

Received October 20, 2020, accepted December 16, 2020, date of publication January 1, 2021, date of current version January 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3048876

Topology Agnostic Bounds on Minimum Requirements for Network Failure Identification

VIVIANA ARRIGONI¹, (Member, IEEE), NOVELLA BARTOLINI¹, (Senior Member, IEEE),
AND ANNALISA MASSINI¹, (Member, IEEE)

Department of Computer Science, Sapienza University of Rome, 00198 Rome, Italy

Corresponding author: Viviana Arrigoni (arrigoni@di.uniroma1.it)

This work was supported by the North Atlantic Treaty Organization (NATO), under the Science for Peace and Security grant, number SPS G4936.

ABSTRACT In Boolean Network Tomography (BNT), node identifiability is a crucial property that reflects the possibility of unambiguously classifying the state of the nodes of a network as 'working' or 'failed' through end-to-end measurement paths. Designing a monitoring scheme satisfying network identifiability is an NP problem. In this article, we provide theoretical bounds on the minimum number of necessary measurement paths to guarantee identifiability of a given number of nodes. The bounds take into consideration two different classes of routing schemes (arbitrary and consistent routing) as well as quality of service (QoS) requirements. We formally prove the tightness of such bounds for the arbitrary routing scheme, and provide an algorithmic approach to the design of network topologies and path deployment that meet the discussed limits. Due to the computational complexity of the optimal solution, We evaluate the tightness of our lower bounds by comparing their values with an upper bound, obtained by a state-of-the-art heuristic for node identifiability. For our experiments we run extensive simulations on both synthetic and real network topologies, for which we show that the two bounds are close to each other, despite the fact that the provided lower bounds are topology agnostic.

INDEX TERMS Boolean network tomography, identifiability, network topology, optimal bounds.

I. INTRODUCTION

With the massive growth of the Internet, localizing node failures has become a crucial task. Single organizations have direct access only to limited portions of the internal nodes of the network, and they hardly collaborate in sharing internal performance observations because of commercial conflicts. Boolean Network Tomography (BNT) provides tools for assessing the state of a network through end-to-end monitoring paths, as they do not rely on administrative access privileges, [1]. Boolean Network Tomography overcomes the limitations faced by traditional network monitoring approaches based on pervasively deployed monitoring agents (e.g., SNMP) or pervasively supported network protocols (e.g., traceroute) caused by the complexity and the heterogeneity of modern computer communication networks. As a matter of facts, bugs and configuration errors in various customer software and network functions often induce "silent

failures" that are only detectable from end-to-end connection states, [2].

Each monitoring path can be modelled as the sequence of nodes that it traverses. When a packet correctly reaches its final destination, we say that its path is in a *working* state, and so are its traversed nodes. On the other hand, when packet losses occur along a path, we say that the path *failed*. The latter situation arises when at least one of the nodes of the path is failed. Observations of the outcome of monitoring paths (working/failed) induce a system of Boolean equations where the unknowns are the Boolean states of the nodes in the network. The challenge related to this approach is that such Boolean systems are commonly under-determined, hence allowing multiple solutions, i.e., multiple failing scenarios that lead to the same observations on the path probe outcome, [3]. When the state of a failed node can be uniquely determined from the Boolean system induced by the outcome of monitoring paths, the node is said to be *identifiable*. Therefore, identifiability is a desirable property that allows unambiguous nodes' state classification.

The associate editor coordinating the review of this manuscript and approving it for publication was Keivan Navaie¹.

In this work, we provide topology-agnostic lower-bounds to the minimum number of measurement paths which are necessary to guarantee identifiability to a desired number of nodes. Such bounds represent the dual solution to the optimization problem studied in [4], where we introduced upper-bounds to the maximum number of identifiable nodes given a number of monitoring paths. In contrast with existing literature, we propose theoretical lower-bounds that cannot be violated, independently of the specific characteristics of the topology. The bounds formulations are only based on the number of nodes to identify, on high level routing consistency properties (*arbitrary* and *consistent routing*), and on QoS requirements, expressed in terms of maximum allowed path length. Motivated by the need to complement the analysis of [4], our bounds are a useful tool to measure the capability of a monitored topology to efficiently identify the status of its components. Implementing a monitoring system comes with the cost of installing monitors on the nodes of a network and of traffic caused by path probing; with this work, we aim at providing fundamental guidelines and minimal requirements for achieving the desired level of network identifiability (e.g., number of identifiable nodes). In addition, we formalize the Incremental Crossing Arrangement (ICA) procedure to generate monitoring schemes and underlying topologies that meet the bounds tightly, giving insights on which topology is the most suitable for failure localization. With ICA we formalize a network engineering approach which is at the basis of the theoretical analysis started in [4] and completed in this article.

We hereby list the major contribution of this work.

- We study theoretical bounds on the minimum number of paths to deploy in a network for identifying a desired number of nodes. The bounds do not depend on specific network topologies (i.e., they are topology agnostic bounds).
- We provide the Incremental Crossing Arrangement (ICA) algorithm, which allows topology design meeting the proposed bounds.
- We evaluate the tightness of our bounds on both synthetic and real network topologies. For this purpose we compare the bounds with the results of a state-of-the-art greedy algorithm, hereby referred to as Greedy for Identifiability (GI), for maximizing network identifiability by means of client-to-server probing paths [5].

II. RELATED WORK

Boolean Network Tomography studies non linear relationships existing between the paths and their components, as in the case of congestion or failure localization. The early works on this topic focused on best-effort inference. For example, Duffield [6], [7] and Kompella *et al.* [2] aimed at finding the minimum set of failures that can explain the observed measurements, and Nguyen and Thiran [3] aimed at finding the most likely failure set that explains the observations by a probabilistic analysis of a set of experiments. More recently

TABLE 1. Notation table.

Notation	Description
P	Set of m monitoring paths $P = \{p_1, \dots, p_m\}$
$p \in P$	Monitoring path as list of nodes
$b(v)$	Binary encoding of node v wrt P
$b(v) _i$	i -th element of $b(v)$ (equal to 1 iff $v \in p_i$, to 0 otherwise)
$\chi(v)$	Crossing number of node v wrt a set of paths P
P_F	Incident set of paths of a failure set F
$\mathcal{I}(p)$	Set of identifiable nodes traversed by path p
$M(p)$	Path matrix of path p
$\mathcal{B}(k)$	$\{b \in \{0, 1\}^m, s.t. \sum_{i=1}^m b _i = k\}, k = 1, \dots, m$
B_V	Set of binary encodings of a set of nodes V

and with a similar goal, the authors of [8] build a Markov Decision Process in light of passive traffic data, and solve the tomography problem with a Q -learning technique.

Lately, Ma *et al.* [9] give characterizations of maximum identifiability of node failure under different end-to-end monitoring systems, and extend this work in [10], where they outline topology-specific properties on the number of nodes whose states can be identified under a given number of failures. In contrast, as specified in [4], we provide general, topology-agnostic bounds. An optimal monitor placement for ensuring node k -identifiability under different routing schemes is described in [11]. This optimization problem was introduced by Bejerano *et al.* in [12], where they prove its NP-hardness. The work by Cheraghchi *et al.* [13] studied graph-constrained group testing with the goal of minimizing the number of monitoring paths needed to identify the state (defective or normal) of all network nodes, under the assumption that the maximum number of defective nodes is given. Differently from this work, we do not investigate on the identification of failed nodes in specific failure scenarios, but rather we focus on the identifiability property.

With this article, we complete the analysis of the fundamental bounds on node identifiability in Boolean Network Tomography introduced in [4].

