Problem Definition

- Given a collection of candidate path sets *P* under all possible designs, how well can we monitor the network using path measurement and which design is the best?
- Monitoring performance is measured by the number of nodes that are *k-identifiable* w.r.t *P*
- The optimal solution is hard due to the exponential number of path sets
- We focus on bounding the number of *1-identifiable nodes*, *since the upper bound on 1-identifiable would be an upper bound on k-identifiable as well*

General Network Monitoring Arbitrary routing

Theorem (Identifiability under arbitrary routing). *Given a network with n nodes and m monitoring paths, the maximum number of identifiable nodes under arbitrary routing satisfies*

$$\psi^{AR}(m, n) \leq \min\{n; 2^m - 1\}.$$

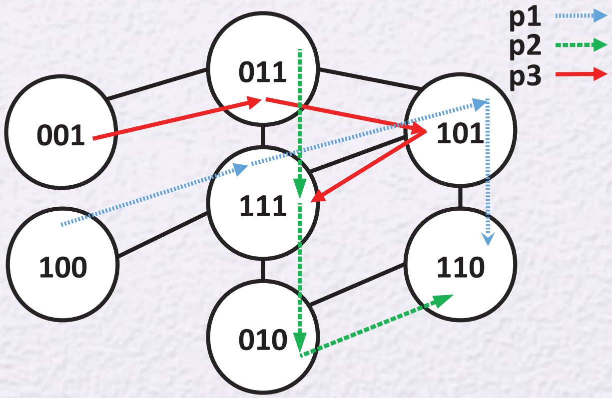
General Network Monitoring Arbitrary routing

Theorem (Identifiability under arbitrary routing). *Given a network with n nodes and m monitoring paths, the maximum number of identifiable nodes under arbitrary routing satisfies*

$$\psi^{AR}(m,n) \leq \min\{n; 2^m - 1\}.$$

The bound is tight since we can construct a topology with m monitoring paths that meets this bound:

1. *Take up to 2^m nodes*
2. *Give binary enumeration*
3. *Construct paths*
4. *Create the edges of the graph*



