计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2022,31(6):93−99 [doi: 10.15888/j.cnki.csa.008533]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

面向企业客户的大型云网监控系统灵敏度优化①

吴　舸, 肖　荣, 张明华, 金道临, 孙毛杰

(上海理想信息产业 (集团) 有限公司, 上海 201315)
通信作者: 肖　荣, E-mail: xiaorong.sh@chinatelecom.cn

摘　要: 面向企业客户的大型云网监控系统在多云/多网环境下为用户提供包括云网监控、应用性能监控等监控服务, 为企业提供一体化的监控运维管理, 对于监控的灵敏度有着很高的要求. 影响云网监控系统的灵敏度的因素存在于数据采集、数据处理、数据存储、数据展现、数据缓存、消息队列等多个环节中, 本文着重对云网监控系统逐层架构进行分析, 通过数据分析、应对方案设计提升系统整体监控灵敏度.

关键词: 云网监控; 监控灵敏度; 故障发现; 数据采集

引用格式: 吴舸,肖荣,张明华,金道临,孙毛杰.面向企业客户的大型云网监控系统灵敏度优化.计算机系统应用,2022,31(6):93–99. http://www.c-s-a.org.cn/1003-3254/8533.html

Sensitivity Optimization of Large Cloud and Network Monitoring System for Enterprise Customers

WU Ge, XIAO Rong, ZHANG Ming-Hua, JIN Dao-Lin, SUN Mao-Jie

(Shanghai Ideal Information Industry (Group) Co. Ltd., Shanghai 201315, China)

Abstract: The large cloud and network monitoring system for enterprise customers provides users with services such as cloud and network monitoring as well as application performance monitoring in a multi-cloud/multi-network environment and provides enterprises with integrated monitoring and maintenance management. It has high requirements for monitoring sensitivity. The factors affecting the system exist in many links such as the data collection, data processing, data storage, data presentation, data cache, and message queue. This study analyses the architecture of the cloud and network monitoring system layer by layer and improves the overall monitoring sensitivity of the system through data analysis and corresponding solution design.

Key words: cloud and network monitoring system; monitoring sensitivity; fault detection; data collection

面向企业客户的大型云网监控系统在多云/多网环境下为用户提供包括云网监控、应用性能监控等监控服务, 为企业提供一体化的监控运维管理, 系统的灵敏度对于提升客户感知、云网资源运维效率、客户黏度有着重要的影响, 监控系统灵敏度依赖于数据采集、数据处理、数据存储、数据展现等多个环节, 云网监控系统通过融合开源成熟解决方案与自研关键技术相结合的方式, 提升系统的整体监控灵敏度.

云网监控系统主要目标是针对客户的网络、设备、应用等进行监控, 为客户提供云网及应用的整体运行状况, 及时发现异常并告警, 确保客户的基础云网资源及在此基础上构建的应用及业务的稳定运行. 云网监控系统的核心是发现异常、定位异常、解决异常, 这些操作均需要在一定的时间内完成, 以符合云网监控系统对企业承诺的 SLA 服务标准, 为了将异常发生时间与运维人员知晓异常的时间之间的间隔 T (简称异常可视化时间) 控制在 SLA 服务标准内, 系统需要尽可能缩短 T, T 越小则云网监控系统灵敏度越高, 使得系统能够在异

① 收稿时间: 2021-08-17; 修改时间: 2021-09-29, 2021-10-19; 采用时间: 2021-10-24; csa 在线出版时间: 2022-05-26

常发生后越快发现异常, 并引导运维人员人工介入处理异常, 提升企业的云网整体运维效率. 原有的云网监控系统的采集频率为 60 s、300 s、600 s 等; 监控页面的刷新频率为 30 s, 异常可视化时间 (*T*) 在 90–630 s 的范围内. 通过对系统灵敏度的优化, 模拟测试结果表明异常可视化时间 (*T*) 最短可以达到 15 s.

1 影响云网监控系统灵敏度的因素分析

1.1 云网监控系统系统逻辑架构

根据云网监控系统的逻辑架构[1], 如图 1 所示, 异常在被监控的资源层发生后, 异常数据会在数据采集层、数据处理层、消息队列、数据缓存、数据存储层、业务及展示层中传输: 资源层异常通过主动上报或主动采集汇聚到数据采集层; 数据采集层对异常进行初步处理后提交到消息队列; 数据处理层从消息队列中获取将异常转换为告警后持久化到数据存储层; 业务及展示层从数据存储层中读取告警信息展示并提醒运营维护人员进行处理. 提升云网监控系统的灵敏度, 就需要缩短异常在各层级中流转、处理的时间.