III. PROBLEM FORMULATION

We represent a network as a undirect graph $G = (V, E)$, where V is the set of the nodes of G and E is the set of its edges. Each node v is either in *working* or *failed* state. In Table 1 we sum up the notation that will be used throughout this article. The state of the nodes is assessed indirectly by a set of *monitoring paths*, $P := \{p_1, \dots, p_m\}$, each being represented as the ordered sequence of nodes it traverses. Node failures cause paths disruption: when a path traverses a failed node, its communication is interrupted. On the other hand, paths traversing only working nodes are working. Each node v may be labeled with a binary encoding of length m , $b(v) \in \{0, 1\}^m \setminus 0^m$, where $b(v)|_i = 1$ if v is traversed by path p_i , $b(v)|_i = 0$ otherwise. We call *crossing number* of a node v , $\chi(v)$, the number of paths that traverse v , i.e., the number of 1s in its binary encoding ($\chi(v) = \sum_{i=1}^m b(v)|_i$). For each path p_i we define a *path matrix* as a binary matrix $M(p_i)$, in which each row is the binary encoding of a node on the path, and rows are sorted according to the sequence p_i . Notice that by

definition $M(p_i)|_{*,i}$ has only ones, i.e., $M(p_i)|_{r,i} = 1, \forall r$. We call the *incident set* of v the set of paths traversing v and denote it with $P_v \subseteq P$.

We call *failure set* of a network, F , the set of all failed nodes. Here, we assume that nodes fail one at a time, and therefore that $|F| = 1$. In such a context, we focus on the property of 1-identifiability. With reference to [14], we give the following definition:

Definition 1: A node v_i is 1-identifiable with respect to $P = \{p_1, \dots, p_m\}$ if $b(v_i) \neq 0^m$ and if for all $v_j \neq v_i, b(v_i) \neq b(v_j)$, i.e., its binary encoding is not null and not identical to that of any other node.

Node identifiability allows non ambiguous node state assessment by means of end-to-end measurement paths. We highlight that in order to be identifiable, a node must be monitored at least by one path. For this reason, a node whose binary encoding is null cannot be identifiable. In this article, we give bounds on the minimum number of paths that are needed for letting $n \leq |V|$ nodes be 1-identifiable under arbitrary and consistent routing schemes.

IV. ARBITRARY ROUTING

In this section we study the minimum number of monitoring paths that can be employed to identify n nodes in a network under deterministic, single-path arbitrary routing. We say that paths follow an arbitrary routing scheme if they do not traverse a node more than once, but they can cross each other non-restrictively.

In Section III, we explained that nodes can be represented with binary encodings depending on what paths traverse them. In addition, we noticed that, in order for nodes in a network to be identifiable, they must have all different encodings. Since the number of different binary encodings of length m , excluding the string 0^m , is $2^m - 1$, the following holds:

Proposition 1: The minimum number of monitoring paths to place in order to identify n nodes under arbitrary routing is $m_{\min}^{AR} = \lceil \log_2(n + 1) \rceil$.

The bound represented by m_{\min}^{AR} does not take into consideration the length of the paths involved. The length of a path p_i is the number of nodes it traverses, d_i ($d_i = |\{v \in V : b(v)|_i = 1\}|$). When constraints to the paths length are given, for instance by defining an upper bound to the maximum length, $d_i \leq d_{\max}$, or to the average path length, $\frac{1}{m} \sum_i d_i \leq \bar{d}$, the bound of Proposition 1 may change. In order to discuss the bound on the minimal number of paths under path length constraints, we observe the following facts:

Observation 1: The number of distinct binary strings in $\{0, 1\}^m$ with k 1s and $m - k$ 0s (with $0 < k \leq m$) is $\binom{m}{k}$. Out of them, there are $\binom{m-1}{k-1}$ strings where the i -th digit is 1. In our context, this means that a path can traverse at most $\binom{m-1}{k-1}$ nodes having crossing number k in order to guarantee identifiability, that is when all encodings are distinct.

Observation 2: The maximum length d_i of a path p_i is $d_i = \sum_{i=0}^{m-1} \binom{m-1}{i} = 2^{m-1}$. In fact, the maximum number of

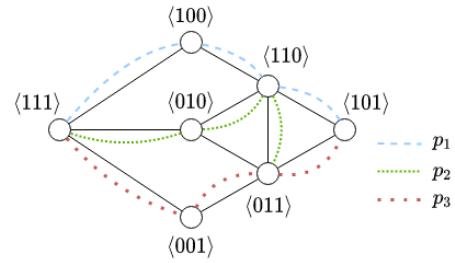


FIGURE 1. Arbitrary routing example.

identifiable nodes using m path under arbitrary routing is $n = 2^m - 1$ (see Proposition 1). The statement follows from Observation 1.

These observations are illustrated in Figure 1. In this simple example we consider $n = 7$ nodes. Assuming arbitrary routing, $m = \log_2(8) = 3$ monitoring paths are enough to identify all nodes. Each path p_1, p_2, p_3 traverses $\binom{m-1}{k-1} = \binom{2}{k-1}$ nodes v with crossing number $\chi(v) = k \in \{1, \dots, m\}$, and the length of each path is $2^{m-1} = 4$.

In Theorem 2 we provide a lower bound to the number of paths to place in order to identify n nodes in a network, in the case a path length constraint is given.

Theorem 2: The minimum number of paths $m_{\min}^{AR, d_{\max}}$ of maximum path length d_{\max} to identify n nodes under arbitrary routing is the solution of the following problem:

$$\min m \quad s.t. \quad \left\lfloor \frac{l_{\max \tilde{k}} \cdot m}{\tilde{k}} \right\rfloor + \sum_{i=1}^{\tilde{k}-1} \binom{m}{i} \geq n, \quad (1a)$$

where

$$\tilde{k} = \max \left\{ j : \sum_{i=1}^j i \cdot \binom{m}{i} \leq m \cdot D \right\} + 1, \quad (1b)$$

$$D = \min \left\{ d_{\max}, 2^{m-1} \right\}, \quad (1c)$$

and

$$l_{\max \tilde{k}} = D - \sum_{i=0}^{\tilde{k}-2} \binom{m-1}{i}. \quad (1d)$$

Proof: In order to minimize the number of paths, we want to have as many distinct encodings as possible with the minimum number of 1s. This fact translates into a strategy that consists in incrementally increasing the crossing number of the monitored nodes until the fixed average path length d_{\max} allows it or until there is no way that paths traverse more nodes without violating identifiability, equation (1c) (see Observation 2).

The quantity \tilde{k} in Equation (1b) says that paths can be placed in such a way that the nodes they traverse are all distinct nodes with crossing number $o < \chi(v) < \tilde{k}$. The quantity $m \cdot D$ is a loose upper-bound to the maximum number of nodes traversed by m paths, where nodes with crossing number j are counted j times, as j paths traverse them. Depending on the

value $l_{\max_{\tilde{k}}}$ in Equation (1d), some more paths with crossing number \tilde{k} may be traversed. Notice that $l_{\max_{\tilde{k}}}$ represents the number of nodes of crossing number \tilde{k} that each path p can traverse considering that it traversed all distinct nodes with crossing number $< \tilde{k}$ (see Observation 1). Extended to all m paths, $\left\lfloor \frac{l_{\max_{\tilde{k}} \cdot m}{\tilde{k}} \right\rfloor$ in Equation (1a) is the number of distinct nodes with crossing number \tilde{k} that can be traversed by m paths of length $l_{\max_{\tilde{k}}} + \sum_{i=0}^{\tilde{k}-2} \binom{m-1}{i}$. \square

Constraints on paths length are usually imposed by QoS requirements and influence substantially the minimum amount of paths needed to identify a certain number of nodes. In a network where shortest path routing schemes are applied, the value of d_{\max} is the diameter of the network. Differently, multi-service networks serve for more than a single service, to which a number of clients access. Each service is characterised by a different service level agreement (SLA) that regulates the routing scheme to adopt as well as the reserved portion of the network. QoS requirements may also vary for each service. In this scenario, paths lengths may be different from one another depending on what service a path belongs to. Information about paths lengths for different services in multi-service networks justifies the introduction of the notion of average path length of a network, \bar{d} . Furthermore, average path length can be easily computed when the network topology and the routing scheme implemented on it are known.

Corollary 1: The bound of Theorem 2 holds also when we consider the average path length, \bar{d} , or an upper-bound to it, instead of d_{\max} . The statement of Theorem 2 only changes in Equation (1c), where the value of \bar{d} is to be substituted to d_{\max} . We call such bound $m_{\min}^{AR, \bar{d}}$.

