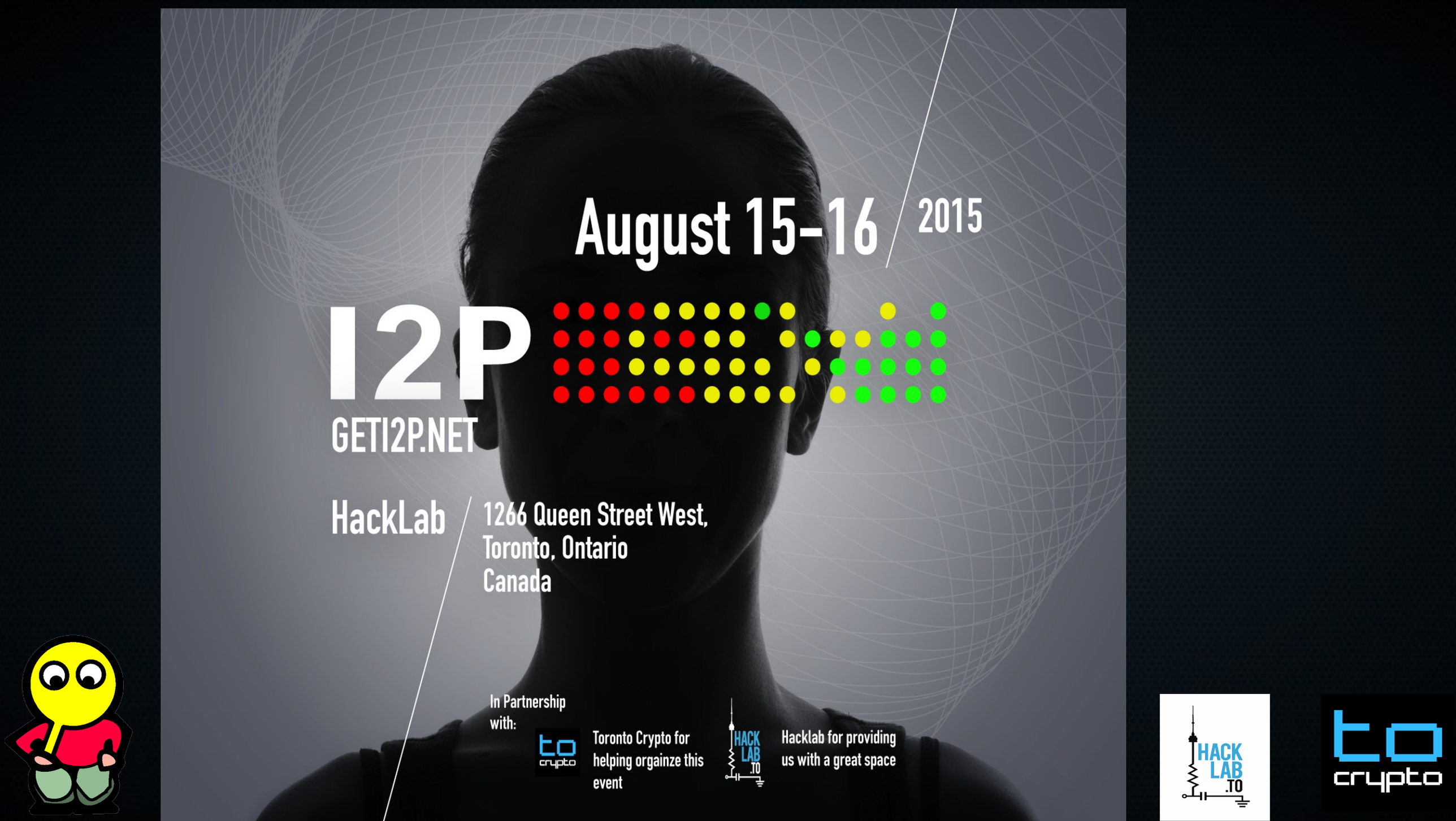

August 15-16
2015
I2P
GETI2P.NET
HackLab
1266 Queen Street West,
Toronto, Ontario
Canada
In Partnership
with:
Toronto Crypto for
helping orgainze this
event
Hacklab for providing
us with a great space

- The Connection Limit Challenge

#I2PCon Toronto
August 15-16, 2015
zzz
zzz@i2pmail.org
geti2p.net
@i2p

HACK LAB .TO

to crypto

Google

HACK LAB .TO

to crypto

From then... to now

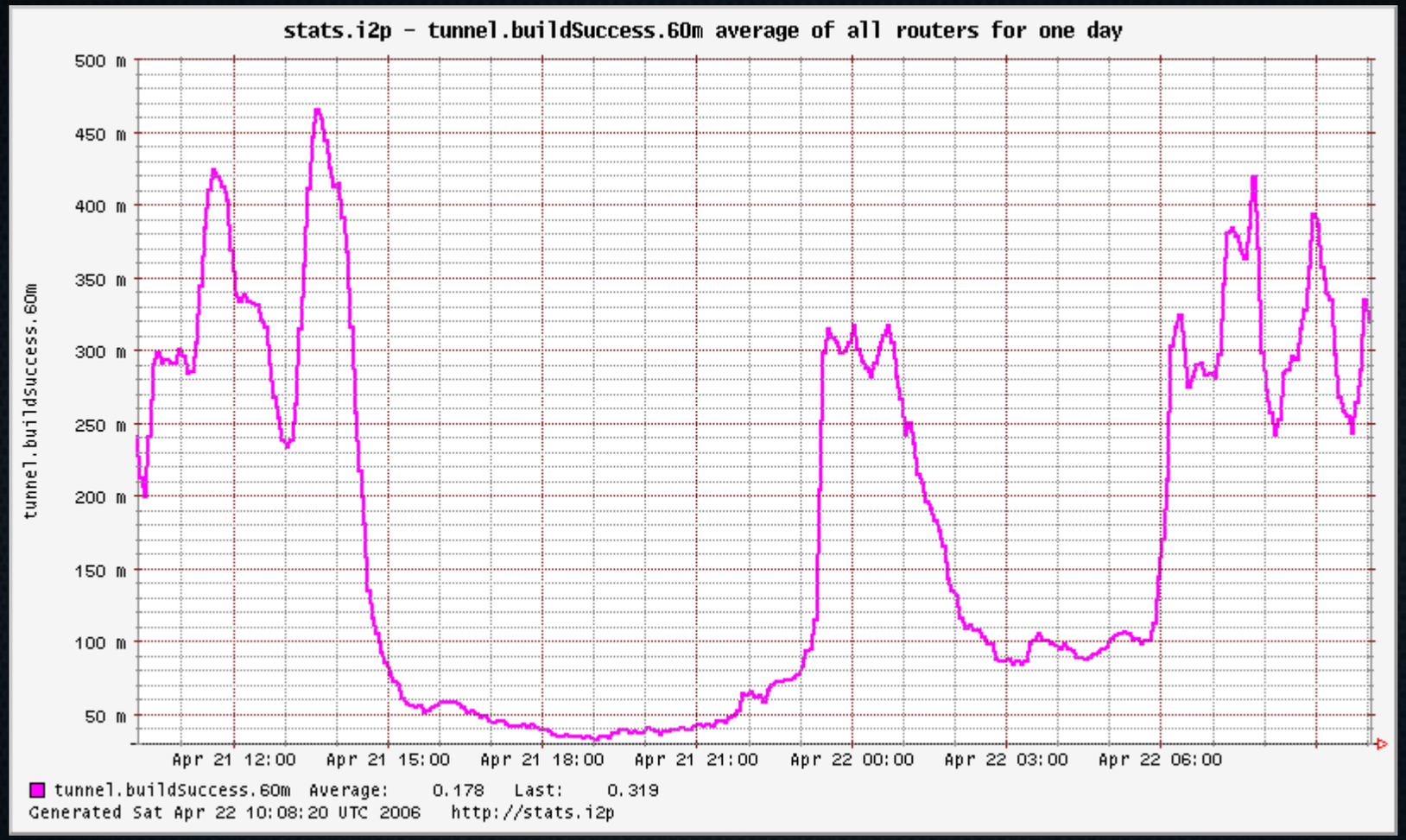
- less than 1000 routers
- 1 hour (!) idle timeout
- 3 floodfills