1.2 云网监控系统灵敏度影响因素

针对系统各层级的功能, 表 1 分析了各层级中影响异常流转、处理时间的主要因素. 以下为影响云网监控系统灵敏度的应用层方面的主要因素, 消息队列、数据存储、数据缓存方面的影响因素, 可以通过成熟开源软件的性能调优方式解决, 其他因素需要通过云网监控系统本身的优化进行解决.

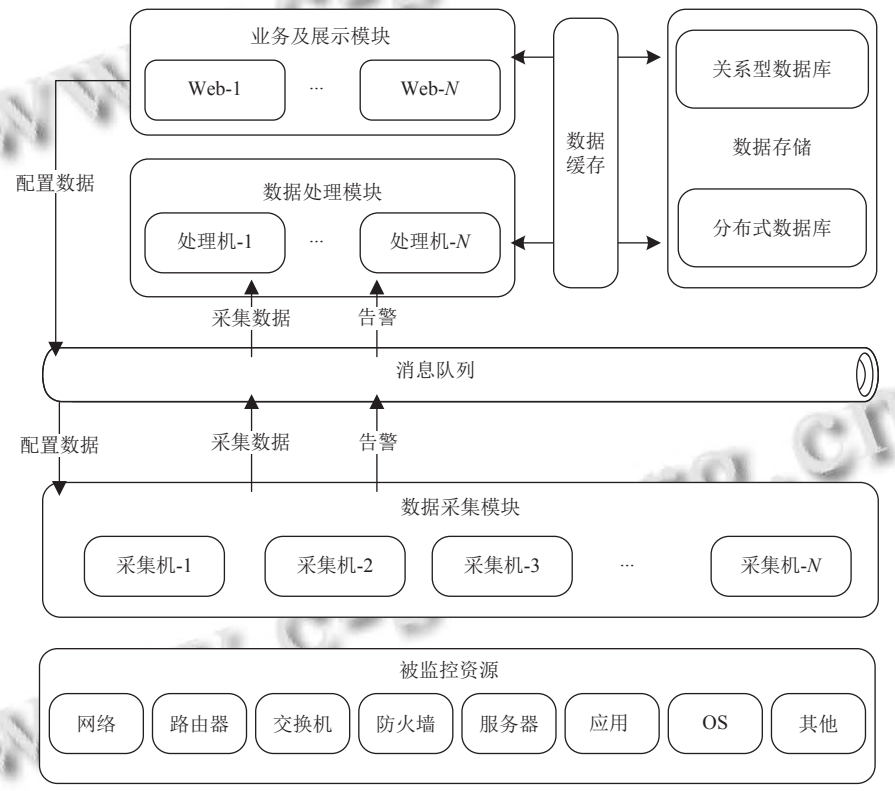

图 1 云网监控系统逻辑架构

2 云网监控系统灵敏度优化

2.1 网络层面的优化分析

云网监控系统使用 SD-WAN 服务, 通过广域网动态路径控制选择延迟最小的链路, 优化采集机与被监控设备之间的网络链接, 缩短资源层异常到达采集机的时间.

2.2 硬件层面的优化分析

云网监控系统采用私有云构建, 在兼顾成本的前提下, 尽量购买硬件性能好的资源, 分布式数据库尽量安装在物理机上, 以提升系统的基础性能, 缩短异常数据在计算、IO、传输等环节的时间.

2.3 中间件、数据存储、数据缓存的优化分析

云网监控系统基于开源的 ActiveMQ、MySQL、HBase、Redis 构建了消息队列、数据存储、数据缓存. 系统集成 ActiveMQ 时, 数据采集侧 (生产者) 采用异

步发送及 DeliveryMode=NON_PERSISTENT 的模式提升数据写入速度; 数据处理侧 (消费者) 采用 Prefetch、事件驱动、多消费者协同等方式提升数据消费速度[2]. 系统采用 MySQL 存储相对静态的基础配置数据, 采用 HBase 存储大数据量的动态采集数据, 在 HBase 的存储模型设计时将设备 ID、服务 ID、时间戳等通过翻转处理组合为 RowKey, 分散存储动态告警、性能数据, 提升数据持久化、查询的速度[3]. 系统采用 Redis 缓存相对静态的配置数据及最近的动态采集数据, 提升数据汇聚、处理、查询的速度[4]. 通过以上优化措施, 系统可以达到近 1 400 条/s 的数据处理、存储速度, 更快的处理速度可以使系统在更短的时间内发现异常.

