

压缩感知的多参数链路故障定位算法

王汝言^① 吴晴^{*①} 熊余^{①②} 谢雨^① 赵莹^①

^①(重庆邮电大学重庆市光纤通信技术重点实验室 重庆 400065)

^②(重庆大学计算机学院 重庆 400030)

摘要: 为了提高故障定位性能,降低单一判别参数在单位过程中的约束,该文提出一种基于压缩感知和信息熵差的多参数链路故障定位算法。该算法首先利用贝叶斯网络进行快速故障预测,其次引入参数故障覆盖范围,利用压缩感知进行故障筛选,最后定义参数故障信息熵差完成根源故障定位。仿真结果表明,该算法预测出的故障集合具有可压缩性,筛选后的故障集合保留了真实故障,定位时具有较高的故障检测率和较低的故障误检率。

关键词: 压缩感知; 故障定位; 贝叶斯网络; 信息熵

中图分类号: TN915

文献标识码: A

文章编号: 1009-5896(2013)11-2596-06

DOI: 10.3724/SP.J.1146.2013.00265

Multi-parameters Link Failure Localization Algorithm Based on Compressive Sensing

Wang Ru-yan^① Wu Qing^① Xiong Yu^{①②} Xie Yu^① Zhao Ying^①

^①(Chongqing Key Laboratory of Optical Fiber Communication, Chongqing University of Posts and Telecommunication, Chongqing 400065, China)

^②(College of Computer Science, Chongqing University, Chongqing 400030, China)

Abstract: To improve the performance and decrease the constraints of fault localization with single distinguish parameter, a multi-parameters link failure localization algorithm is proposed based on compressive sensing and entropy difference. Firstly it makes a fast fault prediction by Bayesian network, then it introduces a parameter named fault coverage and selects probable link failure using compressive sensing, finally defines fault information entropy difference and obtains the root fault based on the parameter. The simulation results show that the predicted fault set can be compressed and the selected probable fault set contains the true fault, meanwhile the proposed algorithm achieves high detection rate and low false positive rate.

Key words: Compressive Sensing (CS); Fault localization; Bayesian network; Information entropy

1 引言

随着网络中承载的数据量的增加,提高网络生存性,减少网络故障造成的损失,具有很大的意义。为了实现网络故障丢失业务的及时恢复,需要快而准的故障监测与定位机制。因此,作为生存性研究中必不可少的一部分,高性能的定位算法一直是国内外研究的热点^[1,2]。

网络中监测模块获取的告警可以视为故障表现出的外部征兆,根据收集到的征兆集合,可以推测出最有可能发生的链路故障集合,从而定位出根源故障。由于一个征兆可能对应多处故障,所以,为

了提高定位性能,必须引入合理的故障传播模型来表达故障和征兆之间的这种不确定因果关系。

基于监测圈、监测迹和监测树等传统的故障定位方法^[3-7]能达到很好的定位性能。但是,需要使用额外的探测波长实现故障监测,定位成本较高。因此,为了降低用于故障定位的网络资源耗费,利用网络中的业务获取征兆的故障定位算法获得了广泛地研究。此类算法中,基于贝叶斯网络的故障定位算法即使在征兆不完全的情况下,仍能实现网络中准确的故障定位,具有很大的优势。文献[8,9]采用基于贝叶斯网络的故障传播模型,利用近似推理算法实现故障定位,定位性能较好,但计算复杂度较高。文献[10-13]采用简化的贝叶斯网络,即二分图作为故障传播模型,定义一个判别参数作为故障判别标准,降低了计算复杂度,取得了较好的定位性能。由于参数具有一定的限制,因此,使用单一

2013-03-04 收到, 2013-05-27 改回

国家自然科学基金(60972069, 61001105), 重庆市自然科学基金重点项目(2011BA2041), 重庆市教委科学技术研究项目(KJ110531)和重庆市高校优秀人才支持计划(2011-29)资助课题

