doi:10.3969/j.issn.1000-436x.2013.08.007

网络异常性指数的一种直推式定量计算方法

张永铮[1]，周勇林[2]，杜飞[1]
(1. 中国科学院 信息工程研究所，北京 100093；2. 国家计算机网络应急技术处理协调中心，北京 100029)

摘 要：针对网络异常性指数的计算问题，基于数量特征指数、成分特征指数、分布特征指数和模式特征指数提出了一种直推式定量计算方法——QCDP 法，通过应用真实网络流数据的 7 个实验验证了该方法的有效性。理论分析与实验结果表明：与传统的基于流量的直推式方法相比，QCDP 法能够更有效地反映出典型网络安全事件对宏观态势产生的影响；与归纳式方法相比，QCDP 法具有更好的客观性、实时性和实用性。
关键词：网络安全；宏观态势；异常性；指数；直推式
中图分类号：TP393.08 **文献标识码**：A **文章编号**：1000-436X(2013)08-0053-09

Transductive quantitative calculation approach of network abnormality index

ZHANG Yong-zheng[1], ZHOU Yong-lin[2], DU Fei[1]
(1. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China;
2. National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing 100029, China)

Abstract: For the problem of network abnormality index calculation, a transductive quantitative calculation approach named QCDP was proposed based on quantitative characteristics index, composition characteristics index, distribution characteristics index and pattern characteristics index. Seven experiments using real network traces were made to validate the effectiveness of QCDP. Theoretical analysis and experimental results show that, compared with the traditional transductive method based on traffic, the QCDP can more effectively reflect the macro situation of typical network security incidents; compared with the inductive methods, the QCDP has better objectivity, instantaneity and practicability.
Key words: network security; macro situation; abnormality; index; transduction

1 引言

互联网宏观安全态势感知是保障国家关键网络基础设施和重要信息系统安全的重要技术手段之一，已成为人们研究的热点，而安全指数作为反映和测度网络安全态势的一种核心方法，具有重要的理论意义和实际价值。为此，前期专门针对网络运行安全指数开展了较为系统的研究工作[1]，给出了网络运行安全指标、指数的新定义，提出了 10 维属性指数分类模型、层次式指数体系模型[2]、可用性指数及其定量计算方法[3]。同时也提出了异常性指数的概念，由于异常性指数可用于互联网总体异常性宏观态势的定量感知，因此，本文将重点针对异常性指数的定量计算问题展开讨论。

网络异常性指数的计算方法可总结为归纳式方法和直推式方法。

归纳式方法主要指以网络原始数据中检测或归纳出的相应安全事件、规则为依据的一类计算方法，该类方法从具体安全事件上间接地反映了网络安全态势，可充分利用已有安全资源，其指数的应

收稿日期：2013-05-02；**修回日期**：2013-06-16
基金项目：国家高技术研究发展计划(“863”计划)基金资助项目(2012AA012803, 2013AA014703)；国家科技支撑计划基金资助项目(2012BAH46B02)；中国科学院战略性科技先导专项基金资助项目(XDA06030200)
Foundation Items：The National High Technology Research and Development Program of China (863 Program)(2012AA012803, 2013AA014703); The National Science and Technology Support Program (2012BAH46B02); The Knowledge Innovation Program of the Chinese Academy of Sciences (XDA06030200)

用效果依赖于事件检测的准确性。现有工作主要集中在网络异常行为检测领域，具有代表性的工作包括：文献[4]基于改进的 TCM-KNN(transductive confidence machines for K-nearest neighbors)置信度机器学习算法，提出了一种网络异常检测的新方法；文献[5]提出了基于熵及熵之间相关系数的流量异常检测方法，比传统的基于流的异常检测粒度更细；文献[6]提出了利用随机森林树算法对网络流量进行分类，寻找异常值的方法；文献[7]提出了采用直方图和关联规则挖掘骨干网流量异常的方法；文献[8]提出了端到端定位异常事件的方法，使用 Holt-Winters 预测法预测下一周期的往返时间，通过对比与测量值的误差确定异常行为；文献[9]基于构造网络行为的有限状态机，通过状态转化关系识别异常行为；文献[10]提出了基于盲源分离技术的异常检测算法；文献[11]提出了基于非广度熵诊断网络流量异常的算法；文献[12]提出了针对网络性能的异常检测和定位的主动式探测算法框架，该框架包含 3 个算法，比传统的算法计算复杂度小。

