



# A Heuristic for IP Summarization in perfSONAR Performance Monitoring Architecture

Distributed and Meta-Systems Laboratory

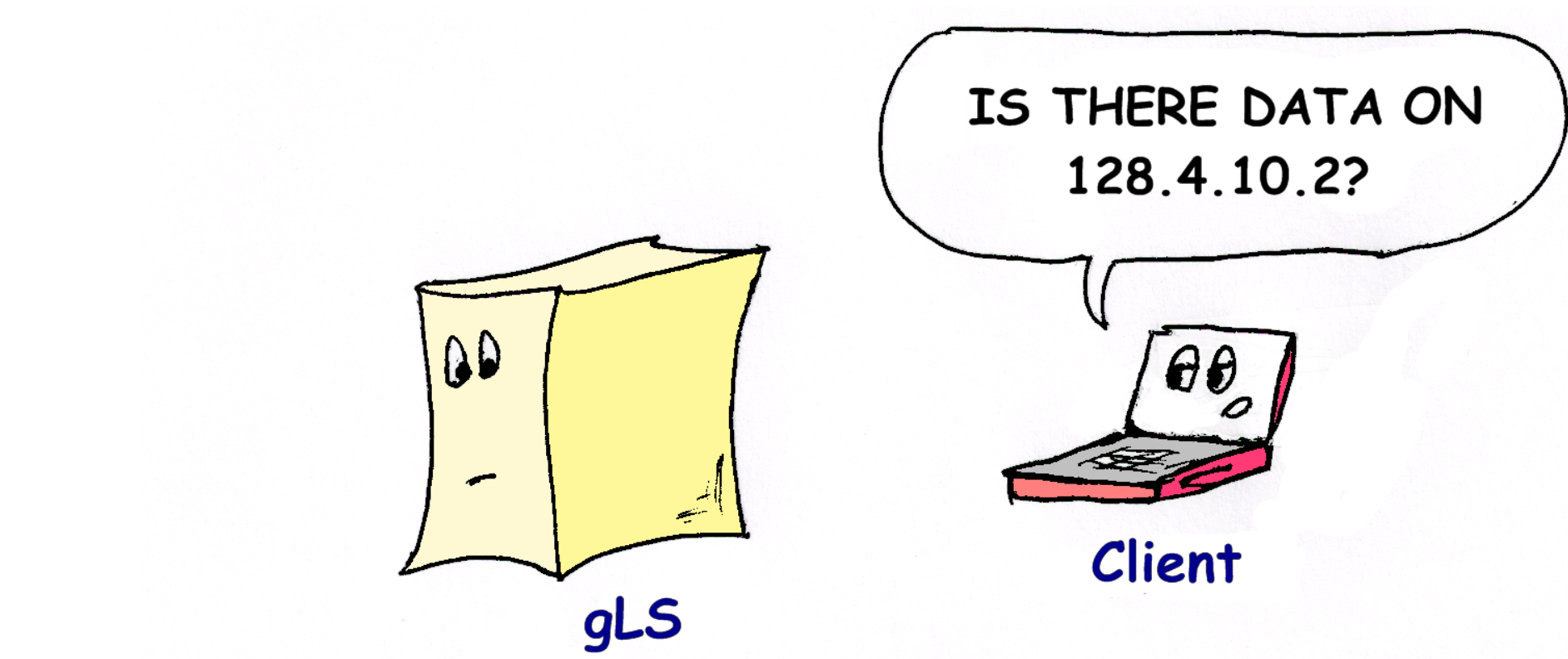
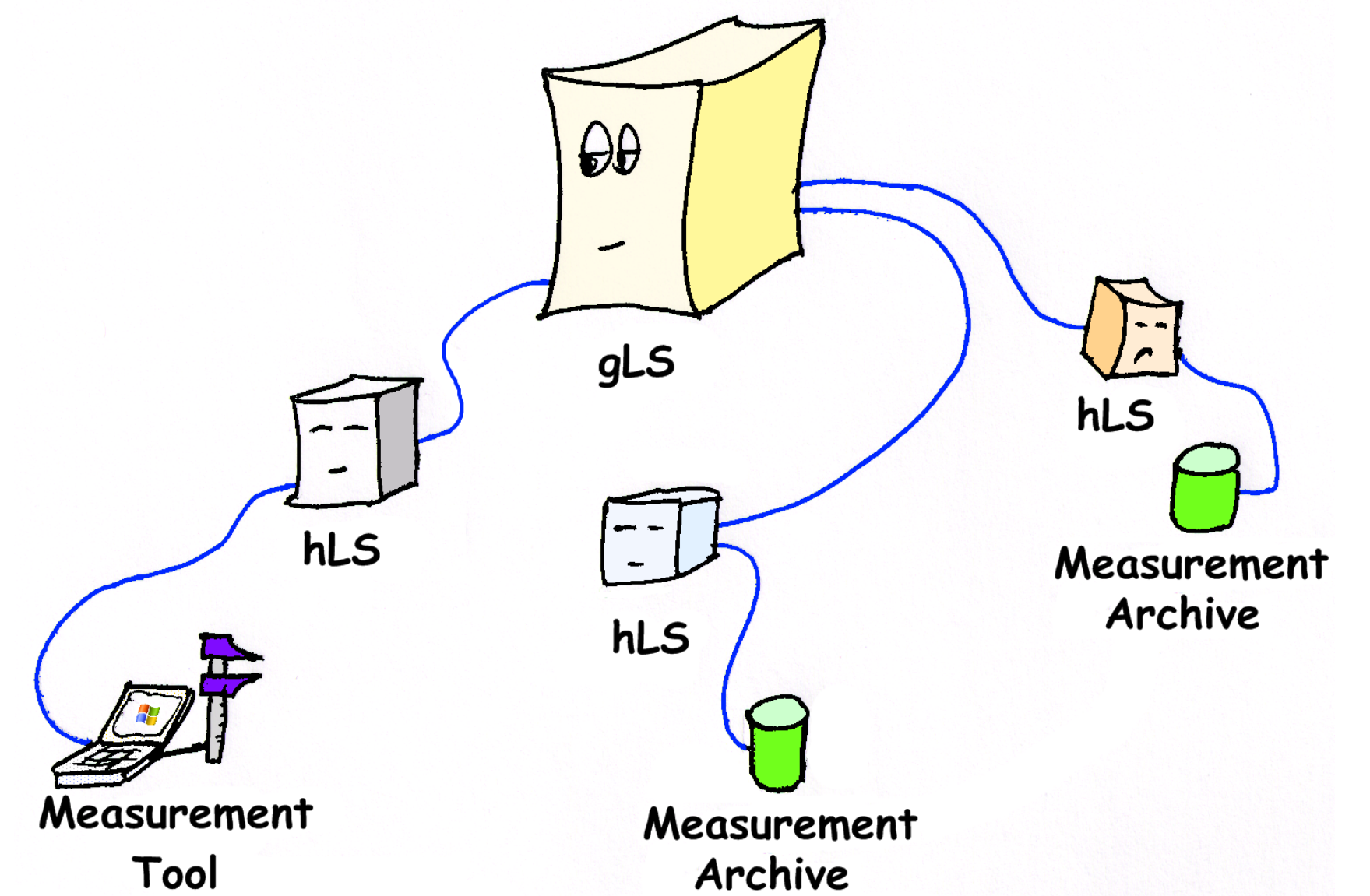
Computer and Information Sciences, University of Delaware, Newark, Delaware