*通信作者: 吴晴 chongqingwuqing@163.com

的判别参数很难实现更精确的故障定位。

为了提高二分图故障传播模型中定位算法的性能，本文提出了一种基于压缩感知和信息熵差的多参数故障定位算法 (Multi-parameters fault localization based on Compressive Sensing and Entropy Difference, MCSED)。该算法将定位过程分为故障预测、筛选和定位 3 个模块。首先，预测模块快速找出征兆对应的最大可能故障集合；然后，筛选模块引入参数故障覆盖范围将最大可能故障集合转化为信号，并利用压缩感知 (Compressive Sensing, CS) 方法消除此信号的冗余信息，得到冗余度较低的可能故障集合；最后，定义参数信息熵差 (Entropy Difference, ED)，判别出根源故障。利用多个参数分别完成故障的筛选和定位，改善了定位性能。

2 故障定位框架

MCSED 算法将故障定位过程分为故障预测、故障筛选和故障定位 3 个模块。故障定位框架图如图 1 所示。

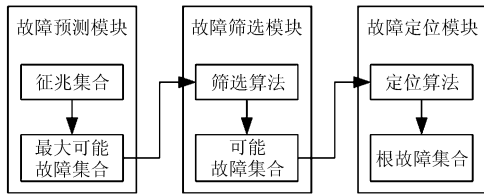


图 1 故障定位框架图

故障预测模块功能是根据观察到的征兆集合，快速找出最大可能故障集合。故障预测的步骤如下：第 1 步，选取表达网络中故障和征兆之间的关系故障传播模型；第 2 步，收集所有网络中的征兆，得到征兆集合；第 3 步，找出所有可能产生征兆集合中征兆的故障，得到最大可能故障集合。

故障筛选模块的功能是根据筛选算法对最大可能故障集合进行筛选处理，得到筛选后的可能故障集合。其中，筛选算法的设计是关键。

故障定位模块的功能是根据定位算法对筛选后的可能故障集合进行处理，得到相应的根故障集合，实现故障定位。

3 故障预测模块

为了实现快速地故障预测和精确地故障定位，选择概率加权的二分图 (Probabilistic Weighted Bipartite Graph, PWBG) 作为故障传播模型^[13]，如图 2 所示。根据 PWBG，可以快速准确地找出与征兆集合中征兆相关的所有可能故障，得到最大可能

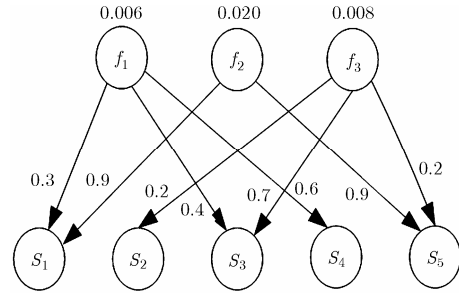


图 2 概率加权的二分图(PWBG)

故障集合。最大可能故障集合是指与征兆集合相关的所有可能故障组成的集合。

定义参数冗余度 R (Redundancy rate) 来表示冗余故障数量在最大故障集合中所占的比率，其定义如式(1)所示。

$$R(H_{Max}) = \frac{|H_R|}{|H_{Max}|} = 1 - \frac{|F_C|}{|H_{Max}|} \quad (1)$$

其中 H_{Max} 表示最大可能故障集合， H_R 表示 H_{Max} 中冗余故障组成的集合， F_C 表示网络中真实故障集合。

最大可能故障集合包含故障的数目总是大于网络中实际发生的故障数目，即其中大部分故障是实际未发生的。因此，必须对集合中的故障进行筛选，过滤发生可能性较小的故障，得到具有较少元素数目的可能故障集合，最大可能地消除冗余故障对定位算法的影响，实现更精确的故障定位。