- 30,000 routers
- 2 minute idle timeout
- 750 floodfills

HACK LAB .TO

to crypto

2006 disaster



HACK LAB .TO

to crypto

Congestion Collapse

- Positive feedback drives you to disaster
- Not clear where the limits are
- Can happen with little warning
- Hysteresis makes it hard to recover
- Need multiple strategies for backoff

HACK LAB .TO

to crypto

Reasons why it's hard

- Source routing
- Full-mesh network, no restricted routes
- DHT only a small portion of the traffic
- iterative, not recursive, DHT lookups
- Keyspace rotation
- 10-minute tunnels
- Must have limits
- Home firewalls get confused / lose mappings
- Fast tier / ff full-mesh
- ff increase leads to more conn limits
- Expensive conn setup in cpu (DH), bandwidth, and latency

HACK LAB .TO

to crypto

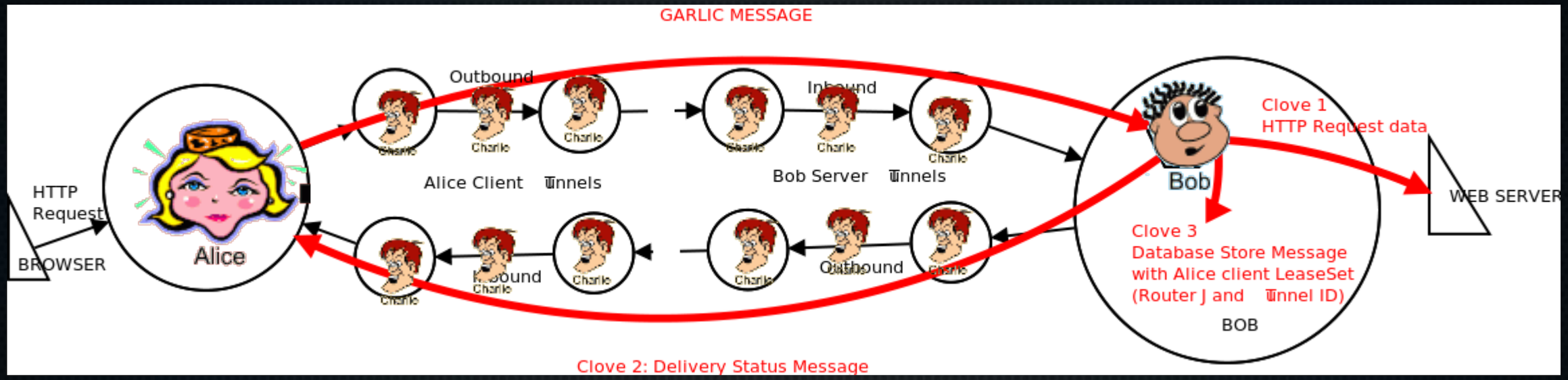
Floodfill challenge

- Thought that we'd get to the point where the ffs didn't know all the other ffs
- Actually the opposite - they know all and want to connect to all

HACK LAB .TO

to crypto

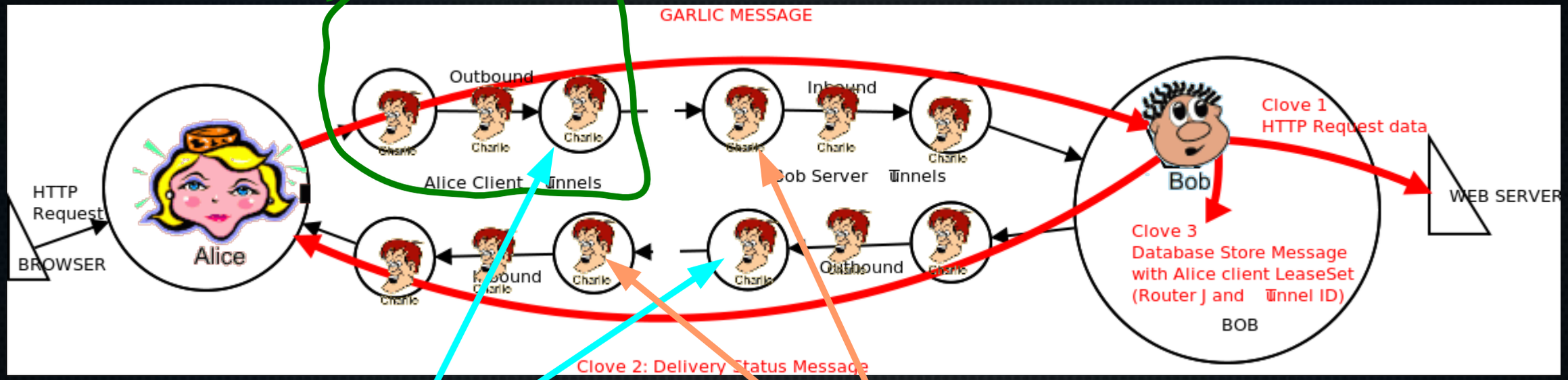
One-way path (with tagset ack)



HACK LAB .TO

to crypto

Key Roles



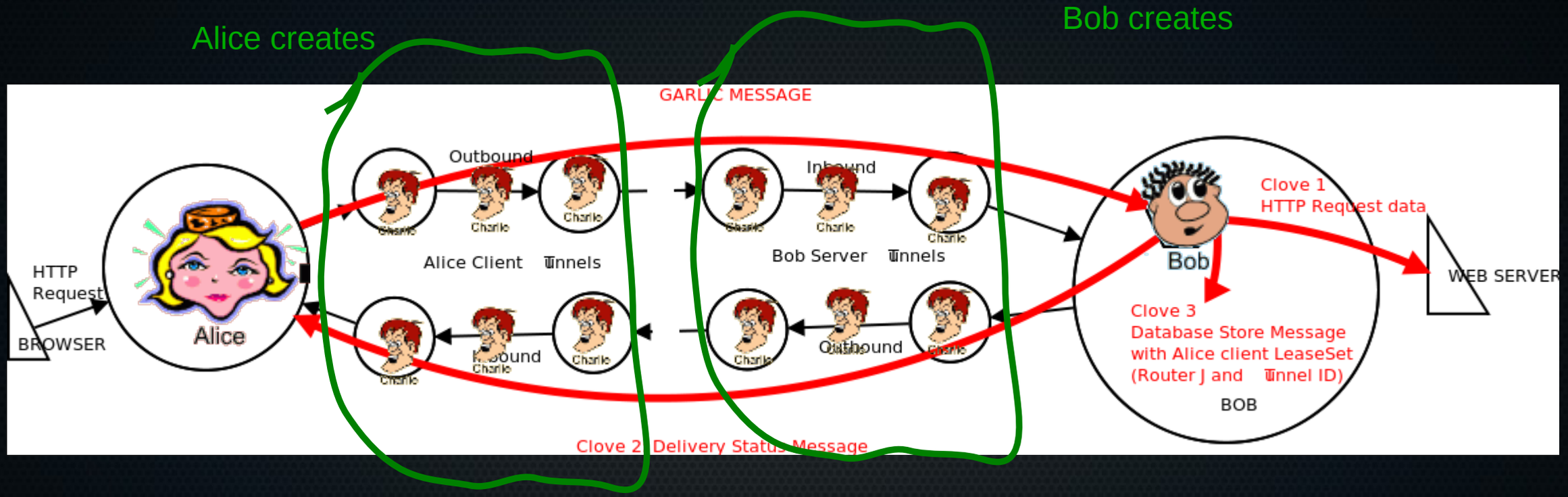
Outbound Endpoints (OBEP)

