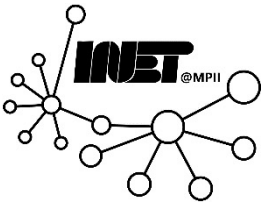
Networking at scale

Prof. Anja Feldmann, Ph.D.

Balakrishnan Chandrasekaran, Ph.D.



Designs for Scale

How to deal with large numbers (millions) of entities in a system?

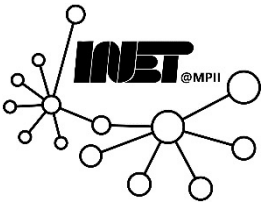
- IP devices in the Internet (*billions!!*)
- Users in P2P network (*millions*)

More generally …

- *Are there advantages to large scale?*

- *"For every type of animal there is a most convenient size, and a large change in size inevitably carries with it a change of form."*

  — On Being the Right Size, J. B. S. Haldane



Designs for Scale

How to deal with large numbers (millions) of entities in a system?

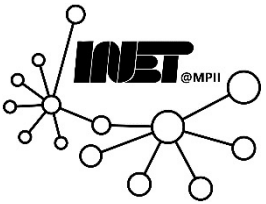
- IP devices in the Internet (*billions!!*)
- Users in P2P network (*millions*)

More generally …

- *Are there advantages to large scale?*
- *"For ev… …rge change in size …*

  — Or…

Is there a "right size" for networks?
What aspects determine this right size?



Dealing with scale: Hierarchical routing

Scale: > 500 million destinations

- *Cannot* store all destinations in routing tables!
- Routing table exchange would *swamp* links!!

Administrative autonomy

- Internet: Network of networks
- Each network admin may want to control routing in its own network

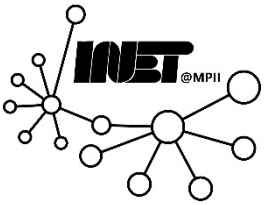
Hierarchical routing

Aggregate routers into regions, ***"autonomous systems" (AS)***

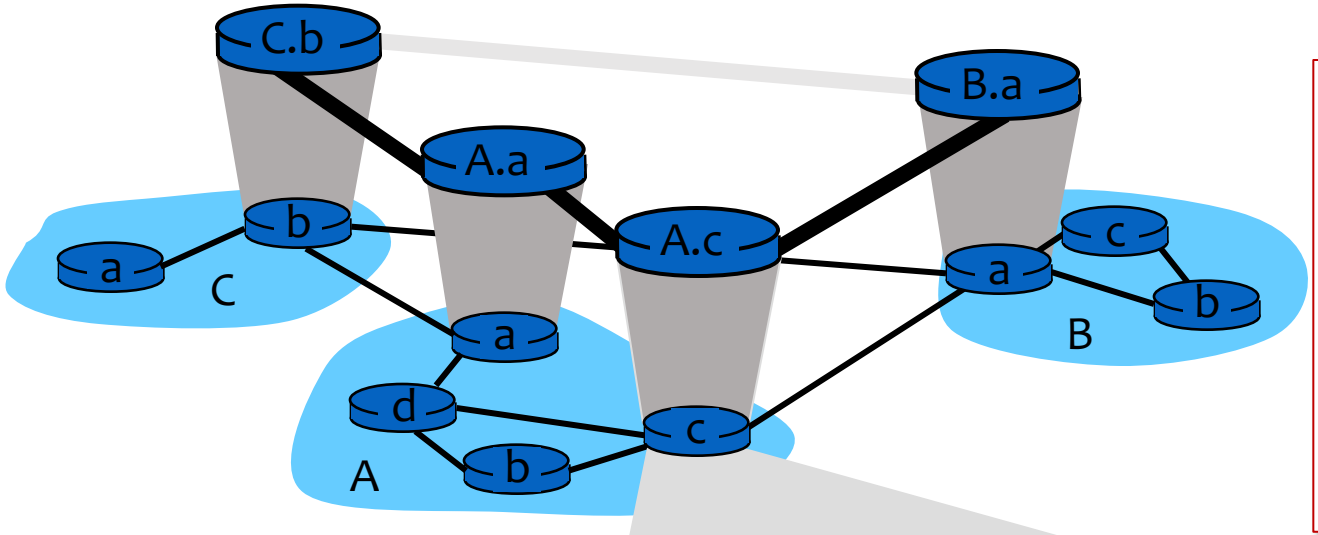
- Routers in same AS run same routing protocol
  - ***"Intra-AS"*** routing protocol
- Routers in different AS can run different intra-AS routing protocol