4 故障筛选模块

故障筛选模块的目的是在保留真实故障的前提下，尽可能多地将最大可能故障集合中的冗余故障剔除，得到冗余度较小的可能故障集合。

将所有可能故障视为信号，定义参数信号强度 (Signal Intensity, SI)，使得发生可能性越大的故障，其对应信号的信号强度越大。因此，故障筛选问题就转化为信号处理问题，即保留信号强度较大的信号。

文献[10]中定义参数覆盖范围表达故障发生的可能性，其定义过程如式(2)所示。

$$C(f) = |\{s | s \in d(f) \cap S_N\}| \quad (2)$$

其中 $d(f)$ 表示所有故障 f 可以解释的征兆的集合， S_N 表示所有观察到的征兆的集合， $C(f)$ 表示故障 f 的覆盖范围。

故障的覆盖范围表示故障可以解释的观察到的征兆数目，能定性地说明故障发生的概率大小。故障对应的覆盖范围值越大，其发生的概率也就越大。因此，采用故障的覆盖范围作为其对应信号的信号强度，可将最大可能故障集合转化为一组具有不同

信号强度的信号。由于冗余故障对应的信号是冗余信息，因此，引入压缩感知方法，处理后的信号包含较少的冗余信息，其对应的故障集合具有较少的冗余故障，实现故障筛选。

4.1 压缩感知理论

压缩感知(Compressive Sensing, CS)的主要思想是，将具有稀疏特性的原始信号投射到指定的感知矩阵上，得到观测向量，对此观测信号进行重构，得到包含原始信号中重要信息的重构信号。由于观测向量包含的原始信号的重要信息，且其包含较少的冗余信息，所以，可以对观测向量进行重构，得到既包含原始信号重要信息又具有较小冗余信息的重构信号。

压缩感知的研究主要有信号的稀疏表示即可压缩性验证、感知矩阵的设计和信号的重构 3 方面^[14,15]。

(1)压缩感知的先验条件 信号的稀疏表示是压缩感知的先验条件。如式(3)所示， \mathbf{x} 表示原始信号($\mathbf{x} = (x_1, x_2, \dots, x_n)$)， Ψ 为 \mathbf{x} 的稀疏基， θ 为稀疏系数。

$$\mathbf{x} = \Psi\theta \tag{3}$$

(2)压缩感知的观测模型 压缩感知是将原始信号 \mathbf{x} 投影到一组与变换基不相关的感知矩阵 Φ 上，从而得到观测向量 \mathbf{y} ，观测向量可根据式(4)求取：

$$\mathbf{y} = \Phi\mathbf{x} \tag{4}$$

根据式(3)和式(4)可知，观测向量和稀疏系数的关系可以表示为

$$\mathbf{y} = \Phi\mathbf{x} = \Phi\Psi\theta \tag{5}$$

(3)压缩感知的信号重构 压缩感知的主要目的是通过稀疏度小于原始信号 \mathbf{x} 的观测向量 \mathbf{y} 重构出与原始信号相同或相近的恢复值 \mathbf{x}' 。如式(6)所示，满足 $\mathbf{y} = \Phi\mathbf{x}$ 的所有情况中找出最具有稀疏特性的信号 \mathbf{x}' 即为所求。其中 $\|\mathbf{x}\|_0$ 是 \mathbf{x} 的零范数，表示 \mathbf{x} 中非零元素的个数。

$$\min \|\mathbf{x}\|_0, \text{ s.t. } \mathbf{y} = \Phi\mathbf{x} \tag{6}$$

4.2 压缩感知的故障筛选算法

感知预测模块得到的最大故障集合具有很大的冗余度，为了降低冗余度，减少冗余故障对故障判定时的影响，必须尽可能多地剔除最大故障集合中的冗余故障，筛选出发生可能性较大的故障。将故障视为信号，引入覆盖范围作为故障的信号强度，则将故障筛选问题转变成信号处理问题。压缩感知方法可以作为一种信号处理方法，达到保留信号重要信息并去掉信号部分或全部冗余信息的目的，实