Inbound Gateways (IBGW)

HACK LAB .TO

to crypto

Tunnel Creation

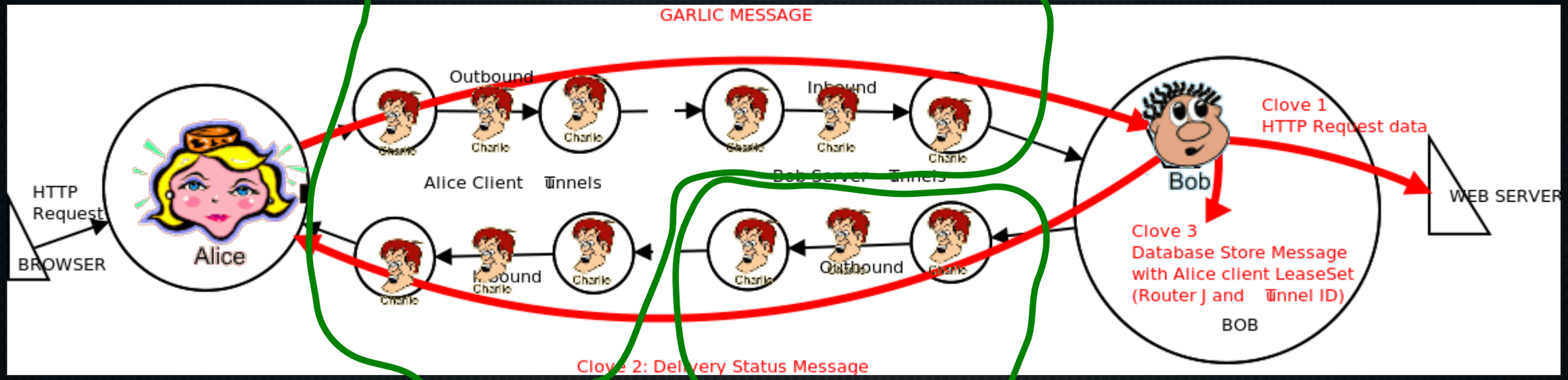


HACK LAB .TO

to crypto

Source Routing

Alice selects



Bob selects

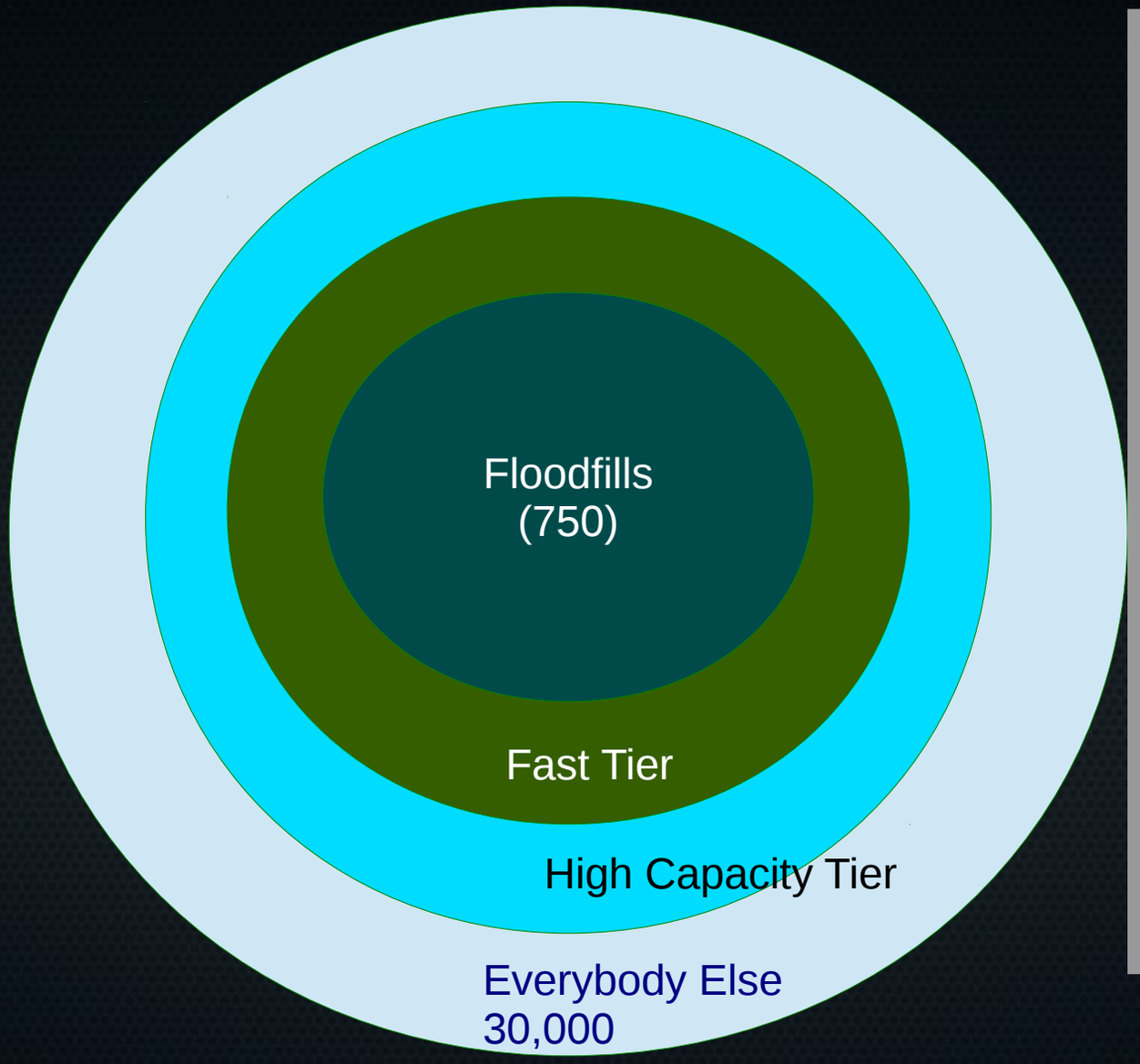
HACK LAB .TO

to crypto

DHT Proximity

- Floodfills are close to other floodfills, and flood to them, proximity doesn't change
- Due to rotation, you aren't closest to yourself
- Closest floodfill changes every day

Tiers



Each router decides if it is floodfill and publishes that in the network database.

Fast and high capacity tiers are the individual view of each router, and change over time, but the views probably converge to a rough consensus.