The bound of Theorem 2 suggests that the number of nodes that m monitoring paths can identify grows with the path length, (Equation (1b)). Nevertheless, we can show that the growth stops for $d_{\max} > 2^{m-1}$.

Corollary 2: The number of nodes that m paths can identify grows with d_{\max} as long as $d_{\max} \leq 2^{m-1}$.

Proof: In Observation 1 we point out that, given a set of m paths, each of them can traverse at most $\binom{m-1}{k-1}$ nodes with crossing number $\chi(v) = k$. The maximum value for $\chi(v)$ is m , and therefore the maximum path length for a path is $\sum_{i=0}^{m-1} \binom{m-1}{i} = 2^{m-1}$. This fact motivates the expression in Equation (1b). \square

We highlight that knowledge of path length does not necessarily imply explicit knowledge of the paths - in terms of what nodes they traverse.

A. DESIGN VIA INCREMENTAL CROSSING ARRANGEMENT (ICA)

The proof of Theorem 2 suggests a technique to build a network topology $G = (V, E)$ and related monitoring paths P with maximum identifiability, where $|P| = m$. We call this technique *Incremental Crossing Arrangement (ICA)*.

We shall describe this strategy in detail when the average path length \bar{d} is given, as the case with d_{\max} is more general.

ICA, the idea. The technique works by generating n node encodings in increasing order of crossing number with respect to the m monitoring paths in use, where m is the optimal solution of the problem in Theorem 2. The set of encodings defines a design for the monitoring paths deployment, that must traverse nodes according to the generated encodings: path p_i traverses any node v for which $b(v)|_i = 1$, $\forall i \in \{1, \dots, m\}$. The network topology is then constructed by considering a node for each of the generated Boolean encodings, and adding links between any pair of nodes appearing sequentially in any path.

ICA in details. Algorithm 1 formalizes the incremental crossing arrangement design, used to determine the binary encodings of the identifiable nodes.

As we consider m paths, the node encodings will be sequences of m bits in $\mathcal{B} \triangleq \{0, 1\}^m$. We also denote with $\mathcal{B}|_i \subset \mathcal{B}$ the set of m -digits binary encodings having a 1 in the i -th position, i.e., $\mathcal{B}|_i = \{b \in \mathcal{B} \text{ s.t. } b|_i = 1\}$. The nodes corresponding to encodings of $\mathcal{B}|_i$ will be monitored (at least) by path p_i . Moreover, we denote with $\mathcal{B}(k) \subset \mathcal{B}$ the set of all binary encodings having exactly k digits equal to 1, therefore $\mathcal{B}(k) \triangleq \{b \in \mathcal{B} \text{ s.t. } \sum_{i=1}^m b|_i = k\}$. The nodes corresponding to encodings in $\mathcal{B}(k)$ have crossing number equal to k .

Finally, given a generic set of binary encodings $B \subseteq \mathcal{B}$, we denote with $\ell_i(B)$ the number of encodings of B having a one in the i -th position: $\ell_i(B) \triangleq |B \cap \mathcal{B}|_i$. The value of $\ell_i(B)$ represents the length of a path p_i traversing all the nodes in $B \cap \mathcal{B}|_i$, exactly once.

Without loss of generality, we consider paths of balanced length, i.e. we set the length d_i of path p_i to a value $d_i \in \{\lfloor \bar{d} \rfloor, \lfloor \bar{d} \rfloor + 1\}$ (lines 2 - 4).

The incremental crossing arrangement approach incrementally generates the solution set B_V by including all the encodings of $\mathcal{B}(i)$, $i = 1, \dots, \tilde{k} - 1$ corresponding to nodes with crossing number lower than or equal to $\tilde{k} - 1$. It then considers some encodings with \tilde{k} digits equal to one. For this purpose it generates a family \mathcal{F} of subsets in $\mathcal{B}(\tilde{k})$, i.e., $\mathcal{F} \subseteq 2^{\mathcal{B}(\tilde{k})}$ (line 7) whose elements B are such that $\ell_k(B \cup B_V) \leq d_k$. The algorithm then looks for a maximal cardinality set B^* in the family \mathcal{F} and adds it to the solution B_V , s.t. $B_V = \bigcup_{k=1}^{\tilde{k}} \mathcal{B}(k) \cup B^*$. Notice that the maximality of the cardinality of B^* implies that no encoding with \tilde{k} digits equal to one can be added to the set B_V without violating the path length constraint $\ell_k(B_V) \leq d_k$ for some path $k = 1, \dots, m$, or without removing at least one encoding already in B_V .

The procedure described so far is sufficient to produce a network topology where all nodes are identifiable and where the number of paths m is the solution of the problem in Theorem 2 given the average path length, \bar{d} . In the produced topology, there can be values of $k \in \{1, \dots, m\}$ for which $\ell_k(B_V) < d_k$ and, more precisely, given the balanced path length, $\ell_k(B_V) = d_k - 1$, corresponding to paths longer than strictly necessary to identify n nodes, i.e., overlength paths. Overlength paths cannot traverse nodes with the same

encoding without compromising the achievement of maximum identifiability. Therefore, in order for the average path distance to meet the value \bar{d} , we proceed as follows, with a procedure that we call *Path Completion*. First, we observe that under ICA, the bound on the minimum number of monitoring paths can sometimes be met tightly even when the average path length is slightly lower than the given \bar{d} . This condition is verified when the ratio inside the floor operator of the bound expression of Theorem 2 is not integer. Nevertheless, the same bound can still be met tightly with the exact average length provided as input, by operating as follows: let $S \subset \{1, \dots, m\}$ be the set of overlength path indexes, namely $S \triangleq \{k, \text{ s.t. } \ell_k(B_V) = d_k - 1\}$. It holds $|S| = \left\lceil m \cdot D - \sum_{i=1}^{\tilde{k}-1} i \cdot \binom{m}{i} \pmod{\tilde{k}} \right\rceil$, hence the number of overlength paths is lower than or equal to $\tilde{k} - 1$.

We choose an encoding $b' \in B_V \cap \mathcal{B}(\tilde{k} - |S|)$ such that $b'|_k = 0, \forall k \in S$, and such that $(\bigvee_{k \in S} \mathbf{e}_k \vee b') \notin B_V$, where \mathbf{e}_k is an m -dimensional identity vector with all zeroes but a one in the k -th position.¹ Then we remove b' from the solution set B_V and replace it with $b'' \triangleq \bigvee_{k \in S} \mathbf{e}_k \vee b'$, i.e., with a new encoding b'' such that $b''|_k = 1, \forall k \in S$, and $b''|_k = b'|_k$ otherwise.

Algorithm 1: Incremental Crossing Arrangement

Input: n and \bar{d} .

Output: A set of encodings B_V which can be mapped onto a topology graph $G = (V, E)$ where all $|V| = n$ nodes are identifiable by using m paths with average length \bar{d} .