直推式方法主要指以网络原始数据为依据的一类计算方法，该类方法尽量规避了主观因素和多重环节的处理，具有简单明了、客观性强、实时性强、计算复杂度低、实用性强等优势。现有工作主要为传统的流量计算法。

鉴于直推式方法的优势，本文将围绕指数的直推式计算问题展开研究。综上所述，尽管现有大量的归纳式方法对指标、指数的选取具有积极的借鉴意义，然而，从公开文献来看，针对异常性指数的直推式计算方法的研究工作很鲜见，现有方法虽简单易用但感知能力很差，为此，本文提出了一种针对网络异常性指数的直推式定量计算方法——QCDP 法，本方法的主要思想和技术特点为：以真实的网络流数据为计算依据，尽可能避免主观因素和多重评判，力图基于网络流数据通过直推式方法对异常性指数实施客观计算，以支撑互联网总体异常性的宏观态势感知。

2 基本概念

为方便后续讨论，分别给出指标、指数、异常性指数、基期和报告期等相关概念：

定义 1 网络运行安全指标(network operation security indicator)是指能够反映网络信息系统运行安全态势的网络数据特征的概念和数量，本文中简称为安全指标或指标。网络运行安全指标用于反映和度量网络信息系统在运行过程中的安全状态及其趋势。

定义 2 网络运行安全指数(network operation security index)是指能够反映网络信息系统运行安全态势的网络数据特征变化程度的相对数，本文中简称为安全指数或指数。网络运行安全指数用于反映和度量网络信息系统在运行过程中安全态势的变化量。

定义 3 基期(base period)和报告期(reporting period)。一个指数通常可由其对应指标在一个基准时期的指标数量和当前考察时期的指标数量计算而得，其中，将所选定的基准时期称为基期，将当前考察时期称为报告期。一般地，可选取网络相对稳定和安全的时期作为基期。

定义 4 网络异常性指数(NBI, network abnormality index)，是指用于反映由安全威胁或攻击所引起的网络通信数据特征异常程度的一类指数，该类指数通过网络数据异常反映和度量安全威胁或攻击行为及其对网络脆弱性的利用程度，例如流量指数、协议成分指数、IP 分布指数、端口分布指数、包间隔分布指数、汇聚模式指数等。

3 直推式计算方法

网络异常性指数将由数量特征指数、成分特征指数、分布特征指数和模式特征指数来组合计算。

3.1 数量特征指数

定义 5 数量特征指数(QCI, quantitative characteristics index)，是指用于反映网络数据在数量特征上的变化程度的一类指数，如流量指数等。其计算公式为

$$QCI(net)=\max\{TI(net),PI(net)\} \tag{1}$$

其中，*net* 表示网络对象，即由路由器组成的网络节点集合，*TI*(*net*)为网络对象 *net* 的流量指数(TI, traffic index)，*PI*(*net*)为网络对象 *net* 的分组速率指数（PI, packet rate index），其计算公式如式(2)和式(3)所示，max 表示取最大值函数。由于流量指数与分组速率指数存在一定程度的关联性，因此，用这 2 个指数的较大值计算数量特征指数。

$$TI(net)=\max\left\{\left.\frac{T_{t_1}(net)\big/N_{t_1}(net)}{T_{t_2}(net)\big/N_{t_2}(net)}\right|\begin{matrix}(t_1=r,t_2=b)\vee\\(t_1=b,t_2=r)\end{matrix}\right\} \tag{2}$$

$$PI(net)=\max\left\{\left.\frac{P_{t_1}(net)\big/N_{t_1}(net)}{P_{t_2}(net)\big/N_{t_2}(net)}\right|\begin{matrix}(t_1=r,t_2=b)\vee\\(t_1=b,t_2=r)\end{matrix}\right\} \quad (3)$$