HACK LAB .TO

to crypto

RI Store

Store to floodfill closest to yourself; he will flood to closest floodfill peers for redundancy

HACK LAB .TO

to crypto

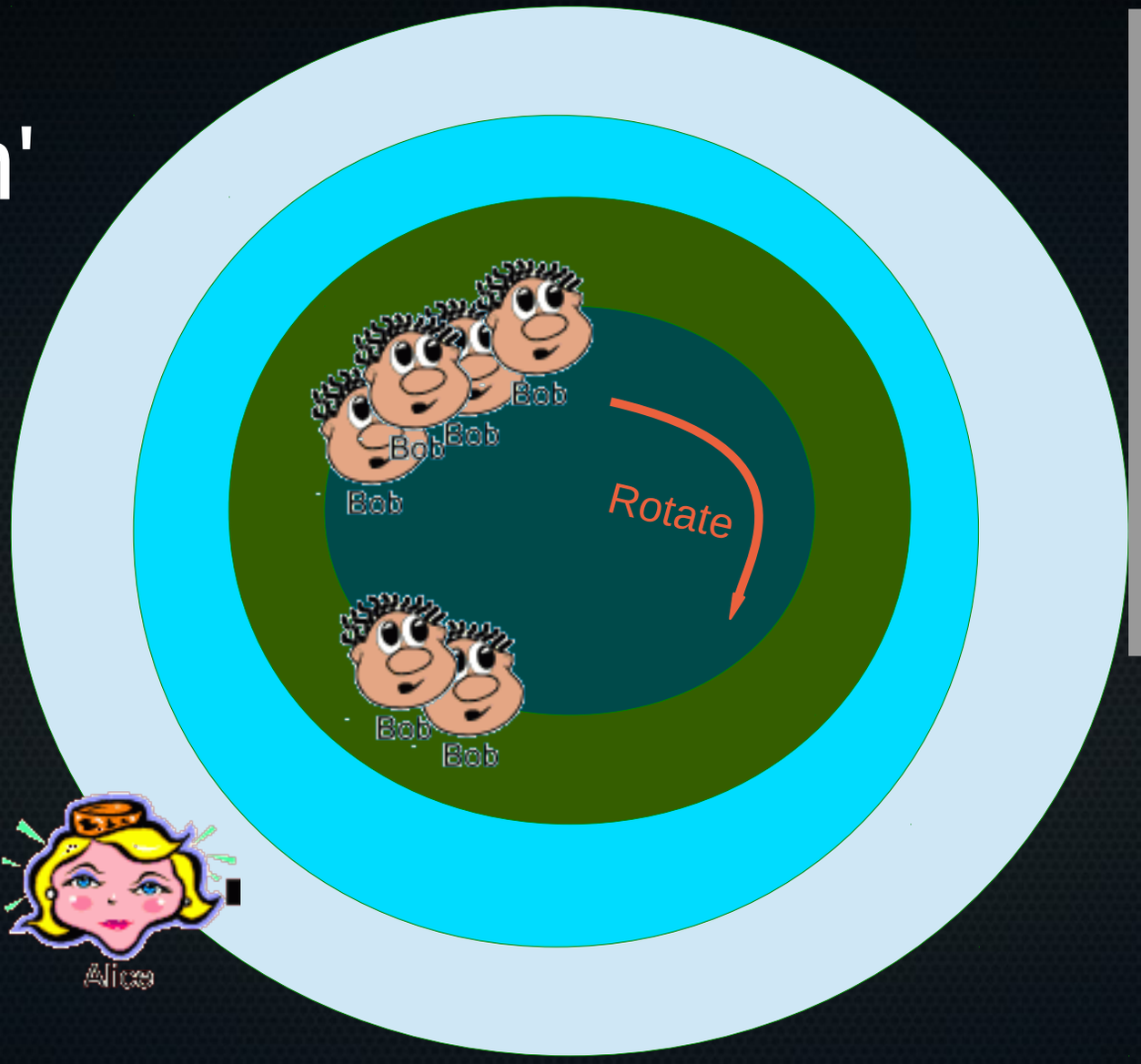
Just before midnight...

Just before midnight, the floodfill will also flood to the floodfill peers about to become closest after the rotation

HACK LAB .TO

to crypto

Midnight DHT 'rotation'



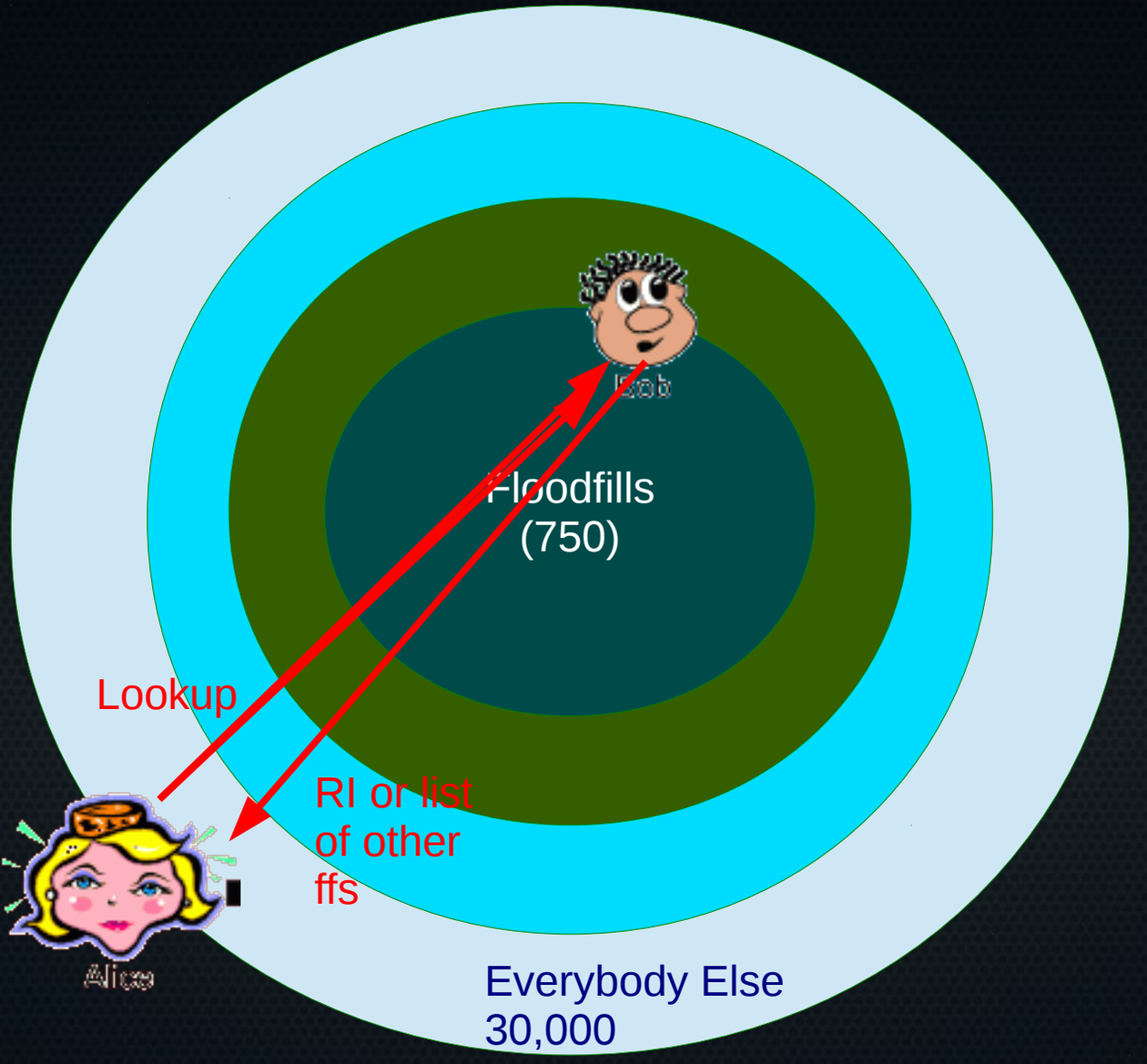
At midnight, the floodfill DHT 'rotates' and a new set of floodfills is now closest to Alice.

