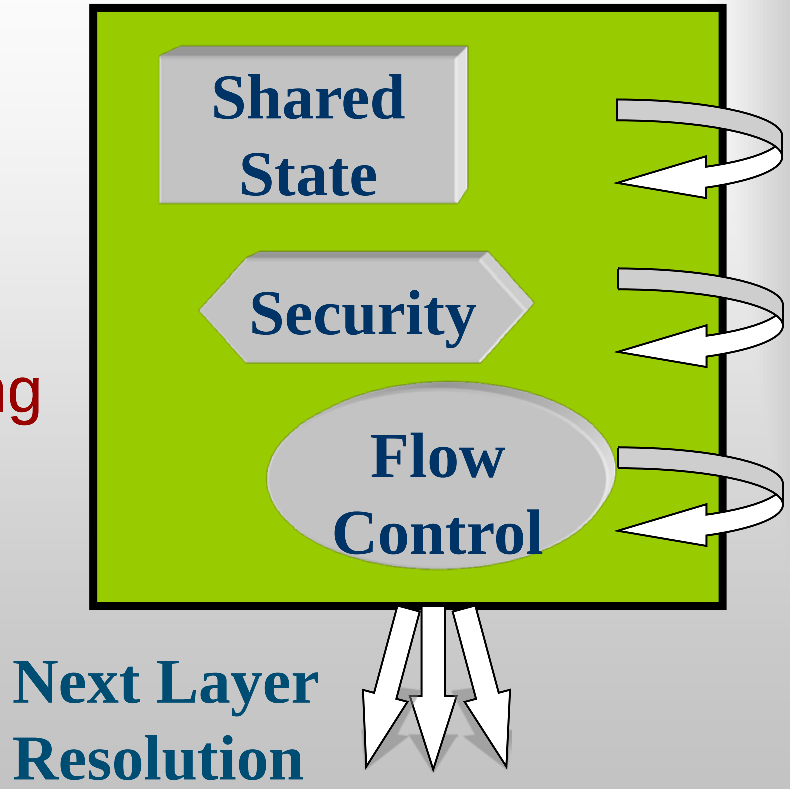
Information Sciences Institute

RNA Design

Information Sciences Institute

RNA Metaprotocol

- Template of basic protocol service:
  - Establish / refresh state
  - Encrypt / decrypt message
  - Apply filtering
  - Pace output via flow control
  - Pace input to allow reordering
  - Multiplex/demultiplex
    - includes switching/forwarding



Information Sciences Institute

Components of RNA MP

Instantiate MDCM's "Process DATA"

- Establish / refresh state
- Encrypt / decrypt message
- Apply filtering
- Pace output via flow control
- Pace input to allow reordering
- Multiplex/demultiplex as indicated
  - includes switching/forwarding

RNA MP Template

```
LAYER(DATA, SRC, DST)
   Process DATA, SRC, DST into MSG
   WHILE (Here <> DST)
      IF (exists(lower layer))
         Select a lower layer
         Resolve SRC/DST to next layer S',D'
         LAYER(MSG, S', D')
      ELSE
         FAIL /* can't find destination */
      ENDIF
   ENDWHILE
   /* message arrives here */
   RETURN {up the current stack}
```

Next-hop Resolution

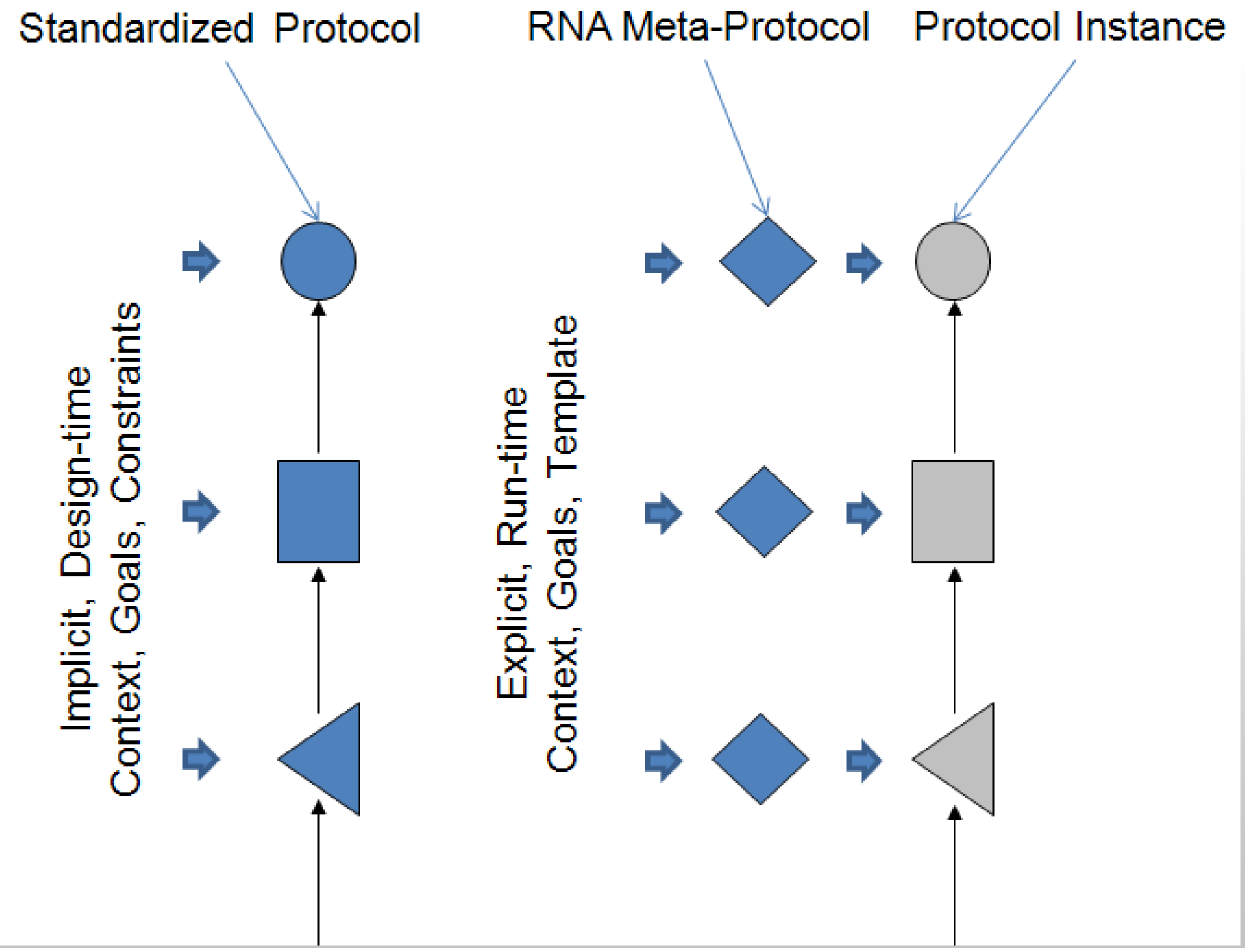
Next Layer Resolution

```
START PATTERN MIN

# This simply specifies a buffer. no reodering etc.
PATTERN MIN
   REQ MUST BUFFER 1
   ARG BUFFER 1 VAR size 1000
   LINK ADD SELF 0 BUFFER 1
   ...
# Next use this pattern if MIN is successful
PATTERN ORDERED_DELIVERY
   FOLLOWS MIN
   REQ MUST REORDERING 1
   LINK DEL ….
   LINK ADD ….
…
# If reordering successful, try more stuff…
PATTERN ENCRYPTED_ORDERED_DELIVERY
   FOLLOWS ORDERED_DELIVERY
   REQ MUST ENCRYPTION 1
   ARG ENCRYPTION 1 VAR algo des
   ARG ENCRYPTION 1 VAR keysize 512
   ....
```

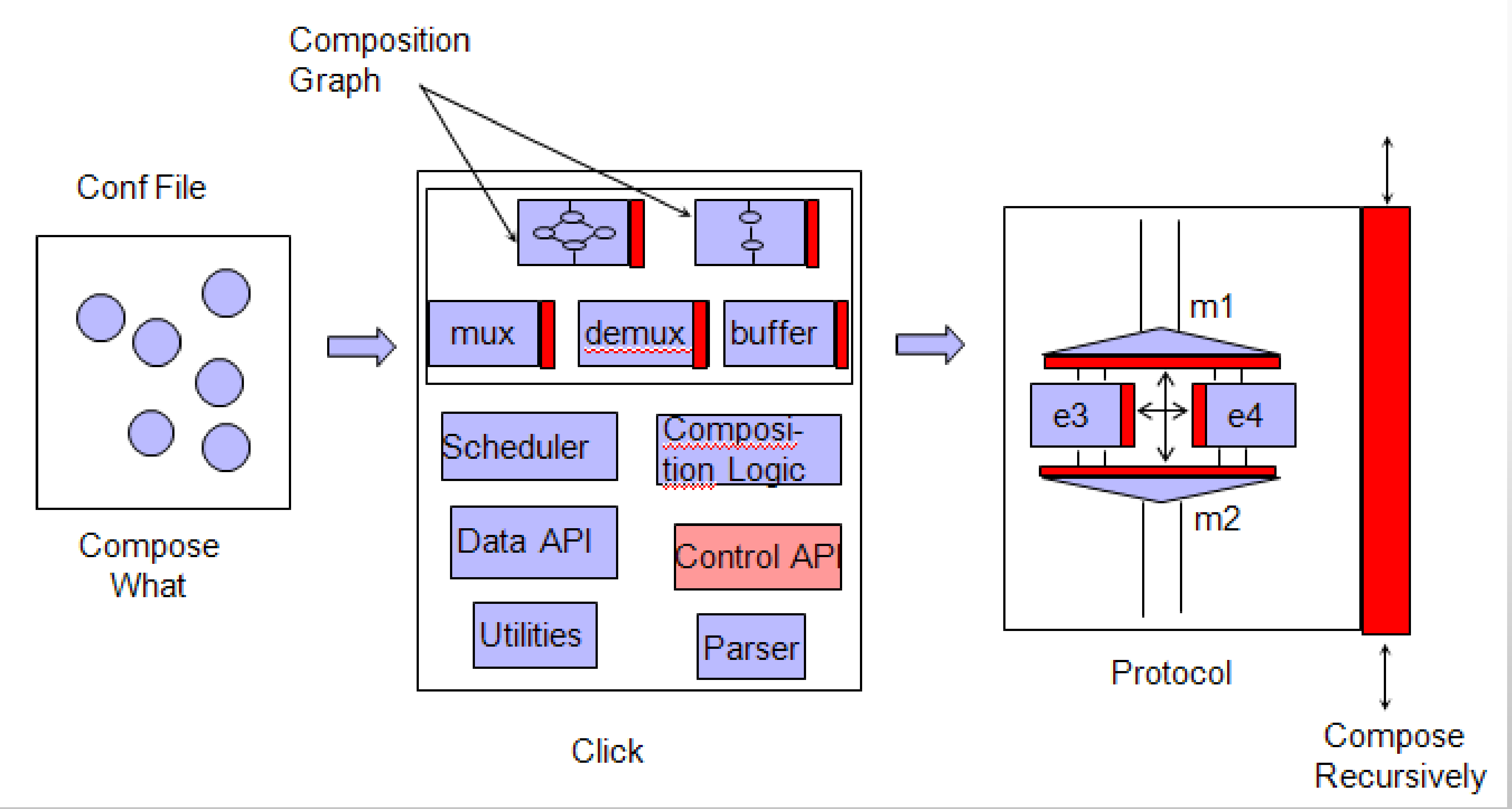
Information Sciences Institute

Instantiation



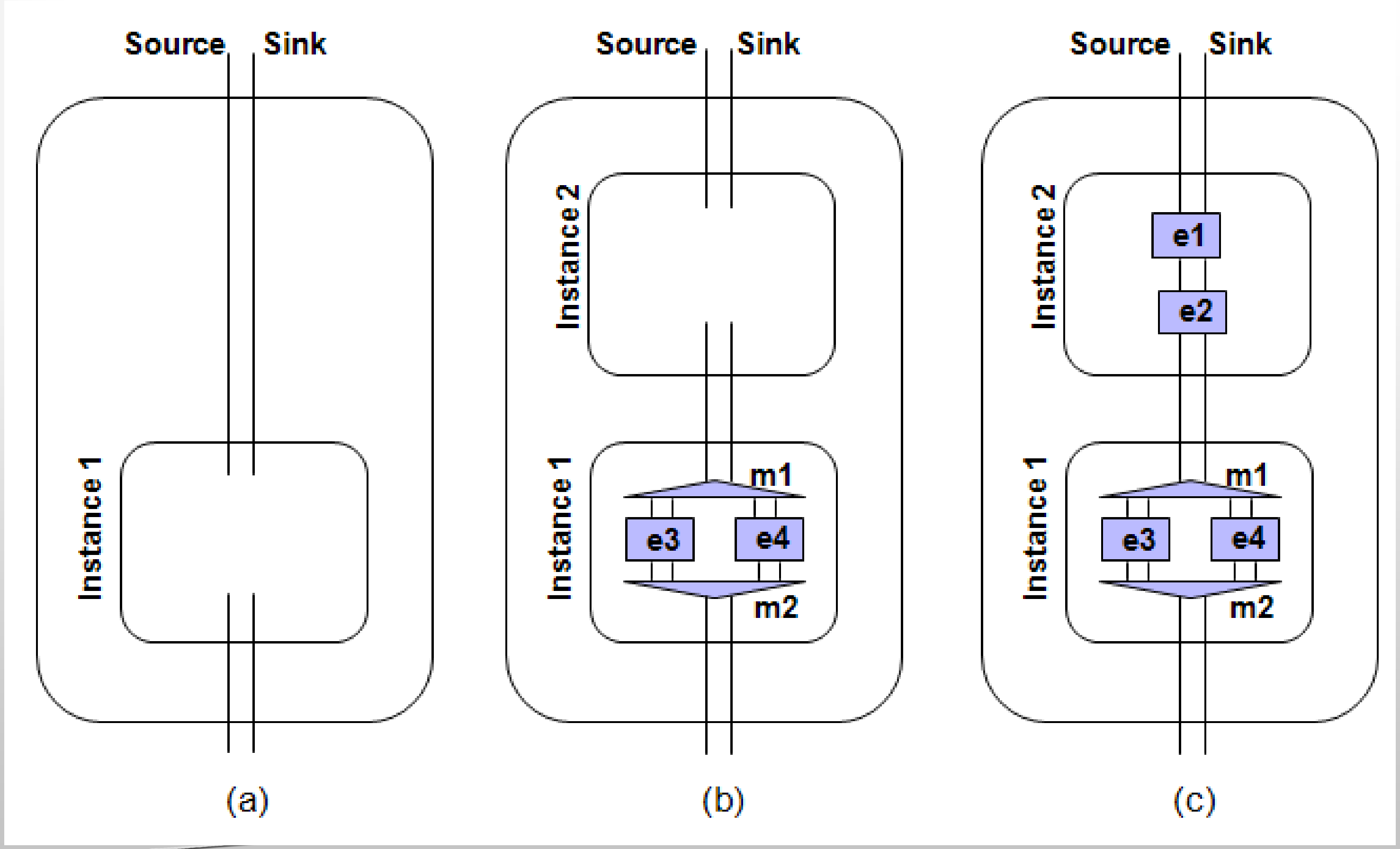
Information Sciences Institute

Click Implementation



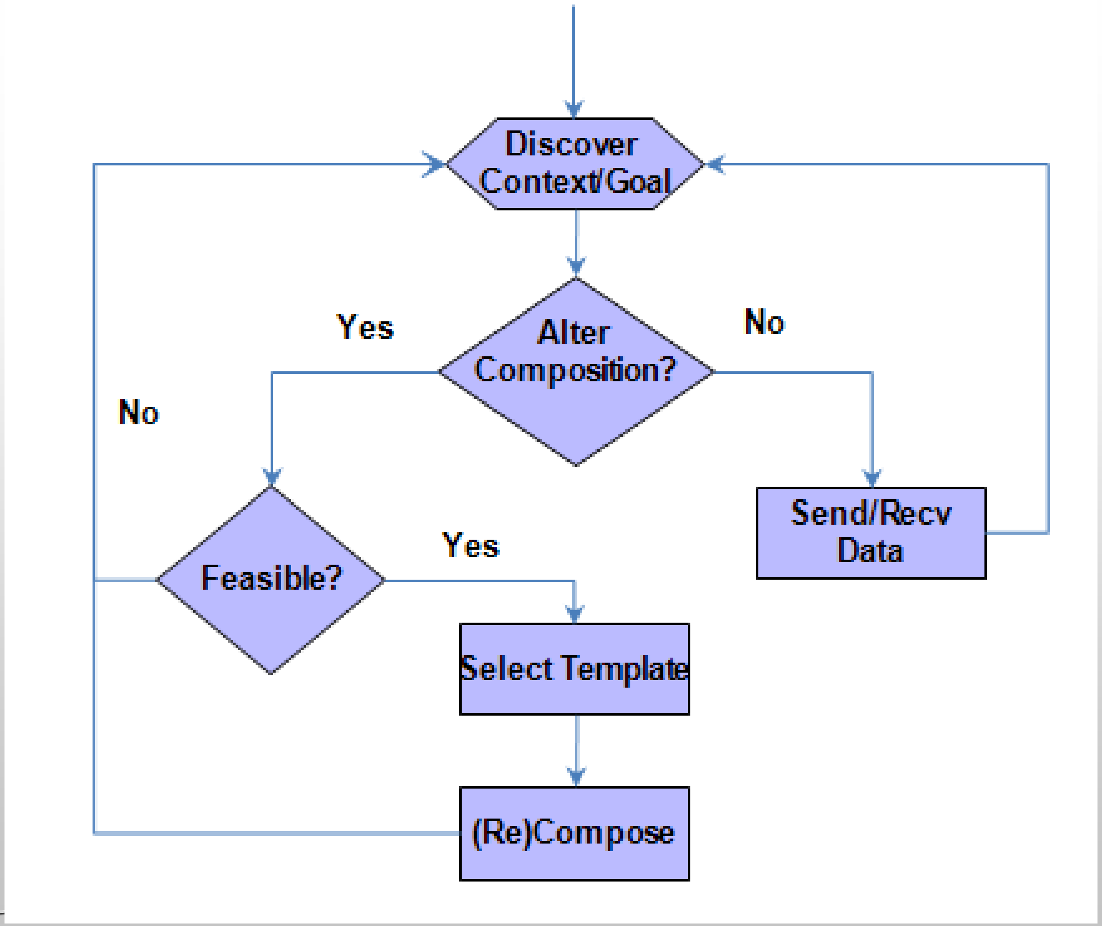
Information Sciences Institute

Building a Stack



Information Sciences Institute

Composition Process



Information Sciences Institute

Other Components

- MP design
  - What’s inside the “box”r
  - Interlayer coordination
  - Context sensitivity, environment tuning
- Dynamic negotiation protocol
  - Cross-layer negotiation, IETF TAE
- Composeable/recursive extensions
  - Network management/SLAs
  - Security (user/infrastructure)
  - Non-comm services (storage, computation)
- Integrated optimization
  - Caching, precompute/prefetch
  - Pinning, dampening

Information Sciences Institute

Related Work

Information Sciences Institute

Related Work Summary

- Recursion in networking
  - X-Bone/Virtual Nets, Spawning Nets, TRILL, Network IPC, LISP
  - *RNA natively includes resolution and discovery*
- Protocol environments
  - Modular systems: Click, x-Kernel, Netgraph, Flexible Stacks
  - Template models: RBA, MDCM
  - *RNA adds a constrained template with structured services*
- Context-sensitive components
  - PEPs, Shims, intermediate overlay layers, etc.
  - *RNA incorporates this into the stack directly*
- Configurable über-protocols
  - XTP, TP++, SCTP
  - *RNA makes every layer configurable, but keeps multiple layers.*

Information Sciences Institute

RNA and Network IPC

- Similarities
  - Recursive protocol stack
  - Unified communication mechanism
  - Focus on process-to-process interaction
- Differences
  - RNA uses MDCM to define IPC as combining a Shannon-style channel with namespace coordination
  - RNA provides a detailed (and demonstrated) mechanism that achieves unification and recursion
  - RNA supports both recursion and forwarding in a single mechanism

