*基于可编程交换机的高效键值存储监控系统*

作品简介

1. 需求与背景

键值存储在现代互联网服务中扮演着关键角色，作为存储基础设施的核心组成部分。这种存储方式通常以键值格式存储可访问的项目，用户通过向系统发送带有特定键的请求来获取相应的项。

实际应用中，对键值存储的服务质量有着严格的要求，即服务水平目标（SLO，Service Level Objective）。然而，热点问题成为满足SLO的一大挑战，即极少数热门项相较于其他项被过度访问，导致性能下降。例如，实时突发新闻事件可能导致90％的访问集中在仅占1％的项上 [5]，从而导致存储服务器超载和尾部延迟增加 [6–8]。热点问题逐渐严重，其影响巨大，每100毫秒的加载延迟会导致亚马逊销售下降1％，Google搜索每增加500毫秒的加载延迟会损失20％的流量 [9]。

热点监控成为解决热点问题的关键，它通过收集工作负载信息并确定热点及其特征，为快速全面的热点缓解提供了必要的信息。

然而，现有方法在实现可扩展性即面对大量请求，能够迅速监测热点和多功能性即能够检测不同的工作负载类型和过载位置方面存在限制。如主机端遥测方案缺乏对交换机的可见性，基于通用CPU方案存在着性能瓶颈。为了解决这些问题，项目提出了**SpotMon，一个可扩展且多功能的热点监控系统**。SpotMon借助**可编程交换机在数据平面聚合监控数据**，由于交换机具备天然的机架级可见性，缓解了网络链路和控制器过载的问题。同时，引入**基于sketch的数据结构**，支持多功能热点查询抽象，使得计数和查询监控数据的子集变得高效而准确。通过**分层查询算法**，系统操作员可以轻松地表达他们关心的热点特征，如热点项、工作负载类型以及过载位置。本项目同时提供了四种优化方法，可以对热点引起的问题进行有效的缓解。

1. 设计思路

SpotMon模型主要由以下几个模块组成，并且每一模块所解决的问题如下。

**1.数据平面模块:** 这个模块位于可编程交换机的数据平面，实时监控通过网络传输的数据流量。通过在数据平面进行监控，可以及时捕捉到网络中的热点负载情况。

**2.控制平面模块:** 控制平面模块负责接收来自数据平面的监控数据，并进行数据处理和负载分析。它通过分析监控数据，识别出热点负载、负载类型和负载位置，为后续的查询和优化提供基础。

**3.Lake模块:** Lake是SpotMon的核心组件，使用了Bloom Filter和Count-Min Sketch等数据结构。它负责对监控数据进行处理，将数据流划分为重负载和轻负载，并利用精确统计和估计的方法进行热点负载的监测。Lake的设计使得系统可以准确地感知热点负载情况，为后续的查询和优化提供了准确的基础。

**4.多功能热点查询模块:** 这个模块为系统操作员提供了多类型负载感知查询的接口。操作员可以根据需求查询不同类型的热点负载信息，从而实现精细的负载监控和管理。在其中实现了一种层次化查询算法，用于识别热点负载。算法从粗粒度到细粒度逐步进行查询，同时避免了不必要的查询操作，从而提高了查询的效率和精确度。

**5.热点优化模块:** SpotMon采用了四种优化技术来针对热点负载进行综合性的优化。这些技术包括通道平衡、服务器平衡、进程平衡、**消除写入项目**。通过根据热点负载信息进行系统性能的优化，可以提高系统的吞吐量和响应时间。

1. 成果描述

SpotMon的引入为键值存储系统带来了显著的成果。通过实时监控和多类型感知，SpotMon能够精准地识别系统中的性能瓶颈，并采取相应的优化策略来应对这些问题。其核心模块Versatile Hotspot Query实现了多类型负载感知，使系统能够区分不同类型的负载热点，从而更有针对性地进行优化和调整，保障关键任务的响应时间和吞吐量。同时，SpotMon采用了负载均衡、请求重定向、负载迁移和热点数据过滤等四种综合性优化技术，相互协作地优化了系统的性能，有效降低了热点引起的SLO问题。此外，SpotMon在可编程交换机上实现，硬件资源的消耗相对较低，同时在保证监控精度的同时，最大程度地降低了硬件资源的使用。总体而言，SpotMon为键值存储系统的优化和性能提升做出了重要贡献，通过实时监控、多类型感知和综合性优化方案，有效地应对了热点引起的SLO问题，提高了系统的稳定性和性能表现。

1. 创新与特色
2. SpotMon基于可编程交换机在网络内部对键值存储系统实时监控，具备网络内的可见性，全面监控并检测键值存储系统热点。
3. SpotMon利用可编程交换机高速聚合网络流量，大大减少了上报至控制平面的流量信息，具有具有高度的可扩展性和鲁棒性。
4. SpotMon基于高效的概要数据结构测量网络流量信息，同时支持多种负载类型信息查询，具备可证明的测量误差保证。
5. SpotMon 基于测量信息提出了多种负载感知算法，同时支持多种查询接口，以满足系统操作员的不同测量需求。
6. 已有知识产权成果

专利已进入实质审查阶段。

1. 应用价值或产业化前景

本项目主要应用于电商、搜索引擎、社交网络等需要高性能在线服务的键值存储系统。本项目可以检测工作负载类型和过载位置的热点，旨在快速和全面地进行热点缓解。

1. 团队简介

吴春明：浙江大学系统结构与网络安全研究所副所长、教授，长期从事互联网体系结构、柔性可重构网络、软件定义网络与安全、网络业务试验床和网络空间主动防御等领域的科研工作，主持、参加了二十余项国家自然科学基金，国家973、863 计划项目，国家科技基础条件平台建设等重大项目。发表学术论文八十余篇，专利二十余项，合作著作三部。曾任国家863 信息领域网络与通信技术主题专家组专家，现担任国家“十三五”重点研发计划“网络空间安全”重点专项科技规划组成员、编制组专家，国家“网络空间安全”专项专家分委会委员，国家2030 科技专项“国家网络空间安全”重大项目实施方案编制组专家，中国电子学会委员、《电子学报》编委，浙江大学-中兴通讯联合创新中心执行主任，拟态技术与产业创新联盟理事会副理事长。曾荣获国家科技进步二等奖，公安部颁发的网络安全保障工作突出贡献个人一等奖章。

祖敬涵：浙江大学计算机科学与技术学院信息安全专业2022级本科生，目前从事eBPF与linux内核方面的研究，加入蚂蚁集团隐语开源社区共建计划，现有一篇eBPF优化键值存储系统尾延迟论文在投。在竞赛中负责文稿撰写、PPT制作、视频制作、网站搭建和实验测试。

王子扬：浙江大学控制学院自动化2022级本科生，加入蚂蚁集团隐语开源社区共建计划。在竞赛中负责文稿撰写、创新思考。

黄恩浩：2022级本科生，目前从事可编程交换机与隐私计算方面的研究，参与《基于可编程交换机的Sketch网络测量重配置》项目以及《多模态边缘网络关键技术》国家重点研发计划重点专项项目，加入蚂蚁集团隐语开源社区共建计划，在竞赛中负责创意提出、文稿撰写。

声 明

本作品团队 □有 / □没有 成果向企业转移或进行创业的意向，并 □愿意 /□不愿意 由“中国高校计算机大赛-网络技术挑战赛”组织委员会提供牵线搭桥的机会。

本作品团队同意授权“中国高校计算机大赛-网络技术挑战赛”组织委员会使用本文档中的内容制作宣传资料！

指导教师签字: 

 学生代表签字: