RC23599 (W0505-040) May 4, 2005
Computer Science

IBM Research Report

Fault Diagnois in Scale-free versus Random Networks

Natalia Odintsova, Irina Rish
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Fault Diagnosis in Scale-Free versus Random Networks

Natalia Odintsova, Irina Rish
IBM T.J. Watson Research
Hawthorne, NY 10532, USA
nodints,rish@us.ibm.com

Abstract

Cost-efficient problem detection and localization is one of the key requirements to a self-managing system. In this paper, we consider detection and diagnosis (localization) of faults in large-scale computer networks and distributed systems. Particularly, we investigate the effects of network topology (e.g., scale-free versus random graphs) on the cost-efficiency of detection and diagnosis in terms of the number of tests required and the resulting accuracy of diagnosis. Recent studies suggest that the topology of the Internet, world wide web, peer-to-peer (e.g., Gnutella) and many other real-life networks is quite far from being a random graph (i.e., classical Erdos-Renyi model). Instead, such networks exhibit scale-free properties, following power-low degree distributions (e.g., a small subset of nodes - hubs - are connected to a very large number of nodes while the majority of nodes have quite small degree). Our studies show that the network topology dramatically influences the cost-efficiency of diagnosis, such as the number of tests necessary for problem detection and diagnosis. In our studies, we use as tests end-to-end measurements, or probes (e.g., ping or traceroute command). Scale-free networks appear to be much harder to diagnose than random networks of the same size, as they require significantly larger number of probe stations and end-to-end probes. We investigate the effect of the probe stations' location on the necessary number of probes. We also estimate the computational complexity of multi-fault diagnosis in scale-free and random networks by Bayesian inference. Our studies provide an important information for planning and deployment of probing-based diagnosis in extremely large-scale practical networks (e.g., large corporate intranets, Internet, and GRID-computing systems).

1. Introduction

One of the key features of autonomic systems is the ability to analyze data in real-time and predict potential problems to avoid disastrous scenarios through prompt execution of remedial actions. For example, in IP network management, we would like to quickly identify which router or link has a problem when a failure or performance degradation occurs in the network. In the e-Commerce context, our objective could be to trace the root-cause of unsuccessful or slow user transactions (e.g. purchase requests sent through a web server) in order to identify whether it is a network problem, a web or back-end database server problem, etc. Another example is real-time monitoring, diagnosis and prediction of the health of a large cluster system containing hundreds or thousands of workstations performing distributed computations (e.g., Linux clusters or GRID-computing systems).

A commonly used approach to problem diagnosis in distributed systems management is *event correlation* [7,8,9], in which every managed device is instrumented to emit an alarm when its status changes. By correlating the received alarms a centralized manager

is able to identify the problem. However, this approach usually requires heavy instrumentation, since each device needs to have the ability to send out the appropriate alarms. Also, it may be difficult to ensure that alarms are sent out, e.g. by a device that is down. Finally, it might be impossible to obtain the event data from all parts of the network, especially if it contains ``black boxes'' such as proprietary components.

To avoid these problems, an alternative diagnostic approach has been developed that is based on *end-to-end probing* technology [6,1,2,3,4]. A *probe* is a test transaction whose outcome depends on some of the system's components; diagnosis is performed by appropriately selecting the probes and analyzing the results. In the context of distributed systems, a probe is a program that executes on a particular machine (called a probe station) by sending a command or transaction to a server or network element and measuring the response. The *ping* and *traceroute* commands are probably the most popular probing tools that can be used to detect network availability. Other probing tools, such as IBM's EPP technology [6], provide more sophisticated, application-level probes. For example, probes can be sent in the form of test e-mail messages, web-access requests, a database query, and so on.

In practice, probe planning (i.e., choice of probe stations, targets and particular transactions) is often done in an ad-hoc manner, largely based on previous experience and rules-of-thumb. Thus, it is not necessarily optimized with respect to probing costs (related to the number of probes and probe stations) and diagnostic capability of a probe set. More recent work [1,2,3,4] on probing-based diagnosis focused on optimizing the probe set selection and provided simple heuristic search techniques that yield close-to-optimal solutions. However, these approaches assume that an initial probe set is always provided, and do not address the problem of constructing a probe set from a given network topology. This problem includes optimal selection of both probe stations and probes, and becomes particularly complicated in large-scale networks (e.g., networks containing at least several hundreds or thousands of nodes, not even talking about practical networks that contain tens of thousands nodes, such as, for example, Internet at AS level).

In this paper, we consider detection and diagnosis (localization) of faults in large-scale computer networks and distributed system. Particularly, we investigate the effects of network topology (e.g., scale-free versus random graphs) on the number of tests necessary for detection and diagnosis. Recent studies suggest that the topology of the Internet, world wide web, peer-to-peer (e.g., Gnutella) and many other real-life networks is quite far from being a random graph (i.e., classical Erdos-Renyi model). Instead, such networks exhibit scale-free properties, following power-law degree distributions (e.g., a small subset of nodes - hubs - are connected to a very large number of nodes while the majority of nodes have quite small degree). Our studies show that the network topology dramatically influences the cost-efficiency of diagnosis, such as the number of tests necessary for problem detection and diagnosis. In our studies, we use as tests end-to-end measurements, or probes. Scale-free networks appear to be much harder to diagnose than random networks of the same size, as they require significantly larger number of probe stations and end-to-end probes. We provide some estimates of the number of probe stations and probes required to diagnose a given network, and investigate the effect of the

probe stations' location on the necessary number of probes. We also estimate the computational complexity of multi-fault diagnosis in scale-free and random networks by Bayesian inference. Our studies provide an important information for planning and deployment of probing-based diagnosis in extremely large-scale practical networks (e.g., large corporate intranets, Internet, and GRID-computing systems).

