用灰色优势分析确定网络安全评估指标

叶 李，王 娟，秦志光

(电子科技大学计算机科学与工程学院 成都 610054)

【摘要】针对如何确定网络安全评估指标的问题，在从NetFlow数据对象中提取关键域数据的基础上，将数据按相关维度进行多种形式的汇总分析，通过灰色关联优势分析的方法，计算网络安全事件与网络数据各种特征的灰色关联度，确定了体现网络安全事件的重点特征因素，提出了一系列评估网络安全的重点指标。实验结果表明，提取的指标能有效地反映网络的安全态势。

关 键 词 优势分析； 灰色关联； 网络安全； 网络安全事件
中图分类号 TP3 文献标识码 A

Net Safety Evaluation Index Based on Grey Advantage Analysis Method

YE Li，WANG Juan，QIN Zhi-Guang

(School of Computer Science and Engineering, University of Electronic Science and Technology of China Chengdu 610054)

Abstract Aiming to solve the problem of net safety evaluation index, the data extracted from NetFlow are analyzed. And the data are aggregated by several relational dimensions. Using the method of grey relation advantage analysis, the grey relation values of network safety affairs and the characters of the data are calculated. The pivot characters representing the affairs were decided. Then a set of network safety evaluation indices are proposed. The experiment results show that the indices can reflect the network situation efficiently.

Key words advantage analysis; grey relation; network security; network safety affairs

随着网络技术的发展，计算机病毒、网络入侵与攻击等各种网络安全事件给网络带来的威胁和危害越来越大，需对网络数据流进行特征分析，得出网络入侵、攻击和病毒的行为模式，以采取相应的预防措施。宏观网络的数据流日趋增大，其特征在很多方面都有体现。为了系统效率，只需对能体现网络安全事件发生程度与危害的重点特征进行分析，并得出反映网络安全事件的重点特征，形成安全评估指标。

NetFlow协议由Cisco公司开发，是一种实现网络层高性能交换的技术[1-2]。NetFlow协议的工作方式是通过网络中的交换设备采集所有当前经过的数据流，并存放于自身的缓存中，然后按一定的格式发送给指定的服务器。利用高性能设备的流缓存(Flow Cache)方式能很好地避免普通采集模式的丢包问题，保证数据采集的完整性。针对宏观网络，系统采用从NetFlow中提取网络特征的方法。

目前有通过相关分析、方差分析、回归分析、主成分分析等分析网络数据流和安全事件重点特征的数理统计方法。这些方法虽然能解决不少实际问题，但大都是较少涉及网络流因素的线性分析，对于网络流所体现出的多因素、非线性问题则难以处理。数理统计分析需要有大量的样本，且分布应服从或近似服从正态分布，在实际的网络安全事件发生程度中经常难以满足，但可通过对网络安全事件与众多网络特征因素建立关联矩阵，分析各因素之间的关系，从而确定影响网络安全的优势因素和非优势因素[3]。

灰色关联分析方法利用灰色系统模型，仅需要贫信息、小样本优势，就可以弥补采用数理统计分析所导致的缺陷。根据序列曲线几何形状的相似程度来判断两序列曲线之间联系是否紧密，曲线越接近，相应序列之间的关联度就越大，反之就越小。因此，通过灰色关联模型可以对影响一个抽象系统的众多因素进行分析，得出哪些是主要因素，哪些是次要因素；哪些因素对系统发展影响大，哪些因

收稿日期：2007– 09 – 07
基金项目：国家242计划(2006C27)
作者简介：叶 李(1977 –)，男，博士生，主要从事信息安全方面的研究.

