Selectively Retrofitting Monitoring in Distributed Systems

Animashree Anandkumar† animakum@mit.edu

Chatschik Bisdikian‡ bisdik@us.ibm.com

Ting He‡ the@us.ibm.com

Dakshi Agrawal‡ dakshi@us.ibm.com

†Dept., of EECS, Massachusetts Institute of Technology, Cambridge, MA 02139.

‡Networking Tech., IBM Watson Research, Hawthorne, NY 10532.

ABSTRACT

Current distributed systems carry legacy subsystems lacking sufficient instrumentation for monitoring the end-to-end business transactions supported by these systems. In the absence of instrumentation, only probabilistic monitoring is possible by using time-stamped log-records. Retrofitting these systems with expensive monitoring instrumentation provides high-granularity, precise tracking of transactions. Given a limited budget, local instrumentation strategies which maximize the effectiveness of monitoring transactions throughout the system are proposed. The operation of the end-to-end system is modeled by a queuing network; each queue represents a subsystem which produces time-stamped log-records as transactions pass through it. Two simple heuristics for instrumentation are proposed which become optimal under certain conditions. One heuristic selects states in the transition diagram for local instrumentation in the decreasing order of the load factors of their queues. Sufficient conditions for this load-factor heuristic to be optimal are proven using the notion of stochastic order. The other heuristic selects states in the transition diagram based on the approximated tracking accuracy of probabilistic monitoring at each state, which is shown to be tight at low arrival rates.

Categories and Subject Descriptors

H.3.4 [**Systems and Software**]: Distributed systems— *Performance evaluation*

General Terms

Performance, Measurements, Theory

Keywords

Probabilistic transaction monitoring, Queuing networks, Stochastic comparison, Bipartite matching.

1. INTRODUCTION

Many enterprise systems in operation have evolved over a long period of time. They started out as small ventures and underwent rapid unplanned expansion in the face of increasing demand and business consolidations. New systems and applications were installed and old systems were never fully upgraded, leading to a lump of heterogeneous systems with varied OS and hardware configurations. Complete documentation of these end-to-end systems is hardly available. Troubleshooting errors are usually handled manually on a case-by-case basis leading to long troubleshooting delays and increased man-hour expenses.

System troubleshooting can be handled in a more organized fashion by continuously monitoring all the transactions supported by these systems. However, in the absence of sufficient instrumentation, e.g., according to ARM guidelines [1], precise and effective monitoring is hard to come by. Instead, troubleshooting can only rely on limited information in the form of log-records (or, transaction *footprints*), which usually contain timestamps for different events in that subsystem. However, in legacy systems, there is no guarantee that the log-records, along with the timestamps, carry any form of identifiers for the transactions. It thus becomes important to find the set of footprints which are most-likely generated by each transaction instance as it travels through different subsystems.

Such probabilistic monitoring based on timestamps can only achieve a certain level of accuracy. The monitoring accuracy can be improved through retrofitting instrumentation locally at various subsystems. For instance, instrumentation at each state in a state-transition model provides precise tracking of progress of transactions in the instrumented state by introducing a local identifier for each transaction entering the state. If all states are instrumented, then starting with the entry point of a transaction into the system, transaction footprints can be tracked through all the applications traversed by the transaction. However, this is an expensive proposition, and when the instrumentation budget cannot cover all the states, we need to decide which states to instrument, see Fig.1. What follows below is a pursuit of easy to use guidelines for spending the instrumentation budget.

2. SUMMARY OF CONTRIBUTIONS

We model the distributed system using a state-transition diagram, where each state represents a subsystem or an application traversed by the transactions. Each state is modeled as an infinite-server queue $./GI/\infty$ with semi-Markovian service times [3]. We assume Poisson arrivals into the system at the unique start state with rate λ. For every transaction that passes through a state, an entry and an exit footprint consisting of timestamp records are available. Crucially, the footprints do not contain the identity of the transaction generating them due to lack of monitoring instrumentation.