- 1: Calculate m according to Theorem 2 and Corollary 1
- 2: Calculate $m_1 \triangleq m \cdot (\bar{d} - \lfloor \bar{d} \rfloor)$;
- 3: **For** $i = 1, \dots, m_1$ **do** set $d_i = \lfloor \bar{d} \rfloor + 1$
- 4: **For** $i = m_1 + 1, \dots, m$ **do** set $d_i = \lfloor \bar{d} \rfloor$
- 5: $B_V = \emptyset$
- 6: **For** $i = 1, \dots, \tilde{k} - 1$ **do** $B_V = B_V \cup \mathcal{B}(i)$
- 7: Calculate the family \mathcal{F} defined as $\mathcal{F} \triangleq \{B : B \subseteq \mathcal{B}(\tilde{k}) \wedge \ell_k(B \cup B_V) \in [d_k - 1, d_k], \forall k = 1, \dots, m\}$
- 8: Choose $B^* = \arg \max_{B \in \mathcal{F}} |B|$
- 9: $B_V = B_V \cup B^*$
- 10: **if** $\exists k \in \{1, \dots, m\}$ s.t. $\ell_k(B_V) = d_k - 1$ **then**
 \lfloor Perform *path completion* and update B_V
- 11: Return B_V

ICA: Example A (where path completion is not necessary). Figure 2 shows an example of a topology generated by means of incremental crossing arrangement. We are given $n = 11$ nodes and $\bar{d} = 4.75$. The minimum number of paths satisfying eqs.(1a) to (1d) is $m = 4$. We set $d_i = 5, \forall i = 1, 2, 3$ and $d_4 = 4$ (lines 2 - 4). According to ICA, we first generate all the encodings of $\mathcal{B}(1)$ and in $\mathcal{B}(2)$ and set $B_V = \{1000, 0100,$

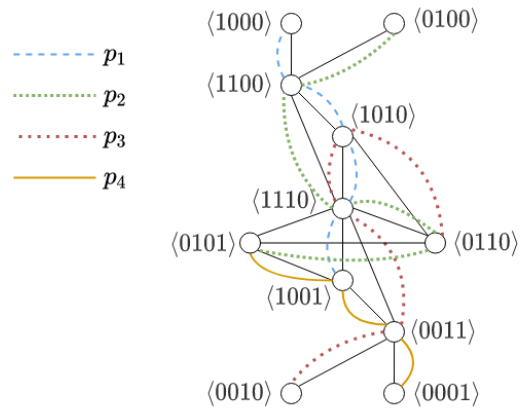


FIGURE 2. ICA execution on Example A. Solid lines represent graph edges.

0010, 0001, 1100, 1010, 1001, 0110, 0101, 0011} (line 6). Then we start generating some encodings in $\mathcal{B}(3) = \mathcal{B}(\tilde{k})$ until no other encoding can be added without violating the path length constraint (lines 7 - 9), obtaining $B_V = \{1000, 0100, 0010, 0001, 1100, 1010, 1001, 0110, 0101, 0011, 1110\}$, where each encoding corresponds to a node of the graph G . Then we define the corresponding monitoring paths, by letting path p_i traverse all the nodes whose encoding has a 1 in the i -th position, in arbitrary order, $\forall i \in \{1, \dots, m\}$. Finally, we design the underlying topology by connecting each pair of nodes appearing in a sequence in any of the paths, as shown in Figure 2.

ICA: Example B (with path completion). Figure 3 shows another example of a topology generated by means of incremental crossing arrangement. We are given $n = 11$ and $\bar{d} = 5$. The minimum number of paths satisfying eqs. (1a) to (1d) is again $m = 4$. To meet the requirement on average length, we set $d_i = 5 \forall i = 1, \dots, 4$ (lines 2 - 4). According to ICA (line 6), we first generate all the encodings of $\mathcal{B}(1)$ and $\mathcal{B}(2)$ and set $B_V = \{1000, 0100, 0010, 0001, \underline{1100}, 1010, 1001, 0101, 0011\}$. Then we choose one of the possible B^* ((lines 7 - 9)), for instance $B^* = 1110$, obtaining $B_V = \{1000, 0100, 0010, 0001, \underline{1100}, 1010, 1001, 0101, 0011, 1110\}$

Finally, we observe that $\ell_4(B_V) = 4 < d_4$. We then perform the path completion procedure (line 10) and choose one of the encodings $b' \in B_V \cap \mathcal{B}(\tilde{k} - |S|) = \mathcal{B}(2)$ for which $b'|_4 = 0$ and $b' \vee \mathbf{e}_4 \notin B_V$. One encoding that satisfies this condition is $b' = 1100$. We replace b' with $b'' = 1101$ and obtain the set of encodings $\{1000, 0100, 0010, 0001, \underline{1101}, 1010, 1001, 0101, 0110, 0011, 1110\}$, each corresponding to a node of the graph G . Then we define the corresponding monitoring paths, by letting path p_i traverse all the nodes whose encoding has a 1 in the i -th position, in arbitrary order, $\forall i \in \{1, \dots, m\}$. Finally, we design the underlying topology by connecting each pair of nodes appearing in a sequence in any of the paths, obtaining the topology of Figure 3.

It is worth observing the following.

¹We can always find an encoding b' with the described properties because B_V contains all the encodings of $\mathcal{B}(\tilde{k} - |S|)$ and not all the encodings b of the set $\mathcal{B}(\tilde{k})$ for which $b|_i = 1, \forall i \in S$.

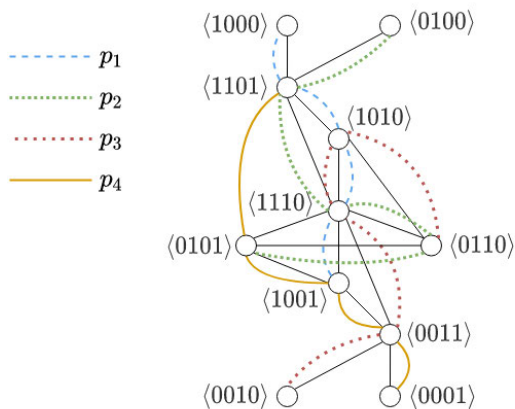


FIGURE 3. ICA execution on Example B. Solid lines represent graph edges.

Observation 3: ICA produces a network topology and related monitoring paths such that all nodes have a crossing number lower than or equal to \tilde{k} .

When knowledge of \tilde{d} is not available, and instead we limit the paths length to be at most d_{\max} , Algorithm 1 can be simplified allowing the following changes: the number of paths m is computed as in Theorem 2 (line 1). To all nodes, we initially assign $d_i = d_{\max}$ (lines 2-4). The algorithm continues as it is until the condition that some paths satisfy $\ell_k(B_V) = d_{\max} - 1$ is met (line 10). When d_{\max} is used, path completion is not required, and for such paths we simply set $d_k = \ell_k(B_V)$. The returned set of encodings B_V can be mapped onto a topology graph where all n nodes are identifiable by using m paths with maximum path length d_{\max} .

1) TIGHTNESS OF THE BOUND UNDER ARBITRARY ROUTING
In this section we show that the bound given by Theorem 2 can be achieved tightly for a specific family of topologies constructed via ICA.

Proposition 2 (Tightness of Theorem 2): For any $n \in \mathbb{Z}^+$ (positive integer) and $\tilde{d} > 0$, there exists a set P of m monitoring paths with average length \tilde{d} , such that m is the solution of the problem in Equations (1a) to (1d).

Proof: We recall that the ICA technique builds a topology by creating nodes with unique encodings, in increasing order of crossing number, up to \tilde{k} .

To prove the proposition, we need to show that the minimum number of monitoring paths required to identify n nodes is provided in Theorem 2. ICA initially generates all the encodings of $\mathcal{B}(i)$, for $i = 1, \dots, \tilde{k} - 1$. As a consequence, it follows from Observation 1 that each path will traverse at least $d(\tilde{k} - 1) \triangleq \sum_{i=0}^{\tilde{k}-2} \binom{m-1}{i}$ identifiable nodes. In fact, the encodings of the nodes of $\mathcal{I}(p_i)$ (identifiable nodes traversed by path p_i), must have a “1” in the i -th position. Therefore the number of distinct encodings corresponding to nodes of $\mathcal{I}(p_i)$ is at least equal to the number of binary sequences of $(m - 1)$ elements, with up to $(\tilde{k} - 2)$ ones.

Under incremental crossing arrangement, each path also traverses other nodes with crossing number equal to \tilde{k} . Each of these nodes will appear in exactly \tilde{k} paths. The number of such nodes is therefore given by $\left\lfloor \frac{\sum_{k=1}^m (d_k - d(\tilde{k}))}{\tilde{k}} \right\rfloor$.

In conclusion, with this construction, ICA generates the following number of node encodings:

- $\binom{m}{i}$ encodings corresponding to nodes with crossing number equal to i , for $i = 1, \dots, \tilde{k} - 1$, and
- $\left\lfloor \frac{\sum_{k=1}^m (d_k - d(\tilde{k}))}{\tilde{k}} \right\rfloor$ encodings corresponding to nodes with crossing number equal to \tilde{k} .

