$\mathcal{P}^5$: A Protocol for Scalable Anonymous Communications

Rob Sherwood, Bobby Bhattacharjee, Aravind Srinivasan

University of Maryland, College Park

www.cs.umd.edu/projects/p5

Outline

- Anonymity
- Naive solution

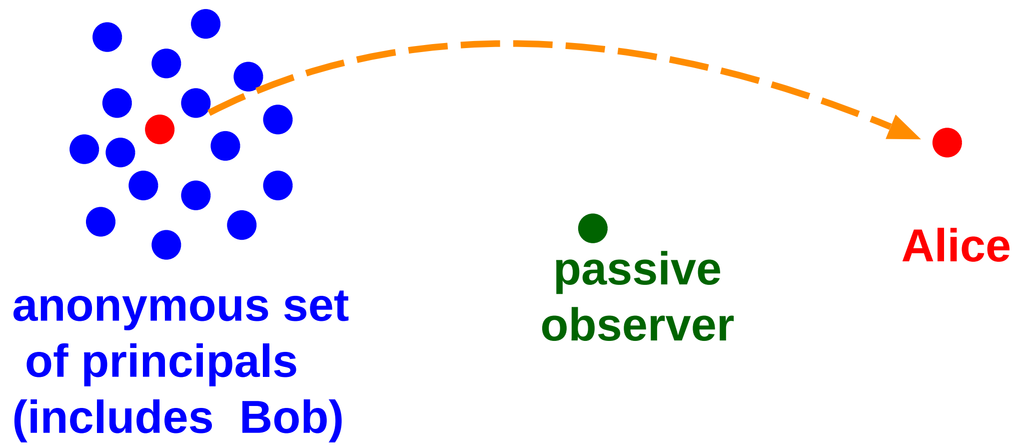
  Refinements
- $\mathcal{P}^5$ protocol
- Attacks and Performance Analysis
- Open Issues

$\mathcal{P}^5$ Goals

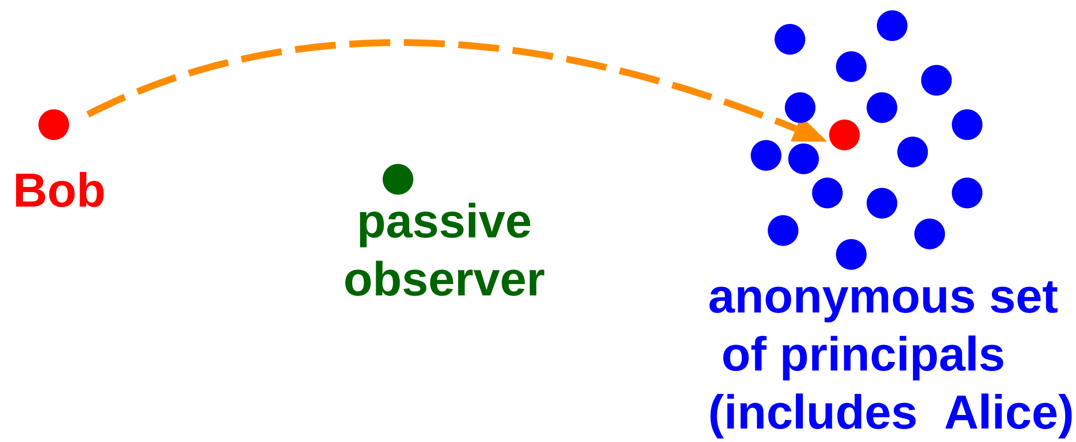
- Sender, Receiver, and Sender–Receiver anonymity
- Scalable
- Implementable under current infrastructure
- “Passive Global Observer” attack model
- Completely decentralized, minimal trust

Different types of Anonymity

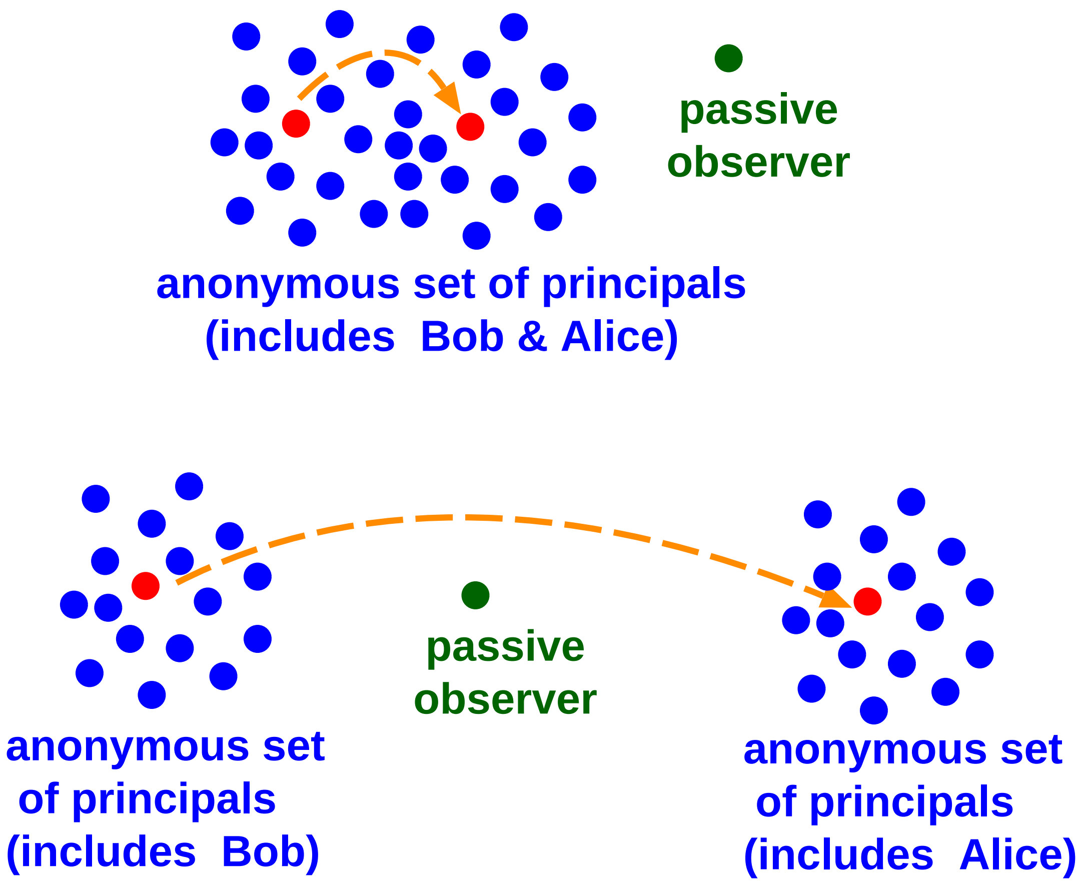
Sender anonymity



Receiver anonymity



Sender-receiver Anonymity



Passive observer cannot determine *who* Bob & Alice are, or whether they are talking

Naive Solution: Broadcast Ring

- Assume all parties are arranged in a logical ring

  upstream and downstream neighbors are known

  . . . but cannot be mapped to their identities

- Messages do not contain any information that distinguishes original sender from last-hop forwarder

- All messages are the same size, and are sent at a constant rate

- All messages are per-hop encrypted

Sending messages on a broadcast ring

Suppose Alice wants to send message m to Bob:

- Alice acquires Bob's public key, p_{Bob}, out of band
- Alice sends $\{m, H(m)\}_{p_{Bob}}$ to her upstream neighbor
- Each incoming packet is decrypted and verified
- Each incoming packet is also forwarded to upstream neighbor

System provides sender- and receiver-anonymity

Sender-receiver anonymity on broadcast rings

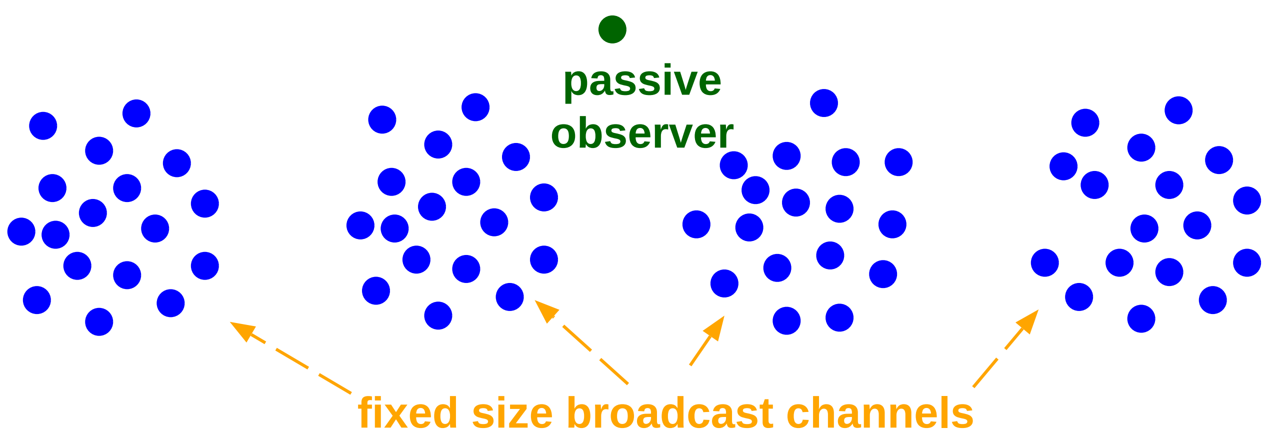
- What if there are only *two* users talking?
- Introduce noise — When Alice does not have a message to send:
  - She creates a “noise” packet . . .
  - . . . and sends it upstream as if it were a real packet
- Achieves sender-, receiver-, and sender-receiver anonymity

Does not scale: high latency and drop rates

Refining the Broadcast Channel. . .

There is a tradeoff between communication efficiency and anonymity

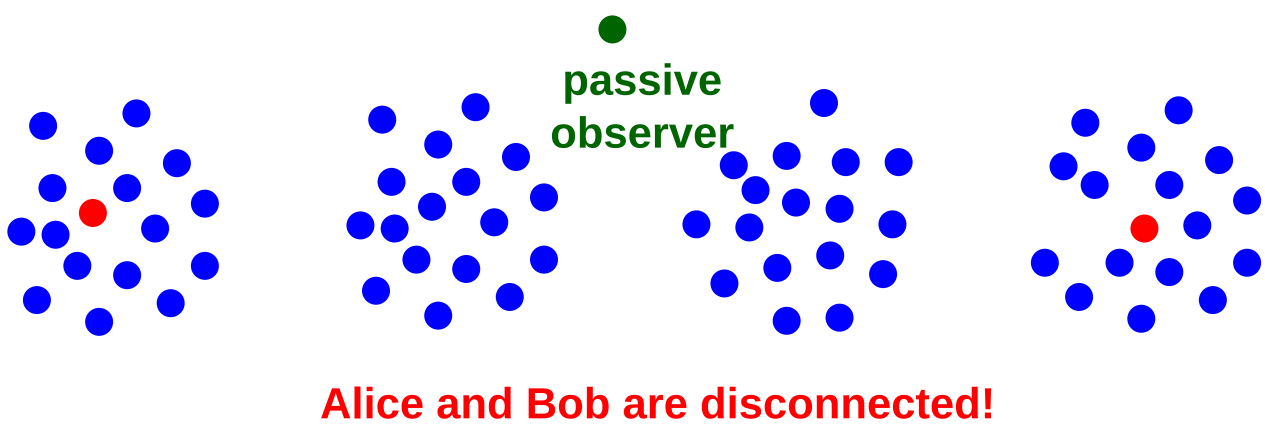
- Users may want to bound anonymity if it means better communications efficiency
- Solution: Use multiple "fixed" sized rings, and map each users to a ring using a well-known function



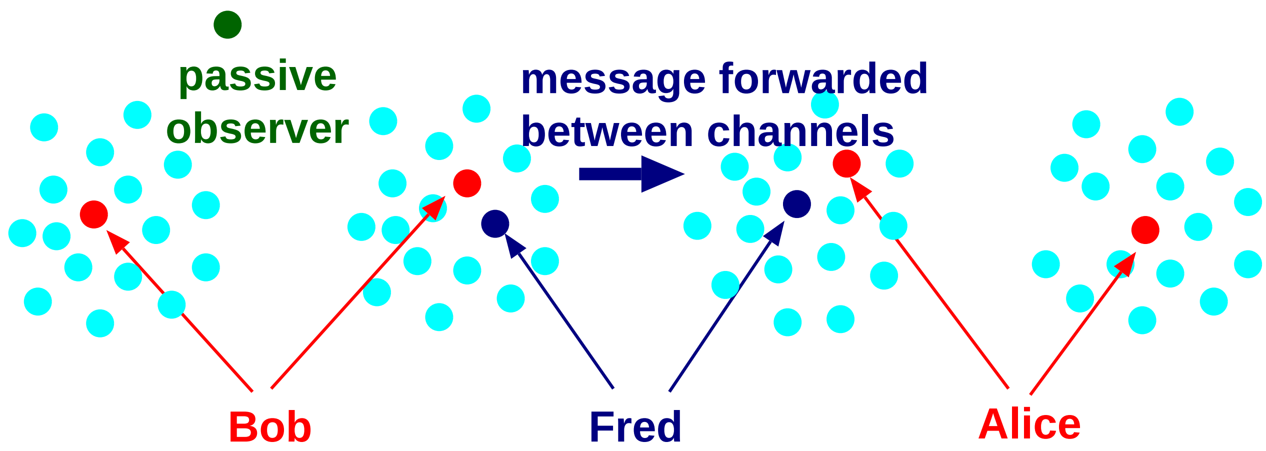
Refining the Broadcast Channel. . .

There is a tradeoff between communication efficiency and anonymity

- Users may want to bound anonymity if it means better communications efficiency
- Solution: Use multiple "fixed" sized rings, and map each users to a ring using a well-known function



Routing between different channels



- Each user joins a small number of channels without replacement
- Users “advertise” the set of channels they can reach and forward between channels to which they are joined
- If Alice and Bob do not have a channel in common, then their messages are relayed between channels by other users

Analysis

With high probability, a path of at most length 2 exists between any two users provided:

- Each user joins 3 channels
- There are at least 100 people/channel on average . . .
- . . . and at most 100 channels
- Path length increases by 1 when number of users increase to 15,000 (150 channels)

Problems

- Real system sizes are not fixed or known a-priori

    Difficult to choose number of channels

- People have different anonymity requirements

    "Globally quantifying paranoia is difficult"

- Good solution still requires some mechanism for:

    Dynamically creating channels

    Extensible addressing

$\mathcal{P}^5$: Extensible Broadcast

- $\mathcal{P}^5$ provides mechanisms for

  dynamic creation of new broadcast channels

  directly addressing named subset of channels

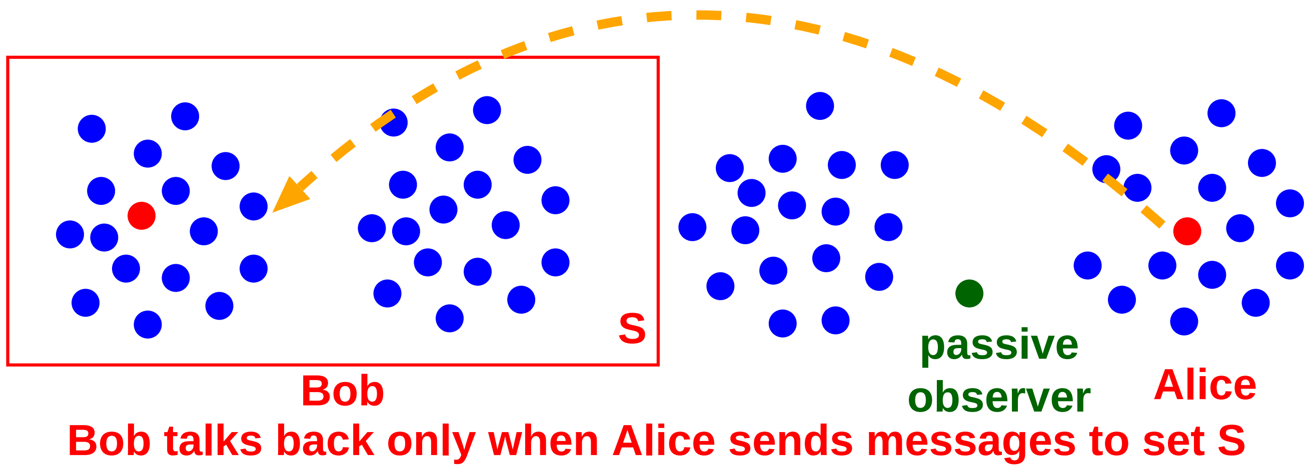
- Dynamic channel creation $\Rightarrow$ don't need to know system size a-priori

Subset addressing

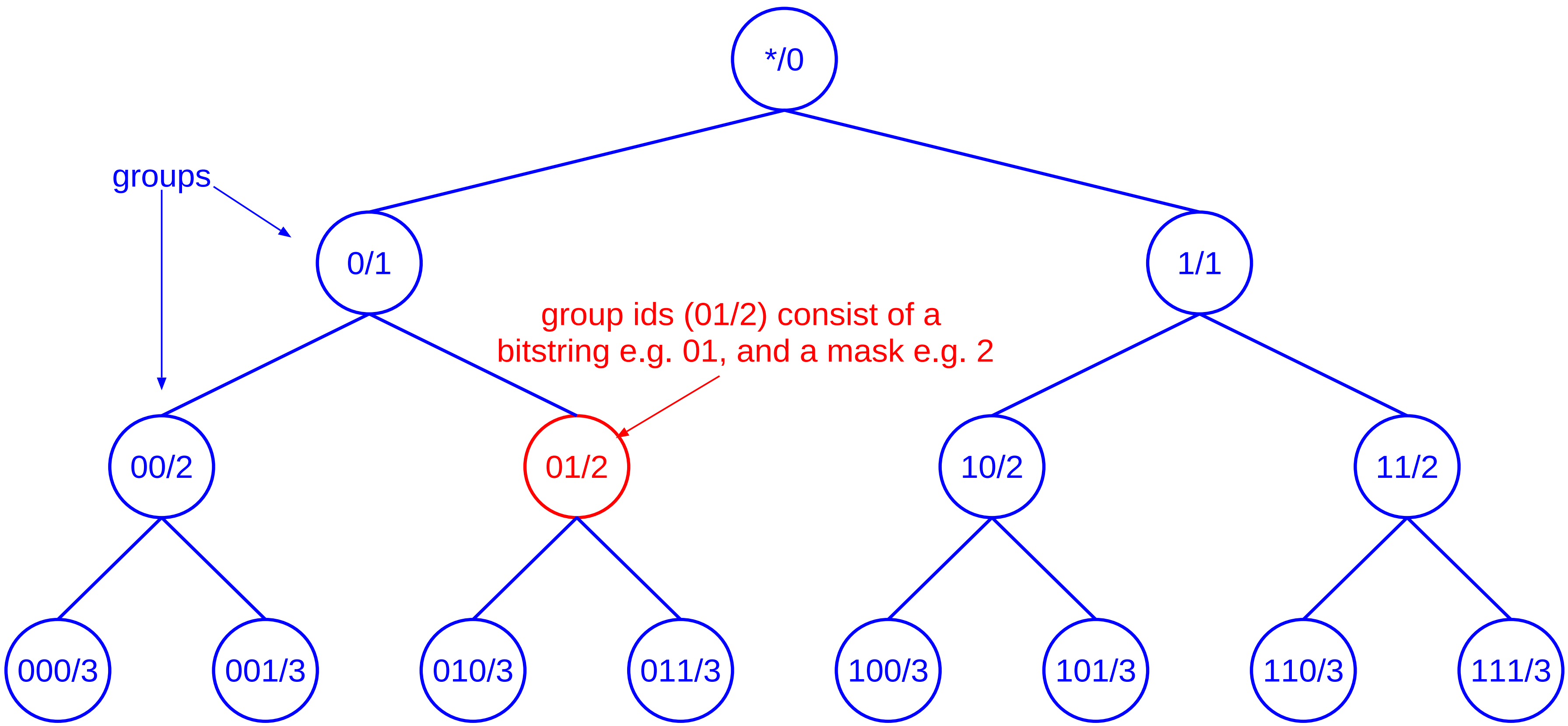
- Users can join any channel but they only talk when they have the anonymity of a "large enough" subset