表 1 影响监控系统灵敏度的因素分析

序号	所属层	主要影响因素
1	被监控资源层	被监控对象与采集机之间的网络速度
2	数据采集层	数据采集频率、采集耗时
3	消息队列	数据采集 (生产者)与数据处理 (消费者)的速度
4	数据处理层	数据处理 (消费者)的速度、数据分析产生告警的速度
5	数据存储	数据持久化的速度
6	数据缓存	缓存数据命中率
7	业务及展现层	页面刷新的速度、服务端与客户端的网络速度

2.4 数据采集层的优化分析

云网监控系统的数据采集层主要通过被动接收或主动采集的方式获取被采集资源的性能、告警、状态数据, 数据采集层影响监控系统灵敏度的两个主要因素是: 采集频率和采集耗时, 采集频率越高、采集耗时越小, 则发现异常的时间越短.

(1) 采集频率优化

在云网监控系统中, 数据采集通过每个采集任务进行, 提升采集系统的频率需要缩短采集间隔, 会大量增加采集任务的数量. 采集任务的数量取决于被监控资源数量、采集指标数量、采集间隔. 采集任务数量=被监控资源数量×采集指标数量/采集间隔. 采集指标的个数取决于客户要求监控的服务内容 (CPU、内存、端口流量、系统状态、风扇等).

在实践过程中发现频率的提升会带来严重的性能问题 (图 2、图 3 分别为 30 s、10 s 间隔单台采集机 CPU 使用率图, 可以明显看到 10 s 间隔下, CPU 使用率频繁触及 100%): 尽管系统通过多采集机分散采集任务, 且通过并行线程方式处理采集任务, 但大量的采集任务 (如表 2 所示) 会竞争采集机有限的线程、网络资源, 未被及时处理的采集任务处于排队状态, 滞留的采集任务不断堆积导致系统无法正常进行采集.

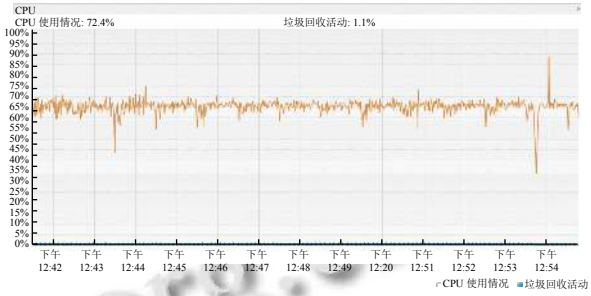

图 2 单台采集机 CPU
(2 000 台设备, 20 个 OID, 30 s 采集间隔)

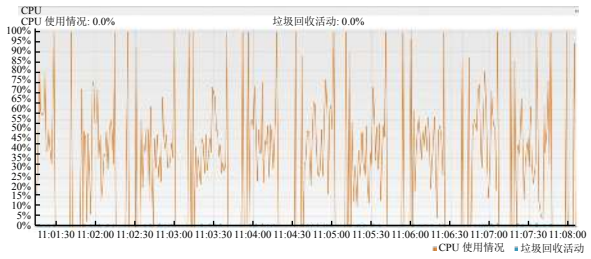

图 3 单台采集机 CPU
(2 000 台设备, 20 个 OID, 10 s 采集间隔)

表 2 采集任务数量举例

序号	被监控资源数量 (个)	采集指标数量 (个)	采集间隔 (s)	采集任务数量 (个/s)
1	10 000	10	300	333
2	10 000	10	60	1 667
3	10 000	10	30	3 333
4	10 000	10	10	10 000

为了解决这一问题, 系统设计了一种多级可变频率网络监控数据采集方法及装置[5] (如图 4 所示): 不是所有设备均需要同时提升采集频率, 只需要针对出现异常的设备提升采集频率即可, 该方法根据最新采集数据与历史数据的离散度识别异常设备, 赋予系统在设备采集数据离散程度加大或已发生告警时临时提高数据采集频率、异常消除或数据离散程度降低后降低采集频率的自适应能力, 可以有效地提高数据采集效率、资源利用效率及系统的整体监控灵敏度.