General Network Monitoring Consistent routing

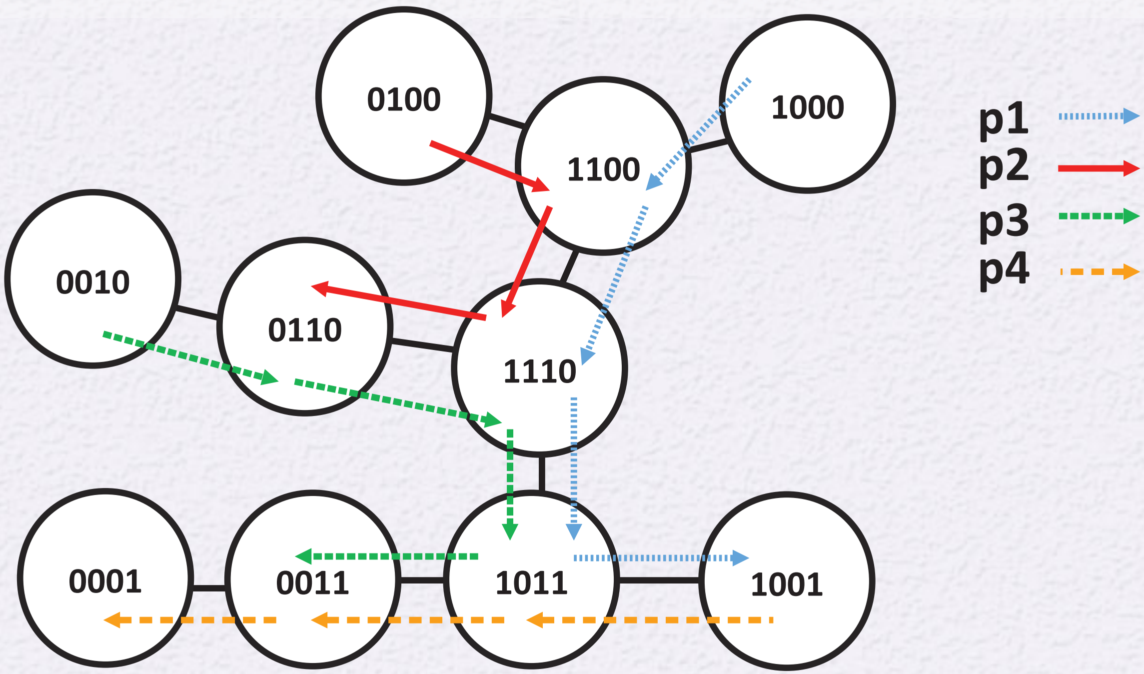
II. Consistent routing

Definition *A set of paths P is* consistent *if* $\forall p,\ p' \in P$ *and any two nodes u and v traversed by both paths (if any), p and p' follow the same sub-path between u and v.*

Definition

We define the *path matrix* of $\hat{p}_i$ as a binary matrix $M(\hat{p}_i)$, in which each row is the binary encoding of a node on the path, and rows are sorted according to the sequence $\hat{p}_i$. Notice that by definition $M(\hat{p}_i)|_{*,i}$ has only ones, i.e., $M(\hat{p}_i)|_{r,i} = 1,\ \forall r$.

Example of path matrix



$$M(\hat{p}_3) = \begin{array}{c} \text{flips} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} b_1 & b_2 & b_3 & b_4 \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \end{array}$$

General Network Monitoring
Consistent routing

Lemma *Under the assumption of consistent routing, all the columns in all the path matrices have consecutive ones.*

Lemma *Given* $m = |P| > 1$ *consistent routing paths, whose length is at most* d^* *(in number of nodes), the maximum number of different encodings in the rows of* $M(\hat{p}_i)$ *is equal to* $\min\{2 \cdot (m-1), d^*\}$, $\forall p_i \in P$.

General Network Monitoring
Consistent routing

Theorem (Identifiability with consistent routing). *Given n nodes, and $m > 1$ consistent routing paths of length at most d^* (in number of nodes), the maximum number of identifiable nodes satisfies:*

$$\psi^{CR}(m,n,d^*) \leq \min\left\{\sum_{i=1}^{i_{max}}\binom{m}{i} + \left\lfloor \frac{N_{max} - \sum_{i=1}^{i_{max}} i \cdot \binom{m}{i}}{i_{max}+1} \right\rfloor ; n\right\},$$

where $i_{max} = \max\{k \mid \sum_{i=1}^{k} i \cdot \binom{m}{i} \leq N_{max}\}$,
and $N_{max} = m \cdot \min\{2 \cdot (m-1); d^*\}$.

General Network Monitoring Consistent routing

Proof

- Each *identifiable* node must have a unique encoding
- For every path matrix, we have 2*(m-1) possible different encodings, so totally m*min{2*(m-1), d*}
- We are counting multiple times the nodes that appear in multiple path matrices
- If encoding b has k digits equals to 1, then b appears among the rows of k different path matrices

