

一种基于增量贝叶斯疑似度的事件驱动故障定位算法

张成 廖建新 朱晓民

(北京邮电大学网络与交换技术国家重点实验室 北京 100876)

(东信北邮信息技术有限公司 北京 100083)

摘要: 现有的故障定位算法大多基于时间窗口,窗口大小设置的合理与否会对算法准确度产生重要影响。为了避免因窗口设置不当造成算法性能的下降,该文以概率加权的二分图作为故障传播模型,提出了一种基于增量贝叶斯疑似度(Incremental Bayesian Suspected Degree, IBSD)的启发式故障定位算法。IBSD 算法采用事件驱动的方式依次分析观察到的征兆,通过增量计算对应故障的贝叶斯疑似度,确定当前征兆前提下最有可能的故障集。仿真实验表明,IBSD 算法具有较高的故障检测率和较低的故障误检率,即使在部分告警无法观察的情况下,算法依然具有较高的故障检测率。算法具有多项式计算复杂度,可以满足大规模通信网故障定位的要求。

关键词: 故障管理; 故障诊断; 故障定位; 故障传播模型

中图分类号: TN915

文献标识码: A

文章编号: 1009-5896(2009)06-1501-04

An Event-Driven Fault Localization Algorithm Based on Incremental Bayesian Suspected Degree

Zhang Cheng Liao Jian-xin Zhu Xiao-min

(State Key Laboratory of Networking and Switching Technology, Beijing University of Posts
and Telecommunications, Beijing 100876, China)

(EB Information Technology Co. Ltd., Beijing 100083, China)

Abstract: Most fault localization techniques is based on time windows. The size of time windows impacts on the accuracy of the algorithms greatly. This paper takes weighted bipartite graph as fault propagation model and proposes a heuristic fault localization algorithm based on Incremental Bayesian Suspected Degree (IBSD) to eliminate the above shortcomings. IBSD sequentially analyzes the incoming symptoms in an event-driven way and incrementally computes the Bayesian Suspected Degree and determine the most probable fault set for the current observed symptoms. Simulation results show that the algorithm has high fault detection ratio as well as low false positive ratio and has a good performance even in the presence of unobserved alarms. The algorithm which has a polynomial computational complexity could be applied to large scale communication network.

Key words: Fault management; Fault diagnosis; Fault localization; Fault propagation model

1 引言

为了尽量减小故障对于通信网的影响,保证通信业务高可用性和高可靠性,需要高效实用的故障定位算法^[1]。当前的故障定位算法^[2-5]大多采用基于时间窗口的征兆分析方法。这种方式征兆处理效率较高。但存在以下问题:(1)故障分析时延较大,无法在收到征兆的时刻起对征兆进行实时分析;(2)时间窗口大小设置的合理与否对定位准确度有重要影响;(3)时间窗口大小难以准确设定,找到不同故障之间的征兆边界几乎是不现实的。为了解决这些问题,本文以概率加

权的二分因果图作为故障传播模型,提出了一种基于增量贝叶斯疑似度的启发式故障定位算法 IBSD(Incremental Bayesian Suspected Degree)。

2 相关工作

Katzela 等采用分而治之的思想,提出了 DCMMDA (Divide and Conquer Maximum Mutual Dependency Algorithm)算法^[6]。Yemini 等提出了基于代码簿的故障定位方法^[7]。Steinder 等引入贝叶斯网作为故障传播模型,通过 3 个近似推理算法^[8,9]: BE(Bucket Elimination)算法,IBU(Iterative Belief Updating)算法,IMPE(Iterative Most Probable Explanation)算法对模型进行求解。黄晓慧等基于贪婪思想提出了 MCA/MCA+(Max Covering Algorithm)算法^[10,11]。以上这些故障定位算法都基于时间窗口。

基于事件驱动的故障定位算法较少。Steinder 等提出了

2008-05-15 收到,2008-09-18 改回

国家杰出青年科学基金(60525110),国家 973 计划项目(2007CB307100,2007CB307103),新世纪优秀人才支持计划(NCET-04-0111)和电子信息产业发展基金项目(基于 3G 的移动业务应用系统)资助课题