  In $\mathcal{P}^5$, individual users can choose specfic subsets

  Sender anonymity is at least as large as receiver anonymity

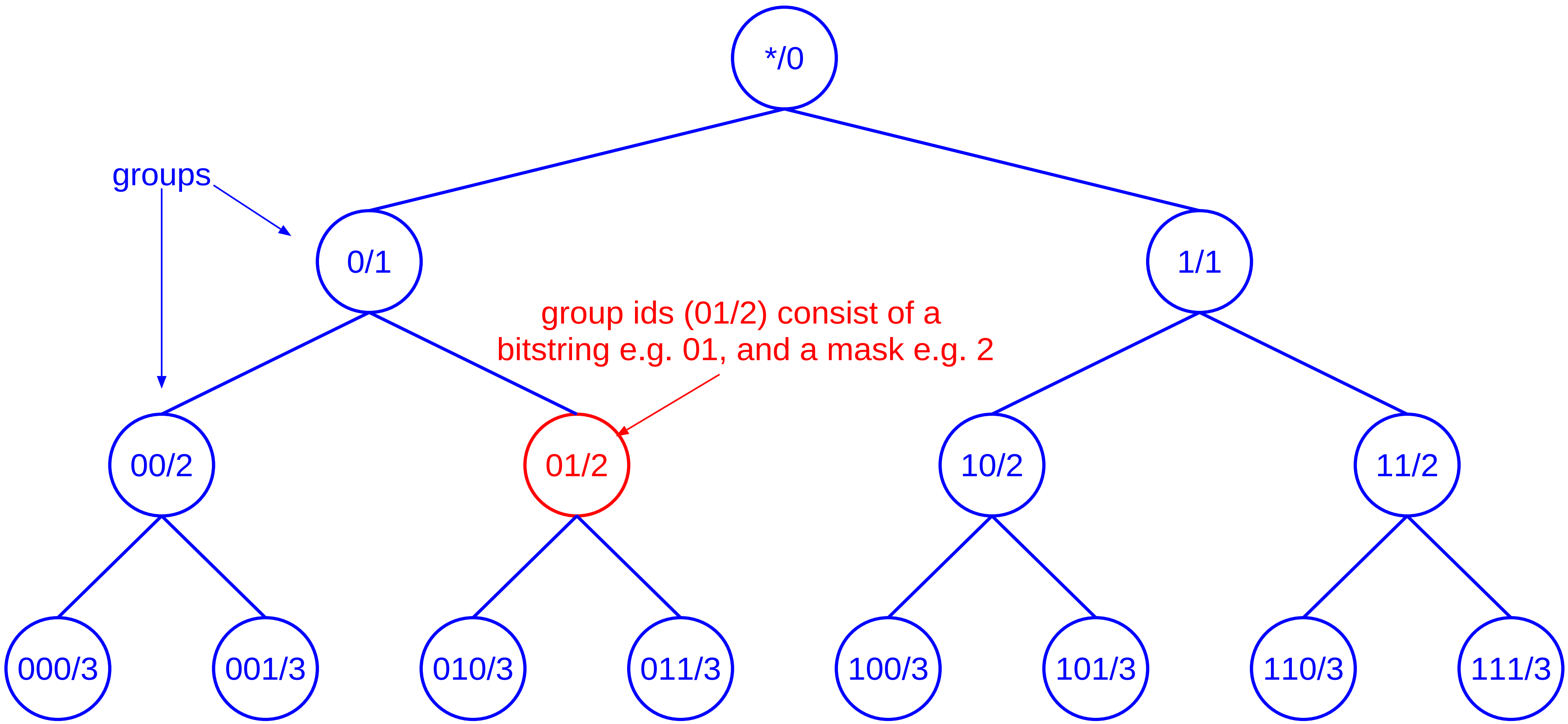


The $\mathcal{P}^5$ Logical Broadcast Hierarchy



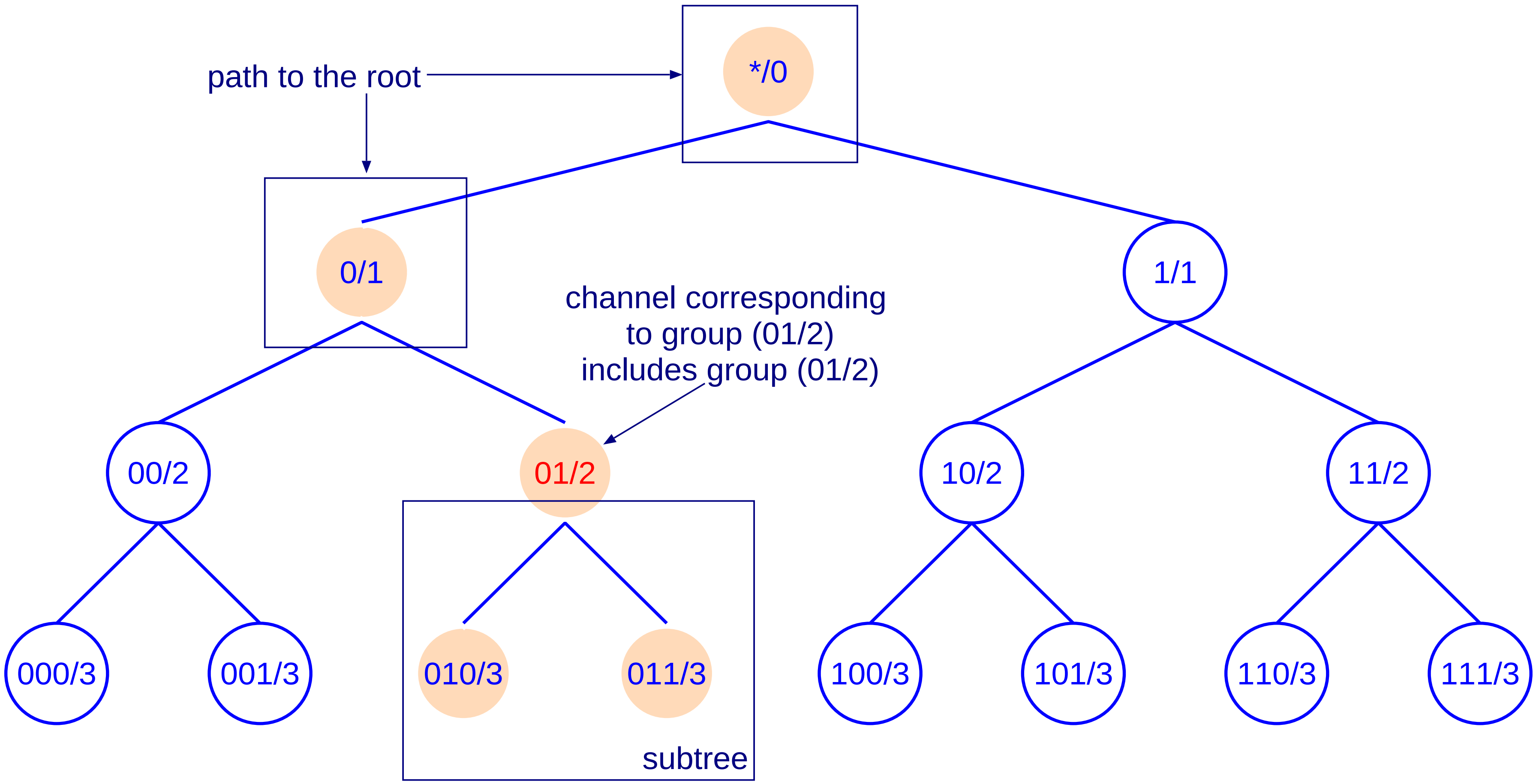
- This virtual structure is overlaid onto the underlying topology