2. Experimental Setup

The experiments were conducted using networks built by INET generator [13], a generator which simulates scale-free properties of Internet topology at Autonomous System (AS) level. The original Internet topologies (BGP tables) were obtained from the National Laboratory for Applied Network Research (NLANR) and The University of Oregon Route Views project (BGP table information was transformed into adjacency lists describing 51 Internet topologies collected between Nov. 1997 and Feb. 2002). We experimented with networks of size from 500 to 1500, and compared them to random graphs of same size, obtained from corresponding scale-free graphs by a random edge rewiring process. Probe stations (or sources) were selected randomly (10 trials per each network, from 1 to 100 sources per trial). The probe set for detection was selected using a greedy algorithm that chooses next a probe covering the largest number of previously uncovered nodes.

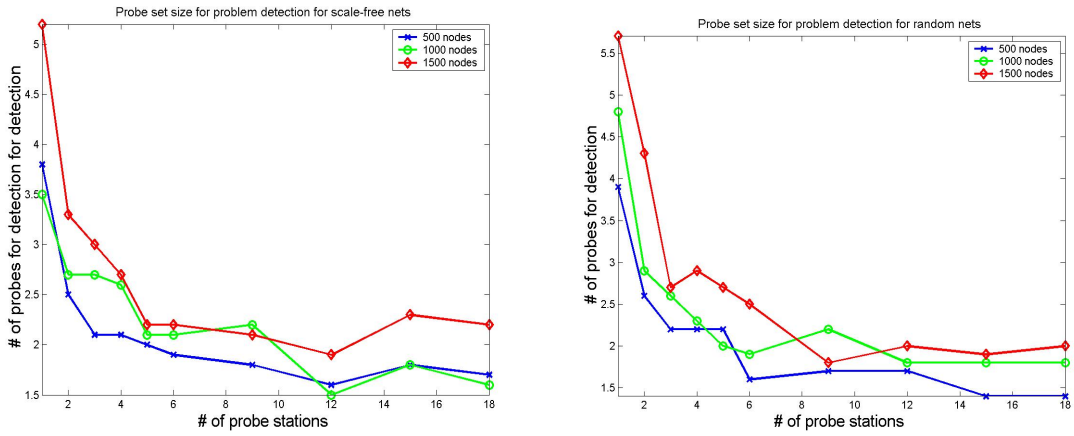
For simulating the probes' paths, we used two methods – random walks and shortest paths. We present results for both cases, however, since the shortest-path algorithm is more realistic as the one that is close to the network routing algorithm, we made emphasis on the shortest-path probes. However, even though random-walk probes are not supported in typical TCP/IP networks, it becomes possible in different types of distributed systems, such as peer-to-peer networks and also in application level probes where nodes denote software components rather than physical nodes and/or links, and a probe is a complicated transaction constructed out of software components in variety of ways.

2.1. Random-walk Probes

For random walks, each trial involved choosing randomly 20 targets among all the nodes of the network, and then choosing sources randomly among the rest of the nodes. The number of sources varied from 1 to 20. The resulting probe set was optimized to find the near-minimal probe set that allows to detect if there is any fault in the system (that is, the minimal number of probes necessary to cover all the nodes), and to find the minimal probe set that provides unique diagnosis of every single fault. For each network size and the number of sources the results were averaged over 10 trials.

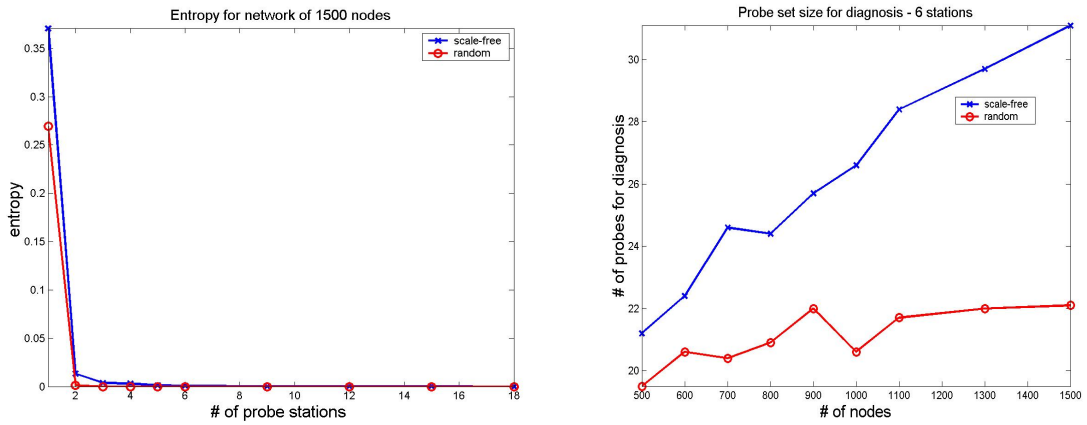
The probe set needed to cover the whole network in the case of random walks consists of just several probes (as could be expected given the property of random walks), and varies but slightly with network size, and with topology type. Choosing randomly as little as 4

sources and 20 targets yields the probe set that can be optimized to get the probe set for detection (that is, full network coverage) consisting of just 2-3 probes for networks as large as1500 nodes (Fig.1a,b). More importantly, such relatively small initial probe set (80 probes) is enough for unique diagnosis of a single fault in the system, which can be seen by analyzing conditional entropy of fault distribution given the probe outcomes incurred by the corresponding probe set. Entropy drops abruptly with the increase in number of sources (Fig. 2a), and becomes 0 for as little as 4 sources for network sizes of 1500 nodes. The initial probe set can be optimized (retaining its diagnostic power) to get near-optimal probe set for unique diagnosis of a single fault consisting of 20-30 probes, depending on network topology type. Fig.2b shows the size of (near) optimal probe set for diagnosis as a function of network size for scale-free and random graphs, for a fixed number of sources (4). Unlike probe set size for detection, which doesn't particularly depend on the topology type, probe set size for diagnosis is larger for scale-free graphs than for random – 32 versus 23 probes for the network of 1500 nodes, and the difference grows with the network size. That suggests that randomization of scale-free graphs will yield the network that is more tractable for diagnosis by random-walk probes.