Gateway Routers

- Special routers in AS
- Run intra-AS routing protocol with all other routers in AS
- Also responsible for routing to destinations outside AS
  - Run inter-AS routing protocol with other gateway routers



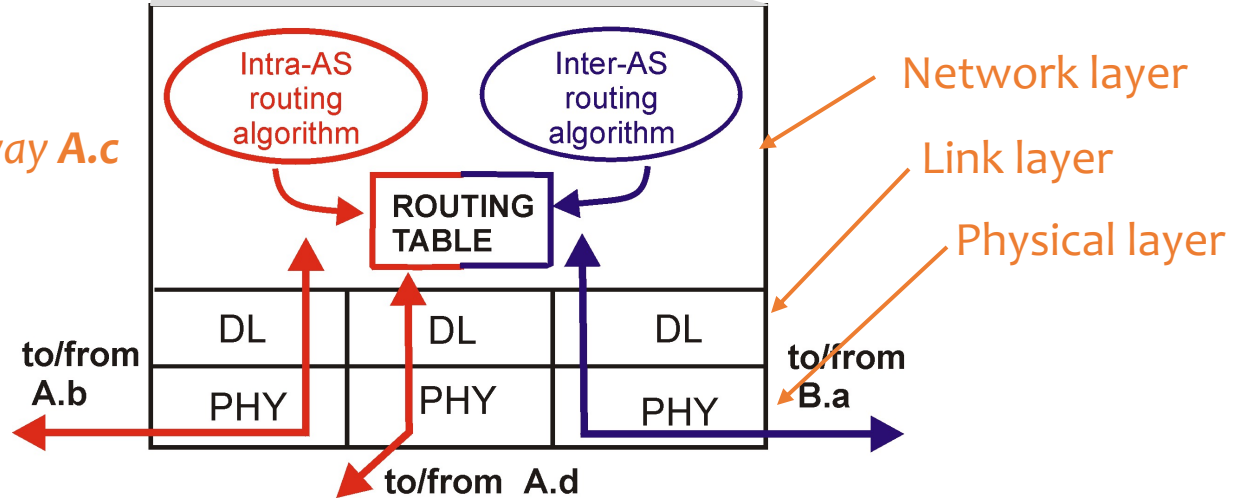
Inter-AS & Intra-AS routing



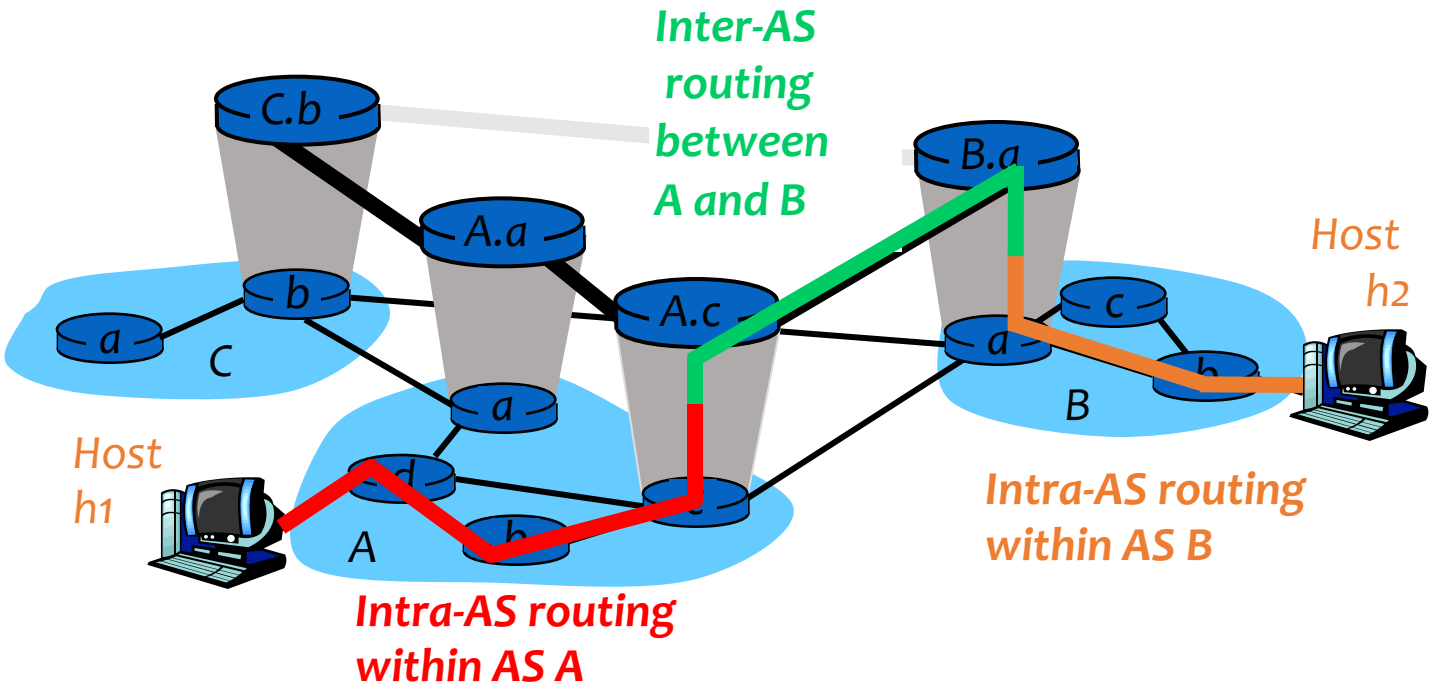
Gateways

- Perform inter-AS routing amongst themselves.
- Perform intra-AS routers with other routers in their AS.

Inter-AS, Intra-AS routing in gateway ***A.c***

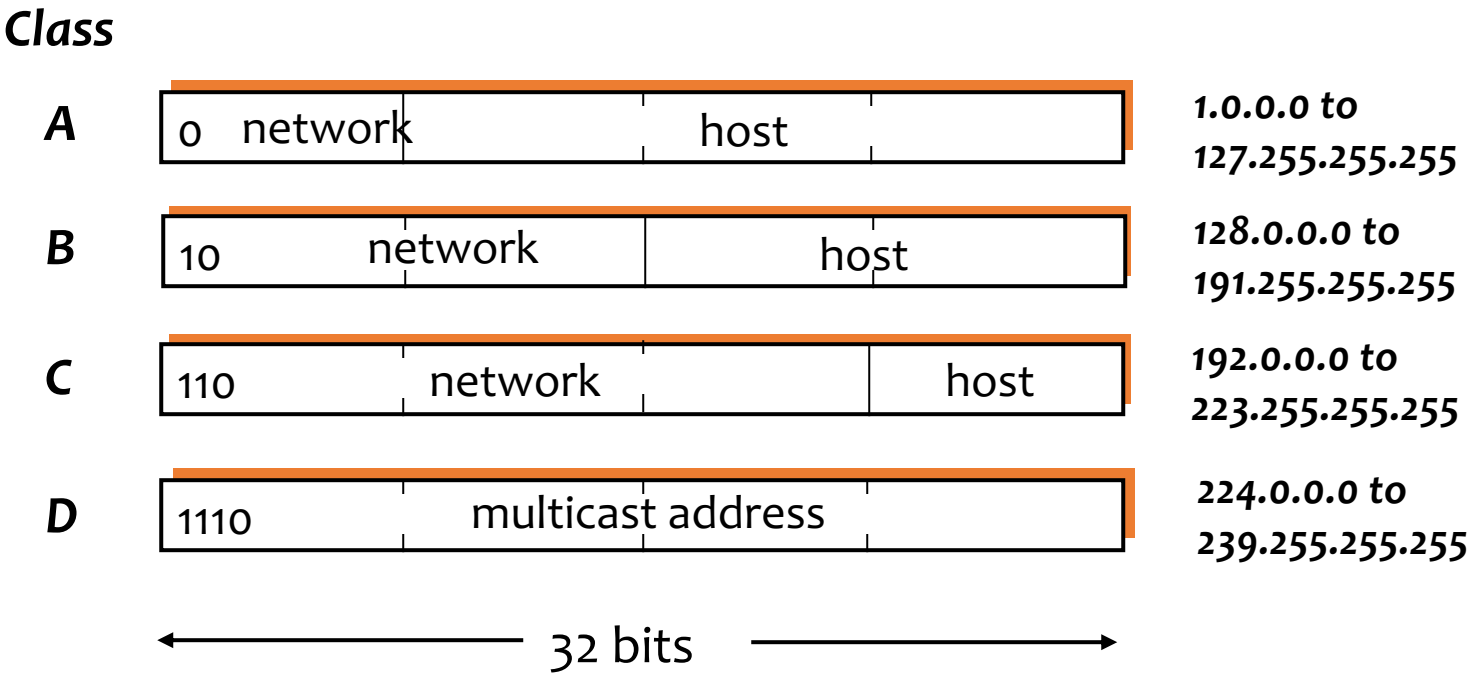


Inter-AS & Intra-AS routing



Dealing with scale: Addressing

- Old fashioned *"classful"* addressing



IP addressing: CIDR

Classful addressing

- *Inefficient* use of address space, address space *exhaustion*

  e.g., class B net allocated enough addresses for 65K hosts, even if only 2K hosts in that network