The $\mathcal{P}^5$ Logical Broadcast Hierarchy



- Assume $H(p_{Alice}) = 0101\cdots$; Alice may join any group with id. of the form $(\star/0)$, $(0/1)$, $(01/2)$, $(010/3)$, etc.

$\mathcal{P}^5$ Channels



- Messages are sent to channels, which are addressible subsets of groups

$\mathcal{P}^5$ Hierarchy operations

- Forwarding using a “min-common-prefix” check
- Algorithm for constructing and maintaining hierarchy in paper

  In practice, $\mathcal{P}^5$ hierarchy overlaid on single spanning tree
- Users join a few groups (more than 1) for routing purposes

User actions on the $\mathcal{P}^5$ tree

- Users are mapped to a group using a hash of their public key
- Local security parameters L_{min} and L_{max}:

  L_{min} : the smallest acceptable anonymous channel size

  L_{max} : the largest acceptable anonymous channel size

- When joining the system, Alice computes $b_{Alice} = H(PK_{Alice})$, and picks an m_{Alice} value such that:

  $$L_{min} \leq k \leq L_{max}$$

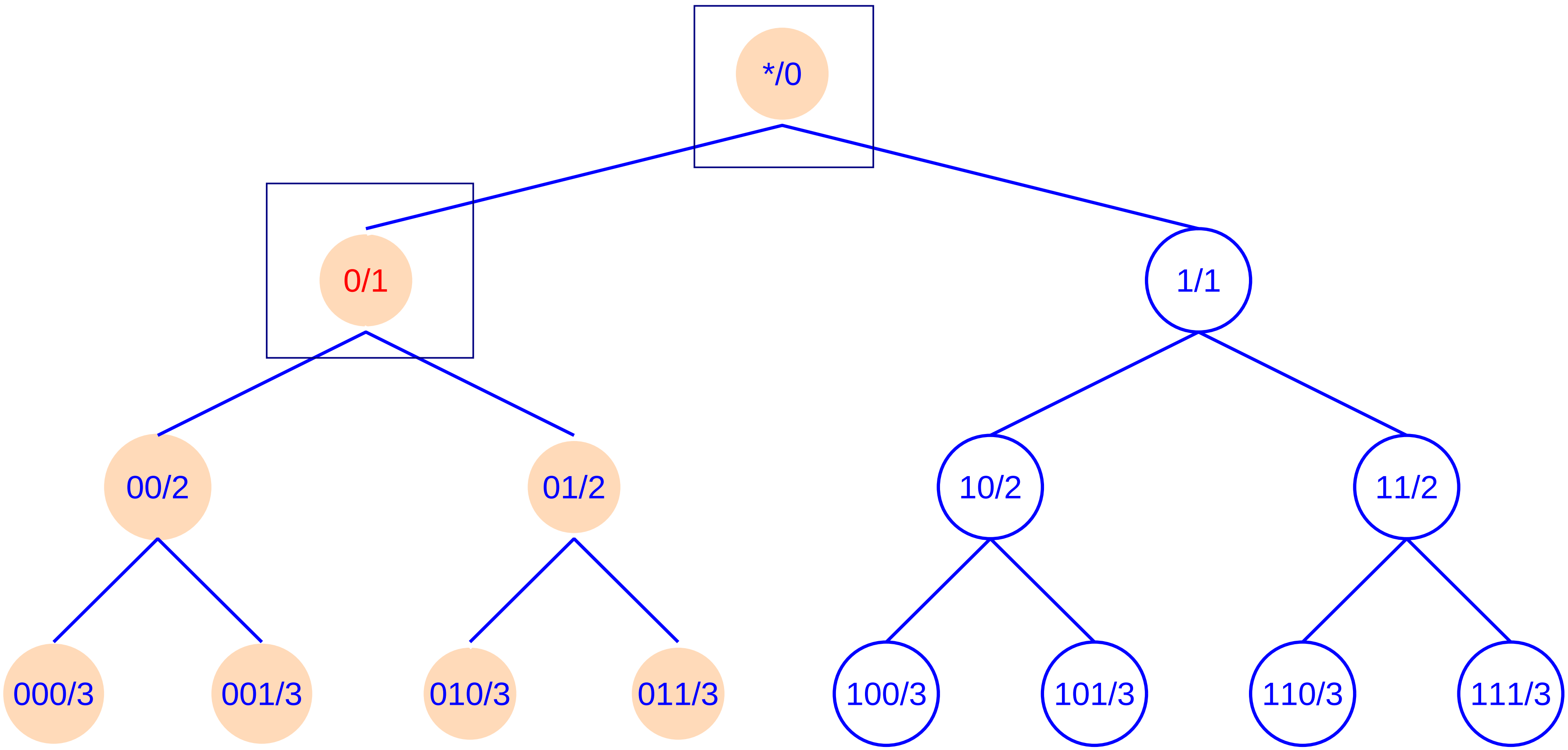
  where k is the number of people in channel (b_{Alice}, m_{Alice})