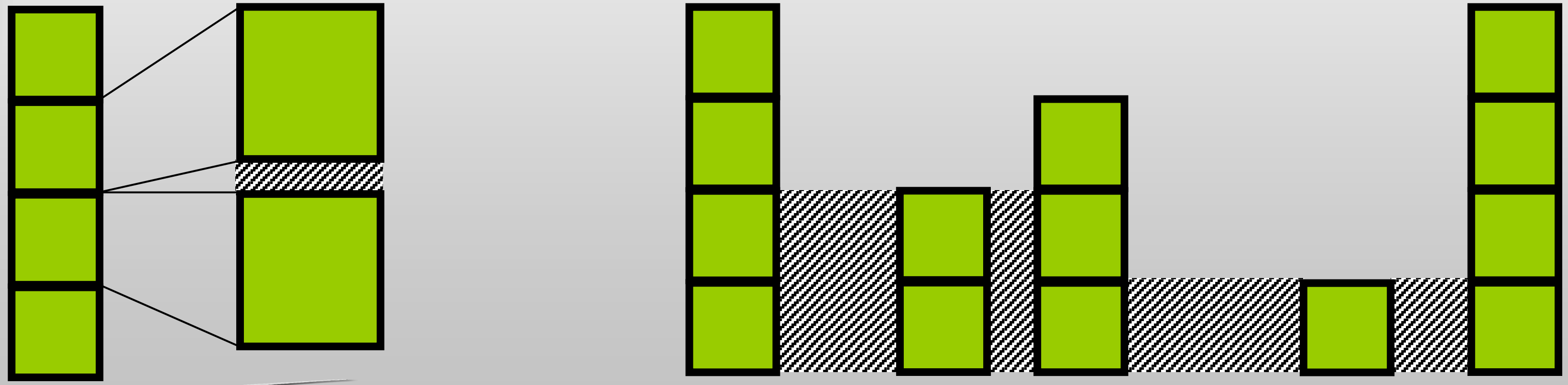
Information Sciences Institute

Implications of Recursion

Information Sciences Institute

Fills the gaps

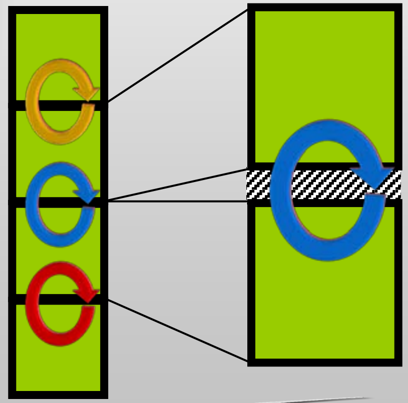
- Between layers (left, from *Choices*)
  - Affects next-layer
- Between stacks (right, from Padlipsky)
  - Affects next-hop

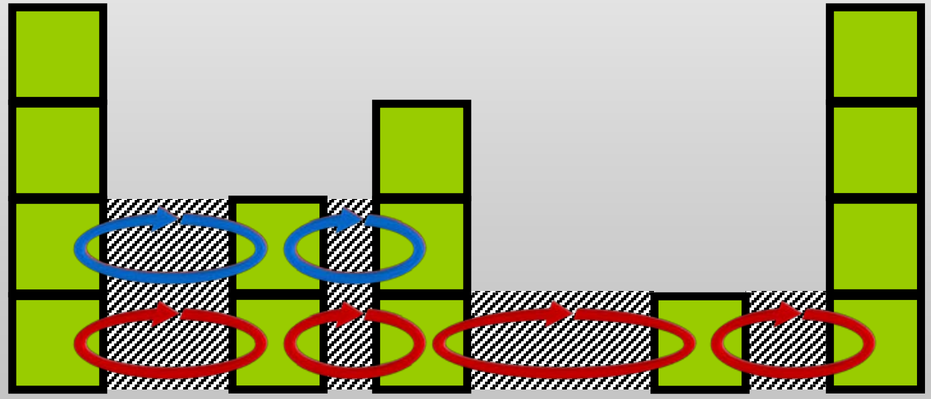


Information Sciences Institute

Integrates Layering and Forwarding

- Layering (left)
  - Heterogeneity via O(N) translators
  - *Requires successive recursive discovery*
- Forwarding (right)
  - N^2 connectivity via O(N) links
  - *Requires successive iterative discovery*





Information Sciences Institute

Uniquely enables…

- Integrates data, control, mgt, security
  - All are different ways of managing state inbetween resolutions
  - State can be shared – TCP RTT, NM liveness, BGP timers, etc. are all the same info.
- Integrates routing and resolution
  - Both are just ways to manage the tables
- Integrates provisioning and conn. mgt
  - Provisioning is at layer N is just a new connection at layer N-1

Information Sciences Institute

Summary

- Recursion is an integral part of networking
  - Falls out of multiparty communication
- Recursion is a native part of layering
  - Whether IP/ethernet, or LISP (IP/IP), or TRILL (ether/ether)
- Recursion allows us to keep layering
  - Layering is critical to constrain scope

Information Sciences Institute

Conclusions

- Virtualization requires recursion
- Recursion supports layering
- Recursion supports forwarding

One recurrence to bind them all…

- *Recursion is a native network property*
  - Integrates and virtualization, forwarding and layering **in a single mechanism**

Information Sciences Institute

Credits

- ID tree and related issues
  - Christos Papadopolous and Dan Massey CSU
- MDCM
  - Yu-Shun Wang
- RNA
  - Yu-Shun Wang, Venkata Pingali
- Naming unification
  - Venkata Pingali
- Virtual networking (X-Bone *et al.*)
  - Lars Eggert, Yu-Shun Wang, Greg Finn, Steve Hotz, Oscar Ardaiz-Villanueava, Norihito Fujita