CIDR: Classless InterDomain Routing

- Network portion of address of *arbitrary* length
- Address format: ***a.b.c.d/x***, where x is #bits in network portion of address

IP addresses: How to get one?

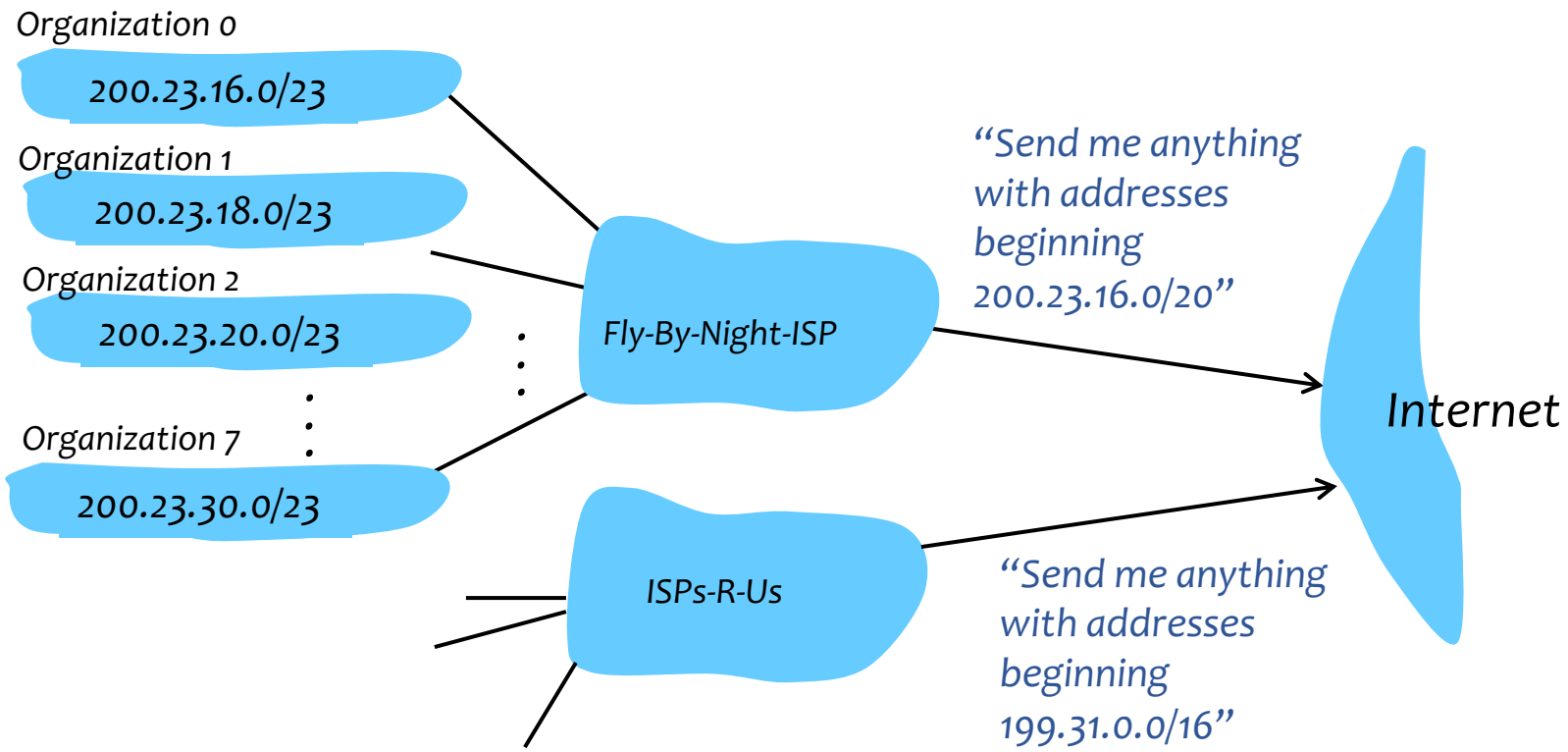
How does a network get the network part of IP address?

