

引用格式:程军. 基于数据融合分析的网络监控系统框架构建与关键技术研究 [J]. 微电子学与计算机, 2019, 36(2): 97- 100.
[CHENG J. Struct construction and key technology research of network monitoring system based on data fuzzing analyze [J].
Microeletronics & Computer, 2019, 36(2): 97- 100.]

基于数据融合分析的网络监控系统 框架构建与关键技术研究

程 军

(新时空科技有限公司, 山东 济南 250001)

摘 要: 针对增强复杂网络环境下监控预警能力的需要, 研究构建了一种基于数据融合分析的网络监控系统框架, 探讨了 IP 分组缓存管理、通联特征提取、目标关联识别、关系结构分析等关键技术, 对网络流量实现整体监控、综合挖掘和融合利用, 为网络监控技术手段的完善拓展提供方案借鉴。

关键词: 网络监控; 融合利用; 特征提取; 关联识别; 关系分析

中图分类号: TP393

文献标识码: A

文章编号: 1000—7180(2019)02—0097—04

Struct Construction and Key Technology Research of Network Monitoring System Based on Data Fuzzing Analyze

CHENG Jun

(New horizon Technology Ltd, Jinan 250001, China)

Abstract: Aiming at enhancing the ability of monitoring and reflection under complex network conditions, a new network monitoring structer based on data ALL analyse is established, and key skills such as IP packet buffer management, communication character abstract, correlated identify, and relationship structer analyze are discussed. The structer can improve overall monitoring, full dig, and all use of network flows, and important reference are offered for improving and extending network monitoring skills.

Key words: network monitoring; fuzzy usage; character extract; correlation identifaction; relationship analyze

1 引言

随着网络技术的蓬勃发展和快速应用, 社会生产生活方式发生了巨大而深刻的变化. 与此同时, 网络违法犯罪行为数量激增、手法多样复杂, 对网络安全监控工作提出了严峻挑战. 因此, 在复杂的网络环境下, 如何运用大数据等新技术手段提高网络监测预警能力是亟待研究解决的问题。

本文将数据融合分析技术用于网络监控领域, 构建了一种新的网络监控系统框架, 可提高网络流量的整体监控、综合挖掘、融合利用等能力, 对于促

进网络监控手段完善与拓展具有重要意义。

2 系统框架构成

基于数据融合分析的网络监控系统总体上由网络数据实时采集单元和网络数据融合分析单元两部分组成. 前者实时抓取出入被监控网络的 IP 分组, 后者对抓取到的网络数据进行融合分析, 获取被监控网络目标综合通联信息, 系统总体结构如图 1 所示。

网络数据实时采集单元功能结构如图 2 所示, 由 2 个 DM9000 模块, 1 片 FPGA 芯片, 1 个

PMC3386 千兆网络接口等四部分组成. DM9000 模块由网络接口、接收 FIFO、发送 FIFO、MCU 接口等四部分组成,主要功能是与被监控网络 and 核心网建立通信链路,进行数据缓存和发送、接收. PMC3386 千兆网络接口由 PMC3386 芯片和网络接口组成,功能是为所采集网络数据向网络数据融合分析单元的实时传送建立千兆通信链路. FPGA 芯片内部由仲裁控制、数据收发接口、发送分发、接收分发、千兆 MAC 芯片接口等模块组成,功能是协调控制网络数据的发送、接收和分发,在完成网络数据正常收发的基础上实时采集并传送给网络数据融合分析单元.