现故障集合的筛选。

压缩感知的工作流程如图 3 所示。定义真实故障对应信号的信号强度为重要信息，冗余故障对应信号的信号强度为冗余信息。

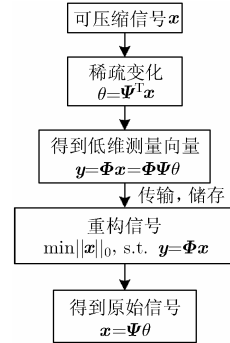


图3 压缩感知工作流程

(1)压缩感知的第 1 步是验证信号的可压缩性，即证明所处理的信号包含冗余信息。最大故障集合中包含冗余故障，具有很高的冗余度，其对应信号包含较多的冗余信息，具有可压缩性。因此，利用压缩感知方法处理故障集合对应信号 \mathbf{x} 具有合理性。

(2)压缩感知的第 2 步是根据观测模型得到信号的观测向量值。观测模型主要是利用感知矩阵 Φ 对信号进行投影，得到信号的观测向量值。设计合适的感知矩阵，使得到的观测向量值既包含原始信号的重要信息，又尽可能多的消除冗余信息，是提高感知筛选方法故障筛选性能的关键。

设定信号强度阈值 α_{SI} ，设计感知矩阵，保留原始信号 $\mathbf{x} = (x_1, x_2, \dots, x_n)$ 中信号强度高于阈值的故障信号。其中， α_{SI} 由式(7)得出，其中， $\mu(0 \leq \mu \leq 1)$ 为尺度因子，灵活控制 α_{SI} 的大小。

$$\alpha_{SI} = \mu \cdot \text{Max}\{x_1, x_2, \dots, x_n\} \tag{7}$$

由于对角矩阵与目标矩阵相乘，可实现目标矩阵内元素大小的伸缩。因此，可引入对角矩阵 \mathbf{A} ($\mathbf{A} = \text{diag}(a_1, a_2, \dots, a_n)$) 作为原始信号 $\mathbf{x} = (x_1, x_2, \dots, x_n)$ 的感知矩阵 Φ 。对角矩阵元素值可由式(8)得出。根据设计的感知矩阵 $\Phi = \mathbf{A}$ ，由式(4)可以得出原始信号 \mathbf{x} 的观测向量 \mathbf{y} 。

$$a_i = \begin{cases} 1, & x_i \geq \alpha_{SI} \\ 0, & x_i < \alpha_{SI} \end{cases} \tag{8}$$

(3)压缩感知的第 3 步是根据观测向量值 \mathbf{y} 重构出原始信号值。如式(6)所示，满足 $\mathbf{y} = \Phi\mathbf{x}$ 的所有情况中找出最具有稀疏特性的信号 \mathbf{x}' 即为所求，即满足要求的集合中， \mathbf{x}' 中非零元素个数最少。

将重构信号 \mathbf{x}' 中的非零元素对应的故障组成的集合称为筛选后的故障集合 H_s 。使用筛选后的故障

集合 H_S 的真实故障覆盖率 $\eta(H_S)$ 和冗余度 $R(H_S)$ 来观察筛选算法性能。 $\eta(H_S)$ 和 $R(H_S)$ 的计算如式(9)和式(10)所示, 其中, H_{R_S} 为 H_S 中冗余故障组成的集合, F_{C_S} 为 H_S 中真实故障组成的集合。

$$\eta(H_S) = \frac{|H_S| - |H_{R_S}|}{|F_C|} = \frac{|F_{C_S}|}{|F_C|} \quad (9)$$

$$R(H_S) = \frac{|H_{R_S}|}{|H_S|} = 1 - \frac{|F_{C_S}|}{|H_S|} \quad (10)$$

$\eta(H_S)$ 越大, 表示 H_S 保留原始故障集合 H_{Max} 的重要信息越多。 $\eta(H_S)=1$ 时表示 H_S 保留了 H_{Max} 中所有重要信息。相比 $R(H_{\text{Max}})$, $R(H_S)$ 越小, 表示 H_S 中冗余故障越少, 即筛选算法性能越好, $R(H_S)=0$ 时表示 H_S 中没有冗余故障。