The number of generated encodings does not change if ICA applies the path completion procedure, which consists in a replacement of an encoding $b' \in \cup_{i=1}^{\tilde{k}-1} \mathcal{B}(i)$ with an encoding $b'' \in \mathcal{B}(\tilde{k})$. In both cases, ICA constructs the set B_V in a way that each encoding corresponds to a unique node, and the nodes are traversed by paths of average length \tilde{d} , guaranteeing identifiability of all the nodes corresponding to the generated encodings.

In order to show that the number of paths provided by Theorem 2 is enough to identify at least n nodes in the network, we need to prove that $\left\lfloor \frac{\sum_{k=1}^m (d_k - d(\tilde{k} - 1))}{\tilde{k}} \right\rfloor = \left\lfloor \frac{l_{\max_{\tilde{k}} \cdot m}{\tilde{k}}} \right\rfloor$. This holds because $\sum_{k=1}^m (d_k - d(\tilde{k} - 1)) = \sum_{k=1}^m d_k - m \cdot d(\tilde{k} - 1)$, that is equal to $l_{\max_{\tilde{k}}} \cdot m$ in Equation (1d), being $\tilde{d} = \frac{1}{m} \sum_{k=1}^m d_k$. \square

Notice that Proposition 2 requires $\tilde{d} \leq 2^{m-1}$ as having longer paths would require at least a path to traverse different nodes with duplicate encodings, losing identifiability with respect to the bound value.

While Proposition 2 gives a characterization of sufficient conditions for building a network topology achieving the bound, we note that there exist topologies that do not meet the conditions, but still achieve the bound.

Observation 4: The statement of Proposition 2 holds also when d_{\max} is given instead of \tilde{d} . In fact this simply translates into the more general scenario where initially $d_i = d_{\max}$ is assigned to all paths and where path completion is not performed. Indeed, path completion does not serve for identifiability increase, but only to meet the input condition on the average path length.

V. CONSISTENT ROUTING

As we have seen in Theorem 2, given a number of nodes to identify, the number of required paths can be logarithmic in the number of nodes. Nevertheless the bound of Theorem 2 is achieved only when the routing scheme allows paths to traverse arbitrary sequences of nodes.

If routing needs to meet additional requirements, the theoretical bound given by Theorem 2 can be increased.

We now consider the impact of the routing scheme on the identifiability of nodes via Boolean tomography. In the sequel, we assume that paths satisfy the following property of routing consistency.

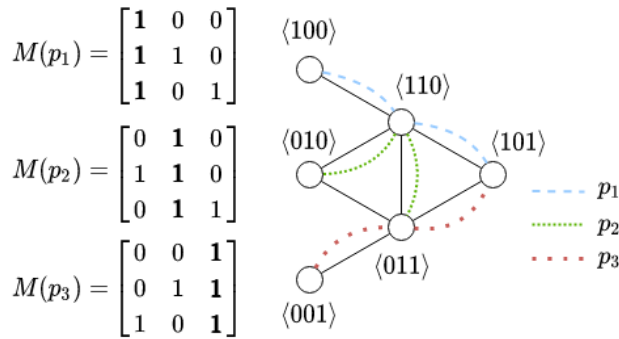


FIGURE 4. Consistent routing example.

Definition 3: 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 .

We remark that routing consistency is satisfied by many practical routing protocols, including but not limited to shortest path routing (where ties are broken with a unique deterministic rule). Note that routing consistency implies that paths are cycle-free. In the following, we recall from [4] necessary for the proof of Theorem 15, where we provide the bound on the minimum number of paths under consistent routing.

Lemma 1: Under the assumption of consistent routing, if any two different rows of the matrix $M(p_i)$ are equal, then the corresponding nodes are not 1-identifiable.

Definition 4: A column $M(p)|_{*,k}$ ($k = 1, \dots, m$) of a path matrix $M(p)$ has consecutive ones if all the "1"s appear in consecutive rows, i.e., for any two rows i and j ($i < j$), if $M(p)|_{i,k} = M(p)|_{j,k} = 1$, then $M(p)|_{h,k} = 1$ for all $i \leq h \leq j$.

Lemma 2: Under the assumption of consistent routing, all the columns in all the path matrices have consecutive ones.

Lemma 3: Given $m = |P| > 1$ consistent routing paths, each path p_i having length d_i , the maximum number of different encodings in the rows of $M(p_i)$ is upper-bounded by $\min\{d_i; 2 \cdot (m - 1)\}$.

Formal proofs to Lemma 1 to 3 are given in [4]. In order to ease their comprehension, we show them in the simple example of Figure 4, where all nodes are 1-identifiable under consistent routing. Routing consistency is verified as the i -th column of all path matrices $M(p_i)$ is 1.

Theorem 5: The minimum number of paths $m_{\min}^{CR, d_{\max}}$ of maximum path length d_{\max} to identify n nodes under consistent routing is the solution of the problem of Theorem 2, where Equation (1c) reads:

$$D = \min \{d_{\max}, 2 \cdot (m - 1)\}. \quad (2)$$

Proof: The proof is analogous to the one of Theorem 2. The only difference in this case is that $D = \min \{d_{\max}, 2 \cdot (m - 1)\}$, because of Lemma 3. \square

The same considerations on the knowledge of the average path lengths for Theorem 2 hold for Theorem 5:

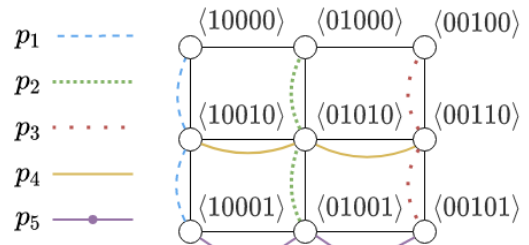


FIGURE 5. 3×3 grid network.

Corollary 3: When \bar{d} is known, the optimal solution $m_{\min}^{AR, d_{\max}}$ of the problem in Theorem 2, is a lower bound to the minimum number of paths to identify n nodes, if we substitute d_{\max} with \bar{d} in Equation (2). We call $m_{\min}^{CR, \bar{d}}$ the bound computed with \bar{d} .

Corollary 4: The bound provided in Theorem 5 may be achieved by allowing the maximum value for the crossing number of a node to be 3.

Proof: We need to prove that the maximum value of \tilde{k} is 3 under the assumption of consistent routing. Let us assume that $\bar{d} \geq 2 \cdot (m - 1)$, and therefore that $D = 2 \cdot (m - 1)$. Recall that $\tilde{k} = \max \left\{ j : \sum_{i=1}^j i \cdot \binom{m}{i} \leq m \cdot D \right\} + 1$. For $j = 2$ and $\forall m \geq 2$, it holds that

$$\sum_{i=1}^2 i \cdot \binom{m}{i} = m + 2 \frac{m(m-1)}{2} = m^2 < 2m \cdot (m - 1),$$

whereas for $j = 3$:

$$\begin{aligned} \sum_{i=1}^3 i \cdot \binom{m}{i} &= m^2 + 3 \frac{(m-2)(m-1)m}{6} \\ &= \frac{m^3 - m^2}{2} + m > 2m \cdot (m - 1) \quad \forall m. \end{aligned}$$

Since $\sum_{i=1}^N i \cdot \binom{m}{i}$ is a growing function of N , \tilde{k} is at most 3. \square

A. CASE OF STUDY: GRID NETWORKS

The bound provided in Theorem 5 is tight on square grid networks with n^2 nodes, using $2n - 1$ paths of maximum length $d_{\max} = n$. By contradiction, assume $d_{\max} = n$ and $m_{\min}^{CR, d_{\max}} = 2n - 2$. It is easy to see that $\tilde{k} = 2 \forall n \in \mathbb{N}, n > 2$. Therefore $l_{\max \tilde{k}} = n - 1$ and the number of nodes that may be identified with $m = 2n - 2$ paths is $(n - 1)^2 + 2n - 2 = n^2 - 1 < n^2$. An example of such topology and of paths placement is in Figure 5.