The same floodfills are still closest to each other

HACK LAB .TO

to crypto

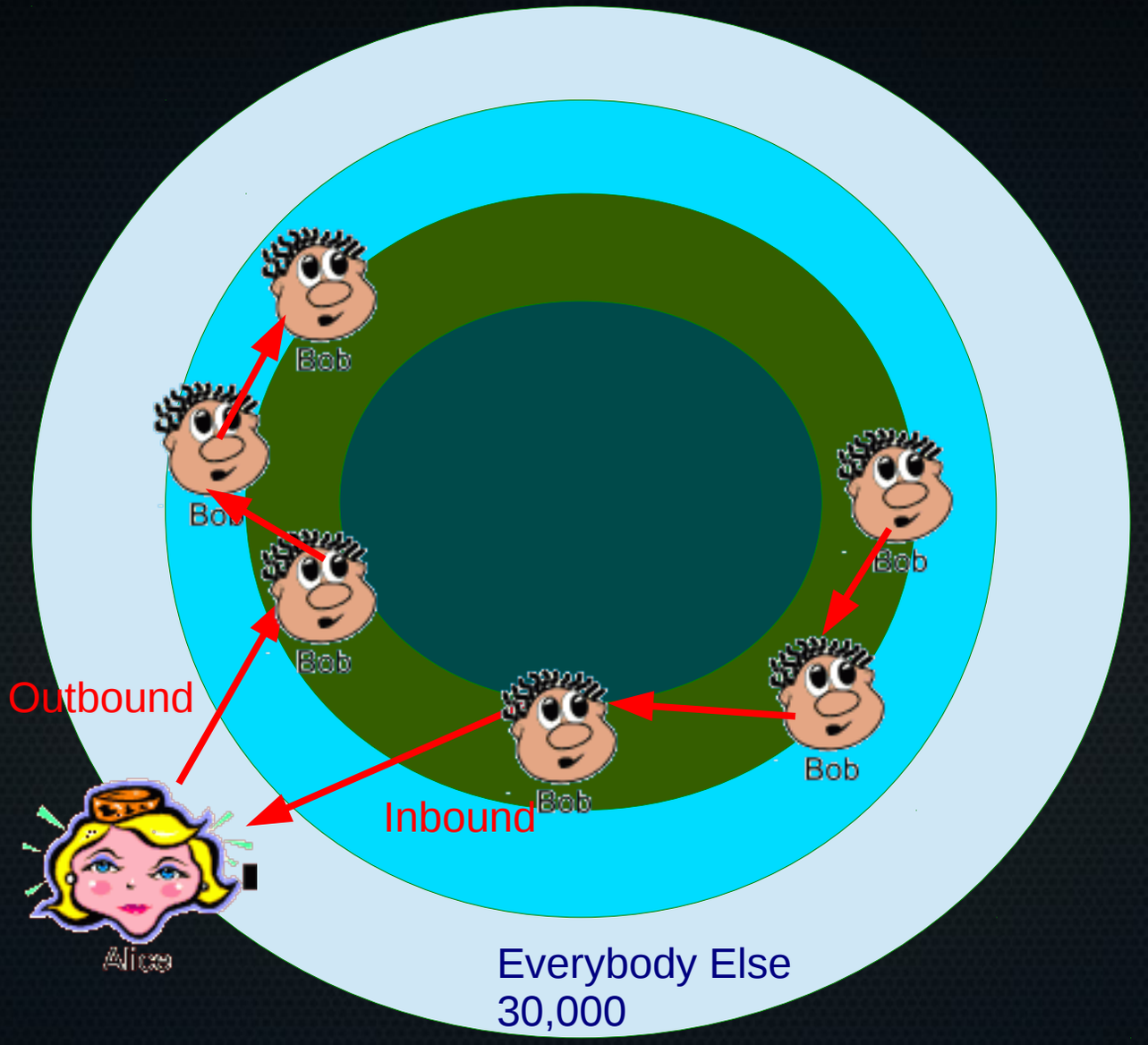
RI Lookup



HACK LAB .TO

to crypto

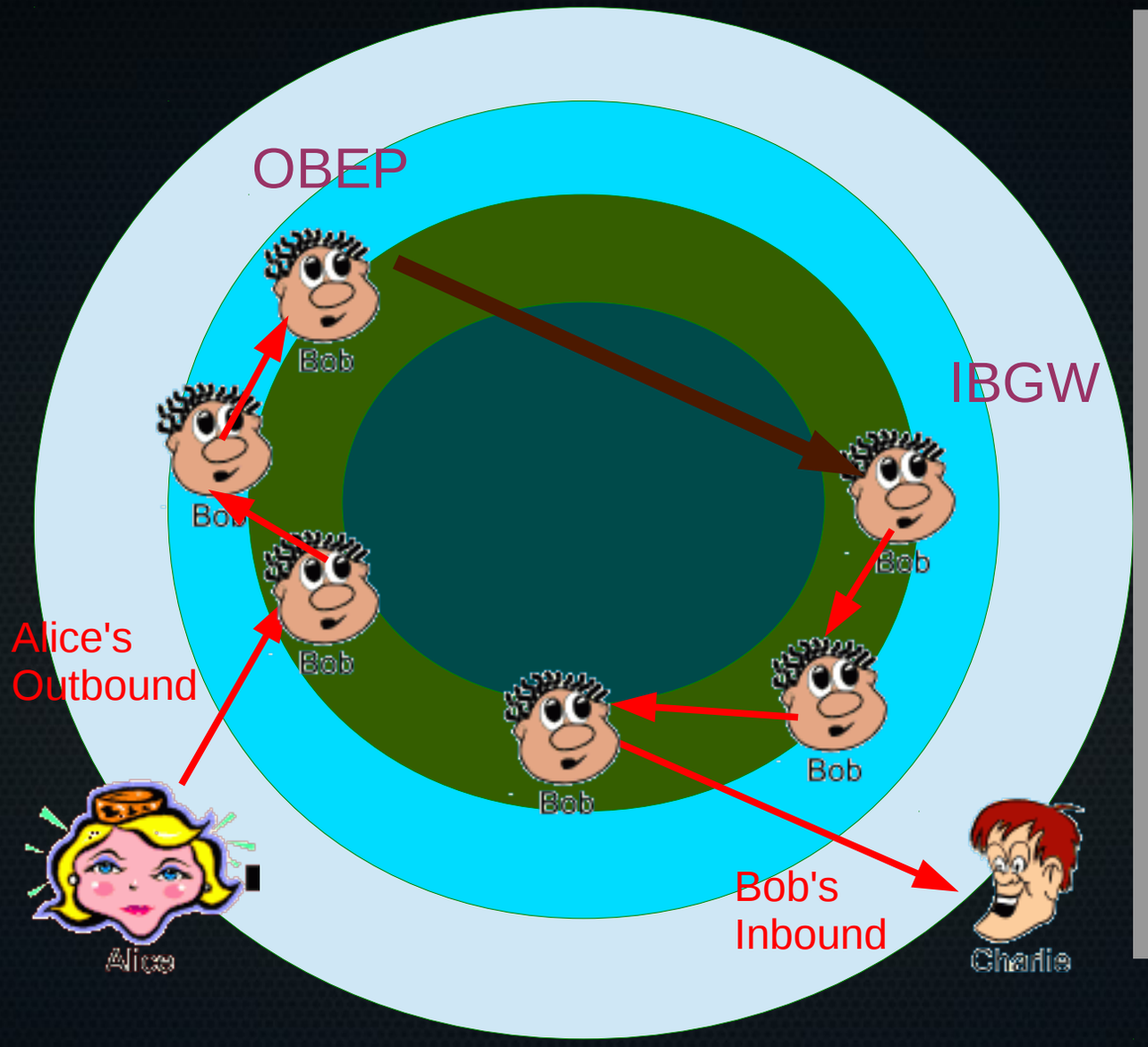
Client Tunnels



HACK LAB .TO

to crypto

OBEP / IBGW Proximity and pair selection



Alice selects OBEP from her pool and IBGW from his leaseset

This connection failure is the cause of most so-called 'tunnel failures'.