素对系统发展影响小；哪些因素对系统发展起推动作用需强化发展，哪些因素对系统发展起阻碍作用需加以抑制。

通过灰色关联优势对NetFlow网络流量数据的网络行为进行特征分析，可确定其中最能体现网络安全事件的重点特征，形成安全评估指标，从理论上支持所选取的特征因素的合理与有效性，进而支持系统快速有效地反映宏观网络的安全态势。

1 相关技术

1.1 灰色关联与优势分析[3-4]

设Y_1，Y_2，…，Y_s为系统特征行为数据序列，$Y_i=(y_i(1),y_i(2),\cdots,y_i(n))$；$i\in(1,2,\cdots,s)$；$X_1$，$X_2$，…，$X_m$为相关因素行为序列，$X_j=(x_j(1),x_j(2),\cdots,x_j(n))$，$j\in(1,2,\cdots,m)$；$Y_i$与$X_j$长度相同。

对于分辨系数$\xi\in(0,1)$，给定实数$\gamma(y_i(k),x_j(k))$，定义为：

$$\gamma(y_i(k),x_j(k))=\frac{\min\limits_j\min\limits_k\left|y_i(k)-x_j(k)\right|+\xi\max\limits_j\max\limits_k\left|y_i(k)-x_j(k)\right|}{\left|y_i(k)-x_j(k)\right|+\xi\max\limits_j\max\limits_k\left|y_i(k)-x_j(k)\right|} \tag{1}$$

从而定义Y_i对X_j的灰色关联度记为γ_{ij}：

$$\gamma(Y_i,X_j)=\frac{1}{n}\sum_{k=1}^{n}\gamma(y_i(k),x_j(k)) \tag{2}$$

式中 $i\in(1,2,\cdots,s)$，$j\in(1,2,\cdots,m)$。计算得出的所有的γ_{ij}构成$s\times m$的灰色关联矩阵：

$$\boldsymbol{\Gamma}=(\gamma_{ij})=\begin{bmatrix}\gamma_{11}&\gamma_{12}&\cdots&\gamma_{1m}\\\gamma_{21}&\gamma_{22}&\cdots&\gamma_{2m}\\\vdots&\vdots&\vdots&\vdots\\\gamma_{s1}&\gamma_{s2}&\cdots&\gamma_{sm}\end{bmatrix}$$

此灰色关联矩阵中第i行的元素是系统特征数据序列$Y_i(i\in(1,2,\cdots,s))$与相关因素序列X_1，X_2，…，X_m的灰色关联度；第j列的元素是系统特征数据序列Y_1，Y_2，…，Y_s与$X_j(j\in(1,2,\cdots,m))$的灰色关联度。

若存在$l,j\in(1,2,\cdots,m)$，满足$\gamma_{il}\geq\gamma_{ij}$，$i\in(1,2,\cdots,s)$，则称因素$X_l$优于$X_j$，记为$X_l\succ X_j$。

若对任意的$j\in(1,2,\cdots,m)$，$j\neq l$，恒有$X_l\succ X_j$，则称X_l为最优因素。

一般情况下，可以直接通过式(1)计算得到的各相关因素序列与系统特征行为序列的灰色关联度值比较多个相关因素对系统影响的重要性。

1.2 NetFlow技术[5-6]

网络的流(Flow)是一个特定来源和目的端的单向数据报文序列。NetFlow以流作为数据统计的采集单位，也就是将来源IP、目地IP、来源Port、目地Port和传输协议五个属性相同的报文整合成一个流。NetFlow协议的核心是对流缓存进行组织、管理，最终可提供遵循某种汇聚方法而得到流的统计数据。NetFlow数据采集的工作原理是由路由器、交换机或者其他任何支持NetFlow 的软硬件维持一个保存流的统计数据的缓存，每一个活动的流在缓存中都占有一项记录。当一个不同于现有记录特征的数据包进入时，就自动地为该数据包在缓存中开辟新的流记录。后续进入缓存的数据包，如果和已有的记录具有相同的特征，其统计信息就会被加到相应的记录中去。NetFlow会不停地刷新缓存，将合适的记录移出缓存，并将所有被移出的记录聚合到UDP包中，发送给网络上指定的接收者。