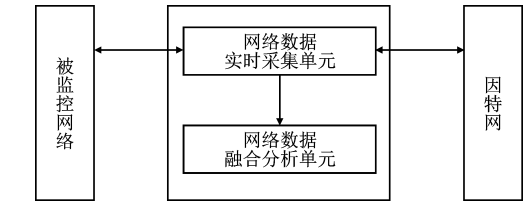


图 1 系统总体结构

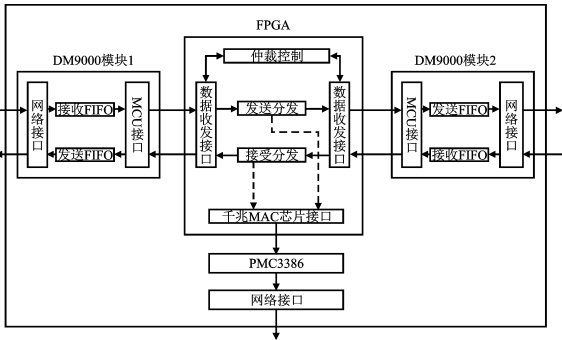


图 2 网络数据实时采集单元功能结构

网络数据融合分析单元的硬件基础为带有千兆网卡的 PC 机或工控机;软件功能结构如图 3 所示,由千兆网络接口、网络数据收取、IP 分组缓存与管理、通联特征提取、IP 分组重组、应用业务解析还原、异常数据筛选、网络监控数据库、网络关键节点发现、目标关联识别、通联关系分析、网络组织结构描绘等模块组成. 千兆网络接口模块和网络数据收取模块完成网卡管理控制和 IP 分组收取. IP 分组缓存与管理模块对所收取的 IP 分组按照协议五元标识组进行分类和存贮,使同一端对端通信的 IP 分组毗邻存放,并对缓存区进行高效管理调度,以提高存贮空间利用率和数据存取速度. IP 分组重组模块对缓存区中的 IP 分组按照 TCP/IP 协议规定进行

碎片组合、序列号先后排序、IP 头剔除、TCP/UDP 头剔除, TCP/UDP 段重组等操作,恢复出应用业务数据段. 通联特征提取模块完成五元组标识、端对端数据流量、通联时间等信息的提取. 异常数据筛选模块对应用业务数据中的异常信息等进行识别筛选. 网络监控数据库用于 IP 分组数据、网络通联特征、异常信息等存贮管理. 网络关键节点发现、目标关联识别、通联关系分析、网络组织结构描绘等模块进行网络监控数据的融合分析,比对发现同一类别网络目标,获取网络组织结构、通联关系、活动规律、关键节点属性等.

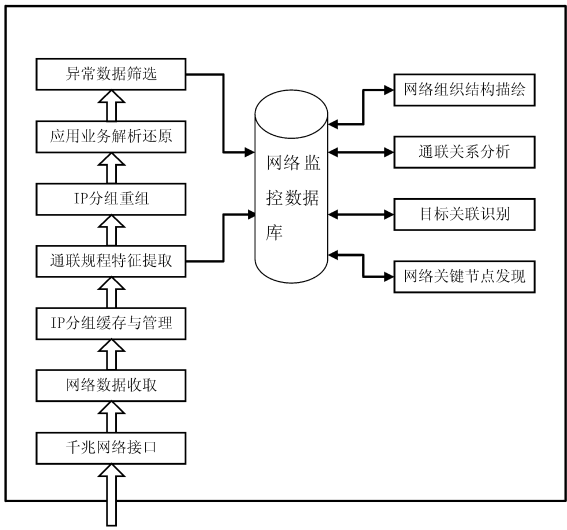


图 3 网络数据融合分析单元功能结构

3 主要关键技术

3.1 IP 分组缓存管理

IP 分组缓存区管理技术的性能优劣直接关系 IP 分组分拣速度,影响缓存区利用率和访问效率. 设计实现了分段索引管理算法,如图 4 所示,其主要过程为:(1)分别提取 IP 分组源地址和目的地址的高 16 位,合并组成 32 位分段字;将所有分段字以链表形式管理;(2)根据分段字对 IP 分组进行分段存储,将分段字相同的 IP 分组置于同一分段下管理. 对同一分段下的每一个 IP 分组,以五元组标识信息组成标识字,将所有标识字组成索引链表进行管理,然后以标识字为指针开辟缓存区,并将 IP 分组写入缓存区. 由于五元组标识信息能唯一确定一个端对端通信,该机制保证了同一端对端通信的所有 IP 分组存储于相同缓存区. 该算法具有运算复杂度小、易于工程实现等优点,可满足 IP 分组分拣、缓存管理

等需要.