VI. EXPERIMENTAL RESULTS

We evaluate the tightness of the bounds proposed in the previous sections in comparison to with the results obtained by a state-of-the-art heuristic ([5]). For this purpose we run experiments on synthetic as well as real network topologies, implemented in Matlab. First, in Figures 6 and 7 we show the

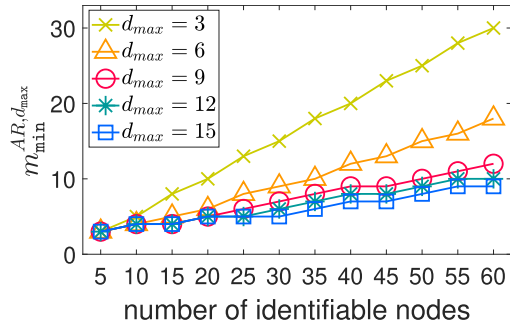


FIGURE 6. Bound of Theorem 2 $m_{\min}^{AR,d_{\max}}$, varying paths lengths.

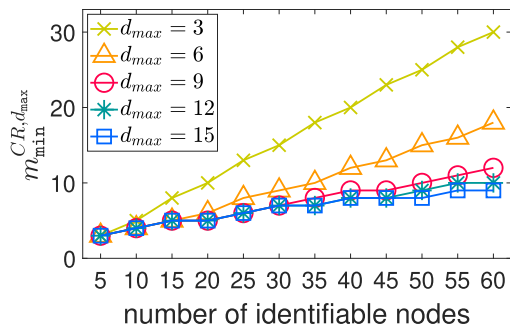


FIGURE 7. Bound of Theorem 5 $m_{\min}^{CR,d_{\max}}$, varying paths lengths.

trend of the bounds on the minimum number of paths for the identification of n nodes, for the two cases of arbitrary and consistent routing, respectively (i.e., bounds of Theorems 2 and 5), under varying n , and path length d_{\max} . Observe that the dependence on the path length of the values of $m_{\min}^{AR,d_{\max}}$ and $m_{\min}^{CR,d_{\max}}$ is stronger for smaller values of d_{\max} . This is an expected behaviour, as in Equation (1d), it holds that $D = 2^{m-1}$ for all values of $d_{\max} \geq 2^{m-1}$. As a result, for all such values of d_{\max} , the minimum number of paths needed to identify n nodes is the same. This phenomenon is more evident in the case of consistent routing (Figure 7) because $D = \min\{d_{\max}, 2 \cdot (m - 1)\}$ (see Equation (2)), and therefore $D = 2 \cdot (m - 1)$ for all $d_{\max} \geq 2 \cdot (m - 1)$.

A. TOPOLOGIES

We hereafter list the topologies (synthetic and real) used in our evaluation:

- 1) **Random Geometric (RG) graphs.** RG graphs are synthetic topologies [15] built as follows: n nodes are placed in a unit square and a link is added between any pair of nodes whose distance is lower than or equal to a threshold parameter $\rho > 0$. In our experiments, we generate random coordinates (x_i, y_i) for each node v_i and we vary the value of ρ . This model well simulates ad-hoc wireless networks.
- 2) **Jellyfish topology.** Introduced in [16], Jellyfish are emerging data centre topologies which offer high throughput and capacity, high scalability and failure resiliency. The internal nodes of the Jellyfish (nodes

with degree strictly greater than one) are network switches, whereas leaf nodes are servers.

- 3) **US Signal.** This is the real topology of a fiber optical network in the USA. This topology was made available in the Topology Zoo archive [17], and is composed of 63 nodes and 133 edges.
- 4) **Uninett.** This is an existing Internet topology located in Norway. It has 69 nodes and 98 edges. It was also taken from the Topology Zoo archive [17].

B. BENCHMARK HEURISTIC

In order to evaluate the tightness of our bound, we use a state-of-the-art greedy for identifiability (GI) proposed in [5], as a benchmark for comparisons, and we adapt it to our scenario. GI was originally proposed as an algorithm to place servers for addressing Quality of Service (QoS) and failure identifiability requirements in a joint manner. Given multiple services, and related client sets, the algorithm finds the most suitable server location, among those satisfying QoS requirements, to optimize failure identifiability. The selected server locations are such that the client-to-server paths form several intersecting trees, one for each service, where servers are located at the roots and clients are located at the leaves. GI uses a greedy approach that iteratively selects a number of server positions, one for each service, such that the identifiability obtained by the adopted client-to-server paths, is maximized at each iterative step. The authors prove that the number of paths placed by this heuristic is a constant approximation of the optimal solution.

In our experiments we modify the GI approach to obtain an upper bound to the number of paths that are necessary to uniquely identify the state of a given number of nodes n . In particular, in order to ensure maximum flexibility to the choice of the set of monitoring paths we consider only one client for each server, and a number of paths that is equal to the number of deployed servers. Moreover, we relax the quality of service requirements to obtain all the server locations which are at a distance lower than or equal to d_{\max} from the client. The algorithm ends as soon as the selected paths are sufficient to identify the desired number of nodes n .

In our implementation of GI communication between any two endpoints is obtained through a shortest path routing algorithm. The adoption of a deterministic tie breaking rule ensures that the obtained routing scheme is deterministic. In order to prevent the use of degenerate paths (i.e., paths traversing only one node), servers are never located on the same node as the related client.

In order to boost identifiability of the greedy approach, we consider a preliminary phase where a set of paths is deployed according to a greedy for node coverage approach. This coverage phase is also common to [18], and is motivated by the observation that greedy for identifiability approaches select short paths in the early iterations, to obtain maximum identifiability, which prevents further identification in the later steps of the algorithm, due to insufficient node coverage,

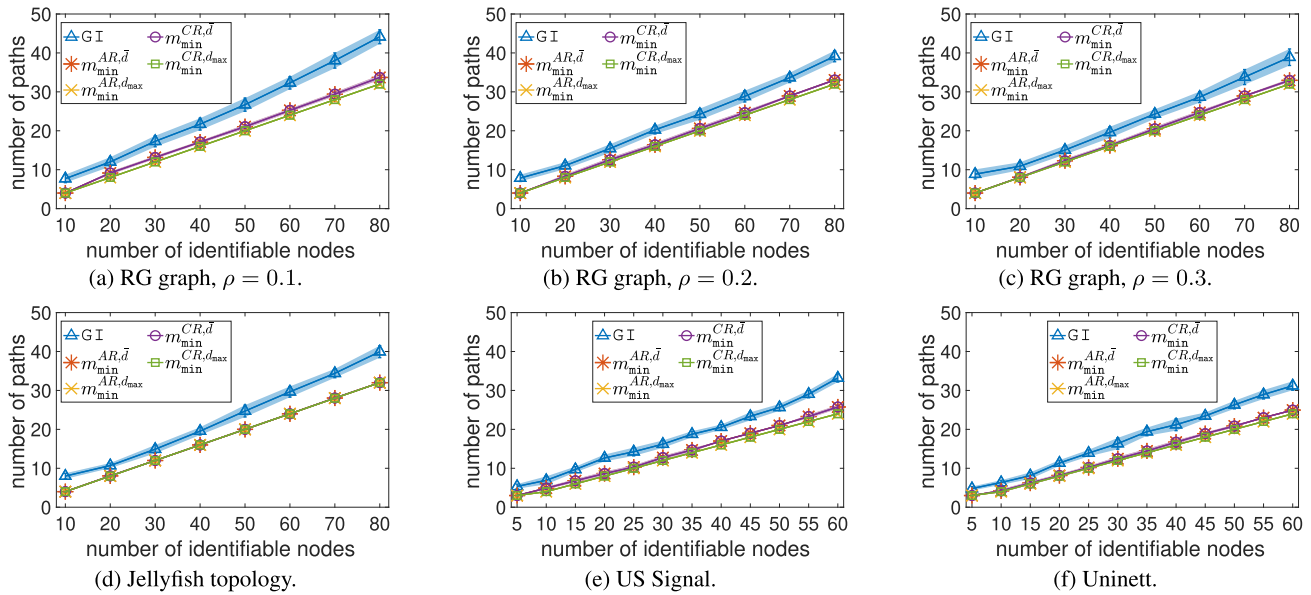


FIGURE 8. Number of paths to identify variable numbers of nodes on different topologies. $d_{max} = 4$.

an issue that is easily solved by letting the algorithm use longer paths in the early execution steps.