NetFlow协议目前包括多个版本，版本之间的差异主要表现在对流采用的汇聚方法不同。以网络安全监控为目的而部署NetFlow，要求获得流的较多细节，因而常采用NetFlow V5。该版本所采集到的流量数据可以支持不同维度的统计分析。从采集数据的原始记录中提取以下八个关键域定义进一步分析，包括：(1) 源IP地址(Source IP Address)；(2) 目的IP地址(Destination IP address)；(3) 源端口号(Source Port Number)；(4)目的端口号(Destination Port Number)；(5) 协议类型(Layer 3 Protocol Type)；(6) 流内数据包数量(Packets)；(7) 数据流的大小(Octets)；(8) 数据流结束时间(Time)。

2 网络评估安全指标的确定

将系统所采集的NetFlow数据流和对应的监控平台报警信息数据按照固定时间段分片，分片后的不同类型的报警信息数据形成系统特征行为数据序列Y_i；从NetFlow中提取的分片数据按照不同维度汇总形成相关因素行为序列X_j。

由于存在有大量实时的NetFlow数据需要处理，所以具体汇总维度和对象根据数据仓库建设的方法进行设计，并考虑系统的后续扩展性，以达到性能与开发的最优结合[7]。

汇总的维度是多种可能影响网络安全行为的预定义的网络评估因素，汇总对象根据汇总维度也有不同。

汇总对象主要有：(1) 数据包字节数；(2) 数据流目的子网个数；(3) 数据流来源子网个数；(4) 数据包数量；(5) 源端口个数；(6) 目的端口个数；(7) 源IP个数；(8) 目的IP个数。

汇总维度主要有：(1) 子网中各种协议；(2) 子

网中数据包大小；(3) 数据流目的子网；(4) 数据流来源子网；(5) 源端口；(6) 目的端口。

在汇总维度(3)～(6)中，由于原始数据中子网的个数与开放端口数很多，所以采取TOPN分析方法，只提取前N个维度汇总后的数值，以减少系统处理的开销。

对NetFlow数据按恰当的时间分片，并对上述汇总对象在各不同维度汇总。形成相关因素行为序列X_j以后，将各相关因素行为序列X_j分别对系统特征行为数据序列Y_i进行灰色关联度计算，得出各网络评估因素对指定系统特征行为的灰色关联度，并进行比较排序，从而得出指定系统特征行为下影响网络安全的主要因素。

综合计算多种系统特征行为下网络评估因素的灰色关联度，即可确定网络安全事件主要的安全评估指标。

3 实　验

下面以DDoS攻击为例，描述重点特征因素确定的过程。

DDoS的攻击行为可以协调多台计算机上的进程发起攻击，在这种情况下，就会有一股拒绝服务的洪流冲击网络，可能使被攻击目标因过载而崩溃。

根据采集到的NetFlow数据和检测到的报警信息数据，按照时间分片构成原始数据集合，具体数值如表1所示。

表**1**　原始数据

分析时段	报警数	数据包个数	数据字节数	源端口数	目的端口数	源IP数	目的IP数
T_1	0	6 144	2 035 392	5 760	5 888	5 016	6 016
T_2	100	6 376	2 071 520	6 776	6 820	7 076	4 276
T_3	150	6 888	2 048 272	6 804	7 076	7 188	4 760
T_4	155	6 632	1 904 832	6 834	7 120	7 204	4 104
T_5	160	6 452	2 011 296	6 888	7 018	7 516	5 144
T_6	60	6 272	2 061 728	6 844	7 106	7 072	5 972
T_7	20	6 016	1 472 512	6 488	6 760	6 816	6 020
T_8	0	6 018	2 021 312	5 888	5 760	5 516	6 016
T_9	0	5 224	1 381 184	4 096	4 224	4 096	4 196