Sending a message

- Alice $\longrightarrow$ Bob: “Hi, I’m in channel (b_{Alice}, m_{Alice})”
- Suppose Bob wants to divulge two bits (01) of his key

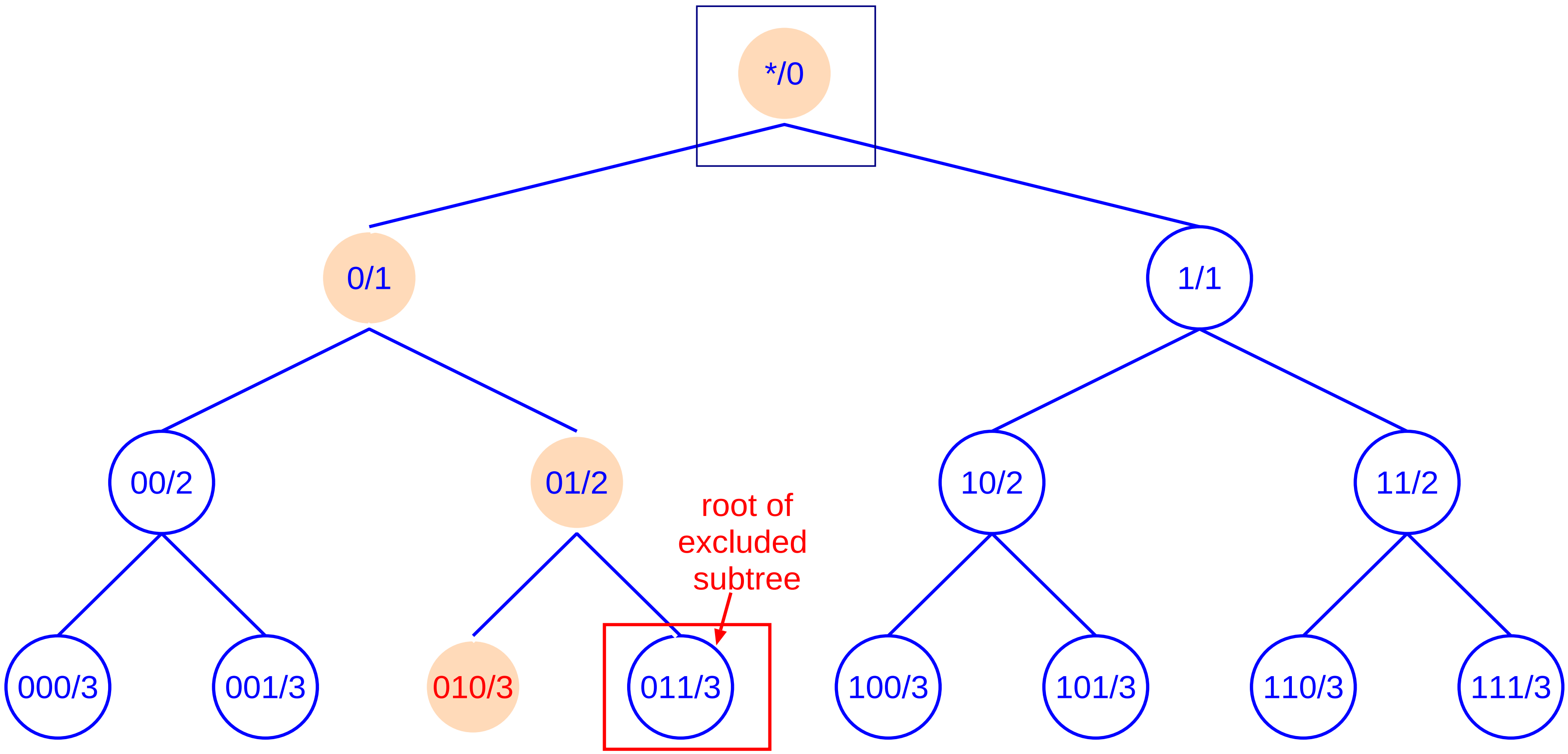
How does Alice know Bob’s mask?

Guessing a Mask: Too broad is fine



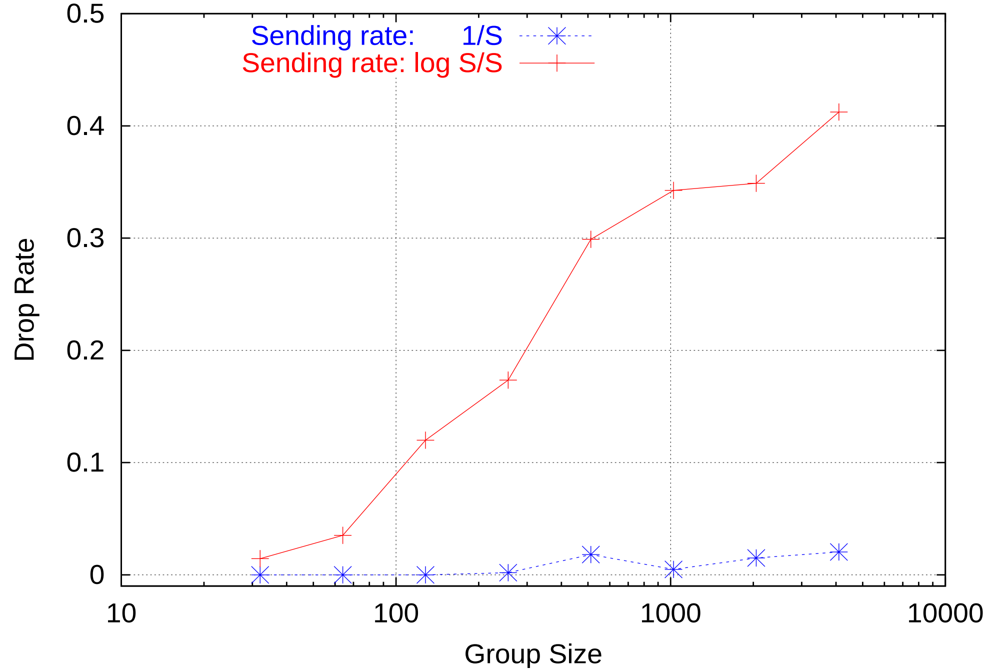
- Alice $\leftarrow$ Bob: "ACK, I'm in channel (b_{Bob}, m_{Bob})"