C. TESTS

The bounds of Theorems 2 and 5 are compared with the results obtained by GI, which provides an upper bound to the minimum number of paths that are necessary to uniquely identify the state of a given number of nodes n . We carry out two different sets of experiments. In the first set, the maximum path length is fixed, $d_{max} = 4$, whereas the number of nodes to be identified is variable. Figure 8 shows the curves of the number of paths necessary to identify an increasing number of nodes for GI with respect to the bounds. The bounds are also evaluated with the average path length resulting from the path choices of GI. Each figure represents the experiment on a different topology. For random topologies (random geometric graphs and jellyfish topologies), we generate graphs of 100 nodes. GI is then run on all such topologies with an increasing number of nodes to identify (from 10 to 80). For the fiber and the internet topologies instead, the number of nodes to identify goes from 10 to 60, being 63 and 69, respectively, the total number of nodes of the networks.

A similar setting was also implemented in the second set of experiments, depicted in Figure 9. For these experiments, curves represent how the number of paths necessary to identify a fixed number of nodes n changes depending on different values of d_{max} . For random topologies, $n = 80$, whereas for US Signal and Uninett, $n = 60$. Also in this case, we generated random topologies having 100 nodes.

Tests on random topologies have been executed by generating 20 different graphs of each type. Shades and bars in the curves of random topologies represent the standard deviation of the mean number of paths used by GI and of the

bounds with variable average path length, \bar{d} (Figures 8a-8d and Figures 9a-9d). In contrast, shades and bars in the curves related to real topologies (US Signal and Uninett, Figures 8e and 8f and Figures 9e and 9f) represent the standard deviation of the mean number of paths used by a randomized version of GI where routing consistency is still satisfied.

We tested on random geometric graphs with different values of ρ (0.1, 0.2, 0.3) in order to analyse the goodness of our bounds on graphs with different properties. When generating random geometric graphs, there is no guarantee that the resulting graph is connected, unless $\rho = \sqrt{2}$ (in this case, the RG graph is a clique, being $\sqrt{2}$ the maximum distance between two nodes in a unit square). Experimentally speaking, we encountered non-connected graphs only for $\rho = 0.1$. When this event occurs, we link together the least number of nodes belonging to different connected components that are the closest (by means of the Euclidean distance), until the graph is connected. Variations of ρ have a great impact on the structure of the resulting graph. The characteristics of the set of 20 random geometric graphs generated with 100 nodes, with respect to different values of ρ are summed up in Table 2. Figure 10 shows an example of how different values of ρ change the structure of a random geometric graph with 100 nodes.

Notice that, when ρ increases, so does the average node degree, whereas the network diameter decreases correspondingly. When the maximum path length is fixed to 4 (Figures 8a-8c), different values of ρ do not imply sensible differences in the performance of GI, whose trend corresponds to the one of our bounds. This confirms the fact that our bounds do not depend on topology structures that can be extremely different. On the other hand, when the maximum path length is variable (Figures 9a-9c), we can observe the

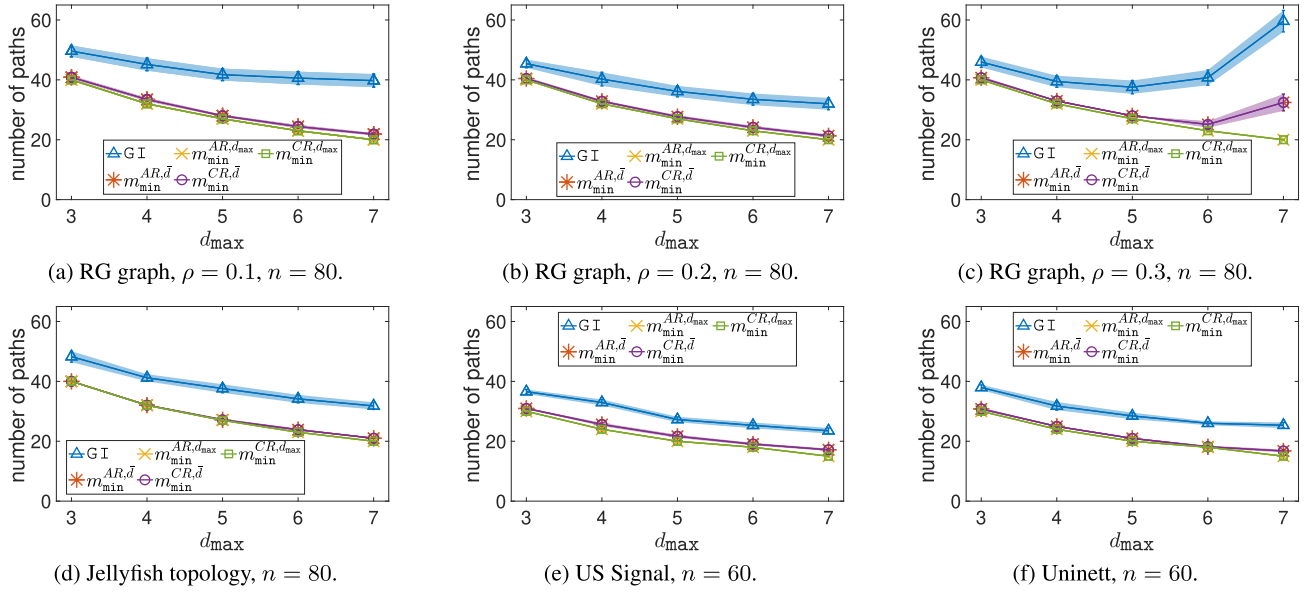


FIGURE 9. Number of paths to identify n nodes with variable values of d_{max} on different topologies.

TABLE 2. Average properties of 20 random geometric graphs of 100 nodes for different values of ρ . Here $\partial_{min/max/avg}$ are the mean minimum, maximum and average degrees of the nodes. δ is the average diameter of the graph.

	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$	$\rho = 0.3^*$
∂_{min}	1	2.1	7.25	6.35
∂_{max}	7.75	18.8	34.2	33.2
∂_{avg}	3.28	10.31	20.99	20.65
δ	39.3	9.75	6.45	7

following facts: first of all, in order to test for $d_{max} = 7$, it is necessary to set the condition that the graph diameter δ is greater or equal to 7. In our experiments, this condition is always met for $\rho = 0.1$ and 0.2 , whereas the same does not hold for $\rho = 0.3$ (see Table 2, column $\rho = 0.3$, where $\delta < 7$). In Table 2, the column $\rho = 0.3^*$ corresponds to random geometric graphs with $\rho = 0.3$ that satisfy the condition $\delta \geq 7$. We generated graphs satisfying this condition for the experiments in Figure 9c. Figures 9a-9c show that our bounds are closer to the curves representing the performance of GI when $\rho = 0.2$. As a matter of fact, when $\rho = 0.3$, the average distance between nodes (in number of hops) is smaller, and the graph diameter is never greater than 7, our maximum path length. As a consequence, nodes are highly connected on average. For this reason, despite the greater path length availability, only a very few paths reach the maximum path length. On the other hand, after the coverage phase, more shorter paths are needed in order to guarantee that $n = 80$ nodes are identifiable, implying that the average path length decreases, as we can observe by the growing trend of the curves $m_{min}^{AR,\bar{d}}$ and $m_{min}^{CR,\bar{d}}$. On the other hand, for $\rho = 0.1$, a few nodes have a high centrality, whereas most of the nodes have degree 2, meaning that the majority of the network

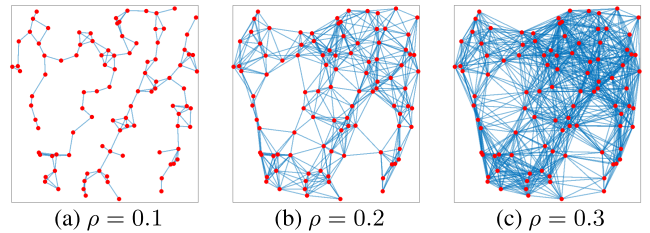


FIGURE 10. Random geometric graphs with 100 nodes having the same geometric coordinates, built with different values of ρ .

nodes are distributed in chains. As node identifiability holds when nodes have unique encodings, e.g., different sets of paths crossing them, chains are structures that are hard to identify by means of monitoring paths, and large path length do not represent an advantage for this specific structure. When $\rho = 0.2$, the resulting graphs are not star-shaped, but at the same time long paths are available, showing their stronger identification power.