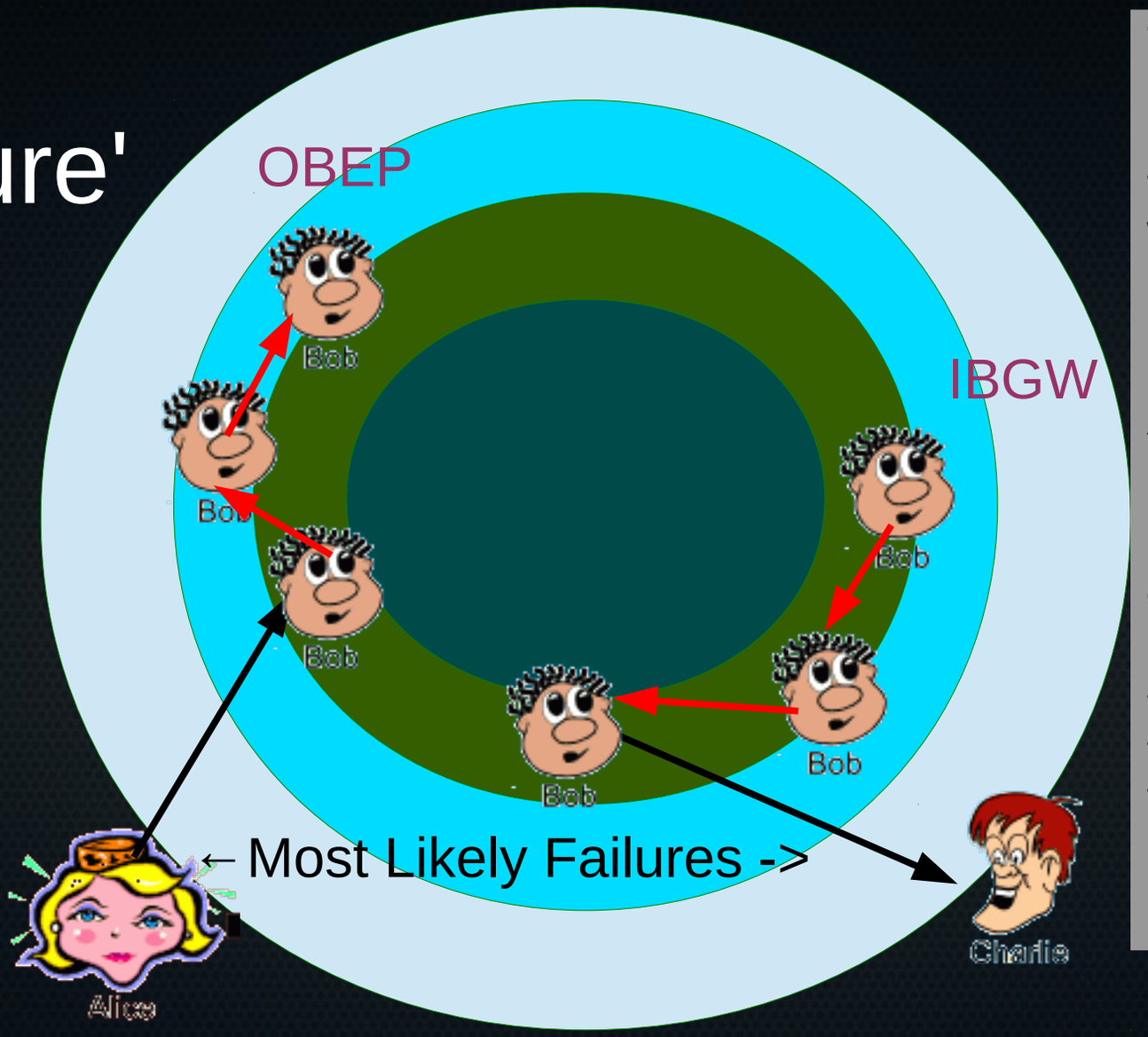
Alice can pick a pair based on DHT closeness but that may hurt reliability compared to random.

Alice could pick pairs at random on ack failure for better reliability but that may cause more congestion.

HACK LAB .TO

to crypto

Actual 'tunnel failure'



Tunnel builds require connections and an acknowledgement, so a tunnel should never fail within the idle timeout (currently two minutes) after creation.

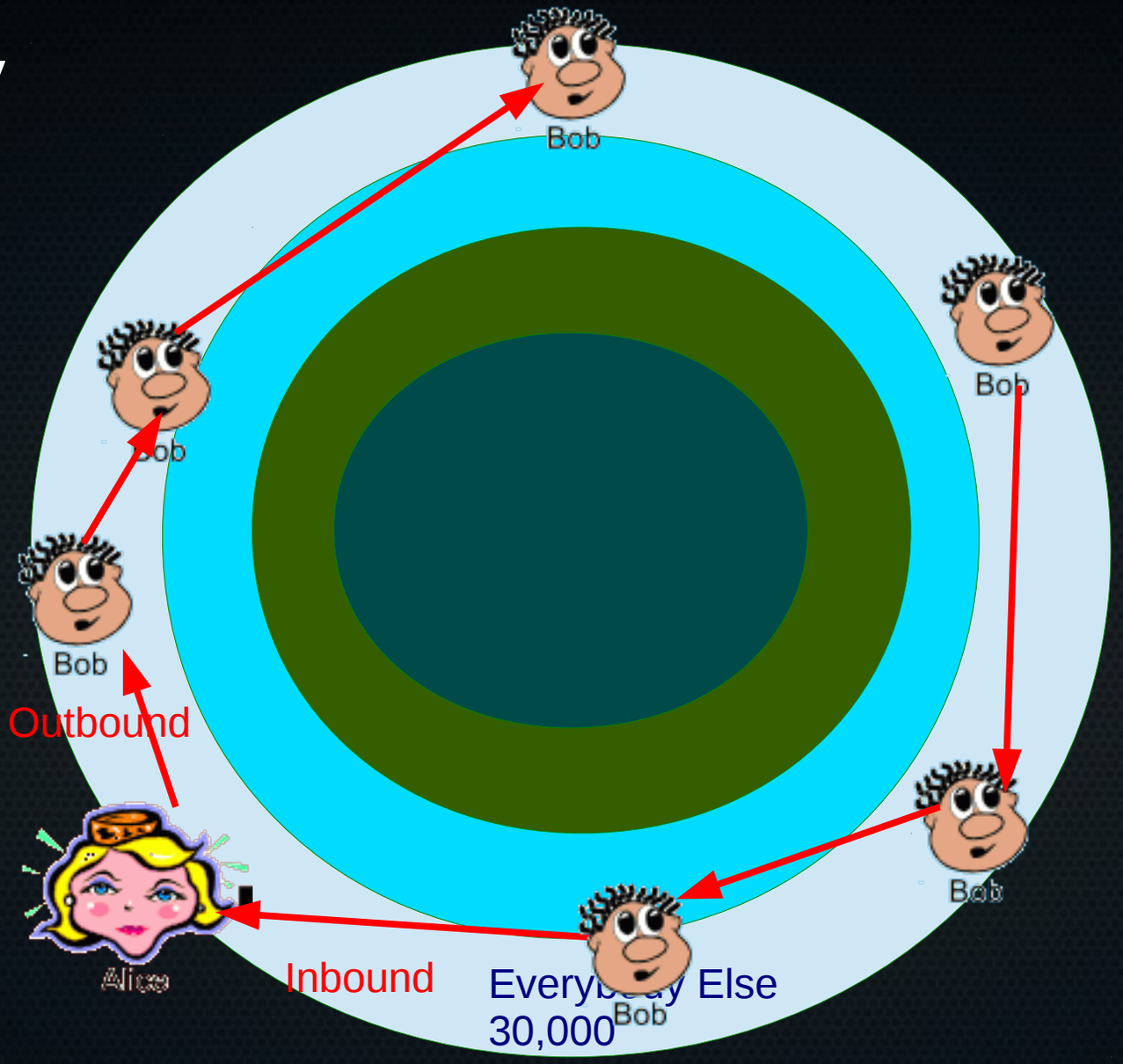
After that, the connection must be reestablished to pass data.