该设计包括定时调度模块、并行采集模块、并行数据处理模块、变频边界计算模块和变频控制模块. 其工作原理如下:

1) 定时调度模块从数据库中加载采集任务信息为

每种采集任务生成高、中、低 3 种频率的定时调度任务, 每种定时调度任务定时为设备采集频率表中对应频率的设备生成采集任务, 加入采集任务队列.

2) 并行采集模块从采集任务队列获取并执行采集任务, 将采集数据加入采集数据队列.

3) 并行数据处理模块处理采集数据队列中的采集数据, 存储到数据库, 同时将最新的采集数据、设备可达状态传递给变频控制模块.

4) 变频控制模块判断设备可达性, 如果设备不可达, 则调整该设备的对应采集任务频率为低频率, 如果设备可达, 则变频控制模块获取变频边界计算模块计算的变频边界.

5) 变频边界计算模块从数据库中加载一段时间内的历史采集数据并缓存, 计算变频边界: 算出缓存的历史数据的算术平均值和标准差, 设置设备的变频边界为算术平均值 K 倍标准差、告警阈值上限、告警阈值下限.

6) 变频控制模块根据最新采集数据与变频边界的关系动态调整采集频率: 若最新采集数据小于等于告警阈值下限或大于等于告警阈值上限, 则调整设备对应的采集任务频率为高频率; 若最新采集数据大于算术平均值 K 倍标准差且小于算术平均值 K 倍标准差, 则调整设备对应的采集任务频率为低频率; 若最新采集数据大于告警阈值下限且小于等于算术平均值 K 倍标准差或者最新采集数据大于等于算术平均值加 K 倍标准差且小于告警阈值上限, 则调整设备对应的采集任务频率为中频率, 其中 $1 \leqslant K \leqslant 3$.

7) 如果设备的采集任务频率与原频率不同, 则更新设备任务采集频率表.

采用自适应调整采集频率的方法, 例如原固定频率为 60 s 的采集, 通过高中低频率 (10 s、30 s、60 s) 的变频控制, 发现异常的时间可以为 10 s (最好情况)、30 s、40 s、60 s (最坏情况), 同时降低了正常设备的采集频率, 有效降低了资源消耗, 提升了系统的容量和稳定性.