将原始数据集合进行无量纲化、归一化处理，并求取灰色关联矩阵$\boldsymbol{\Gamma}$，各相关因素的灰色关联度值如表2所示。

根据灰色关联矩阵得出各因素与系统特征的灰色关联度。由灰色关联矩阵可以看出，DDoS攻击发生后，检测到最大报警数为T_5的时刻，也即是最能体现系统特征行为的时刻，源IP个数、源端口个数、目的端口个数的灰色关联度在分析的几个因素中是最大的。

表**2**　各相关因素灰色关联度值

分析时段	数据包个数	数据字节数	源端口数	目的端口数	源IP数	目的IP数
T_1	0.359 2	0.337 3	0.374 2	0.376 8	0.428 3	0.364 1
T_2	0.624 5	0.571 4	0.582 2	0.600 3	0.612 4	0.991 6
T_3	0.888 8	0.907 0	0.908 6	0.898 8	0.963 7	0.669 8
T_4	0.988 3	0.910 4	0.955 3	0.941 2	0.979 9	0.572 7
T_5	0.887 6	0.945 1	1.000 0	0.972 1	1.000 0	0.663 8
T_6	0.482 8	0.446 3	0.447 0	0.445 2	0.469 1	0.504 0
T_7	0.400 5	0.460 5	0.379 7	0.377 5	0.390 1	0.400 3
T_8	0.364 0	0.338 8	0.369 1	0.382 0	0.405 2	0.364 0
T_9	0.397 3	0.428 5	0.456 8	0.457 3	0.478 5	0.450 8

而序列总的分析结果也表明，DDoS攻击的特征主要体现在源IP个数、源端口个数、目的端口个数方面，与DDoS攻击的实际情况相一致。

对更多的针对网络安全事件的分析实验表明，根据灰色关联优势分析确定网络安全评估指标是符合实际情况的，指标的评估值能有效反映网络的安全态势。

综合计算多种系统特征行为下网络相关因素的灰色关联度，从而确定网络安全事件主要的安全评估指标为：

(1) 子网带宽使用率；(2) 子网数据流入量；(3) 子网流入量增长率；(4) 子网内不同协议数据包的分布；(5) 子网内不同大小数据包的分布；(6) 流入子网内的数据包源IP分布；(7) 子网流量变化率；(8) 子网内不同协议数据包分布比值的变化率；(9) 子网内不同大小数据包分布比值的变化率；(10) 子网数据流总量；(11) 流出子网的数据包目的IP的分布。

4 结　论

本文在对国家骨干网络原始NetFlow数据流和对应网络监控平台报警数据按照时间序列分片处理的基础上，通过灰色关联优势分析方法，对网络数据流中的诸多网络安全事件所反映的特征因素进行分析，确定了主要的反映网络安全事件程度与危害的重点特征因素，从而在理论上支持了所选取的特征因素的合理与有效性，并且提高了系统的后续分析的效率，达到了有效反映宏观网络安全态势的目的。

(下转第1252页)

表4 论坛意见领袖

序号	节点	ΔL	发帖数	点出度	点入度
1	洪浩	0.904 584	3	16	33
2	坚决抵制	0.903 564	2	10	23
3	Zzzzzzzzza	0.902 865	2	14	18
4	Qinchenxv	0.902 166	1	8	16
5	Hrb一浦汪	0.902 136	1	6	16

从表4中可以看到，选取出的意见领袖都是网络群体中的核心人物，以成员1来说，他无疑是网络中最重要的人物，他的发帖引起了网络中大多数成员的特别关注，同时他也积极回复其他成员的帖子，与多名成员建立了交流关系，地位在其他意见领袖之上；其他几位意见领袖的发帖也都引起了群体成员的关注。另外，了解到这几位意见领袖都是论坛的资深会员，在过去几年中都发表了大量的帖子，引起论坛成员的广泛关注，知名度都很高，从上文中意见领袖特点来看，他们也完全符合意见领袖的特点。所以，这种基于小世界网络寻找意见领袖的方法是正确的，找出来的作者的确是论坛中的意见领袖。