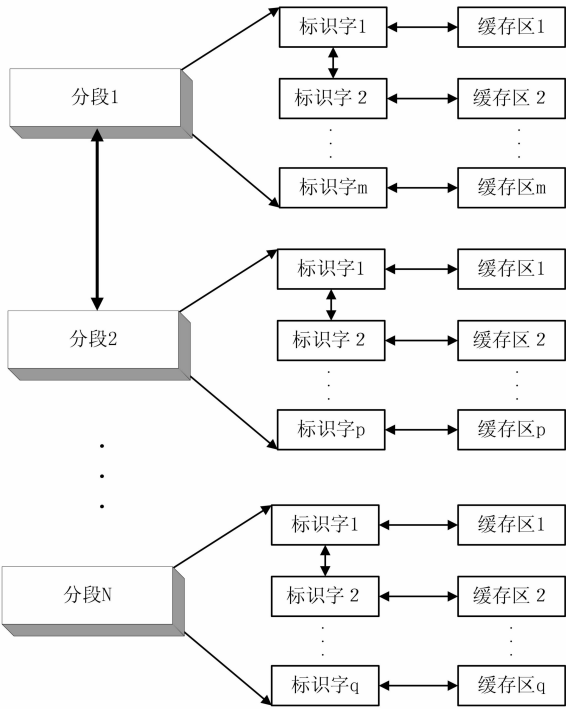


图 4 分段索引缓存管理算法

3.2 通联特征提取

从截获的 IP 分组数据中筛选出通联特征信息,包括 IP 网络通信五元组(源 IP 地址、目的 IP 地址、协议类型、源端口、目的端口)、数据流量、数据传输时间等.提取出的通联特征信息存入网络监控数据库.通联特征信息及敏感数据提取采用特征字查找匹配的方法和 IP 分组首部标识域定位与抽取的技术.特征字查找匹配是从 IP 分组重组后的报文中搜索发现关键词,并提取相应信息. IP 分组首部标识域定位与抽取过程则是从 IP 分组首部中按照 TCP/IP 协议规范解析获取各标识域的取值.

3.3 关联识别

关联识别所采取的主要通联特征:(1)IP 地址网段范围;(2)IP 分组长度;(3)IP 协议类型;(4)端口使用分布;(5)应用特征信息,包括业务类型使用,用户通联行为规律,服务器属性、FTP 站点名称、网站域名信息等.

多维通联特征比对是关联识别过程的重要技术.比对过程中,实际网络目标与已掌握网络目标之间的通联特征,可能完全匹配,可能全不匹配,也可能部分匹配即全部特征中仅有一维或几维特征相符.实际网络监控过程中,最后一种情形居多.对于

前两种情形,直接输出确定的判决结果即可,而对于最后一种情形,则需要区别对待.部分特征相符时,将具有多个种类网络目标特征,与网络目标特征数据库中的多个通联特征具有相似性且相似程度各不相同,有的相似性少一些,有的相似性大一些,此时将输出多个比对结果,形成一对多的映射关系,而不是一对一映射关系.如何科学合理度量通联特征相似性并提出准确有效的判决方法,是需要重点研究解决的问题.为此,提出利用特征距离描述通联相似性,并设计基于累计特征距离的判决方法.

特征距离,即已分析澄清技术规格的网络目标与待识别网络目标之间的通联特征差异或相似性.由于通联特征具有多类别多粒度的特点,因此可描述为 N 个 M 维矢量.当两个通联特征完全相同时,特征距离为 0;当两者完全不同时,特征距离为 $f(N * M)$;当两者部分相同时,特征距离大于 0 且小于 $f(N * M)$.根据特征距离的计算结果判决待识别网络目标属于哪一类网络目标,判决结果不是确定性的,只给出待识别网络目标与已分析澄清技术规格的各类网络目标之间的相似程度.

累计特征距离判决方法的基本原理是:按照 OSI 协议栈模型由下向上的顺序递归进行特征距离计算和累加,既反映某协议层上待识别网络目标与已分析澄清技术规格的各类网络目标之间的相似程度,又确定多个协议层上两者之间的相似程度,从而全面准确地判决待识别网络目标的类别.

3.4 关系结构分析

通联关系结构分析过程中,首先根据用户要求及设定参数,对网络通联特征库中存储的信息进行搜索、发现、筛选和抽取;然后对一个或多个特征信息进行比较、排序、合并、统计及数值运算等处理,获取目标网络拓扑结构、通联关系、关键节点位置、通信规律特点、流量大小、端口使用等信息.其中,通联关系与拓扑结构的自动分析是需要研究的关键技术.具体实现过程包括四步:

(1)对所有通联信息进行搜索和去重处理,使符合条件的通联信息二元组(源 IP 地址和目的 IP 地址)是唯一的.去重过程如下:对搜索出的所有五元组通信信息条目,不妨设为 $1 \sim N$,从条目 $i(i=1,2 \cdots N)$ 中提取源 IP 地址 P_{i1} 和目的 IP 地址 P_{i2} ,然后从 $i+1 \sim N$ 条目中搜索与 P_{i1} 和 P_{i2} 相同的条目,将相应通信流量相加并进行标识,表明已作去重

处理. 如此类推, 迭代上述操作直到所有条目处理完毕.