**perfSONAR** (PERformance Service Oriented Network monitoring Architecture)

• Services-oriented infrastructure for network performance monitoring.

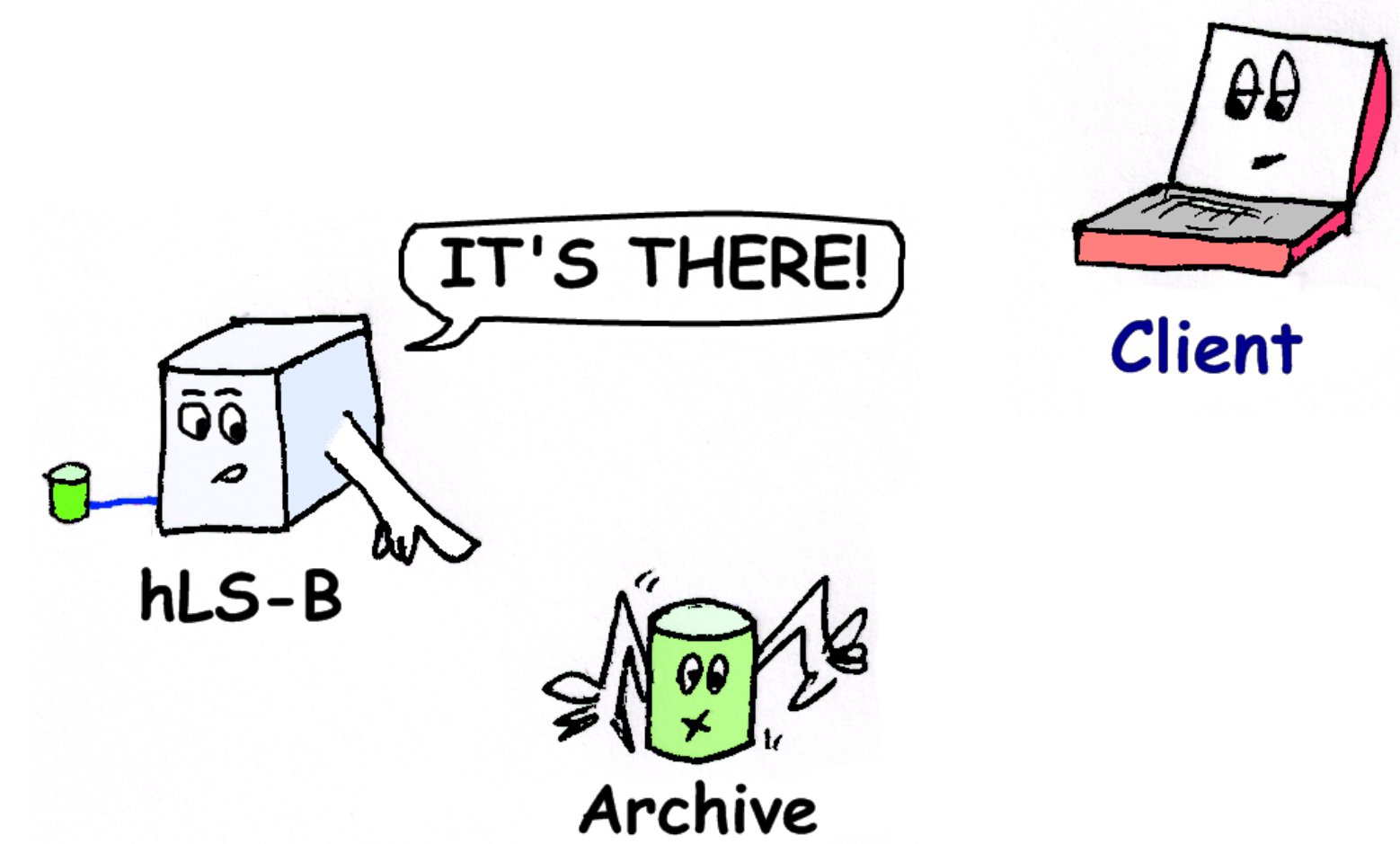