The first OB hop or last IB hop may be the most likely to idle-disconnect and then fail to reconnect, as the fast tier will tend to stay connected to each other.

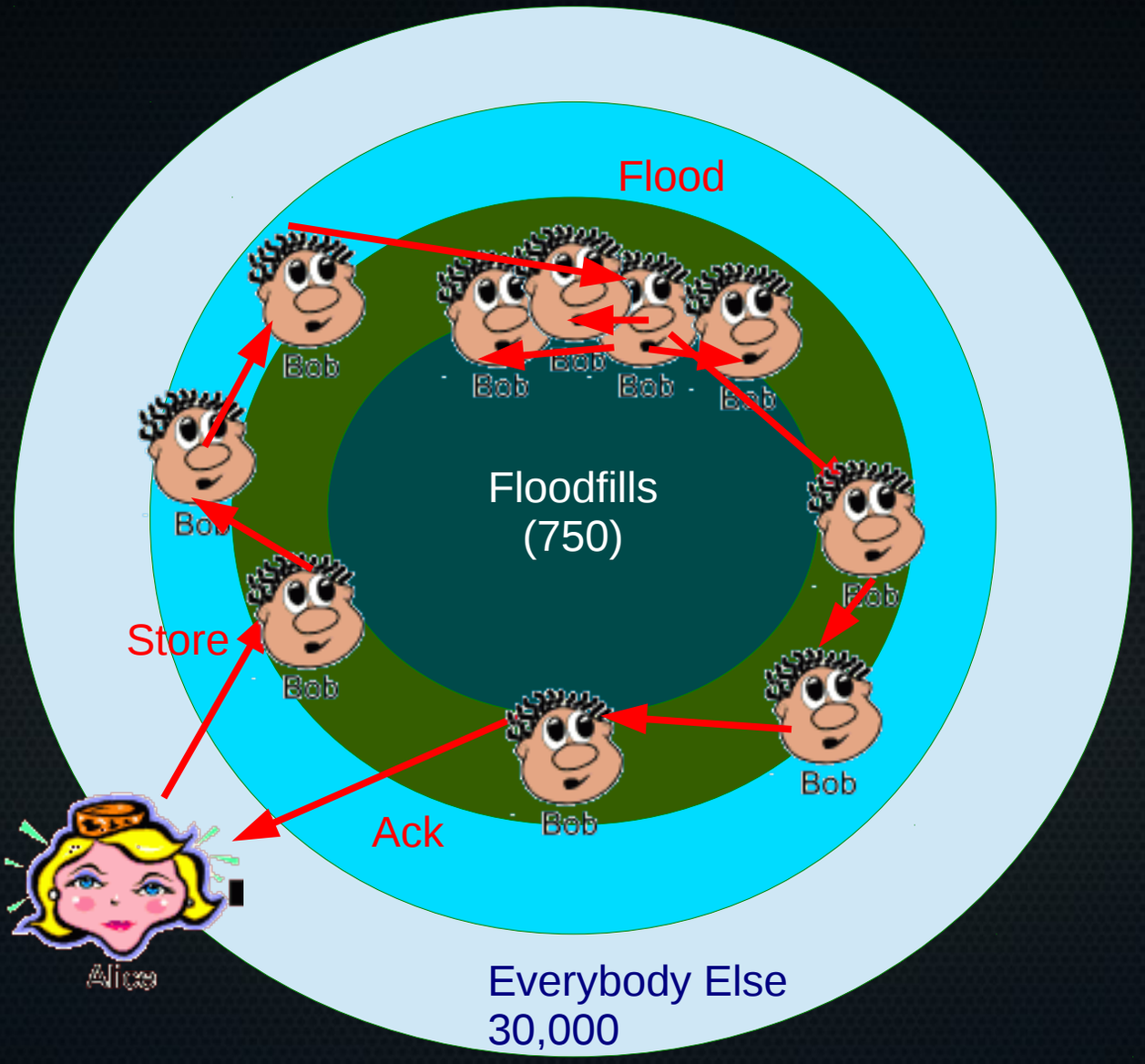
HACK LAB .TO

to crypto

Exploratory Tunnels



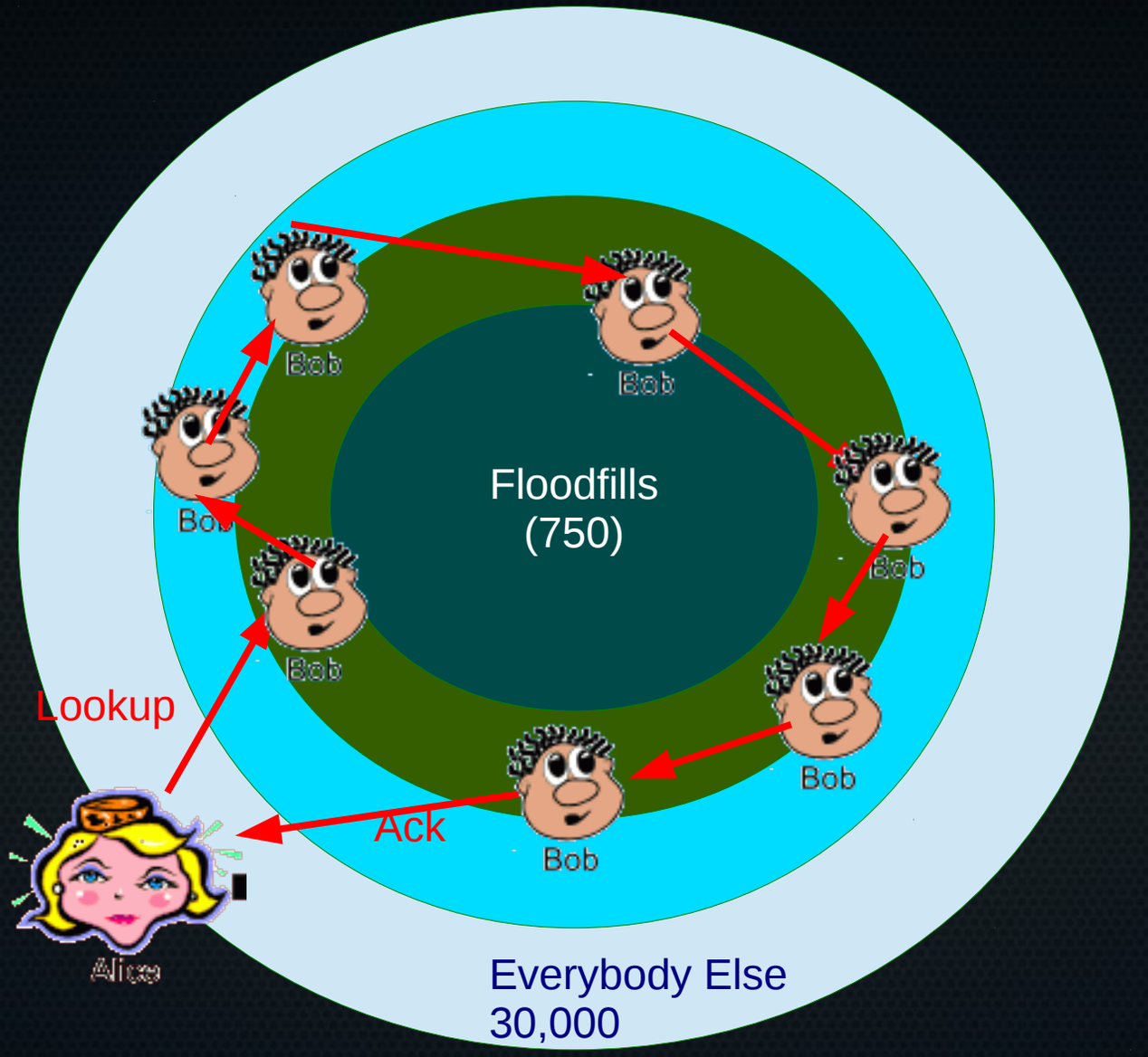
LS Store



HACK LAB .TO

to crypto

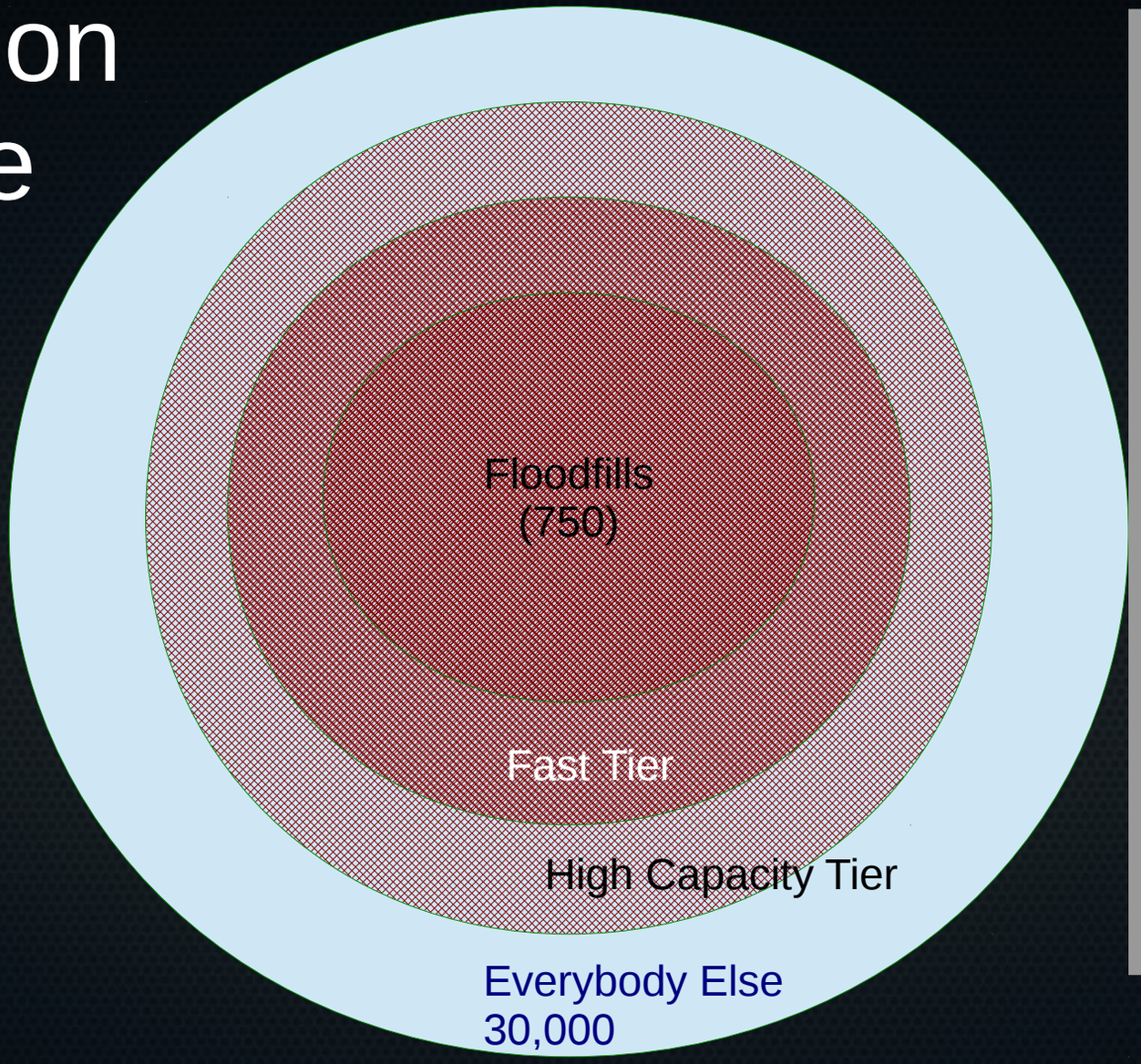
LS Lookup



HACK LAB .TO

to crypto

Connection lifetime



Full Mesh:
The inner tiers tend to stay connected to each other for long times (never idle timeout), whether DHT-close to each other or not.

Meanwhile, the everybody-else tier connects into the inner tiers for short times. These connections fail when the initial data is after the idle timeout from the tunnel build.

Classes seen from a M floodfill

1486 low-bandwidth

- 2 KR
- 4 KU
- 5 L
- 1 LfR
- 437 LR
- 183 LU
- 9 M
- 30 MfR
- 1 MfU
- 582 MR
- 232 MU

- 1 N
- 106 NfR
- 1 NfU
- 118 NR
- 56 NU
- 138 OfR
- 6 OfU
- 98 OR
- 13 OU

640 floodills

- 204 POfR
- 9 POfU
- 145 POR
- 32 POU
- 1 XOf
- 137 XOfR
- 7 XOfU
- 71 XOR
- 17 XOU

HACK LAB .TO

to crypto

Solutions already implemented

- Increase conn limits
- New bandwidth tiers with higher conn limits
- Reduce idle timeout when approaching limit
- Reject tunnels when approaching limit
- Drop tunnel requests when approaching limit and reply peer isn't connected
- Stop exploration when approaching limit
- Rebuild same tunnel at expiration
- Disable tunnel testing