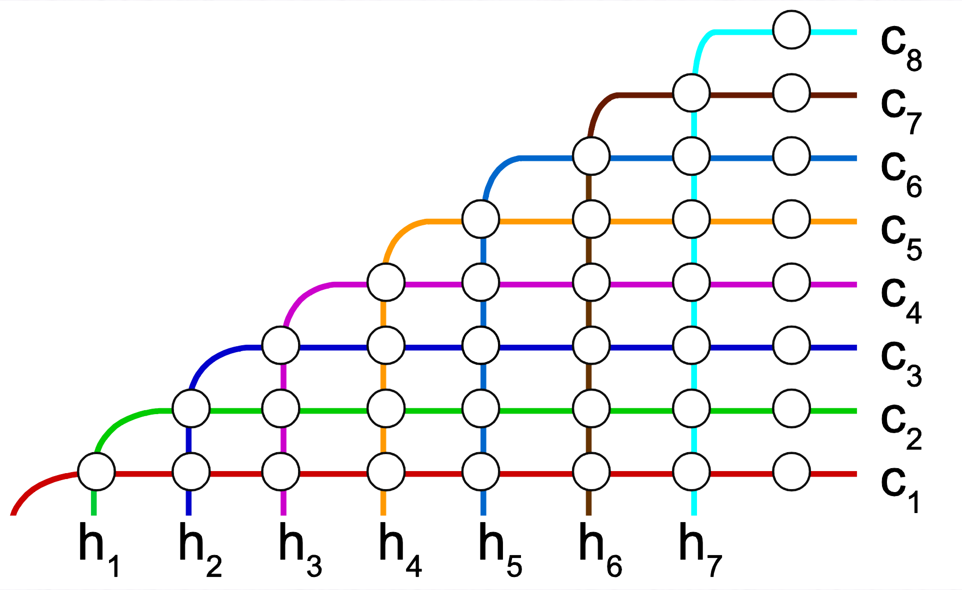
General Network Monitoring Consistent routing

Proof

- Number of distinct encoding is maximized when the number of duplicate encodings is minimized, therefore their number of ones is minimized
- Minimum number of duplicate is achieved when we have $\binom{m}{1}$ different encodings with only one digit equal to one, $2\binom{m}{2}$ with two digits equal to one appearing in two path matrices and so forth until total number of encodings is equal to N_{max}

General Network Monitoring
Consistent routing

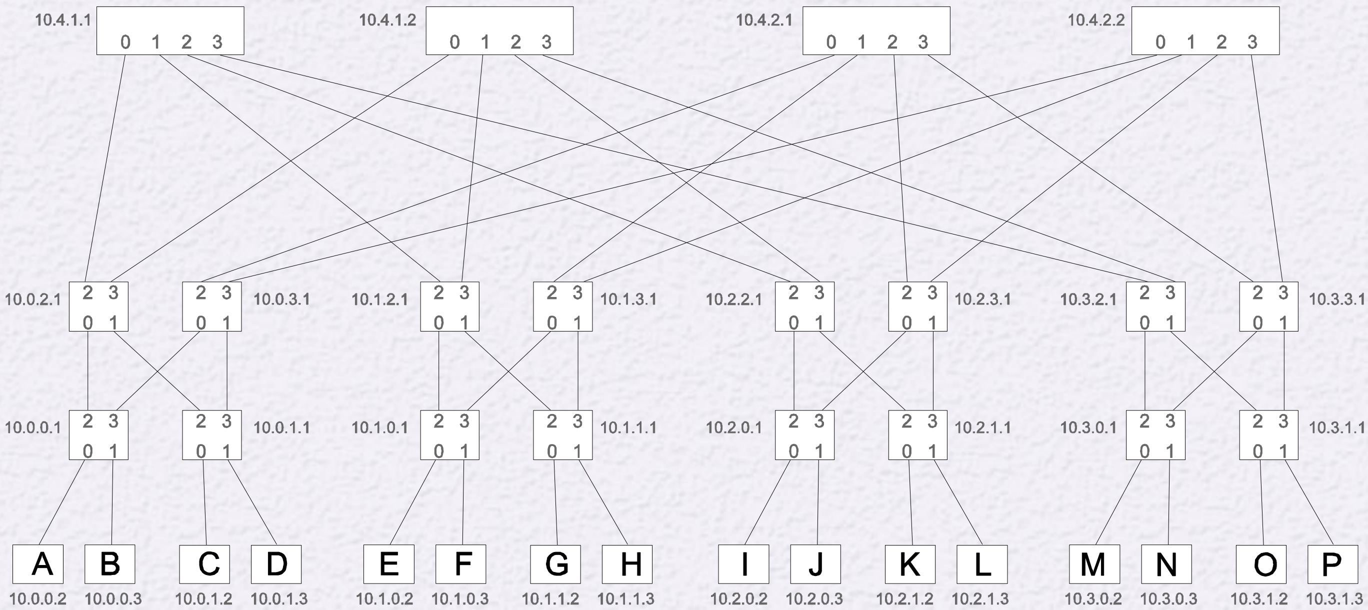
Tightness of the bound on number of identifiable nodes under consistent routing