(a) (b)

Figure 1. Random-walk probes: the number of probes needed to cover all nodes (i.e., detect a fault) versus the number of randomly selected probe sources (a) in scale-free networks of different sizes; (b) in random networks of different sizes.



(a) (b)

Figure 2. Random-walk probes: (a) conditional entropy of fault distribution given the probe outcomes in scale-free versus random networks of size 1500 nodes; (b) near-optimal probe set size for diagnosis of different network sizes and fixed number of probe sources (6).

2.2. Shortest-path Probes

For shortest paths, at each trial the probe set was constructed as follows. From each of the sources, chosen at random, their number being varied from 1 to 100, shortest paths were built to every node of the network. The resulting probe set was optimized to find minimal number of probes necessary for problem detection. The probe set for diagnosis of a single fault in the system was constructed by extending the probe set for detection: first, the conditional entropy of fault distribution incurred by the probe set for detection was calculated, and then additional probes were chosen greedily among the remaining probes, at each step selecting the probe that yields the maximum reduction in entropy, till the entropy was reduced to 0. For each network size and the number of sources the results were averaged over 10 trials.

It appeared that in case of shortest path probes, network topology drastically influences the number of probes necessary for detection and diagnosis, as well as affects how these numbers change with the change in the number of sources. Scale-free networks require significantly more probes than random ones of the same size (Fig.3.a) The behavior of probe set size for problem detection as a function of number of sources doesn't depend on the network size (the whole curve just shifts along y-axes), but depends on the type of the network topology. In scale-free graphs, the number of probes for detection decreases slowly with the increase in number of sources (Fig. 3.b); in random graphs, the probe set size drops abruptly, reaching saturation point, after which further increase in the number of sources doesn't yield essential reduction in the number of probes (Fig. 3.c). This provides for the good trade-off between the number of sources and the number of probes necessary to cover the whole network.

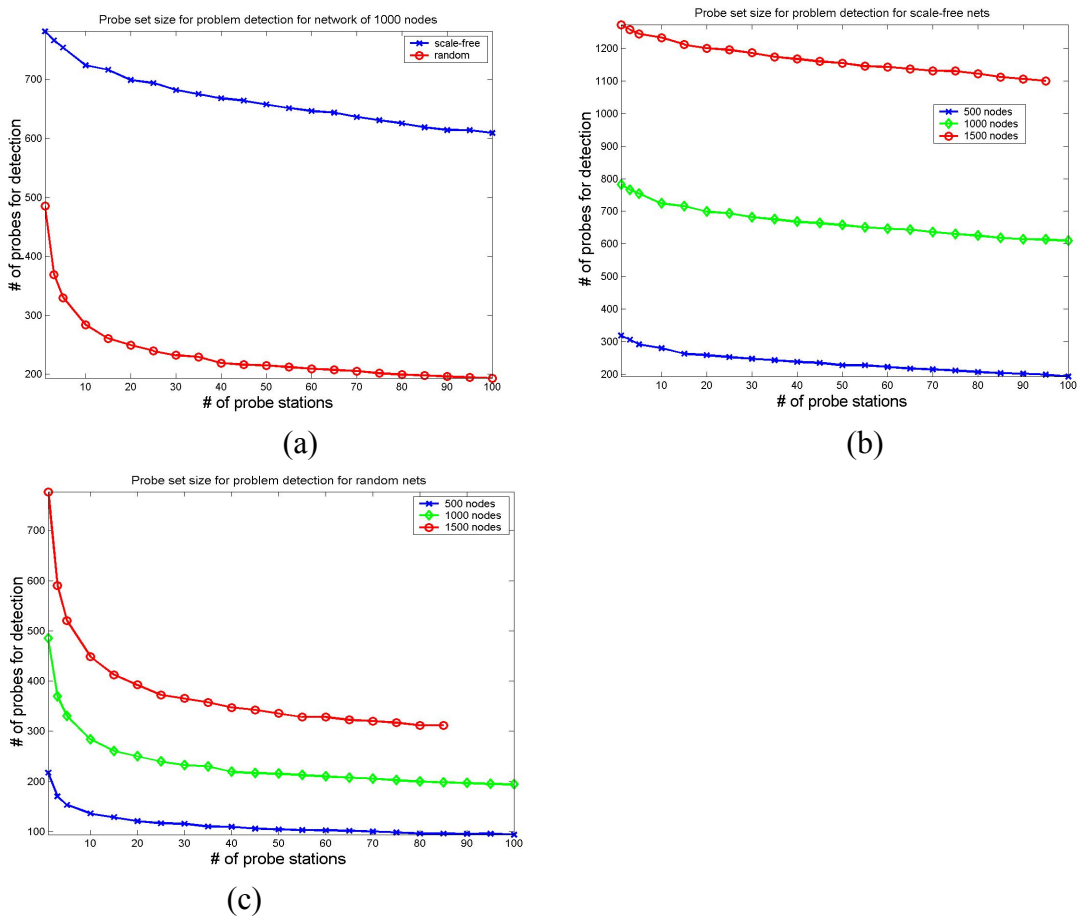
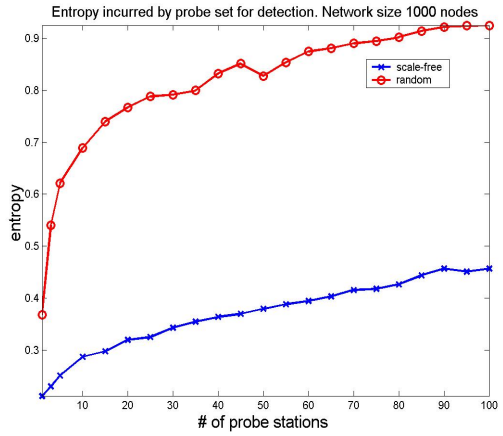


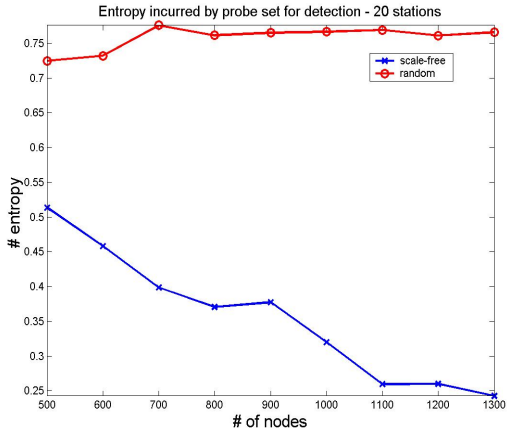
Figure 3. Shortest-path probes: the number of probes needed to cover the whole network (i.e., to detect a fault) versus the number of randomly selected probe sources: (a) in scale-free versus random networks of size 1000 nodes; (b) in scale-free network of different sizes; (c) in random networks of different sizes.