- Adjust published cost based on conn limits, peers will select accordingly
- Reduce expl. tunnel length under severe stress, and other congestion collapse avoidance strategies

HACK LAB .TO

to crypto

Solutions implemented but can/should be tweaked

- Select closest obep/ibgw pair (but can hurt reliability if we pick the same pair after failure)
- Send reply via expl. tunnel
- Send replies direct if connected, else via expl tunnel (longer connection times)
- Min idle time selected so first data in new tunnel doesn't require reconnect

HACK LAB .TO

to crypto

Non-solutions

- Tunnels close to youself
  - no, anon/attack issues
- Tunnels close to each other
  - no, attack issues

HACK LAB .TO

to crypto

Possible improvements

- Bias away from floodfills
- Semi-recursive storage?
- Make shared clients close-on-idle (much slower browsing load)

HACK LAB .TO

to crypto

Current Conn Limits 0.9.21

Share Bandwidth	Class	NTCP	SSU	If floodfill
< 12 KBps	K	50	150	
12-48	L	100	300	
48-64	M	150	450	255 / 765 Not enough!
64-128	N	200	600	340 / 1020
128-256	O	350	1050	595 / 1785
256-2000	P	450	1350	765 / 2295
> 2000	X	600	1800	1020 / 3060

HACK LAB .TO

to crypto

Research

- We still don't know which is more expensive, NTCP or SSU
- What does Tor do?
- What is reasonable capacity for data center routers, VMs, etc.
- Floodfills are also the fast peers - split them? what portion non-floodfill fast?

HACK LAB .TO

to crypto