计算机工程 2012, 38(3) 43-45,48 DOI: 10.3969/j.issn.1000-3428.2012.03.015 ISSN: 1000-3428 CN: 31-1289/TP

软件技术与数据库

基于文本策略和SMCS的海量日志分析方法

张俊峰[1]，冯巧娟[1]，张晓丽[1,2]

(1. 河南城建学院计算机科学与工程系，河南 平顶山 467036； 2. 南京航空航天大学航空科技智能材料与结构重点实验室，南京 210016)

摘要： 现有的海量日志统计分析方法速度慢，且对硬件配置的要求高。为此，提出一种基于文本策略和SMCS的海量日志分析方法。根据文件的软件设计策略，采用日志文件索引方法，将日志文件与日志时间关联，以加快日志提取。SMCS算法采用哈希表、文件归并、堆操作方法对海量日志进行统计分析和内存损耗控制。通过对真实软件进行对比实验，结果表明，该方法的分析速度比传统方法提高4倍。

关键词： Syslog日志 日志分析 SMCS算法 海量日志 文本策略 控制内存

Mass Log Analysis Method Based on File Strategy and SMCS

ZHANG Jun-feng [1], FENG Qiao-juan [1], ZHANG Xiao-li [1,2]

(1. Department of Computer Science and Engineering, Henan University of Urban Construction, Pingdingshan 467036, China; 2. Aeronautical Science Key Laboratory for Smart Material and Structures, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: Now the statistical analysis for the massive logs not only is time-consuming, but also requires higher hard configuration. For these questions, a method based on the file software strategy and SMCS algorithm is proposed in this paper. By the software strategy, the log file indexing and the association between the log file and log time are adopted in this method, so as to quicken the log extraction. In order to analyze the massive logs statistically and control the memory loss, hash tables, file merging and heap operation are used in the SMCS algorithm. By experiment of the real software, results indicate that the proposed method is faster than that of traditional statistical analysis by four times.

Keywords: Syslog log log analysis SMCS algorithm mass log file strategy control memory

收稿日期 2011-08-17 **修回日期** **网络版发布日期** 2012-02-05

DOI: 10.3969/j.issn.1000-3428.2012.03.015

基金项目:

国家自然科学基金资助项目(60907038)；河南省科技攻关计划基金资助重点项目(102102210020)

通讯作者:

作者简介: 张俊峰(1967－)，男，副教授、硕士，主研方向：网络安全，分布式处理；冯巧娟，讲师、硕士；张晓丽，讲师、博士

通讯作者E-mail: pzjf2010@163.com

参考文献：

[1] Lonvick C. The BSD Syslog Protocol[S]. RFC 3164, 2001.

[2] 左利云. 基于网络内容安全的数据流查询优化算法[J].计算机工程.2010, 36(11):45-47 浏览

[4] George R.[J].Randy J, Tim K. Manageing & Using MySQL[M]. 2nd ed. [S. l.]: O’Reilly & Associates Inc.2002,:-

[5] Jun Guzhao.[J].Li Yong, Jing Niuwen. Analysis and Implement of PIX Firewall Syslog Log [C]//Proc. of the 2nd IEEE International Conference on Information Management and Engineering. Piscataway, USA: IEEE Press.2010,:-

[6] Stanley B.[J].Josee L, Barbara E. C++ Primer[M]. 4th ed. [S. l.]: Addison-Wesley Professional.2005,:-

[7] Alsuwaiyel M H. Algorithms Design Techniques and Analysis[M]. [S. l.]: World Scientific Publishing Co..[J]..1999,:-

[8] 严蔚敏, 吴伟民. 数据结构[M]. 北京: 清华大学出版社, 2005.

本刊中的类似文章

1. 朱靖君, 吴海燕, 高国柱, 程志锐.一种基于日志分析的Web负载测试方法[J]. 计算机工程, 2010,36(23): 25-27
2. 窦志成;袁晓洁;何松柏.大规模中文搜索日志中查询重复性分析[J]. 计算机工程, 2008,34(21): 40-41,4
3. 余亚玲;唐红武;杜海霞.基于日志的安全事件管理系统的研究与实现[J]. 计算机工程, 2007,33(16): 128-129,

文章评论

反馈人		邮箱地址	
反馈标题		验证码	3288