(2)源端网络结构分析. 分析过程如下: 首先对搜索出的五元组通联信息的源 IP 地址进行排序, 使网段相同的源 IP 地址聚合, 而网段不同的 IP 地址相互分离. 其次对每一群聚合在一起的源 IP 地址, 按照 A 类、B 类、C 类的 IP 网络分段方法, 依次分类聚合. 其中, 第一步使 B 类网段相同的 IP 地址聚合, 而不同 B 类网段的 IP 地址分离; 第二步使 C 类网段相同的 IP 地址聚合, 而不同 C 类网段的 IP 地址分离. 最终完成网段区分、IP 地址聚合和网络结构分析.

(3)目的端网络结构分析. 与源端网络拓扑结构分析过程相同, 不再赘述.

(4)源端网络与目的端网络之间的通联关系结构分析. 根据相互独立的二元组通联信息和流量大小, 描绘源端网络节点与目的端网络节点之间的连接关系、亲疏程度等.

4 系统部署应用场景

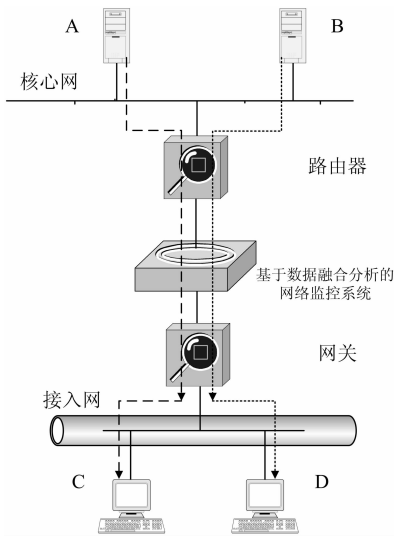


图 5 系统部署使用场景

基于数据融合分析的网络监控系统实施场景如图 5 所示. 该系统架设于核心网路由器与接入网网

关之间, 2 个 DM9000 模块的网络接口分别连接路由器和网关, 网络数据实时采集单元的 PMC3386 千兆网络接口与数据融合分析单元的网卡通过千兆网线连接. 系统运行时, 对于因特网与接入网之间的交互数据, 一方面进行实时转发, 以确保用户对因特网的正常访问使用; 另一方面进行采集分发、存贮管理、通联特征提取、异常数据筛选和融合分析, 既获取监控目标通信内容, 又挖掘网络目标的组织结构、通联关系、行为规律, 发现关键节点位置, 关联识别同类别用户, 实现网络整体运行状态和网络通联情况的严密监控.

5 结束语

网络技术快速发展在便捷资源共享、信息交流和人际交往的同时, 也给网络安全监控技术提出了新挑战. 本文研究提出的基于数据融合分析的网络监控框架能够实现网络流量的整体监控、综合挖掘和融合利用, 对网络监控技术手段的完善拓展提供了方案借鉴. 该框架的原型样机已开发完成, 实网测试效果良好, 具有较好的推广应用前景.

参考文献:

[1] 陈达, 马威, 李晓勇. 一种单向安全隔离与信息交换机制[J]. 信息网络安全, 2014(06): 48-52.

[2] 王锦. 无人值守站点信息单向传输平台的设计与实现[D]. 北京: 北京大学, 2011.

[3] 刘勇燕, 刘勇鹏, 张小红. GAP 技术实现政府内外网隔离的应用研究[J]. 计算机与数字工程, 2009, 37(9): 159-161.

[4] 刘卿. 基于嵌入式 Linux 系统的隔离网闸的设计[D]. 北京: 北京科技大学, 2009.

作者简介:

程 军 男, (1976-), 博士, 高级工程师. 研究方向为网络安全威胁智能识别技术的研究.
E-mail: newhorizon@126. com.