4 总 结

网络的发展迅速地改变着人类这个真实的世界，而虚拟社区就是网络为人类提供的一个崭新空间和交往环境。基于虚拟社区的网络交往特征，人际互动关系是国内外学者关注的热点。本文对在线论坛进行了研究，构建出论坛中的社群网络，并对这个网络进行了特征分析，发现其具备小世界网络的特征，基于小世界网络本文找出了在线论坛中的意见领袖。后继的工作将分析网络成员在网络中的动力学行为，找出他们的网络特征，对复杂网络中人物行为分析研究领域做进一步的研究。

参 考 文 献

[1] OHSAWA Y. Chance discoveries for making decisions in complex real world[J]. New Generation Computing, 2002, 20(2): 143-164.
[2] WATTS D J, STROGATZ S H. Collective dynamics of ‘small-world’ networks[J]. Nature, 1998, 393: 440-442.
[3] WATTS D J. Small World[M]. Princeton: Princeton University Press, 1990.
[4] NEWMAN M E J. Scientific collaboration networks: II. Shortest paths, weighted networks and centrality[J]. Phys Rev E, 2001, 64(016132): 1-7.
[5] NEWMAN M E J. The structure and function of complex networks[J]. SIAM Review, 2003, 45(2): 167 -256.
[6] HOBBS J R. Information extraction from biomedical text[J]. Journal of Biomedical Informatics, 2002, 35(4): 260-264.
[7] 宫 辉, 徐 渝. 高校BBS社群结构与信息传播的影响因素[J]. 西安交通大学学报(社会科学版), 2007, 27(81): 93-96.
[8] ROGERS E M. Diffusion of innovations[M]. [S.l.]: The Free Press, 1962.
[9] KONGACHANDRA R, KIMPANT C, SUWANAPONQ T, et al. Newly-born keyword extraction under limited knowledge resources based on sentence similarity verification[C]// Communications and Information Technology (ISCIT 2004). Sapporo, Japan: IEEE, 2004: 1183-1187.
[10] 高俊波, 张 敏, 王煦法. 一种新的征兆发现算法研究[J]. 小型微型计算机系统, 2006, 27(4): 687-690.
[11] 邹 刚, 刘 洋, 刘 群, 等. 面向Internet的中文新词语检测[J]. 中文信息学报, 2004, 18(6): 1-9.

编 辑 税 红

(上接第1197页)

参 考 文 献

[1] Cisco Systems. Cisco NetFlow introduction[EB/OL]. http://www.cisco.com/warp/public/732/Tech/nmp/netflow/index.shtml, 2007-07-28.
[2] Cisco System. Netflow services solutions guide[EB/OL]. http://www.cisco.com/en/US/products/sw/netmgtsw/ps 1964 /products_implementation_design_guide09186a00800d6a11.html, 2007-07-30.
[3] 刘思峰, 党耀国, 张岐山. 灰色系统理论及其应用[M]. 第3版. 北京: 科学出版社, 2004.
[4] 肖新平, 宋中民, 李 峰. 灰技术基础及其应用[M]. 北京: 科学出版社, 2005.
[5] 杨 嵘, 张国清, 韦 卫, 等. 基于NetFlow流量分析的网络攻击行为发现[J]. 计算机工程, 2005, 31(13): 137-139.
[6] 熊齐邦, 黄明哲. 基于Netflow和异步服务的网络流量监测系统[J]. 计算机工程, 2006, 32(13): 144-146.
[7] INMON W H. 数据仓库[M]. 第4版. 王志海, 译. 北京: 机械工业出版社, 2006.
[8] MORAN J, GRANADA E, MIGUEZ J L, et al. Use of grey relational analysis to assess and optimize small biomass boilers[J]. Fuel Processing Technology, 2006, 87(2): 123-127.
[9] LIN S J, LU I J, LEWIS C. Grey relation performance correlations among economics, energy use and carbon dioxide emission in Taiwan[J]. Energy Policy, 2007, 35(3): 1948-1955.

编 辑 熊思亮