With n=36 nodes, m=8 monitoring paths of maximum length $d^*=8$, we have $N_{max}=\min\{112,64\}=64$, $i_{max}=2$, and

$\psi^{cr}= \binom{8}{1} + \binom{8}{2} + \left\lfloor\frac{0}{3}\right\rfloor=36.$

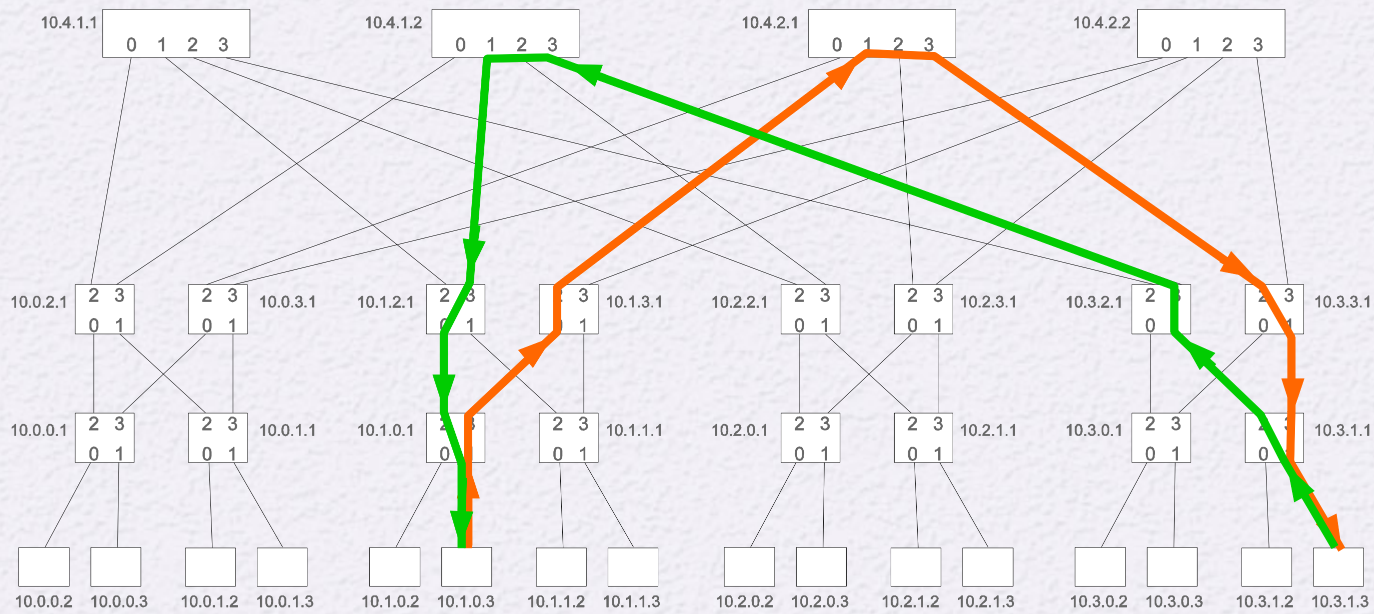
General Network Monitoring
The Case of Half-Consistent Routing



Fat-tree topology (common in data centers), where we assume the routing scheme based on IP addressing of clients and switches as described in