Too restricted is not!



- Bob: ignore message if $m > m_{Bob}$ (or become vulnerable to attack)

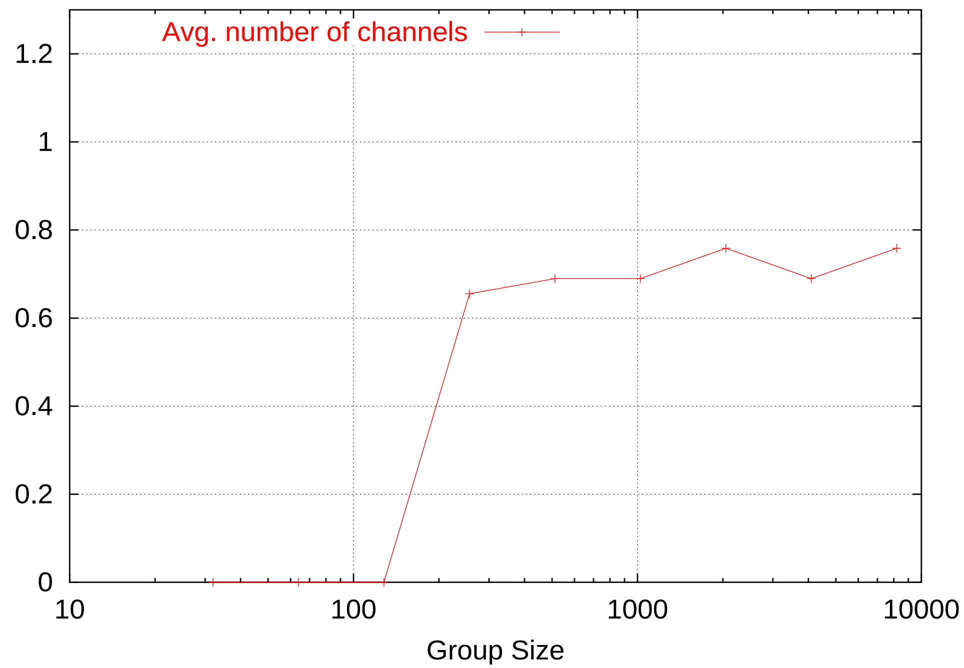
Performance analysis: Packet Loss Rates



- All users, except one randomly chosen pair, send noise
- Security parameters are (100, 300)
- Tree diameter: 13 hops; all users connected to two channels

- For 8192 users, 1.5 Mbps bandwidth, 200 1000 byte pps
  16Kbps with <5% loss, 200 Kbps with 40% loss

Signal Packet Hops



- Matches path lengths predicted via analysis
- For 8192 users, most paths less than 2 hops

  4 out of 6000 sampled paths required 2 hops

Attacks

- Passive Attacks
  - Intersection attack, e.g. joining two channels with same key
    $\Rightarrow$ Different $keys$ for routing
  - Difference attack, e.g. responding to a larger mask
  - Correlation attack, e.g. $\mathcal{P}^5$ without noise packets
- Active attacks
  - Mob attack, e.g. multiple attackers join group to provide "fake" anonymity

Issues

- Per-packet public-key decoding at receivers

  Not feasible for high packet rates (100s pps) using RSA

- Implementation: topology discovery

  - How does Alice know the number of users in her group/channel?
  - Use a set of “topology servers”; topology servers maintain and distribute maps of the form IP Addr $\longrightarrow$ Group
  - At least one server must be uncompromised

Major Issues

- When users leave, they lose anonymity

  Not specific to $\mathcal{P}^5$

- When new users join, communication efficiency decreases, but anonymity does *NOT* increase

  Move to different channel $\longrightarrow$ another difference attack

- Thus, in $\mathcal{P}^5$, users should not change groups after joining

$\mathcal{P}^5$ Generalization/Extension

- Users belong to a single named subset that never changes
- New subsets with configurable efficiency can be added as new members join system
- Routing to subsets is topology-specific

  Rings, trees: special cases

  Topologies with higher-connectivity are possible

Related Work

System	Sender-Anonymity	Receiver-Anonymity	Send.-Recv. Anonymity	*Efficient* Implementation
DC-Net	Y	Y	Y	
Xor-Trees	Y	Y	Y	
Crowds	Y		N	Y
Onion Routing	Y		N	Y
Hordes	Y		N	Y
Tarzan	Y	N	N	Y
$\mathcal{P}^5$	Y	Y	Y	Y

Future Work

- Faster receiver processing
- Better topologies
- Protocol without a topology server
- Formal analysis

Future Work

- Faster receiver processing
- Better topologies
- Protocol without a topology server
- Formal analysis

www.cs.umd.edu/projects/p5