In the absence of any monitoring instrumentation, only probabilistic transaction tracking is possible using the footprint records. The difficulty in matching footprints belonging to the same transaction at entry and exit points comes from uncertainties introduced by out-of-order progress of transactions in the queueing system, especially when the

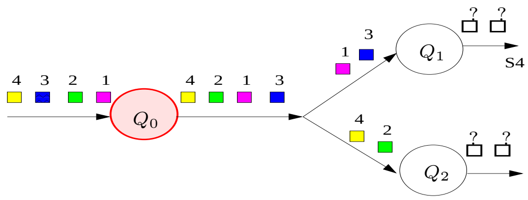

Figure 1: Local instrumentation at queue Q_0 precisely tracks transactions entering and leaving it. But uninstrumented queues Q_1 and Q_2 have to rely on entry and exit timestamps for probabilistic monitoring.

Instrument $E = 2$ out of $|\mathcal{Q}| = 10$ queues, service rates $\mu_k \overset{i.i.d.}{\sim} \text{Unif}[0.5, S_{\max}]$, Weibull shape parameter $w_k \overset{i.i.d.}{\sim} \text{Unif}[0.1, 2]$, 1000 configurations, unit arrival rate and costs.

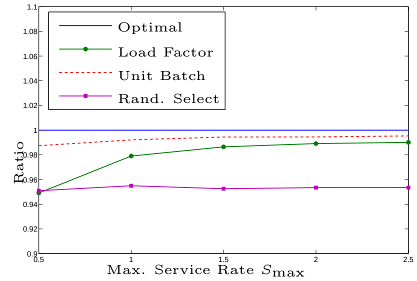

Figure 2: Comparison of instrumentation strategies. Ratio $= \dfrac{\textbf{Performance under heuristic}}{\textbf{Optimal Performance}}$.

queues undertake parallel processing of the transactions. In particular, the optimal maximum-likelihood (ML) rule and the simple sub-optimal first-in first-out (FIFO) algorithms for matching transactions in the absence of instrumentation have been suggested in [2]. These automated probabilistic matching algorithms yield a correct match with a certain probability for a set of entry and exit footprints at each state.

After instrumentation of a particular state, transaction footprints corresponding to the entry and exit of transactions can be matched precisely since instrumentation produces a unique local identifier for each transaction. However, there are monetary and human-capital costs associated with retrofitting instrumentation. Thus, we have an optimization problem where within a given budget constraint, we need to select states in the transition diagram for instrumentation such that the total accuracy of matching transactions at all the states is maximized.

We propose two heuristics for instrumentation which are optimal under some conditions. One of them is based on the *load factors* of the queues at different states in the transition diagram. The load factor of a queue is the ratio of the average arrival rate to the average service rate. The other heuristic is based on an approximate expression of monitoring accuracy at each queue and is tight at low arrival rates.

The load-factor heuristic is based on the intuition that a higher load factor for an infinite-server queue leads to a higher number of transactions being simultaneously serviced, and hence, introduces a higher uncertainty to transaction tracking. Therefore, a simple heuristic is to instrument states in the reverse order of the load factors of their queues. But the service times at queues can follow different distributions and even belong to different distribution families. For example, if one queue has a deterministic service time with a high load factor, then the load-factor heuristic makes a wrong prediction that the particular state needs to be instrumented. However, in reality, we can track transactions precisely with deterministic service times.

One of the main contributions of the paper is to provide sufficient conditions for the optimality of the load-factor heuristic under different service distributions and automated algorithms for matching footprints in the absence of instrumentation. In particular, we consider two simple matching rules, viz., first-in first-out (FIFO) and random matching. Under FIFO matching if the service times, normalized by their arrival rates, and also their absolute *spreads* follow the same stochastic order, then the load-factor heuristic is optimal. Under random matching, a stochastic order on the normalized service times along with a mild condition on their support is sufficient to establish the optimality of the load-factor heuristic. For the special case when all the service distributions belong to the same family (e.g., all are exponentially distributed), the load-factor heuristic is guaranteed to be optimal, and no additional conditions are needed.

We propose another heuristic based on the approximate expressions of accuracy of the probabilistic matching algorithms. In general, computing the exact expressions of matching accuracy is NP-hard and our approximations are tight in the regime of low arrival rates. Intuitively at low arrival rates, over all possible busy-period sizes of the queue, the dominant event is having a busy period of size unity. Hence, we approximate the matching accuracy by the probability of having a unit-sized busy period, which is simple to evaluate. We then propose instrumentation strategies based on this approximation.