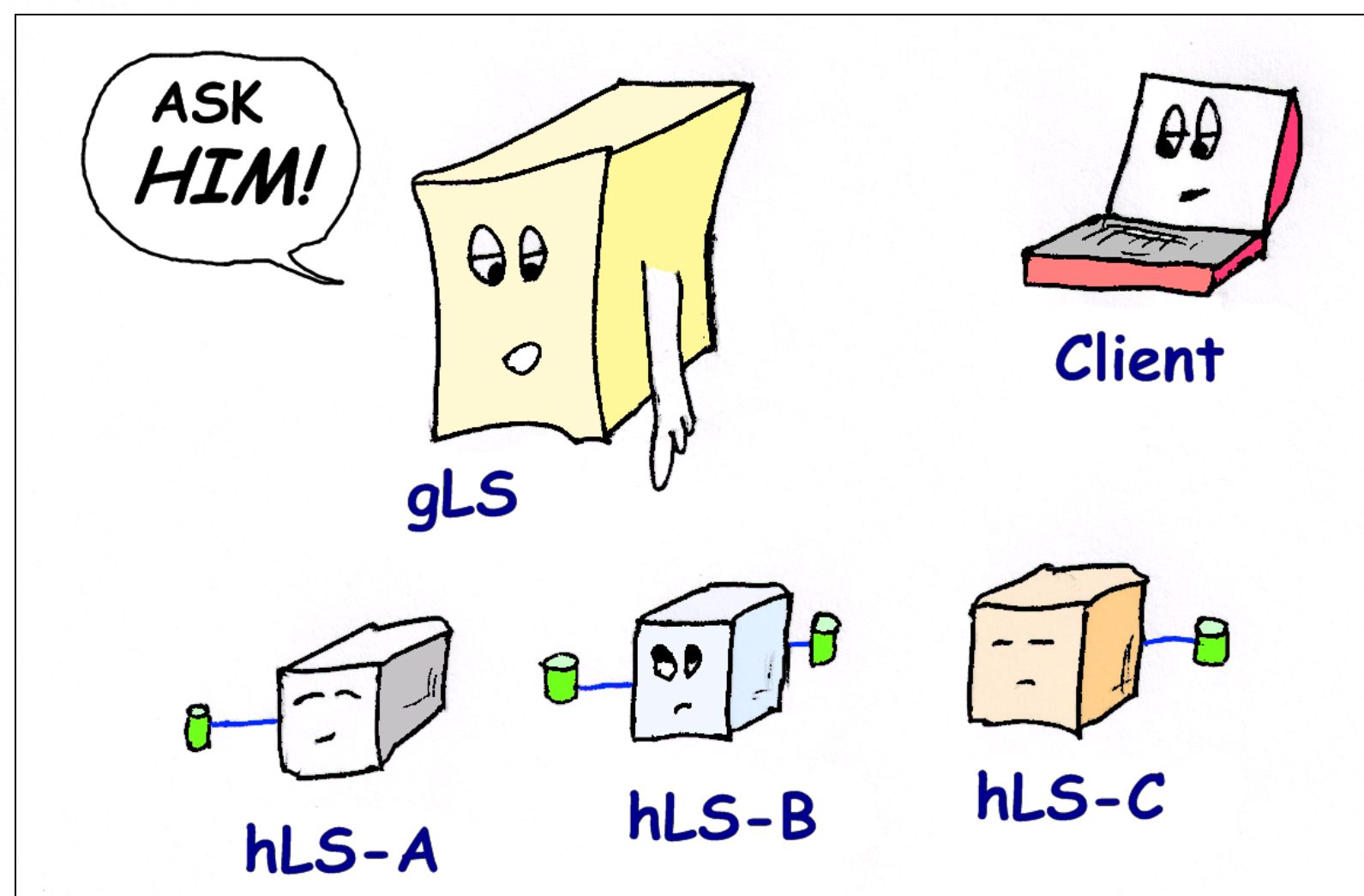
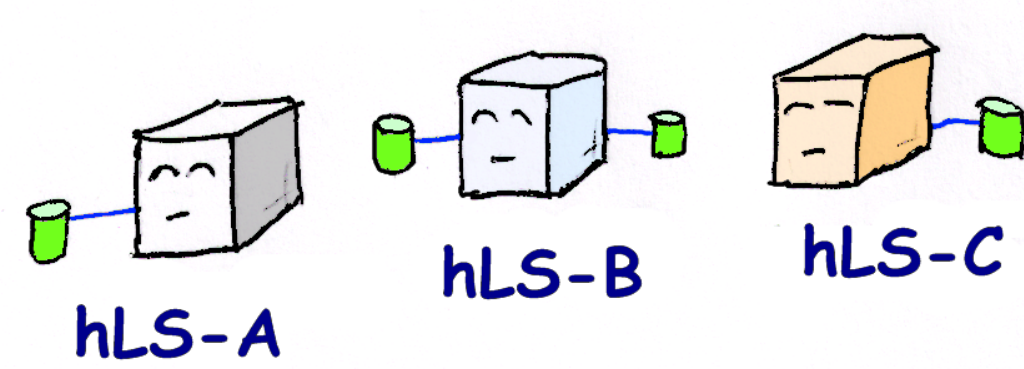
The **Lookup Service** is a directory service within perfSONAR. Composed mainly by:

- **hLS**: local directory of local services (measurement tools, measurement archives, etc.);
- **gLS**: global directory of local directories.

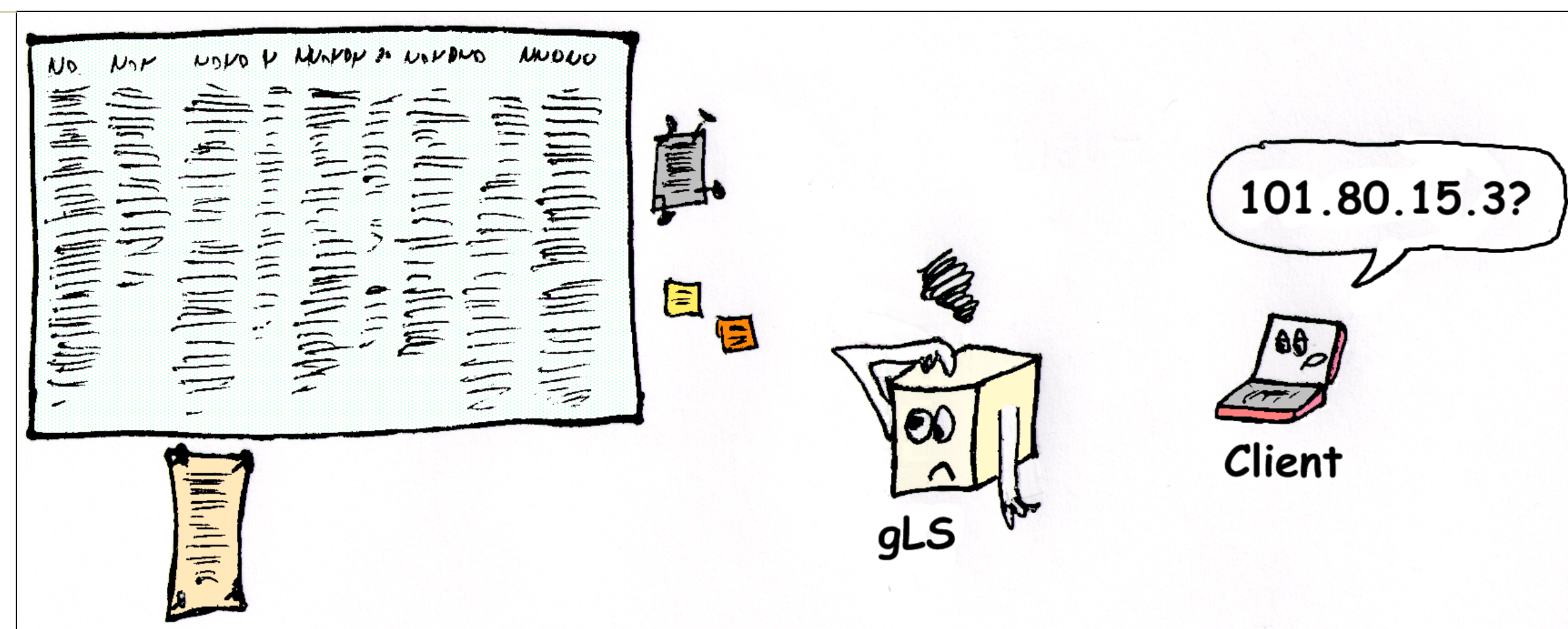
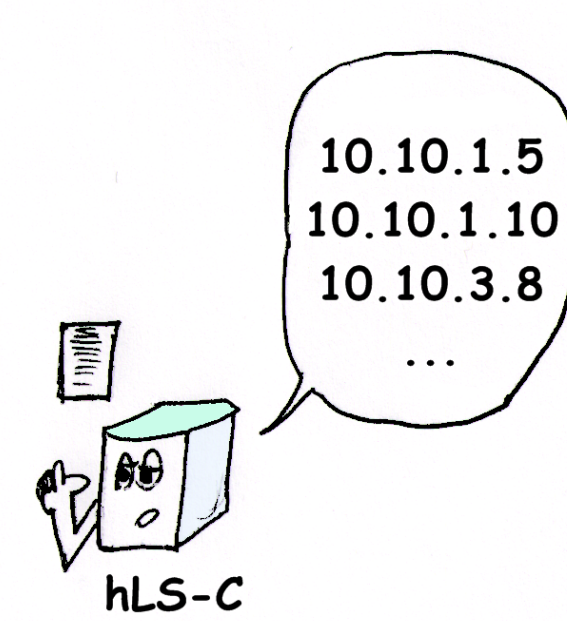
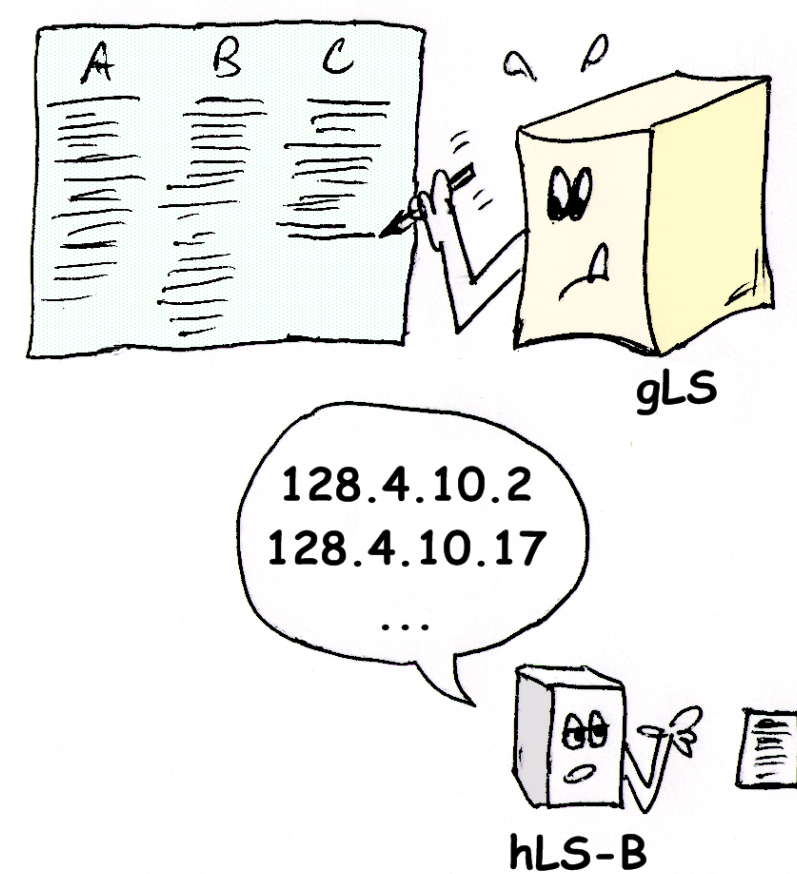