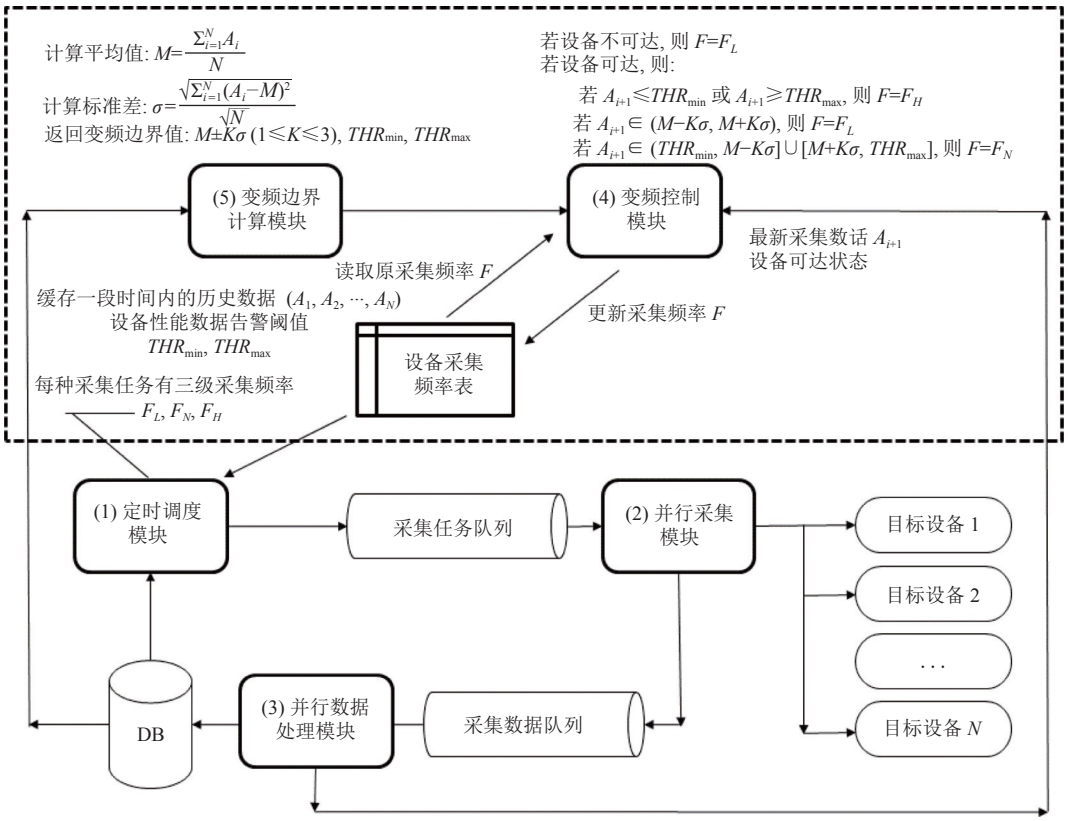

图 4 一种多级可变频率网络监控数据采集方法及装置

(2) 采集超时时间优化

影响数据采集时间的还有一个隐性的因素: 采集超时时间, 在被采集设备可达的情况下, 采集超时时间对系统是无影响的, 当被采集设备由于某种原因不可

达时, 采集任务会一直等到采集超时时间结束才会释放资源, 当大量的被采集设备出现不可达的情况时, 会导致有限的网络、内存资源被长时间占用, 采集任务不断堆积导致系统无法正常采集. 为了消除这一影响, 系统设计了一种基于超时因子的大规模网络数据采集方法及装置[6], 如图 5 所示.

该方法及装置包括定时调度模块、并行数据采集模块、并行数据处理模块、采集超时自适应模块、设备采集耗时表、设备采集超时表. 该装置的定时调度模块、并行数据采集模块、并行数据处理模块功能与变频设计中的对应模块功能相同, 其主要工作原理如下:

1) 定时调度模块为超时因子不等于 0 的设备生成采集任务, 并加入采集任务队列.

2) 并行数据采集模块从采集任务队列中获取并执行采集任务, 根据超时因子设置采集超时时间=超时因子×默认采集超时时间.

3) 判断采集任务是否成功执行, 若成功执行则将采集的数据加入数据队列, 若失败则并行数据采集模块更新超时设备信息表中该设备的超时因子=原超时因子−固定值, 超时因子最小值为 0.

4) 超时因子控制模块以固定频率探测超时因子为 0 的设备, 如果设备恢复上线, 则调整该设备的超时因子为最大值.

通过引入超时因子及延时反馈机制, 赋予云网监控系统面对大规模数据采集的自适应能力, 能够在有效的时间范围内完成大量异构监控对象的数据采集, 有效提高数据采集效率及监控系统的容错能力, 提高云网监控服务的异常发现效率.