基于事件驱动的增量信度更新的 IHU(Incremental Hypothesis Updating)算法^[12]。算法依次分析观测到的征兆逐步更新故障假设。为避免假设集数目膨胀过快,算法限定了可能的故障假设数目,一些故障假设在被更新为最优解之前可能被丢弃。针对 IHU 算法的不足,本文详细分析了故障与征兆的非确定性关系以及后验概率,采用启发式策略,以概率加权的二分图故障传播模型,提出了一种基于增量贝叶斯疑似度的事件驱动故障定位算法 IBSD。

3 故障定位模型与算法

3.1 故障传播模型

基于概率加权的二分图可以更好地表达故障与征兆之间的关系。概率加权二分图 $BG = (V(F \cup S), E(F \times S), P_{F \times S})$, F 为故障集, S 征兆集, $V(F \cup S)$ 为由故障和征兆组成的二分节点集, $E(F \times S)$ 为由故障指向征兆的有向边集合,边权重的集合为 $P_{F \times S}$,表示当故障 f 发生时发出征兆 s 的概率,其中 $f \in F$, $s \in S$ 。征兆域 $\text{domain}(s_i) \subseteq F$,代表与征兆 s_i 关联的所有故障的集合。故障域 $\text{symptom}(f_i) \subseteq S$,代表与故障 f_i 关联的所有征兆的集合。征兆簇 $\text{Cluster}(s_i) = \bigcup_{s_j \in S} D(s_j)$,代表征兆 s_i 所在的最大连通子图。 $S_O \subseteq S$ 为系统中能够观测到的征兆集。 $S_i \subseteq S_O$ 由依次观测到的 i 个征兆组成的征兆集。 $p(f_i | s_j)$ 为出现 s_j 的情况下, f_i 发生的概率。

3.2 故障定位算法

假设: Noisy-OR 模型,即引起某个征兆的多个故障相互独立,任何一个故障发生都将引起该征兆发生;故障独立假设,即不同故障相互独立。

IBSD 算法引入贝叶斯疑似度来度量可能性。设第 i 个征兆 s_i 到达时,累计观察到征兆集 $S(i) \subseteq S_O$,故障 f 发生的贝叶斯疑似度为

$$\text{bsd}_i(f) = \frac{\sum_{m=1}^i p(f | s_m)}{\sum_{s_j \in \text{symptom}(f)} p(f | s_j)} \quad (1)$$

其中 $p(f | s) = \frac{p(f)p(s | f)}{\sum_{f_i \in F} p(f_i)p(s | f_i)}$ 。

令 $\alpha(f) = 1 / \sum_{s_j \in \text{symptom}(f)} p(f | s_j)$,则有

$$\text{bsd}_i(f) = \text{bsd}_{i-1}(f) + \alpha(f)p(f | s_i) \quad (2)$$

对 $\text{domain}(s_i)$ 中每个故障 f ,用公式(2)求解 $\text{bsd}_i(f)$,形成集合 $\bigcup_{f \in \text{symptom}(s_i)} \text{bsd}_i(f)$,将其按大小顺序插入有序集合 $B_{\text{bsd}}(i-1)$,形成 $B_{\text{bsd}}(i)$;对 $F(i-1) = \text{domain}(S(i-1))$ 中的每个故障 f , $\text{bsd}_i(f) = \text{bsd}_{i-1}(f)$ 。

当 $B_{\text{bsd}}(i)$ 中的前 n 个故障,完全覆盖了当前观测到的征兆集时,就认为找到了最有可能的故障假设。即:满足 $S(i) \subseteq \sum_{f_{i=1}^n \in B_{\text{bsd}}(i)} \text{symptom}(f_i)$ 的最小值 n ,最优故障假设集为 $H = \{f_1, f_2, \dots, f_n\}$ 。