5 故障定位模块——基于故障信息熵差的定位算法

由于可能存在多个故障同时具有最大覆盖范围的情况, 导致筛选后的可能故障集合 H_S 包含多个可能故障。此时, 使用覆盖范围作为判别参数进行故障定位极大可能会产生故障误判的情况。因此, 必须引入新的合理的参数来进行故障定位。本模块定义了故障 f 的信息熵差 $\Delta H(f)$, 并以此作为判别参数, 提出了一种基于信息熵差的故障定位算法。以下理论分析表明, 采用信息熵差 $\Delta H(f)$ 作为判别参数是合理的。

网络中征兆发生时, 将会对与其相关的故障提供一定的信息量。信息熵是指与某个故障相关的多个征兆对该故障提供的平均信息量。信息量越大, 表明变量的不确定性越大, 在故障判定时将认为故障发生的可能性越小。

为了分析故障信息熵差作为故障判定参数的合理性, 定义以下两个参数: 理想信息熵和实际信息熵。

故障 f 的理想信息熵是指在理想的情况下, 即与故障 f 相关的所有征兆都发生的情况下, 故障 f 的信息熵。理想信息熵 $H_1(f)$ 的计算如式(11)所示。

$$H_1(f) = - \sum_{s_i \in S(f) \cap S_O} p(f | s_i) \log p(f | s_i) \quad (11)$$

其中 $S(f)$ 是 f 对应的征兆集合, S_O 是网络中全部能够在显示出的征兆集合, $p(f | s_i)$ 可由贝叶斯公式即式(12)得出。

$$p(f | s_i) = \frac{p(f)p(s_i | f)}{\sum_{f_j \in F} p(f_j)p(s_i | f_j)} \quad (12)$$

故障 f 的实际信息熵是指在实际的情况下, 收集到征兆集合 S_N 对故障 f 提供的信息熵。实际信息熵 $H_2(f)$ 的计算如式(13)所示。

$$H_2(f) = - \sum_{s_i \in S(f) \cap S_N} p(f | s_i) \log p(f | s_i) \quad (13)$$

故障 f 的信息熵差 $\Delta H(f) = H_1(f) - H_2(f)$ 越小, 表明故障 f 实际引发的征兆与理想情况下越接近, 可以认为故障 f 发生的可能性越大。因此, 故障 f 的信息熵差 $\Delta H(f)$ 可以作为故障定位判定时的参数。

基于故障信息熵差的定位算法选择具有最小信息熵差的故障 f 加入故障假设集合 H , 并删除故障 f 可以解释的征兆, 重复以上步骤直至所有征兆都被解释为止, 最后故障假设集合 H 即为所得定位结果。算法具体步骤如下:

(1) 初始化故障假设集合为空, $H = \emptyset$ 。

(2) 获取征兆集合 S_N 和筛选后的可能故障集合 H_S 。

(3) 对 H_S 中每一个故障 f , 求取信息熵差 $\Delta H(f)$, 并将 H_S 中故障依照信息熵差 $\Delta H(f)$ 从小到大排列, 得到更新后的 H_S 。

(4) 初始化可被解释的征兆集合为空, $S_{\text{ep}} = \emptyset$ 。

(5) 依次取出 H_S 中的故障 f_i 执行, 直至 $|S_{\text{ep}} \cap S_N| / |S_N| = 1$ 或 H_S 中所有故障都被取完: 如果 $S(f_i) \cup S_{\text{ep}} - S_{\text{ep}} \neq \emptyset$, 则 $H = H \cup \{f_i\}$, $S_{\text{ep}} = S(f_i) \cup S_{\text{ep}}$; 如果 $S(f_i) \cup S_{\text{ep}} - S_{\text{ep}} = \emptyset$, 则 H 和 S_{ep} 不变。