When a client wants to **find** available **services** regarding some network:

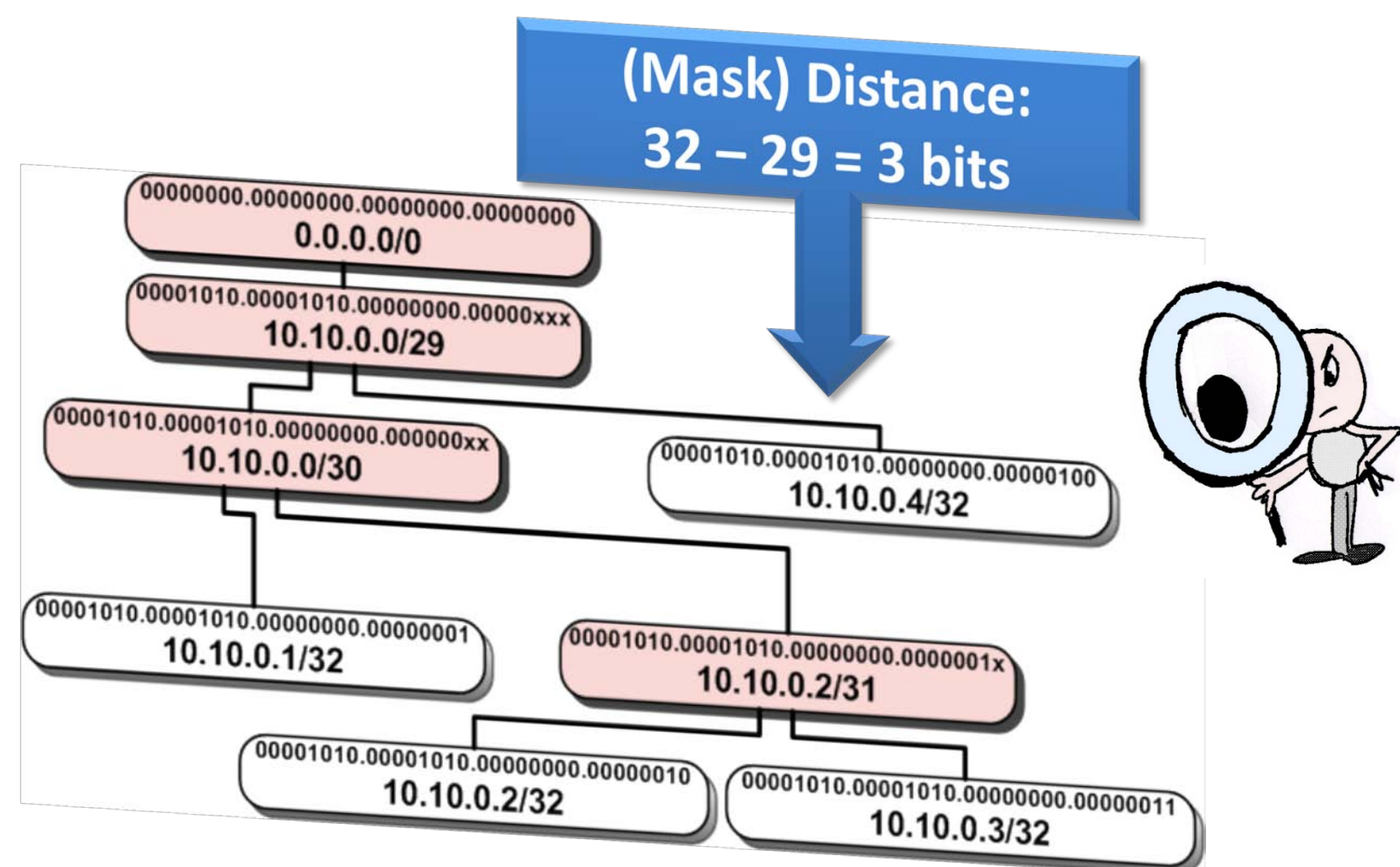
- Client queries gLS;
- gLS informs which hLS controls the desired network;
- Client queries the hLS and receives list of services in that network.



For this scheme to work, each hLS must periodically **publish** their directories to the gLS. But publishing the entire content of the directories may result in **network overload** and demand **excessive resources** from the gLS.



hLSs then **summarize** their directory data before publishing to avoid pushing large amounts of data into the network, and to optimize gLS resources and query time. Specifically for IP addresses, our heuristic accomplishes summarization by employing **IP subnet addresses** to represent the actual host IP addresses controlled by a hLS. The heuristic constructs a special data structure – a **PATRICIA tree** – within which the **inner nodes** are the subnet addresses, and the **leaves**, the actual host IP addresses.