- Typically it gets allocated portion of its provider ISP's address space

ISP's block	**11001000 00010111 0001**0000	00000000	200.23.16.0/20
Organization 0	**11001000 00010111 0001**0000	00000000	200.23.16.0/23
Organization 1	**11001000 00010111 0001**0010	00000000	200.23.18.0/23
Organization 2	**11001000 00010111 0001**0100	00000000	200.23.20.0/23
…	…		
…	…		
Organization 7	**11001000 00010111 0001**1110	00000000	200.23.30.0/23

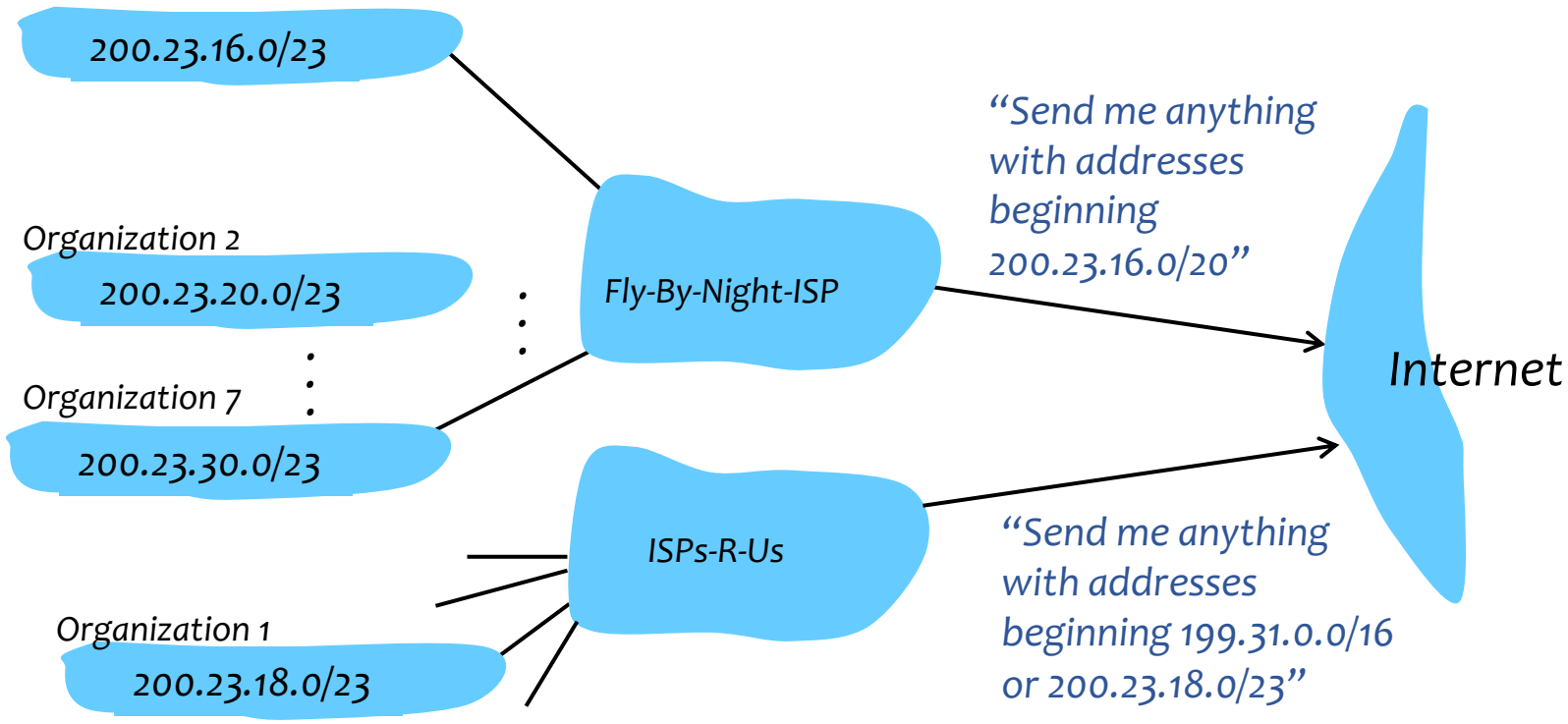
Hierarchical addr.: Route aggregation

Hierarchical addressing allows efficient advertisement of routing information



Hierarchical addr.: Route aggregation

ISPs-R-Us has a more specific route to Organization 1

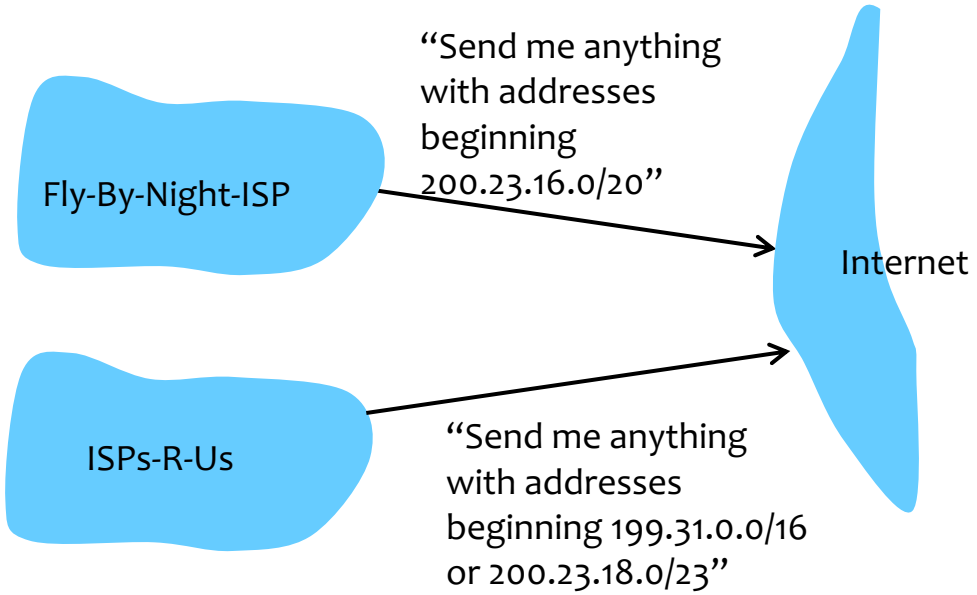


Hierarchical addr.: More specific routes

Multiple advertised routes can contain the same destination, e.g.,

- *200.23.16.0/20*
- *200.23.18.0/23*

both contain *200.23.18.7*



- Always route to most specific destination!
  (longest prefix match)

Dealing with scale: Advs. of large scale?

Take advantage of having to do similar things for others (caching)

Fault tolerance

- Large number of servers
- We have redundancy; multiple routes between sites

Metcalfe's law

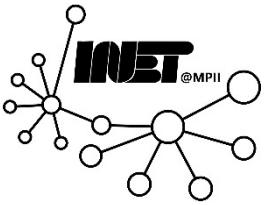
- *"Value"* of a network is proportional to square of number of things connected (bigger is better)

Law of large numbers

- Allocation of resources based on average usage rather than peak
- Amortizing upgrade maintenance over large population
- Popular network and services likely to be upgraded/improved

Denial of service

- Size/replication makes attack harder