Although (and largely due to the fact that) the probe set size for problem detection is smaller for random graphs, the conditional entropy of fault distribution provided the probes outcomes is higher for such graphs as compared to scale-free graphs (Fig. 4a), which means that more additional probes are needed to provide for unique diagnosis of every single fault in the system.

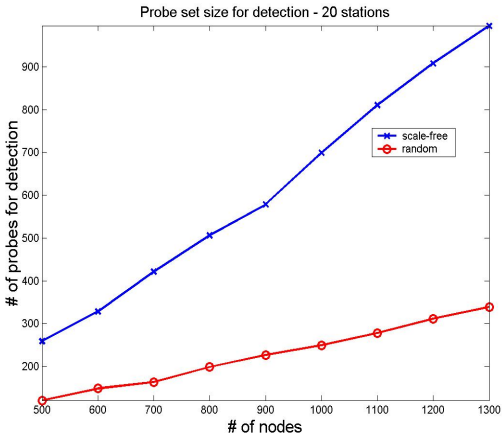
Fig.4b shows the number of probes for detection versus the network size for the fixed number of sources (20) for scale-free and random graphs, and Fig. 4c – conditional entropy of fault distribution incurred by corresponding probe sets. The difference between scale-free and random graphs in the number of probes necessary to cover the network can be explained by the fact that the average length of the shortest path is smaller for scale-free graphs (Fig.5.a), and so more probes are needed to cover the network of the same size.



(a)

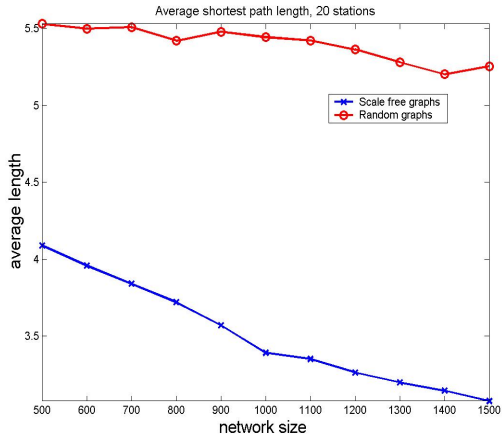


(b)

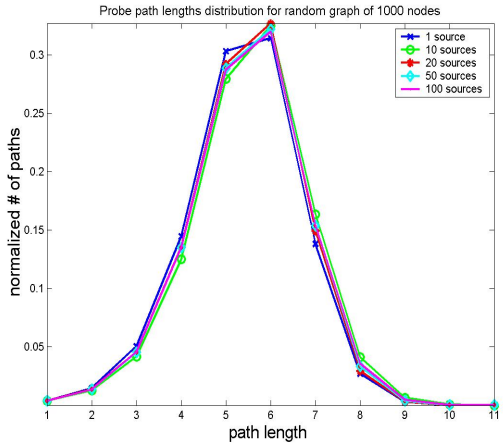


(c)

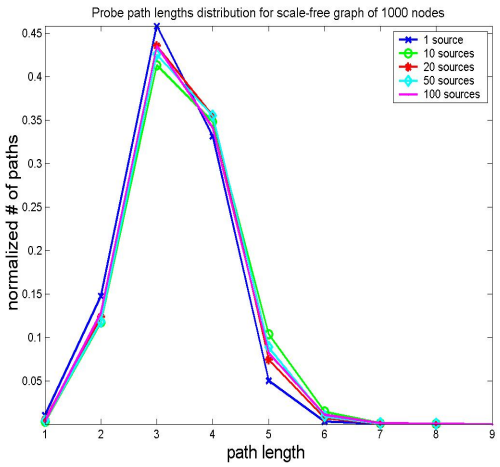
Figure 4. (a) Conditional entropy of fault distribution given the probe outcomes in scale-free versus random networks of size 1000 nodes; (b) Conditional entropy of fault distribution given the probe outcomes for different network sizes and fixed number (20) of probe sources; (c) The number of probes needed to cover all nodes (i.e., detect a fault) for different network sizes and fixed number (20) of probe sources.



(a)



(b)



(c)

Figure 5. Shortest-path probes: (a) average probe length versus network size; probe lengths distribution (recall that a probe follows the shortest path between the source and the destination) for (b) random and (c) scale-free networks of 1000 nodes.

The number of probes needed to extend probe set for detection to provide for unique diagnosis of any single fault practically doesn't change with the number of probe sources, for both scale-free (Fig. 6a) and random (Fig. 6b) networks – the probe set size for diagnosis curve is just shifted along y-axes.

As with probe set for detection, probe set for diagnosis for random graphs first decreases rapidly with the number of sources, until it reaches some saturation point, after which no significant reduction in the probe set size can be achieved. For both, random and scale-free graphs, probe set size for diagnosis increases linearly with the network size (Fig.6c) for a fixed number of sources.

The trade-off between number of sources and number of probes for detection and diagnosis that is observed in random graphs, allows to estimate the optimal number of sources for networks of random topology – 10-20 sources, independent of the network size.