Fig.2 highlights some evaluation results of our work. Specifically, in Fig.2, we plot the ratio of performance under the two heuristics and the optimal instrumentation strategy. As a benchmark, we also plot the performance ratio under ad hoc instrumentation (i.e., randomly selecting states for instrumentation). The service rates are drawn i.i.d. uniformly between a minimum and a maximum service rate. Similarly, the shape parameters for the Weibull-distributed service times are also drawn uniformly from a given interval. We vary the maximum service rate to obtain more diverse queues with faster average service. We observe that both the load-factor heuristic and the unit-batch approximation approach optimality as the service rates increase which is not true for random selection. Hence, our proposed schemes are simple to implement and are efficient, especially at high service rates (or low load factors).

2.1 Problem Formulation

Let γ be the particular matching policy employed at a queue[1] Q_k in the state transition diagram, such as the first-in first-out (FIFO) match or the random match in the absence of instrumentation. Let $P_B^\gamma(k)$ be the probability of correctly matching *all* the footprints generated on entry and exit to queue Q_k that belong to the same busy period of size B, and let $\mathbb{E}[P_B^\gamma(k)]$ denote its expectation over the arrival and service distributions at queue Q_k, assuming initially empty queue.

[1]We use terms state and queue interchangeably.

One possible selective instrumentation strategy is to maximize the total expected matching accuracies along different queues in the transaction model. Note that after a queue has been instrumented, matching entry and exit footprints for that queue is precise and fail-proof. For each queue Q_k, let $z_k \in \{0,1\}$ be the indicator if queue Q_k is selected for instrumentation. Then, the monitoring accuracy at queue Q_k after instrumentation is given by

$$\mathbb{E}[P^{\text{EFF}}(k); z_k] = z_k + (1 - z_k)\mathbb{E}[P_B^{\gamma}(k)], \tag{1}$$

since there is unit accuracy after instrumentation and $\mathbb{E}[P_B^{\gamma}(k)]$ is the accuracy based only on the timestamped footprints.

Therefore, given a set of initially uninstrumented queues $\mathcal{Q}$ in the queueing network, a total budget of E for instrumentation, and costs C_k to instrument the queue $Q_k \in \mathcal{Q}$, we have the following 0-1 integer program

$$\mathbf{z}_*(E; \mathcal{Q}) := \arg\max_{\mathbf{z}} \sum_{Q_k \in \mathcal{Q}} \mathbb{E}[P^{\text{EFF}}(k); z_k], \tag{2}$$

$$s.t. \sum_{Q_k \in \mathcal{Q}} z_k C_k \leq E, \quad z_k \in \{0,1\}, \mathbf{z} := \{z_k, Q_k \in \mathcal{Q}\}.$$

3. SUMMARY OF OUR SOLUTIONS

The optimal instrumentation problem in (2) is the classical NP-complete knapsack problem [5, p. 68]. A greedy solution of (2) is based on the decreasing order of the ratios

$$\rho(k) := \frac{1 - \mathbb{E}[P_B^{\gamma}(k)]}{C_k}. \tag{3}$$

However, in general, computing $\rho(k)$ for each queue Q_k is itself NP-hard, since $\mathbb{E}[P_B^{\gamma}(k)]$ involves computing the permanent of a matrix [4].

We propose two approaches for efficiently computing the order of $\rho(k)$ in (3). One approach avoids computation of $\mathbb{E}[P_B^{\gamma}(k)]$ altogether and instead infers their order through simple queueing parameters such as their load factors. The other approach computes $\mathbb{E}[P_B^{\gamma}(k)]$ approximately by considering only small busy periods B.

3.1 Approach 1: Load-Factor Heuristic

We propose the load-factor heuristic for selection of queues for instrumentation in (2) under equal costs of instrumentation ($C_k \equiv C$). The heuristic selects queues in the decreasing order of their load factors ($L_k = \frac{\lambda_k}{\mu_k}$) until the budget constraint is met. We can see that the load-factor heuristic is the optimal selection strategy for instrumentation whenever the monitoring accuracies $\mathbb{E}[P_B^{\gamma}(k)]$ at different queues are in the reverse order of their load factors L_k.

