Active probing in Grids

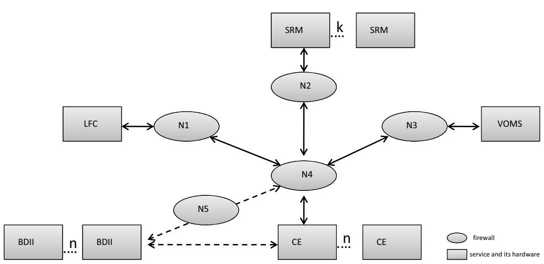
Dawei Feng, Cecile Germain, Michele Sebag
TAO. L.R.I. University Paris-sud

Introduction

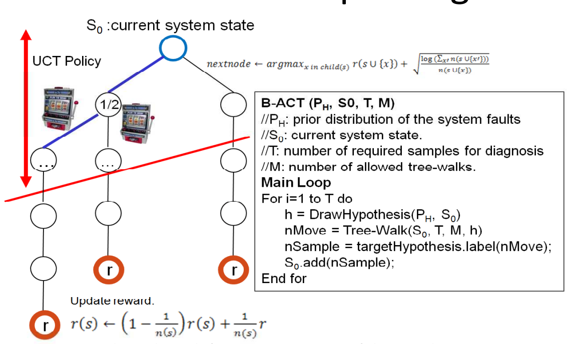
- End to end probing, gathers information about system hidden state via test transaction and analyzes result through probabilistic inference.
- Active probing, selects next most informative probe based on previous tested probes' results.
- UCT based algorithm, models the selection of next probe as a bandit based game, instead of selecting next probe greedily it tries to foresee some steps further through simulation.

Active fault diagnosis in Grids

- Main goal: self-regulation and maintenance of the grids.
- First step: test the availability of all the SEs (storage element) from each CE (computing element) in the biomed grids.
- Currently 10 probes in total, e.g. voms-proxy-init, nmap-bdii,nmap-lfc,lfc-ls,srm-ls,nmap-srm,lcg-cr,lcg-cp,lcg-infosites,nmap-voms.
- A dependency matrix describes the relationship between probes and tested components is given by grid expert.
- Two steps for diagnosis, 1st select a small set of probes which cover all the components in the system, if a problem detected then, 2nd diagnose the faulty component by launching probes progressively.

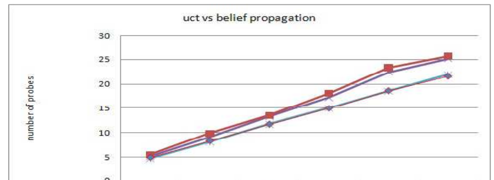


UCT based active probing



- draws a surrogate hypothesis h from version space of the initial set S_0.
- uses the Upper Confidence Bounds(UCB) criterion to select some probes X_{t+1} to be labeled, next state is $S_{t+1} = S_t \cup (X_{t+1}, h(X_{t+1}))$.
- tree walk proceeds until arriving in a tree leaf. some additional probes are selected uniformly, labeled after h and adding to S_0.
- main loop stops until find a faulty component or reach sample requirements and return the most visited child of S_0.
- each node's reward is the information gain with respect to state S_0.

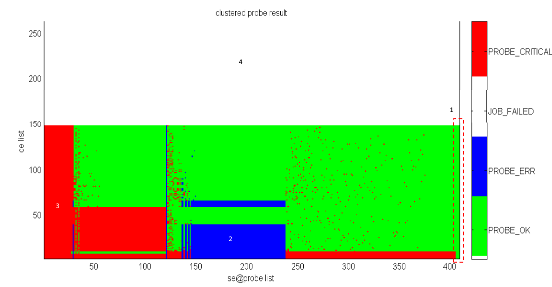
Simulated result of B-ACT



	20	40	60	80	100	120
powerlaw-bp	5.1	9.05	13.31	17.24	22.48	25.15
uct-powerlaw	5.47	9.78	13.47	18.07	23.29	25.64
random-bp	4.74	8.17	11.73	15.06	18.53	21.94
uct-random	4.88	8.29	11.7	14.96	18.65	21.62

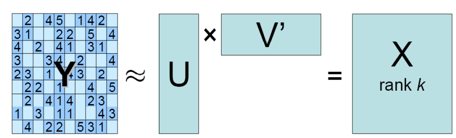
- Two types of simulated graphs: random graph and power-law graph.
- A top k strategy (next top k probes which cover the most number of unknown components in the system) is used in the UCT's random simulation.
- Average number of needed probes to diagnose single fault is linearly increasing with the size of the network.
- The performance of UCT based active probing is almost the same as that of belief propagation.

Result classification & analysis



- Currently, there are 10 testing probes executed from 260 computing elements to test the their accessibility to 120 storage elements in the grid each day.
- Conclusions: 1) basic services such as lfc, bdii, voms always functioned correctly. 2) some SEs were not accessible for all CEs while some SEs were not accessible for some CEs all the time. 3) high rate of job failure lead to the inaccurate diagnosis of the whole system.

Prediction using MMMF

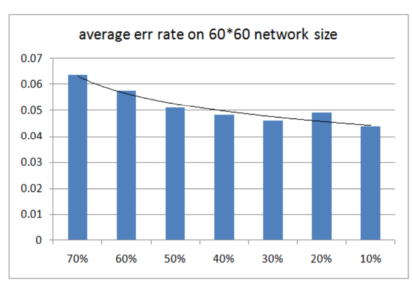


- Bounded norm constraint for the replacement of bounded rank constraint in other matrix factorization method.

$$\text{minimize } \|X\|_{\text{tr}} + c\sum \max(0, 1 - Y_{ia}X_{ia}).$$

Where $\|X\|_{\Sigma}$ is the trace-norm of matrix X, c is a trade-off constraint, S is the set of all non-missing entries in the given matrix Y, $\sum_{ia\in s} \max(0, 1 - Y_{ia}X_{ia})$ is the hinge-loss.

- Convex function and can be solved by standard optimization techniques.



- About 95% prediction accuracy when 50% percent of the test entries are missing.

Reference

1. Rish, I., Brodie, M., et al. (2005). Adaptive Diagnosis in Distributed Systems Adaptive Diagnosis in Distributed Systems. Technology, 23698.
2. Philippe Rolet, Michele Sebag, Olivier Teytaud. (2009) Boosting active learning to optimality: the BAAL algorithm.
3. Nathan Srebro, Jason D.M. Rennie, et al. Maximum-Margin Matrix Factorization, nips 04.