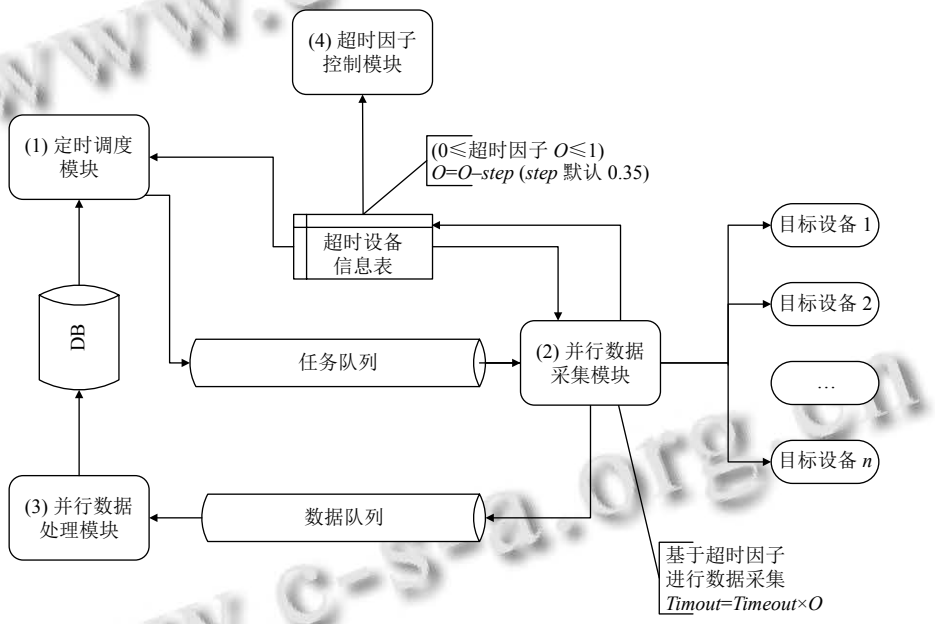

图 5 一种基于超时因子的大规模网络数据采集方法及装置

2.5 数据处理层的优化分析

云网监控系统的数据处理层负责处理消息队列中的采集数据并持久化到数据存储中. 通常情况下, 对于单值类的数据告警一般采用预设阈值的方式, 当采集数据超出阈值一定次数后, 系统发出告警信息. 实践中采集数据一般有一个逐渐趋于阈值并超越阈值的过程. 在数据处理层, 系统在原有阈值告警的基础上, 增加了随采集数据动态变化的预告警机制[7]: 计算缓存的历史数据的算术平均值和标准差, 动态预告警阈值为算术平均值±两倍标准差, 预告警规则如下:

1) 若$A_{i+1} \in (M-2\sigma, M+2\sigma)$且已发出预告警, 则发出原有预告警取消信息.

2) 若 $A_{i+1} \in (THR_{\min}, M-2\sigma] \cup [M+2\sigma, THR_{\max})$ 连续 3 次且未发出预告警, 则发出预告警信息.

其中, A_{i+1}为最新采集数据, M 为最近一段时间内的历史数据的算数平均值, σ 为标准差, $THR_{\min}$,$THR_{\max}$ 分别为手工预设阈值的上下限.

图 6 中的阈值告警控制模块即为原有的通过阈值判断告警的模块, 系统增加了动态调整边界的预告警控制模块后, 赋予了系统在设备采集数据离散程度加

大时产生预告警, 使得运维工程师可以提前进行设备的异常排查和干预, 避免了传统的必须超过固定阈值才能告警的方式, 既能够有效预防异常的发生, 又可以提高系统的整体监控灵敏度.