We now provide conditions for optimality of load factor heuristic when the monitoring strategy, in the absence of instrumentation, is FIFO matching. Let S_k denote the random service time along a queue Q_k normalized by the arrival rate λ_k to the queue. Note that $\mathbb{E}[S_k] = L_k$, the load factor. Let the "spread" V_k of S_k be given by

$$V_k := S_k(1) - S_k(2), \tag{4}$$

where $S_k(1)$ and $S_k(2)$ are independent samples drawn from the distribution of S_k.

THEOREM 1 (FIFO COMPARISON). *For two queues Q_1 and Q_2 with normalized service times S_1 and S_2, we have*

$$|V_1| \overset{st}{\geq} |V_2|, S_1 \overset{st}{\geq} S_2 \Rightarrow \mathbb{E}[P_B^{FIFO}(1)] \leq \mathbb{E}[P_B^{FIFO}(2)]. \tag{5}$$

Hence, if the normalized service times and their absolute spreads at different queues satisfy the same stochastic order, then the load factor heuristic is optimal under equal instrumentation costs and FIFO matching rule at queues which are not instrumented.

We now provide optimality conditions for the load-factor heuristic when the monitoring strategy, in the absence of instrumentation, is random matching, where we uniformly pick a match over all possible valid matches.

THEOREM 2 (COMPARISON UNDER RANDOM MATCH). *Under random matching rule at queues Q_1, Q_2 with normalized service times S_1 and S_2 and their supports $[a_1, b_1]$ and $[a_2, b_2]$,*

$$S_1 \overset{st}{\geq} S_2, a_1 \leq a_2 \Rightarrow \mathbb{E}[P_B^{RAND}(1)] \leq \mathbb{E}[P_B^{RAND}(2)]. \tag{6}$$

Hence, if all normalized service times satisfy a stochastic order and their supports satisfy the above condition, then the load factor heuristic is optimal under equal instrumentation costs and random matching rule.

3.2 Approach 2: Unit-Batch Approximation

The sufficient conditions to guarantee optimality of load-factor heuristic may not always hold. Moreover, the costs for instrumenting different queues may vary and the arrivals may be non-Poisson. For these cases, we devise an alternative strategy for instrumentation based on approximate evaluation of accuracy. The simplest approximation is the unit-batch approximation where we only consider the event that there is a single transaction in the busy period. At low arrival rate to queue Q_k, this approximation becomes tight

$$\lim_{\lambda_k \to 0} \frac{\mathbb{P}[B_k = 1]}{\mathbb{E}[P_B^{\gamma}(k)]} = 1. \tag{7}$$

Using this approximation, we provide a greedy solution for (2) based on the decreasing order of the ratios

$$\rho'(k) := \frac{1 - \mathbb{P}[B_k = 1]}{C_k}, \quad \forall Q_k \in \mathcal{Q}. \tag{8}$$

4. CONCLUSION

In this paper, we formulated the problem of selectively retrofitting monitoring instrumentation to maximize the effectiveness in tracking transactions throughout the end-to-end business system. In contrast to making ad hoc decisions, our strategy enables an efficient allocation of the instrumentation resources. We proposed two simple heuristics which are shown to be optimal under some conditions. Our heuristics are intuitive and base their instrumentation decisions on simple queueing parameters such as the load factor or the probability of a unit-sized busy period. We formally proved the optimality of the load-factor heuristic for a wide-range of service distributions through stochastic comparison.

5. REFERENCES

[1] Application Response Management (ARM). `http://www.opengroup.org/tech/management/arm/`.

[2] A. Anandkumar, C. Bisdikian, and D. Agrawal. Tracking in a Spaghetti Bowl: Monitoring Transactions Using Footprints. In *Proc. of ACM SIGMETRICS*, Annapolis, Maryland, USA, June 2008.

[3] R. Gallagher. *Discrete Stochastic Processes*. Springer, 1996.

[4] L. Valiant. The complexity of computing the permanent. *Theoretical Computer Science*, 8(2):189–201, 1979.

[5] V. Vazirani. *Approximation Algorithms*. Springer, 2001.