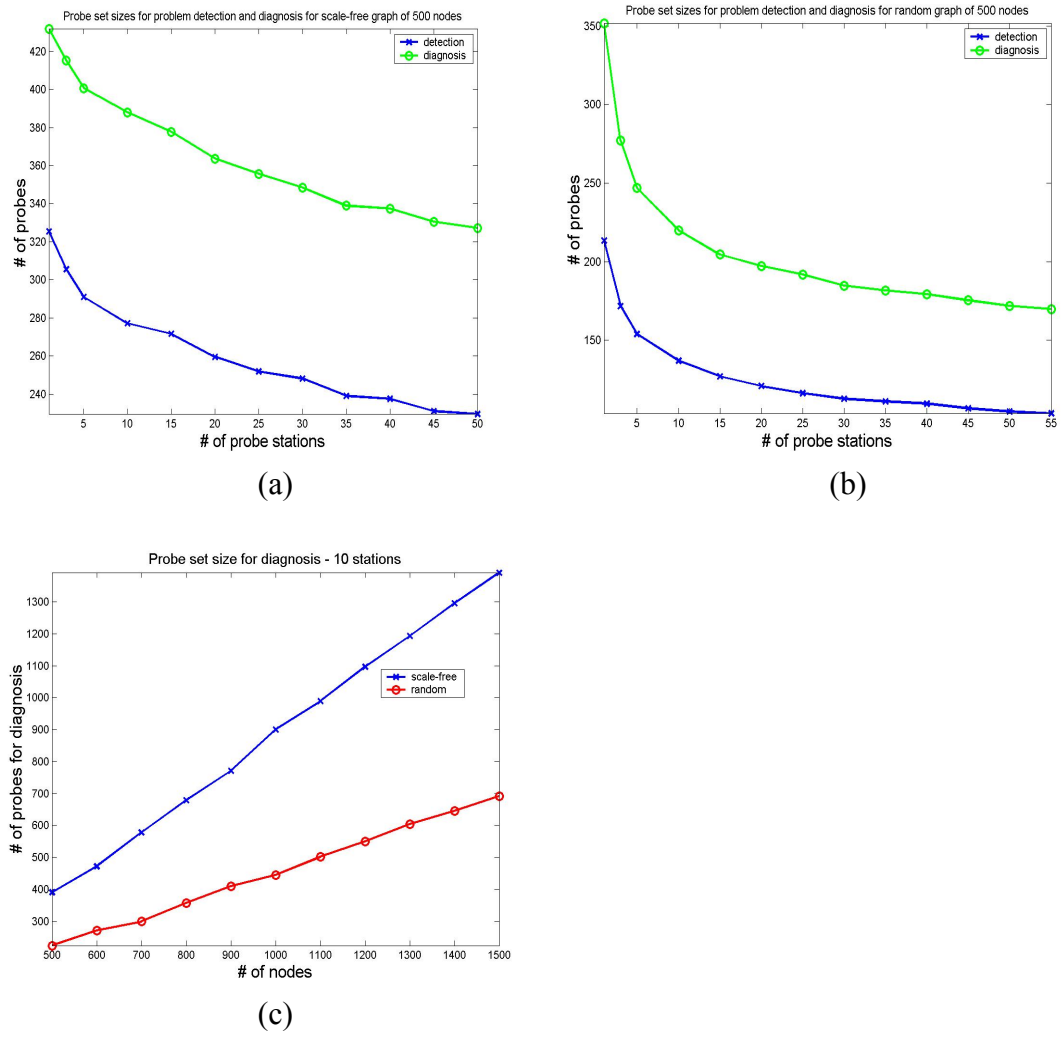


Figure 6. Probe set for diagnosis, obtained by greedily extending probe set for detection for (a) scale-free and (b) random networks of 500 nodes. (c) linear growth of the probe set size for diagnosis with increasing network size and a fixed number (10) of probe stations.

3. Computational Complexity Issues

We now estimate complexity of multi-fault diagnosis by Bayesian inference when no restriction on the number of faults in the system is imposed. As a method of inference, variable elimination method is considered. Complexity of variable elimination method is known to be exponential in the induced width of the probabilistic network's moral graph along the given variable ordering. The algorithm uses min-degree heuristic to find suboptimal node ordering (the induced width of suboptimal ordering is an upper bound on the induced width of the graph). Since probes are observed, we can effectively remove them from the moral graph, and consider the induced width of the remaining graph. The quality of diagnosis depends on the quality of the probe set. The full probe set, consisting of all available shortest-part probes (recall that we construct this set by sending shortest-path probes from randomly selected sources to every node of the network) has large

induced width even for a small number of probe sources (Fig. 7a), which makes variable elimination algorithm intractable. Since reducing the probe set may result in reduction in accuracy of the diagnosis algorithm, we're looking for complexity vs. accuracy trade-off. We investigate complexity and accuracy of diagnosis of probe sets for problem detection, and probe sets that provide unique diagnosis of any single fault. Although the sizes of both probe sets decrease with the number of probe sources, the induced width for these sets increases with the number of sources because the probes have more intersections. In case of scale-free networks, induced width grows slowly (almost logarithmically) with the number of sources. For random networks, we observe the abrupt growth, until it reaches a saturation point (Fig. 7b, c). Induced width in scale-free networks practically doesn't depend on network size; in random graphs, it grows linearly with the network size (Fig.8 a, b). For scale-free networks, the induced width of probe set for problem detection (that is, the probe set that covers every node) is relatively low (5-12) for all network sizes, which makes multi-fault diagnosis by such set of probes tractable, although the accuracy of this diagnosis need to be determined.