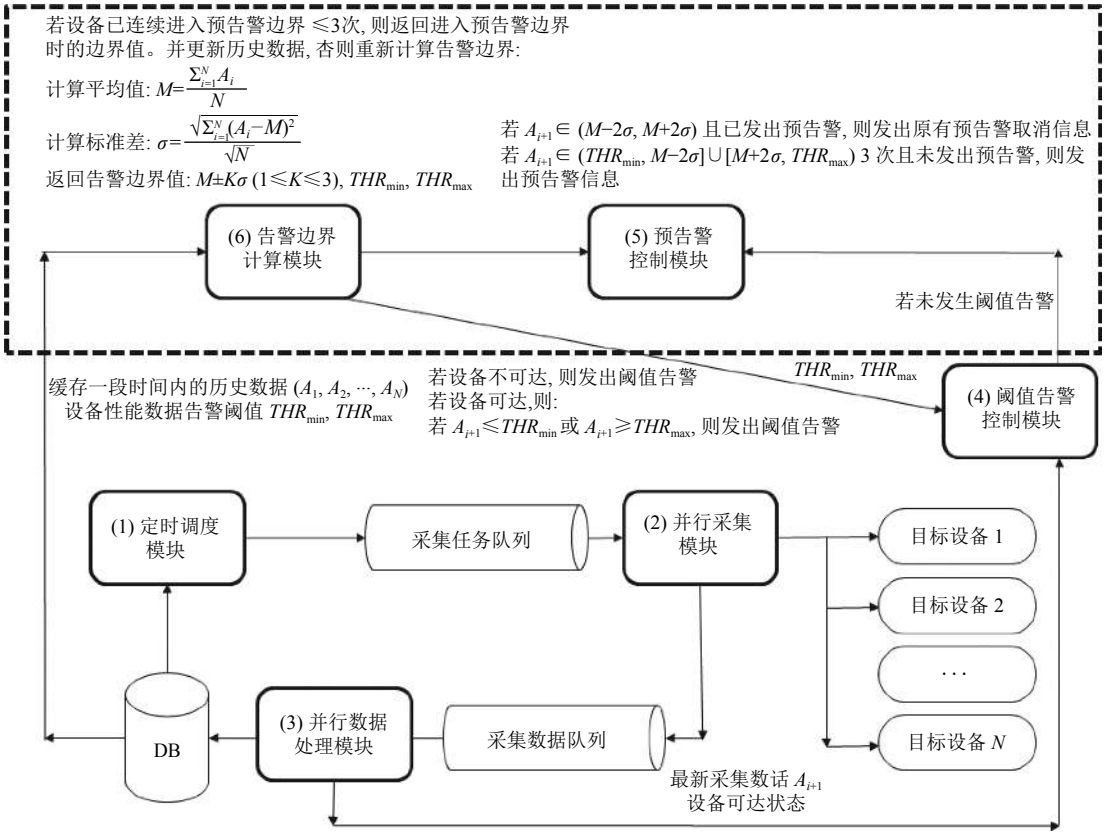

图 6 增加了预告警的云网监控系统

对于告警灵敏度要求较高的企业可以通过调整标准差倍数告警边界及超过告警边界次数进行灵活的预告警设置, 表格 3 为常用标准差倍数及连续 3 次越过告警界限情况下的预告警的概率值. 预告警的概率值=$\left(1-\text{标准正态分布概率}\right)^{\text{超过告警边界次数}}$, 如表 3.

表 3 预告警的概率

序号	标准差	超过告警边界次数 (次)	预告警概率 (%)
1	一个标准差	3	3.2
2	两个标准差	3	0.009
3	三个标准差	3	0.000 002 7

2.6 业务展现层的优化分析

云网监控系统的告警信息进入数据存储后, 一方面可以通过邮件、短信、语音等方式通知运维人员, 另一个重要的途径是通过业务及展现层通知给监控人员, 为了实现实时的通知, 云网监控系统采用 WebSocket 长链接代替客户端定时刷新[8], WebSocket 的服务端在消息队列注册一个消费者, 数据处理层遇到高优先级的告警时, 会将告警信息在持久化前放入消息队列中, WebSocket 的服务端接收到告警信息后, 立即将告警信息推送给客户端, 在客户端页面的显著位置进行告警提示 (同时伴有声音提示). 相比客户端定时刷新机制, 基于 WebSocket 的消息机制能够在更短的时间内将告警信息推送给监控人员, 尽快引导运维人员介入异常处理, 提升了系统的监控灵敏度.

3 结论

基于以上分析及优化, 影响云网监控系统灵敏度的因素中, 除了网络、硬件层面的因素外, 其他因素系统均采取了对应的设计方案进行了优化, 其中系统可控的且对系统监控灵敏度影响最大的两个因素: 采集频率、采集超时时间, 采用了自适应的动态应对方案, 相对于未优化前整个系统的灵敏度有了极大的提高

(最优情况下可以达到秒级). 灵敏度的提升往往伴随着成本的提升, 目前企业的运维需求基本是在分钟或小时级别的异常处理标准, 监控系统的灵敏度需要在客户需求与成本投入之间找到一个结合点, 达到双赢的结果.

参考文献

1 吴舸, 袁守正, 孙鼎. 运营商网络监控系统高可用性设计及应用. 计算机系统应用, 2020, 29(11): 87–91. [doi: 10.15888/j.cnki.csa.007502]

2 戴俊, 朱晓民. 基于 ActiveMQ 的异步消息总线的设计与实现. 计算机系统应用, 2010, 19(8): 254–257, 215. [doi: 10.3969/j.issn.1003-3254.2010.08.059]

3 贺正红, 周娅, 文缔尧, 等. 面向 HBase 的大规模数据加载研究. 计算机系统应用, 2016, 25(6): 231–237. [doi: 10.15888/j.cnki.csa.005194]

4 李翀, 刘利娜, 刘学敏, 等. 一种高效的 Redis Cluster 的分布式缓存系统. 计算机系统应用, 2018, 27(10): 91–98. [doi: 10.15888/j.cnki.csa.006576]

5 吴舸, 袁守正, 孙鼎, 等. 一种多级可变频率的网络监控数据采集方法及装置: 中国, CN113114508A. 2021-07-13.

6 袁守正, 吴舸, 曹征, 等. 一种基于超时因子的大规模网络数据采集方法及装置: 中国, CN111769982A. 2020-10-13.

7 黄冬, 施跃跃, 刘震, 等. 一种确定资源监控阈值的方法及装置: 中国, CN106713029B. 2020-05-01.

8 李道震, 张长生, 郎向伟, 等. 基于 WebSocket 和 ArcGIS Server 的高铁基础设施在线监测系统. 计算机系统应用, 2016, 25(2): 38–44.