- More generally, a system with replicated components is more survivable



Dealing with scale

"For every type of animal there is a most convenient size, and a large change in size inevitably carries with it a change of form." — Is it *true* for networks? Why? How so? Examples?

Ethernet doesn't scale up

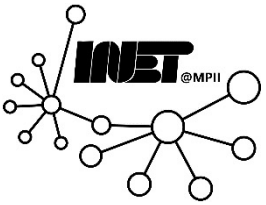
- Geo. distance, speed of light delays degrade perf. of random-access protocols (geographic scaling)
- Maybe scale with #users in geographically narrow net. if bandwidth scales with users

As number of communicating entities grows, need to change/improve manner in which to access communication channel

- Example: Small number of students vs. 500-class lecture, keeping bandwidth fixed as # users scales

Email versus HTTP

- Push systems work ok when small number of sender (email)
- Pull is better with large number of senders (http)



Dealing with scale

"For every type of animal there is a most convenient size, and a large change in size inevitably carries with it a change of form." — Is it *true* for networks? Why? How so? Examples?

Routing

- Large # of users and optimal routes => requires lots of info to compute routes, etc...; Doesn't scale

Certain services become necessary when you get big

- Name storage/translation: DNS, phone books

A single centralized site eventually breaks

- Need replication or other form of distribution

As network gets bigger flooding breaks

- Use limited flooding, caching

Switched vs. routed networks

- Change from layer 2 switched networks to layer 3 routed networks as # users gets bigger