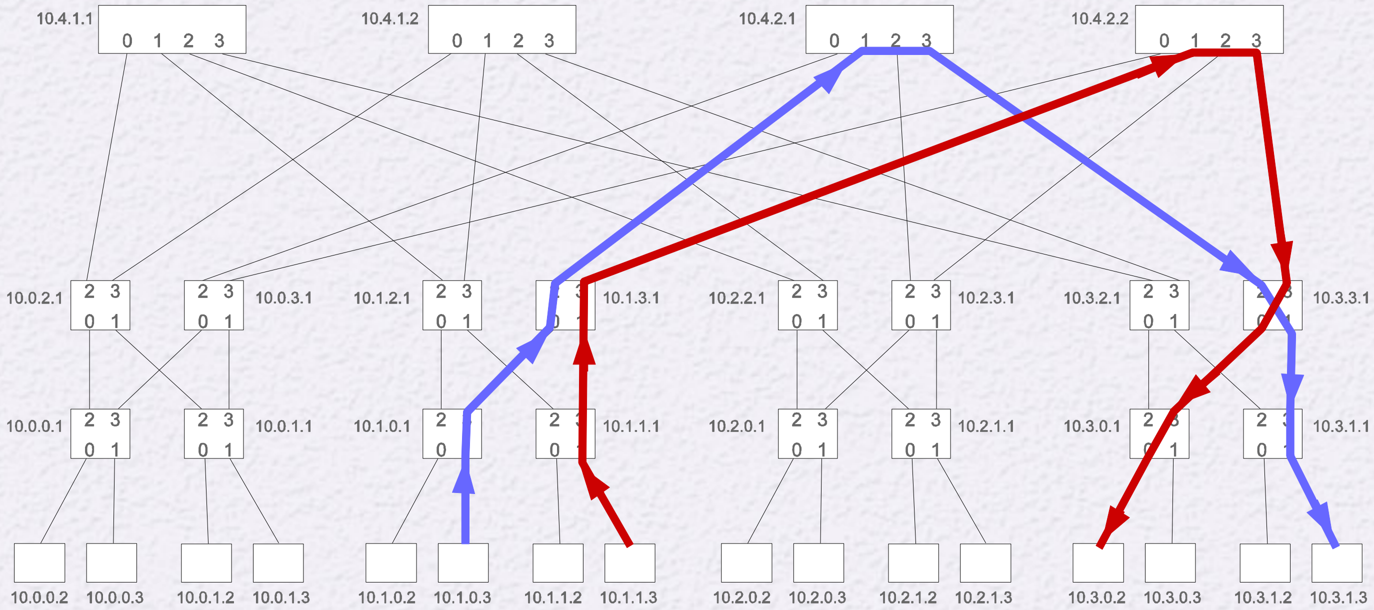
M. Al-Fares, A. Loukissas, A. Vahdat, "A Scalable, Commodity Data Center Network Architecture", ACM SIGCOMM 2008

General Network Monitoring
Half-Consistent routing



Example of half-consistent routing in a fat-tree (based on IP address masks)

General Network Monitoring
Half-Consistent routing



Example of half-consistent routing in a fat-tree (based on IP address masks)