(6) 输出故障假设集合 H 。

6 仿真及结果分析

6.1 仿真实验

本文仿真在文献[13]所提模型的基础上, 针对所提算法, 增加了故障预测模块和故障筛选模块。其中, 故障预测模块快速预测出最大可能故障集合 H_{Max} , 并计算其冗余度。故障筛选模块输出筛选后的可能故障集合 H_S 和筛选算法性能。当尺度因子 $\mu = 0$ 时, 认为 H_{Max} 实际未作筛选处理; 当 $0 < \mu \leq 1$ 时, H_{Max} 被作相应的筛选处理, 得到筛选后的可能故障集合 H_S 。本文算法为 MCSED, 对比算法为 MCA^[10]和 BSD^[13]。

为了观察算法在不同随机网络中的性能, 生成 10 个随机网络, 每个网络产生 50 个有效单故障案例。每个案例算法输入都为 S_i , 输出为故障假设集合 H , 故障检测率(Detection Rate) $DR(S_i)$, $DR(S_i) = |H \cap F_C| / |F_C|$, 故障误检率(False Positive Rate) $FPR(S_i)$, $FPR(S_i) = |H - F_C| / |H|$ 。另外, 为了观察筛选模块性能, 本文算法输出最大可能故障集合冗余度 $R_{S_i}(H_{\text{Max}})$ 、不同尺度因子下筛选后可能故障集合冗余度 $R_{S_i}(H_S)$ 和故障覆盖率 $\eta_{S_i}(H_S)$ 。

设每个网络中案例数为 n , 输出故障检测率

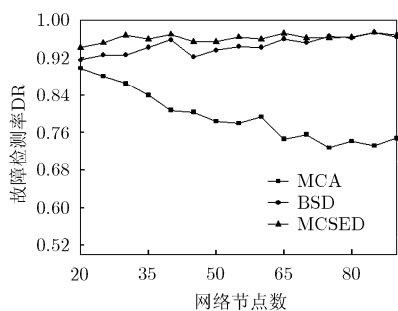


图 5 故障检测率

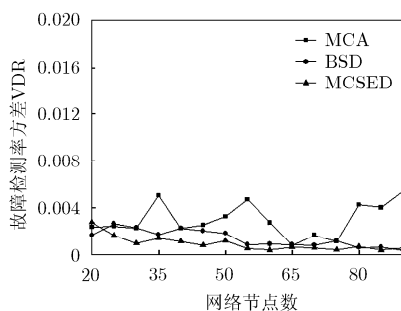


图 6 故障检测率方差

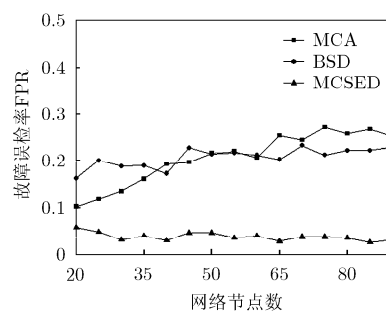


图 7 故障误检率

并利用压缩感知消除信号的冗余信息，得到冗余度较低的故障集合；最后定义参数信息熵差作为故障判别标准，定位出根源故障集合。仿真表明，该算法在不同随机网络中能稳定地表现出较高故障检测率，且很大程度地降低了故障误检率。多参数的定位算法中，为了得到更精确的定位性能，寻找更优的参数组合将是进一步的研究工作。