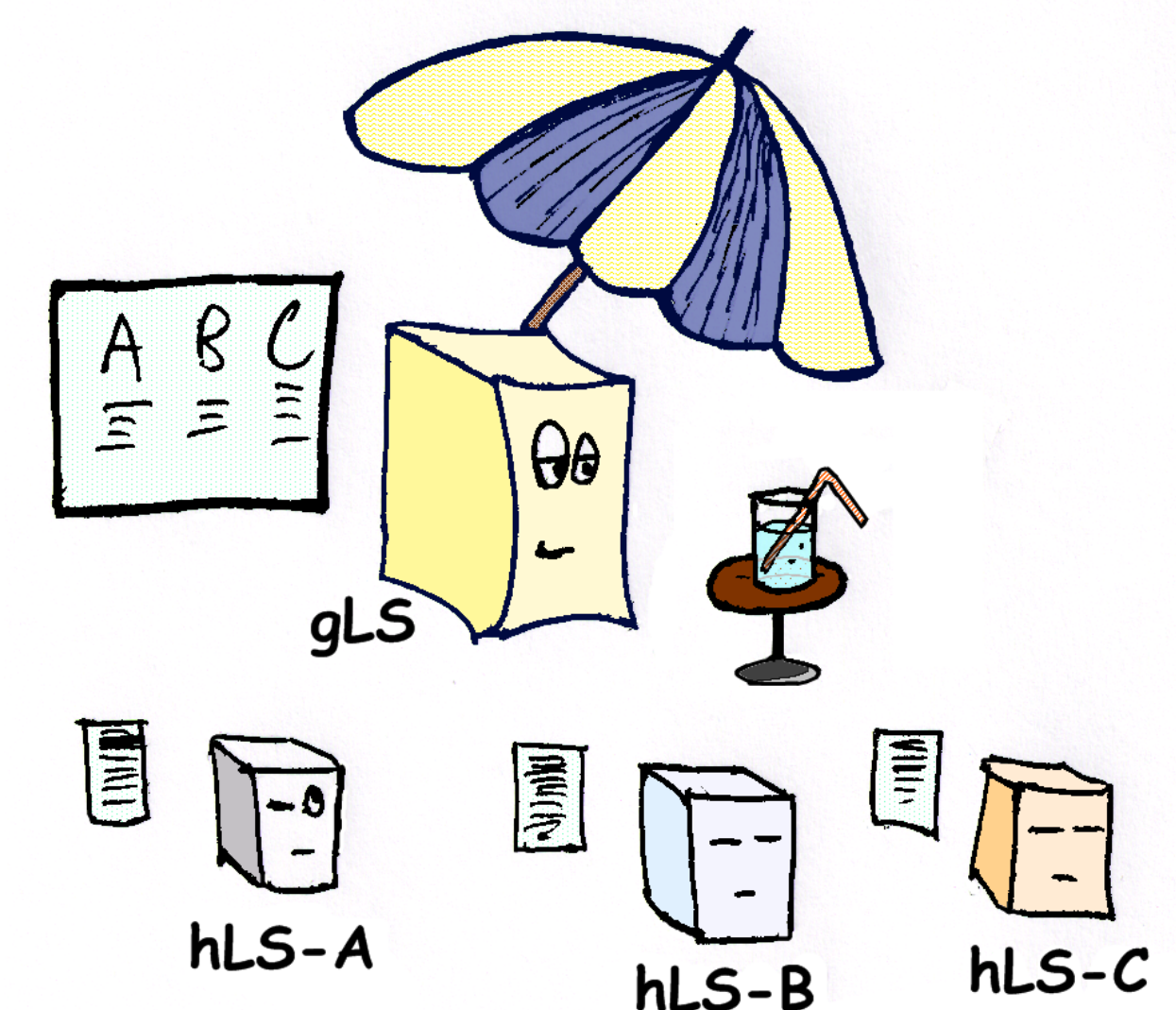


The heuristic uses three **metrics** to decide which inner nodes to pick:

- **Distance**: notion of how many IPs a subnet claims, but do not actually exist in the network;
- **Density**: number of actual IP addresses over total number of possible IPs in a subnet;
- **Minimum Subnet Mask**: avoids too large subnets.

The heuristic uses the metrics and tries to select subnets that do not summarize too much (imagine 0.0.0.0, that represents all possible IPv4 addresses), or summarize too few (less compression efficiency).

After the heuristic selects the summarizing nodes, the hLS publishes the **compressed directory** data. The gLS can perform optimized queries and better manage storage space.



Acknowledgment: Internet2