General Network Monitoring
Half-Consistent routing

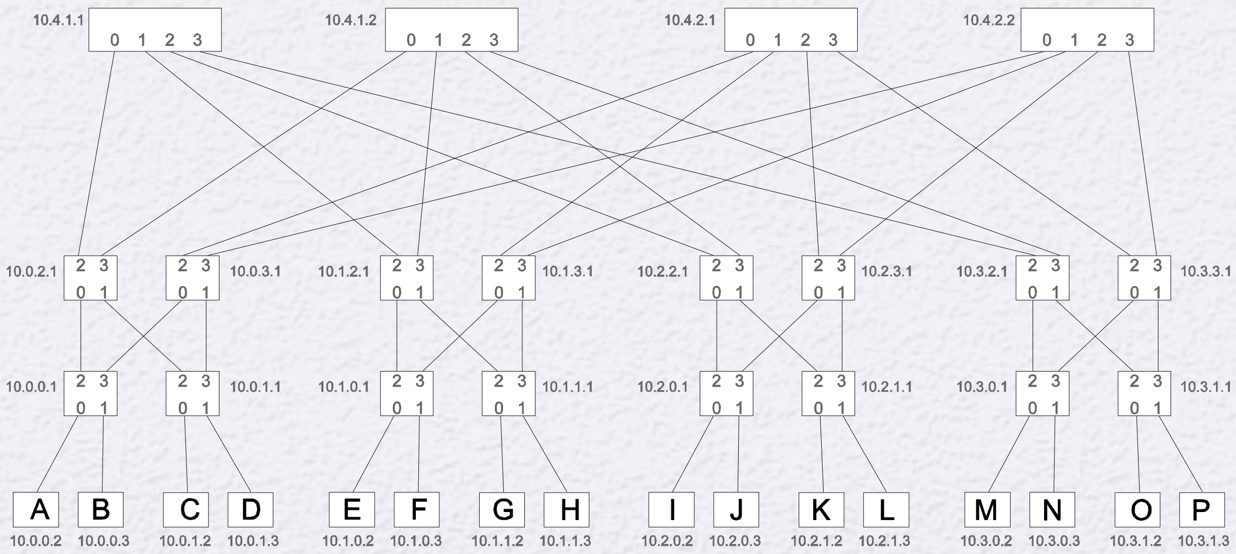
III. Half-consistent routing

Definition: *If a routing scheme guarantees that any path* $p_i \in P$ *can be divided into two segments* $s_1(p_i)$ *and* $s_2(p_i)$*, such that the property of routing consistency holds for the set* $P_{1/2} = \cup_{pi \in P} \{s_1(p_i), s_2(p_i)\}$*, then the routing scheme is called* half-consistent.

General Network Monitoring
Half-Consistent routing

Lemma:

Any shortest-path routing scheme on a fat-tree is half-consistent.



General Network Monitoring
Half-Consistent routing

Lemma. *Given a path* $p_i \in P$ *of maximum length* d^*, *under the assumption of half-consistent routing, with* $m = |P| > 1$ *monitoring paths, the maximum number of different encodings in the rows of* $M(\hat{p}_i)$ *is* $\min\{2^{m-1}, 4*(m-1), d^*\}$.

General Network Monitoring
Half-Consistent routing

Theorem (Half-consistent routing)**.** *In a general network with n nodes, m > 1 monitoring paths, diameter d*, under half-consistent routing, the number of identifiable nodes is upper bounded by*

$$\psi^{hcr}(m,n,d^*) \leq \sum_{i=1}^{i_{\max}} \binom{m}{i} + \left\lfloor \frac{N_{\max} - \sum_{i=1}^{i_{\max}} i \cdot \binom{m}{i}}{i_{\max}+1} \right\rfloor$$

where $i_{max} = \max\{k \mid \sum_{i=1}^{k} i \cdot \binom{m}{i} \leq N_{max}\}$

and $N_{max} = m \cdot \min\{2^{(m-1)}, 4 \cdot (m-1); d^*\}$.

Performance evaluation

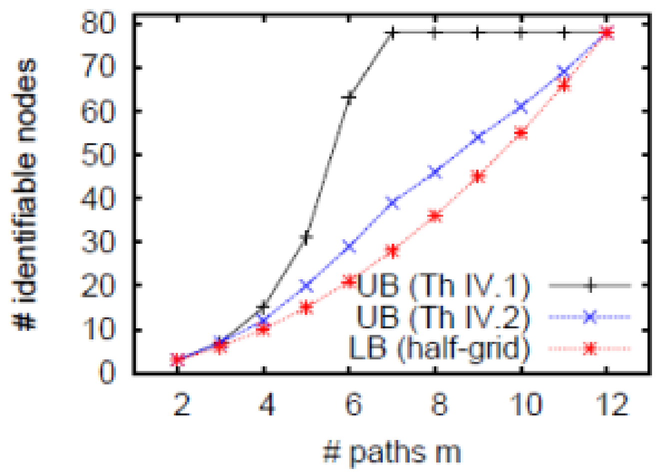


Fig. 14. Upper and lower bounds on the number of identifiable nodes in a half-grid network with $n = 78$, varying m, and $d^* = 12$.