输入为:故障传播模型 $FPM = (F, S, P_f, P_{F \times S})$;可观察

的征兆集合 S_O ;征兆可观察率 OR , $OR = S_O / S$;征兆丢失率 $LR(s)$;征兆虚假率 $SSR(s)$ 。

输出为:能够对事件驱动征兆集 $S(i)$, $1 \leq i \leq \text{length}(S_O)$ 做出最优解释的故障假设集 $H(i)$ 。

AlgorithmIBSD 算法:

(1) 设置 $H = \emptyset$, $B_{\text{bsd}}(0) = \emptyset$, $S(0) = \emptyset$, $F(0) = \emptyset$;

(2) 设置 $\alpha(f) = 1 / \sum_{s \in \text{domain}(f)} p(f)p(s | f)$, $\text{bsd}_0(f) = 0$, $f \in F$;

(3) 按征兆发生的先后顺序依次取 $s_i \in S_O$, $S(i) = \text{union}(S(i-1), s_i)$, $i = 1, 2, \dots, \text{length}(S_O)$ 执行以下各步:

(a) 找出可能产生征兆 s_i 的故障 $\text{symptom}(s_i)$, 构成待选故障子集: $F(i) = \text{union}(F(i-1), \text{symptom}(s_i))$;

(b) 对每个 $f \in F(i-1) - \text{symptom}(s_i)$, $\text{bsd}_i(f) = \text{bsd}_{i-1}(f)$;

(c) 对每个 $f \in \text{symptom}(s_i)$, $\text{bsd}_i(f) = \text{bsd}_{i-1}(f) + \alpha(f) \cdot p(f | s_i)$;

(d) 把 $\bigcup_{f \in \text{symptom}(s_i)} \text{bsd}_i(f)$ 按大小插入有序集合数组 $B_{\text{bsd}}(i-1)$ 中,形成 $B_{\text{bsd}}(i)$;

(e) 初始化征兆集 $S_{\text{bsd}}(i) = \emptyset$, $S_{\text{occupy}}(i) = \emptyset$, $H(i) = \emptyset$;

(f) 依次取出 $\text{bsd}_i \in B_{\text{bsd}}(i)$ 循环执行以下各步,直至 $S_i \subseteq S_{\text{bsd}}(i)$;

(f₁) 获得 bsd_i 所对应的 f_i ;

(f₂) 获得 f_i 所对应的 $S_f = \text{domain}(f_i) \cap S_N$;

(f₃) 如果 S_f 对 $S_{\text{bsd}}(i)$ 增长有贡献,即 $\text{union}(S_{\text{bsd}}, S_f) - S_{\text{bsd}}(i) \neq \emptyset$, 则

$$S_{\text{bsd}} = \text{union}(S_{\text{bsd}}, S_f);$$

$$H(i) = \text{union}(H(i), f_i);$$

(g) $H = \text{union}(H, H(i))$;

(4) 输出故障假设集 $H = \{H(1), H(2), \dots, H(\text{length}(S_O))\}$ 。

3.3 算法分析

本文的对比算法 IHU 算法以故障和征兆的先验概率作为备选故障的选择标准,这种方法常常因为先验概率估计得不够准确,使计算结果难以区分甚至造成混淆。本文定义的增量贝叶斯疑似度,采用贝叶斯后验概率的思想,放大了备选故障之间的区别。IBSD 算法以贝叶斯疑似度作为备选故障的选择标准,具有更加理想的定位准确度。为了避免备选集的膨胀, IHU 算法限定了可能的故障假设的数目,一些故障假设在被更新为最优解之前可能被丢弃。IBSD 算法则采用了另外一种启发式策略,选取征兆集对应的最有可能的前 n 个故障作为故障备选集。4.2 节的仿真结果表明,相比之下这种策略更加有效。

IBSD 算法第(2)步需要计算 $|F| \times |S|$ 次;第(3)步主循环执行 $|S_O|$ 次,第(d)步对 $\text{bsd}_i(f)$ 按大小插入 $B_{\text{bsd}}(i-1)$,最多需要 $|\text{bsd}_i(f)| \times |B_{\text{bsd}}(i-1)| \leq |F|^2$ 次循环,第(f)步循环执行最