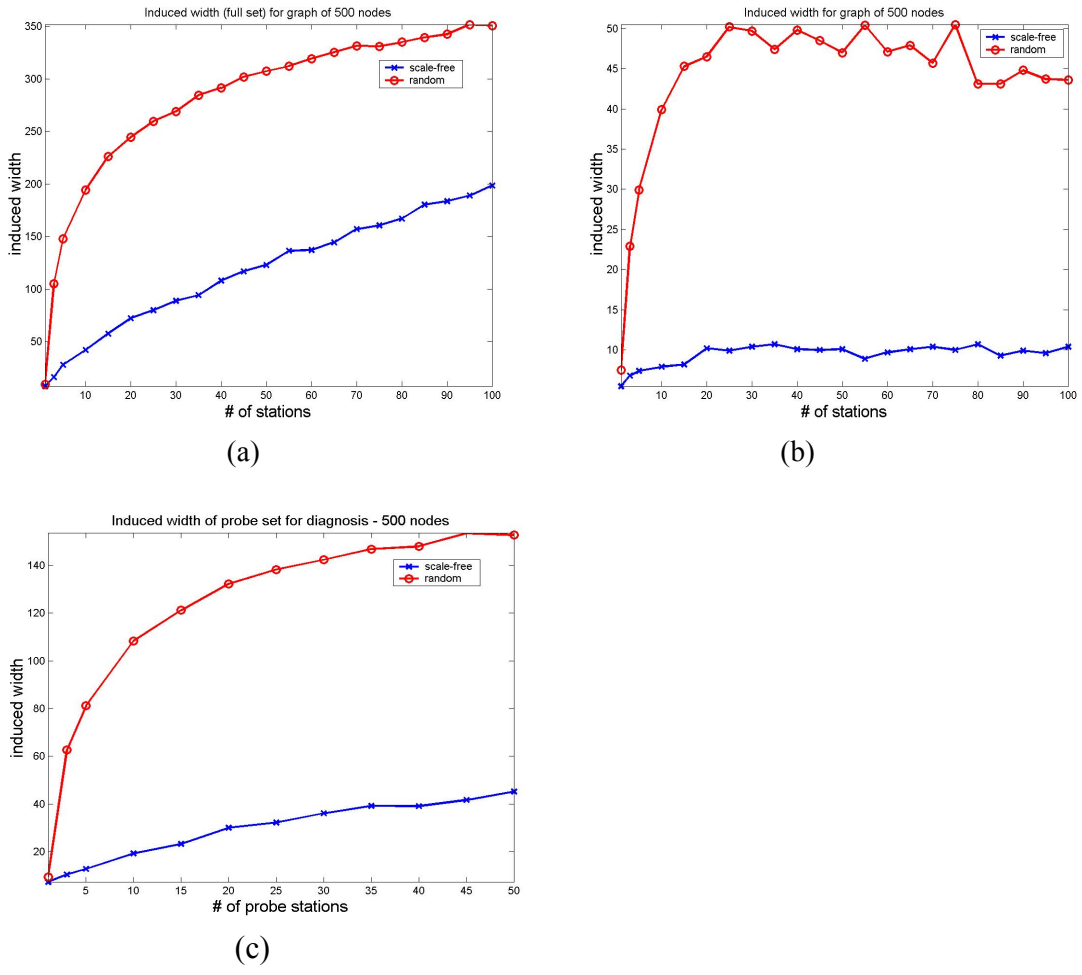
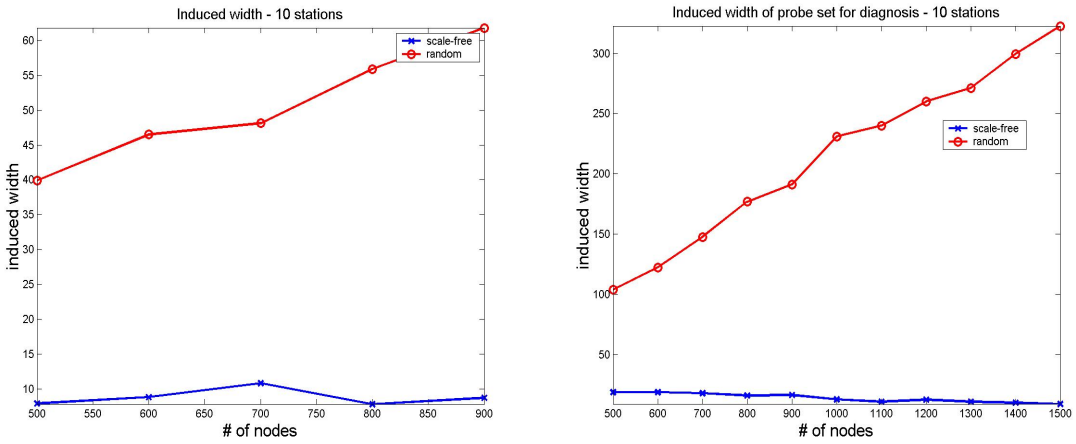


Figure 7. Induced width of the graph underlying the probe set as a function of number of probe sources for scale-free versus random graphs: (a) for the full probe set; (b) for the probe set for problem detection; (c) for the probe set for unique diagnosis of any single fault.

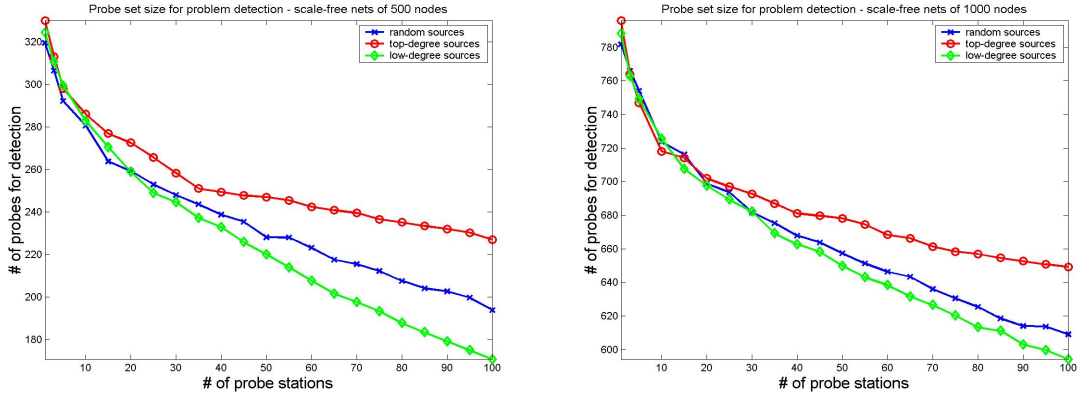


(a) (b)