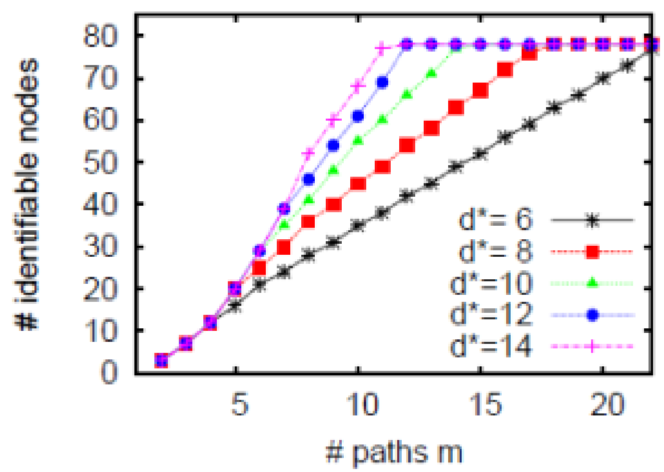
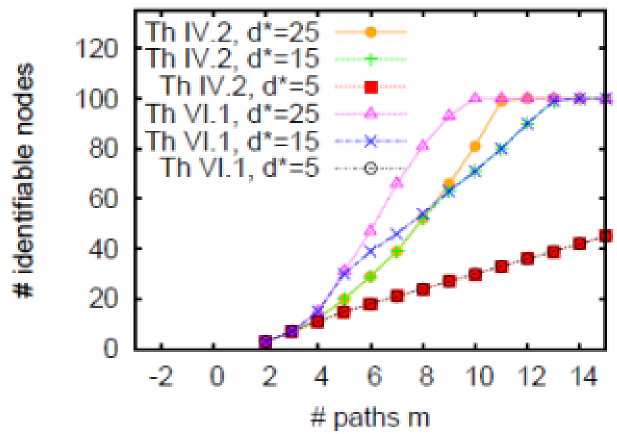


Fig. 15. Bound under consistent routing (Theorem IV.2) with varying number of paths and maximum path length (network as in Figure 14).

Performance evaluation



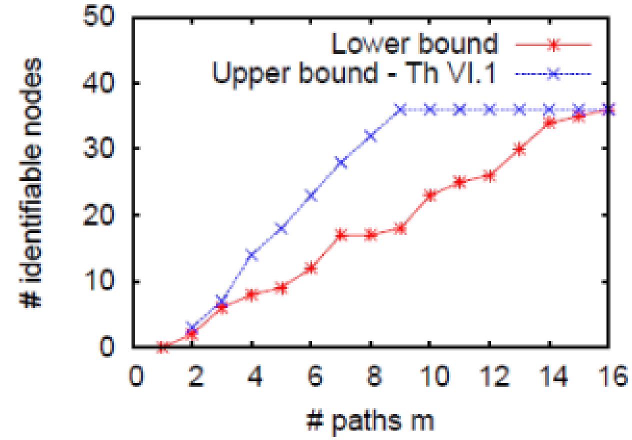


Fig. 20. Comparison between the bounds of Theorem IV.2 (consistent routing) and A.1 (half consistent routing) - 100 nodes, varying d^*.

Fig. 21. Comparison between the bound of Theorem A.1 (half-consistent routing) and a lower bound for a 4-ary fat-tree with three layers.

Conclusions

- The problem of maximizing number of nodes whose states can be identified via Boolean tomography can be seen as graph-based group testing
- Upper bound on the number of identifiable nodes under different routing assumptions has been derived
- Provides insight for the design of topologies and monitoring schemes with high identifiability

Open problems

- Current bounds are topology agnostic. What if we know the adjacency matrix of our network topology?
- Algorithms for monitor deployment and path selection, with the objective to maximize node identifiability.
- We typically have partial knowledge and partial controllability.
  - Some nodes are known to be working, some others are known to be broken. There is a grey area where we want to assess damages. How does this change the algorithms?
  - Monitors can only be placed in our own routers. We don't own the entire network. What is the best we can do with the nodes that we can control?
- Some nodes/paths are more important than others, how can we design algorithms that prioritize identifiability of given nodes?
- Provide further insight for the design of topologies and monitoring schemes with high/low identifiability

Thank You!