参考文献

- [1] Tapolcai J, Ho Pinhan, Ronyai L, *et al.* Failure localization for shared risk link groups in all-optical mesh networks using monitoring trails[J]. *Journal of Lightwave Technology*, 2011, 29(10): 1597-1606.
- [2] Xiong Yu, Xiong Zhong-yang, Wu Da-peng, *et al.* Multi-fault aware parallel localization protocol for backbone network with many constraints[J]. *Photonic Network Communications*, 2012, 24(3): 210-218.
- [3] 王汝言, 常交法, 隆克平, 等. 一种光突发交换网络的故障监测与定位机制[J]. *光电子·激光*, 2006, 17(12): 1477-1481.
Wang Ru-yan, Chang Jiao-fa, Long Ke-ping, *et al.* A fault detection and location mechanism for optical burst switching networks[J]. *Journal of Optoelectronics Laser*, 2006, 17(12): 1477-1481.
- [4] He Wei, Wu Bin, Ho Pinhan, *et al.* Monitoring trail allocation for SRLG failure localization[C]. IEEE Global Telecommunications Conference (GLOBECOM), Houston, TX, USA, 2011: 1-5.
- [5] Doumith Ea, Zahr S Al, and Gagnaire M. Monitoring-tree: an innovative technique for failure localization in WDM translucent networks[C]. IEEE Global Telecommunications Conference (GLOBECOM), Miami, USA: IEEE, 2010: 1-6.
- [6] Hosszu Eva, Tapolcai J, Ronyai L, *et al.* Fast failure localization in all-optical networks with length-constrained monitoring trails[C]. IEEE 4th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops(ICUMT), St. Petersburg, Russia, 2012: 677-683.
- [7] Tapolcai J, Ronyai L, and Ho Pinhan. Link fault localization using bi-directional m-trails in all-optical mesh networks[J]. *IEEE Transactions on Communications*, 2013, 61(1): 291-300.
- [8] Steinder M and Sethi As. Probabilistic fault localization in communication systems using belief networks[J]. *IEEE/ACM Transactions on Networking*, 2004, 12(5): 809-822.
- [9] Steinder M and Sethi As. Probabilistic fault diagnosis in communication systems through incremental hypothesis updating[J]. *Computer Networks*, 2004, 45(4): 537-562.
- [10] Huang Xiao-hui, Zou Shi-hong, Wang Wen-dong, *et al.* Fault management for Internet service: modeling and algorithms [C]. IEEE International Conference on Communications (ICC), Istanbul, Turkey, 2006: 854-859.
- [11] 黄晓慧, 邹仕洪, 褚灵伟, 等. 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, Chu Ling-wei, *et al.* Internet services fault management: layering model and algorithm[J]. *Journal of Software*, 2007, 18(10): 2584-2594.
- [12] 张成, 廖建新, 朱晓民. 一种基于增量贝叶斯疑似度的事件驱动故障定位算法[J]. *电子与信息学报*, 2009, 31(6): 1501-1504.
Zhang Cheng, Liao Jian-xin, and Zhu Xiao-min. An event-driven fault localization algorithm based on incremental Bayesian suspected degree[J]. *Journal of Electronics & Information Technology*, 2009, 31(6): 1501-1504.
- [13] 张成, 廖建新, 朱晓民. 基于贝叶斯疑似度的启发式故障定位算法[J]. *软件学报*, 2010, 21(10): 2610-2621.
Zhang Cheng, Liao Jian-xin, and Zhu Xiao-min. Heuristic fault localization algorithm based on Bayesian suspected degree[J]. *Journal of Software*, 2010, 21(10): 2610-2621.
- [14] 戴琼海, 付长军, 季向阳. 压缩感知研究[J]. *计算机学报*, 2011, 34(3): 425-434.
Dai Qiong-hai, Fu Chang-jun, and Ji Xiang-yang. Research on compressed sensing[J]. *Chinese Journal of Computers*, 2011, 34(3): 425-434.
- [15] Donoho D L. Compressed sensing[J]. *IEEE Transactions on Information Theory*, 2006, 52(4): 1289-1306.

王汝言：男，1969年生，教授，研究方向为全光网络理论与技术、下一代光网络故障管理机制、多媒体信息处理等。

吴晴：男，1989年生，硕士生，研究方向为光通信与网络。

熊余：男，1982年生，副研究员，研究方向为宽带网络可靠性理论及抗毁技术。