On jellyfish topologies (Figures 8d and 9d), our bounds are very close to the results obtained by GI, with negligible differences between the curves. One of the properties of jellyfish topologies is that servers can reach one another with shorter paths with respect to other topologies for data centres (e.g., fat trees), [16], for this reason the curves of our bounds are tighter for smaller values of d_{max} in Figure 9d. Despite this, in both configurations (Figures 8d and 9d) the curves representing the results of GI scale analogously with our bounds.

The results shown in this section highlight that our bounds are very close to the number of monitoring paths that a greedy algorithm would use in order to guarantee nodes identifiability. We stressed our experiments to evaluate our bounds on synthetic networks with very different structures and connectivity properties. Experiments show that the bounds presented

in this work represent a very trustful estimation of the number of paths for node identifiability on Jellyfish topologies, as well as on real internet and fiber optical networks. In addition, we can also observe that knowledge of \bar{d} can be used to provide tighter bounds, specially when there are few paths of length d_{\max} in the network.

VII. CONCLUSION

In this article we provide theoretical lower bounds on the minimum number of monitoring end-to-end paths necessary to achieve the desired level of identifiability in a network in terms of identifiable nodes. We study how the routing scheme affects the bound values by giving two different formulations, for arbitrary and consistent routing, respectively. We also study how requirements on the maximum and average path length affect the bound formulation, highlighting the dependence of the minimal number of required paths on QoS constraints. We proposed a polynomial-time algorithm, called ICA, that takes into consideration such constraints to design a network that meets the bounds for the case of arbitrary routing. We carried out an extensive set of experiments on synthetic and real topologies to evaluate the tightness of the proposed lower-bounds. The synthetic topologies that we used are commonly employed for modelling ad-hoc wireless networks and data centers, whereas the real networks are an internet and a optical fiber network located in Norway and in the USA. We used a state-of-the-art algorithm for network identifiability maximization via path deployment to obtain feasible solutions as upper bounds of the optimum and evaluate the tightness of the proposed lower bound. We show that the provided bounds provide a high approximation in all the performed experiments.

REFERENCES

- [1] A. Adams, T. Bu, T. Friedman, J. Horowitz, D. Towsley, R. Caceres, N. Duffield, F. L. Presti, S. B. Moon, and V. Paxson, "The use of end-to-end multicast measurements for characterizing internal network behavior," *IEEE Commun. Mag.*, vol. 38, no. 5, pp. 152–159, May 2000.
- [2] R. R. Kompella, J. Yates, A. Greenberg, and A. C. Snoeren, "Detection and localization of network black holes," in *Proc. 26th IEEE Int. Conf. Comput. Commun. (INFOCOM)*, May 2007, pp. 2180–2188.
- [3] H. X. Nguyen and P. Thiran, "The Boolean solution to the congested IP link location problem: Theory and practice," in *Proc. 26th IEEE Int. Conf. Comput. Commun. (INFOCOM)*, May 2007, pp. 2117–2125.
- [4] N. Bartolini, T. He, V. Arrigoni, A. Massini, F. Trombetti, and H. Khamfroush, "On fundamental bounds on failure identifiability by Boolean network tomography," *IEEE/ACM Trans. Netw.*, vol. 28, no. 2, pp. 588–601, Apr. 2020.
- [5] T. He, N. Bartolini, H. Khamfroush, I. Kim, L. Ma, and T. La Porta, "Service placement for detecting and localizing failures using end-to-end observations," in *Proc. IEEE 36th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jun. 2016, pp. 560–569.
- [6] N. Duffield, "Simple network performance tomography," in *Proc. ACM IMC*, 2003, pp. 210–215.
- [7] N. Duffield, "Network tomography of binary network performance characteristics," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5373–5388, Dec. 2006.
- [8] S. Pan, P. Li, D. Zeng, S. Guo, and G. Hu, "A Q-learning based framework for congested link identification," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 9668–9678, Dec. 2019.
- [9] L. Ma, T. He, A. Swami, D. Towsley, K. K. Leung, and J. Lowe, "Node failure localization via network tomography," in *Proc. ACM IMC*, Nov. 2014, pp. 195–208.
- [10] L. Ma, T. He, A. Swami, D. Towsley, and K. K. Leung, "Network capability in localizing node failures via end-to-end path measurements," *IEEE/ACM Trans. Netw.*, vol. 25, no. 1, pp. 434–450, Feb. 2017.
- [11] L. Ma, T. He, A. Swami, D. Towsley, and K. K. Leung, "On optimal monitor placement for localizing node failures via network tomography," *Perform. Eval.*, vol. 91, pp. 16–37, Sep. 2015.
- [12] Y. Bejerano and R. Rastogi, "Robust monitoring of link delays and faults in IP networks," in *Proc. IEEE INFOCOM*, Mar. 2003, pp. 134–144.
- [13] M. Cheraghchi, A. Karbasi, S. Mohajer, and V. Saligrama, "Graph-constrained group testing," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 248–262, Jan. 2012.
- [14] N. Bartolini, T. He, and H. Khamfroush, "Fundamental limits of failure identifiability by Boolean network tomography," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2017, pp. 1–9.
- [15] P. Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks," in *Stochastic Analysis, Control, Optimization and Applications* (Systems & Control: Foundations & Applications), W. M. McEneaney, G. G. Yin, and Q. Zhang, Eds. Boston, MA, USA: Birkhäuser, 1999, doi: 10.1007/978-1-4612-1784-8_33.
- [16] A. Singla, C.-Y. Hong, L. Popa, and P. B. Godfrey, "Jellyfish: Networking data centers randomly," in *Proc. 9th USENIX Symp. Netw. Syst. Design Implement. (NSDI)*, 2012, pp. 225–238.
- [17] S. Knight, H. X. Nguyen, N. Falkner, R. Bowden, and M. Roughan, "The Internet topology zoo," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 9, pp. 1765–1775, Oct. 2011.
- [18] M. Mukamoto, T. Matsuda, S. Hara, K. Takizawa, F. Ono, and R. Miura, "Adaptive Boolean network tomography for link failure detection," in *Proc. IFIP/IEEE Int. Symp. Integr. Netw. Manage. (IM)*, May 2015, pp. 646–651.



VIVIANA ARRIGONI (Member, IEEE) received the B.Sc. degree in mathematics and the M.Sc. degree in computer science from the Sapienza University of Rome, Italy, where she is currently pursuing the Ph.D. degree with the Department of Computer Science. Her research interests include computational linear Algebra, network topologies, and information theory.



NOVELLA BARTOLINI (Senior Member, IEEE) received the degree (Hons.) and the Ph.D. degree in computer engineering from the University of Rome, Italy, in 1997 and 2001, respectively. She was a Visiting Scholar with The University of Texas at Dallas, in 2000, for one year, and a Research Assistant with the University of Rome Tor Vergata, from 2001 to 2002. She was a Visiting Professor with Penn State University, from 2014 to 2017, for three years. She is currently an Associate Professor with the Sapienza University of Rome. Her research interests include wireless networks and network management. She was the Program Chair and a Program Committee Member of several international conferences. She has served on the Editorial Board for *Computer Networks* (Elsevier) and *ACM/Springer Wireless Networks*.



ANNALISA MASSINI (Member, IEEE) received the degree in mathematics and the Ph.D. degree in computer science from the Sapienza University of Rome, Italy, in 1989 and 1993, respectively. Since 2001, she has been an Associate Professor with the Department of Computer Science, Sapienza University of Rome. Her research interests include hybrid systems, sensor networks, and networks topologies.