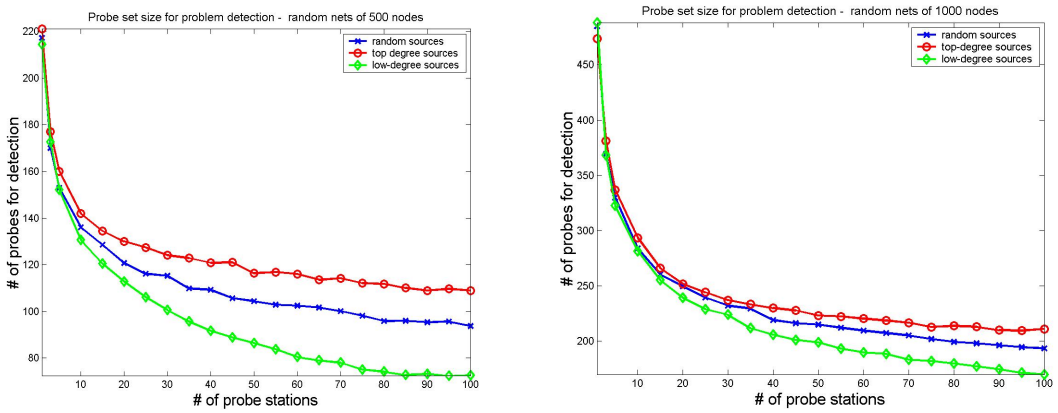
Figure 8. Induced width of the graph underlying the probe set as a function of the network size for the fixed number (10) of probe sources for the scale-free and random graphs:
(a) for probe set for problem detection; (b) for probe set for unique diagnosis of any single fault.

5. Probe Station Selection

The choice of probe stations (sources) influences the size of the resulting probe set. We constructed probe sets for problem detection for various methods of the stations choice – totally random, random among top-degree nodes, and random among low-degree nodes. The difference between these methods is noticeable only for sufficiently large number of sources, in which cases locating probe stations at the leaves rather than hubs seems to be a better choice for both scale-free (Fig. 9 a, b) and random (Fig. 9 c, d) networks.



(a) (b)



(c) (d)

Figure 9. Probe set size for problem detection for various methods of choosing probe stations (totally randomly, randomly among top-degree nodes, randomly among low-degree nodes) for scale-free (a, b) and random (c, d) graphs of different sizes.

5.1 Heuristic for Choosing Probe Stations

For the case when all possible probes can be enumerated, we offer a heuristic for choosing probe stations that helps (on average) reduce the probe set size for detection for a given topology and a given maximum allowed number of sources. The method works as follows: from the given probe set, we greedily choose most informative probes, increasing the source count every time the chosen probe goes from the new source. When the source count reaches the maximum allowed number of sources, we remove from the probe set all probes that don't start in the sources that are already chosen. We then proceed choosing most informative probes from the remaining set. The process ends when the resulting probe set is complete, that is, provide for unique diagnosis, if we are constructing probe set for diagnosis, or covers all nodes, if we are constructing probe set for problem detection. The set of sources is determined by this probe set.

Fig. 10 compares the cost-efficiency of problem detection (in terms of probe set size) for the cases when the sources were chosen using the heuristic and when the sources were chosen randomly for random (a, b) and scale-free (c, d) graphs. The heuristic yields smaller probe sets than the random choice of sources on average. However, for the small number of sources there is practically no difference between random and heuristic source selection. This can be explained by the fact that as the number of sources grows, so does the probability to choose a redundant probe station when choosing the sources randomly. By redundant probe station we mean a station whose probes don't provide additional information about the state of the system, once the probes from other stations in the set are chosen. Such redundant stations will be idle. Unlike that, heuristic choice will include in the source set only such probe stations that are actually needed for suboptimal probe set construction. More efficient use of probe stations results in the decrease of the probe set size. Note also, that since the probability of choosing redundant stations drops with the increase in network size for the fixed number of sources, the difference between

heuristic and random choices becomes less essential, as can be seen by comparison between Fig. 10 (a) and (b), or (c) and (d) .

When no boundary on the number of sources is imposed, the resulting probe set equals to the suboptimal set obtained by the greedy algorithm for probe selection (for the given probe ordering). The number of sources resulting from this probe set is the maximum number of sources needed to optimize the initial probe set by the greedy algorithm. Increase in the number of sources beyond this point does not produce reduction in the probe set size (see the point corresponding to 35 sources on the "heuristic" curve of Fig. 10 a). However, such number of sources can be quite large (35 sources for 200 nodes, as in Fig. 10 a). The trade-off between the number of sources and the number of probes can be determined by constructing suboptimal probe sets for gradually increasing number of sources and then choosing the probe set corresponding to a saturation point. Random graphs provide for a distinct trade-off point (fig. 10 a, b), but scale-free graphs don't (fig. 10 c, d).