多 $|B_{\text{bsd}}(i)| \leq |F|$ 次。所以 IBSD 算法的复杂度最大为 $O(|F|^2 \times |S_o|)$, 与 IHU 算法的计算复杂度相同。

4 仿真结果和分析

4.1 仿真实验

本文采用与文献[12]类似的仿真环境, 并做了适当的增强, 增加了故障生成时间和征兆生成时间。仿真环境主要参数如下: P_F 服从 $(0.001, 0.01)$ 均匀分布, $P_{F \times S}$ 服从 $(0, 1)$ 均匀分布; $OR = |S_o|/|S|$; $\text{symptom}(f)$ 中征兆信息按照参数 LR 的概率丢失; $S_o - \text{symptom}(f)$ 中征兆信息按照参数 SSR 的概率生成虚假征兆。 S_i 定义为观测到的第 i 个征兆 s_i 形成的集合, $S_i = \{s_1, s_2, \dots, s_i\}$ 。Fc 为产生这些征兆的实际故障。设定仿真环境中实际故障个数为 1, 2, 3, 4, 6 的比例分别为 30%, 25%, 20%, 15%, 10%。

4.2 结果分析

图 1(a)为故障检测率比较。当网络节点多于 5 个时, 两种算法的故障检测率趋于稳定。在 $OR=100\%$ 时, IBSD 算法的故障检测率在 89.08%~100%之间, 平均为 98.56%; IHU 算法的故障检测率在 83.91%~99.42%之间, 平均为 96.33%; 在 $OR=50\%$ 时, IBSD 算法的故障检测率在 89.58%~100%之间, 平均为 97.55%; IHU 算法的故障检测率在 84.48%~96.85%之间, 平均为 92.10%。在 $OR=10\%$ 时, IBSD 算法的故障检测率在 68.93%~94.38%之间, 平均为 84.76%; IHU 算法的故障检测率在 51.93%~90.11%之间, 平均为 70.02%。可以看出随着征兆观察率越来越小, 两种算法的故障检测率都有所下降。IBSD 算法的故障检测率始终优于 IHU 算法, 特别是当征兆观察率较低时尤为明显。图 1(b)可以看出, 随着征兆观察率逐步变小, 两种算法故障检测率的稳定性都有很大下降。IBSD 算法的故障检测率比 IHU 算法更稳定。图 1(c)表明, IBSD 算法与 IHU 算法的故障误检率基本相当。图 1(d)可以看出, 随着征兆观察率 OR 的降低, 由于征兆分析的数目减少, 使故障定位时间也有所减少。OR 越低, 故障定位所需的时间就越少。两种算法的单位故障定位时间, 随着网络规模的增长近似于线性方式的增长。尽管 IBSD 算法与 IHU 算法具有相同的计算复杂度, 但是 IBSD 算法的故障定位时间要大于 IHU 算法, 当征兆观察率较低时两者相当。这主要因为, IHU 算法在每次循环中直接计算了备选集的组成及其信度, 而 IBSD 算法则通过两步: (1)增量计算征兆相关的故障的贝叶斯疑似度; (2)对贝叶斯疑似度进行排序, 取出覆盖当前征兆集的最少的前 n 个故障。

5 结束语

本文提出了一种基于增量贝叶斯疑似度的启发式的事件驱动故障定位算法—IBSD 算法。算法摆脱了基于时间窗口的故障定位算法对窗口大小准确设置的依赖, 计算复杂度与 IHU 算法相同, 均为 $O(|F|^2 \times |S_o|)$ 。相比于 IHU 算法, IBSD 算法尽管需要一些额外的故障定位时间, 但其具有更高的故

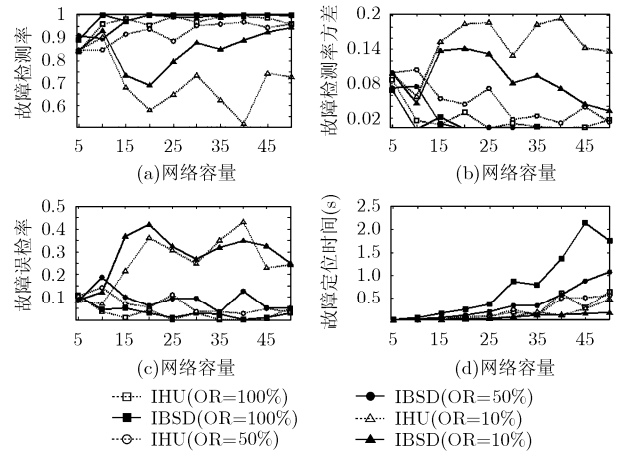


图1 故障定位效果和效率的比较
(OR=100%, 50%, 10%, LR=0%, SSR=0%)

障检测率和基本相当的故障误检率。IBSD 可以满足大规模业务网络故障定位的需要。IBSD 算法没有考虑如何利用 OR, LR, SSR 等信息, 利用这些信息可以进一步提高噪声环境下故障定位的准确率, 特别是有效降低误检率。这将作为进一步的研究工作。

参考文献

- [1] Steinder M and Sethi A S. The present and future of event correlation: A need for end-to-end service fault localization [C]. Proc. World Multi-Conf. Systemic, Cybernetics, and Informatics (SCI), Orlando, FL, 2001: 124-129.
- [2] Steinder M and Sethi A S. A survey of fault localization techniques in computer networks [J]. *Science of Computer Programming*, 2004, 53(2): 165-194.
- [3] Mas C and Thiran P. A review on fault location methods and their application to optical networks [J]. *Optical Networks Magazine*, 2001, 2(4): 73-87.
- [4] Mas C and Thiran P. An efficient algorithm for locating soft and hard failures in WDM networks [J]. *IEEE Journal on Selected Areas in Communications*, 2000, 18(10): 1900-1911.
- [5] Zhao Y, Chen Y, and Bindel D. Towards unbiased end-to-end network diagnosis [C]. Proceedings of the ACM SIGCOMM 2006 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, Pisa, Italy, 2006: 219-230.
- [6] Katzela I and Schwartz M. Schemes for fault identification in communication networks [J]. *IEEE/ACM Trans. on Networking*, 1995, 3(6): 733-764.
- [7] Yemini S and Kliger S. A coding approach to event correlation, integrated network management [C]. Proceedings of the Fourth International Symposium on Integrated Network Management, Santa Barbara, California, USA, 1995: 266-277.
- [8] Steinder M and Sethi A S. End-to-end service failure

- diagnosis using belief networks [C]. Proc. Network Operations and Management Symposium (NOMS), Florence, Italy, 2002: 375-390.
- [9] Steinder M and Sethi A S. Probabilistic fault localization in communication systems using belief networks [J]. *IEEE/ACM Trans. on Networking*, 2004, 12(5): 809-822.
- [10] Huang Xiao-hui, Zou Shi-hong, Wang Wen-dong, and Cheng Shi-duan. Fault management for Internet service: Modeling and algorithms [C]. IEEE International Conference on Communications, Istanbul, Turkey, June 2006, 2: 854-859.
- [11] 黄晓慧, 邹仕洪, 王文东, 程时端. Internet服务故障管理: 分层模型和算法 [J]. *软件学报*, 2007, 18(10): 2584-2594.
Huang Xiao-hui, Zou Shi-hong, Wang Wen-dong, and Cheng Shi-duan. Internet services fault management: Layering model and algorithm. *Journal of Software*, 2007, 18(10): 2584-2594.
- [12] Steinder M and Sethi A S. Probabilistic event-driven fault diagnosis through incremental hypothesis updating [C]. IFIP/IEEE Eighth International Symposium on Integrated Network Management, Colorado, USA, 2003: 635-648.
- 张成: 男, 1976年生, 博士生, 研究方向为业务网络智能化、业务管理等.
- 廖建新: 男, 1965年生, 教授, 博士生导师, 研究方向为移动智能网、下一代网络、业务网络智能化等.
- 朱晓民: 男, 1974年生, 副研究员, 研究方向为移动智能网、业务网络智能化等.