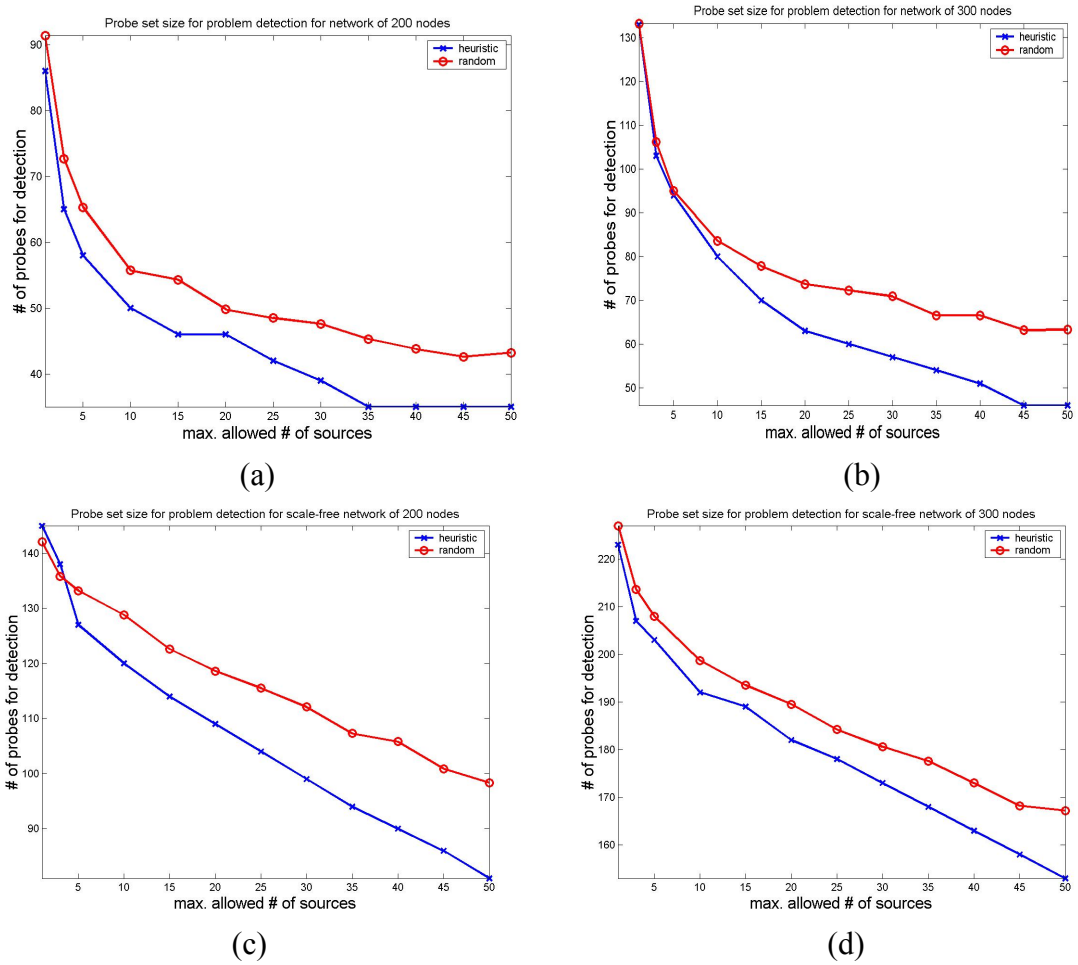


Figure 11. Comparison of the cost-efficiency of the problem detection for the cases when the sources were chosen using heuristic and when the sources were chosen randomly for random (a, b) and scale-free (c, d) graphs of various sizes.

Conclusions

Our experiments show that random graphs require much less probes for problem detection and diagnosis than scale-free graphs do, if the probes follow the shortest-path algorithm. Furthermore, random graphs offer a distinct trade-off point between the number of probe stations and the necessary amount of probes; we estimated that 10-20 sources is the optimal number of sources, independent of the network size. As for scale-free graphs, increase in the number of sources almost does not affect the resulting amount of necessary probes. In addition, we show that constructing probes as random walks instead of shortest-paths, results in significantly smaller (order of magnitude) number of probes necessary for detection and diagnosis, as compared to the shortest-path probes.

We studied the effect of the network topology on the complexity of multi-fault diagnosis by Bayesian inference. Although scale-free graphs require much larger amount of probes for detection and diagnosis, the computational complexity of diagnosis (measured by induced width of the corresponding Bayesian network) is significantly smaller in case of scale-free topology. As our experiments show, this complexity is relatively low and does not depend on the size of the network, if the topology is scale-free. The low complexity makes multi-fault diagnosis in scale-free graphs tractable by Bayesian inference, although the accuracy of such diagnosis depends on the prior probability of fault in each node.

We investigated the effect of probe stations' location on the amount of necessary probes, and showed that the degree of nodes designated as probe stations doesn't really affect the resulting probe set size, especially when the number of sources is small. However, locating probe stations at leaves rather than hubs, yields reduction in the necessary probe set size. For the case when all possible probes can be enumerated, we offer a heuristic for choosing the probe stations that helps reduce the number of necessary probes.

References

1. M. Brodie, I. Rish, and S. Ma. Optimizing probe selection for fault localization. In Distributed Systems Operation and Management, 2001.
2. M. Brodie, I. Rish, and S. Ma. Intelligent probing: a Cost-Efficient Approach to Fault Diagnosis in Computer Networks, in IBM Systems Journal (special issue on AI) 41(3): 372-385.
3. M. Brodie, I. Rish, S. Ma, N. Odintsova. Active Probing Strategies for Problem Diagnosis in Distributed Systems, in Proceedings of IJCAI-03, Acapulco, Mexico.
4. I. Rish, M. Brodie, N. Odintsova, S. Ma, G. Grabarnik. Real-time problem determination in distributed systems using active probing, NOMS-04.
5. T. M. Cover and J. A. Thomas. Elements of Information Theory, New York, John Wiley & Sons, 1991.
6. A. Frenkiel and H. Lee. EPP: A Framework for Measuring the End-to-End Performance of Distributed Applications. In *Proceedings of Performance Engineering 'Best Practices' Conference, IBM Academy of Technology*, 1999.
7. B.Gruschke. Integrated Event Management: Event Correlation Using Dependency Graphs. In *Distributed Systems Operations and Management*, 1998.
8. S. Kliger, S. Yemini, Y. Yemini, D. Ohsie, and S. Stolfo. A coding approach to event correlation. In *Intelligent Network Management (IM)*, 1997.

9. A. Leinwand and K. Fang-Conroy. *Network Management: A Practical Perspective, 2nd Edition*. Addison-Wesley, 1995.
10. J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, 1988.
11. I. Rish, M. Brodie, and S. Ma. Accuracy vs. Efficiency Trade-offs in Probabilistic Diagnosis. In *Proceedings of the The Eighteenth National Conference on Artificial Intelligence (AAAI-2002), Edmonton, Alberta, Canada*, 2002.
12. https://ibm-odyssey.raleigh.ibm.com/Odyssey/
13. J. Winick and S. Jamin. Inet-3.0: Internet topology generator, Tech Report CSE-TR-456-02, University of Michigan, 2002.