其中，$T_t(net)$为 t 时期网络对象 net 的流量指标(NTI, network traffic indicator)，其计算公式如式(4)所示；$P_t(net)$为 t 时期网络对象 net 的分组速率指标(NPRI, network packet rate indicator)，其计算公式如式(5)所示；$N_t(net)$为 t 时期网络对象 net 的网络节点数指标；r 表示报告期，b 表示基期。

$$T_t(net)=\sum\nolimits_{i\in subnet\subseteq net}T_t(i)\left(t=r\vee b\right) \quad (4)$$

$$P_t(net)=\sum\nolimits_{i\in subnet\subseteq net}P_t(i)\left(t=r\vee b\right) \quad (5)$$

其中，$subnet$ 为网络对象 net 的子集，表示在指数计算中所选取的网络节点(即流数据采集点)的集合，i 表示所选取的网络节点；$T_t(i)$表示 t 时期节点 i 的网络流量指标，单位为 bit/s；$P_t(i)$表示 t 时期节点 i 的网络分组速率指标，单位为 packet/s。根据前文所述，计算 $T_t(i)$可由 t 时间内各个路由器节点的流记录字节数累加和的统计结果得出；$P_t(i)$可由 t 时间内各个路由器节点的流记录分组数累加和的统计结果来表示。

3.2　成分特征指数

定义 6　成分特征指数(CCI, composition characteristics index)，是指用于反映同度量因素下网络数据在比例特征上的变化程度的一类指数，如协议成分指数等。其计算公式为

$$\begin{aligned}CCI(net)=&\max\left\{ICI(net),TCI(net),UCI(net)\right\}+\\&\max\left\{SCI(net),RCI(net)\right\}+\\&OCI(net)\end{aligned} \quad (6)$$

其中，$ICI(net)$为网络对象 net 的 ICMP 协议成分指数(ICI, ICMP composition index)，$TCI(net)$为网络对象 net 的 TCP 协议成分指数(TCI, TCP composition index)，$UCI(net)$为网络对象 net 的 UDP 协议成分指数(UCI, UDP composition index)，$SCI(net)$为网络对象 net 的 TCP_SYN 成分指数(SCI, TCP_SYN composition index)，$RCI(net)$为网络对象 net 的 TCP_RST 成分指数(RCI, TCP_RST composition index)，$OCI(net)$为网络对象 net 的端口成分指数(OCI, port composition index)，其计算公式分别如公式(7)~式(12)所示。在相同报文数的前提下，由于 TCP、UDP、ICMP 协议之间以及 TCP_SYN 报文与 TCP_RST 报文之间都存在一定程度的关联性，因此同样采用取大值的方法进行指数的综合计算。

$$ICI(net)=\max\left\{\left.\frac{ICN_{t_1}(net)\big/N_{t_1}(net)}{ICN_{t_2}(net)\big/N_{t_2}(net)}\right|\begin{matrix}(t_1=r,t_2=b)\vee\\(t_1=b,t_2=r)\end{matrix}\right\} \quad (7)$$

$$TCI(net)=\max\left\{\left.\frac{TCN_{t_1}(net)\big/N_{t_1}(net)}{TCN_{t_2}(net)\big/N_{t_2}(net)}\right|\begin{matrix}(t_1=r,t_2=b)\vee\\(t_1=b,t_2=r)\end{matrix}\right\} \quad (8)$$

$$UCI(net)=\max\left\{\left.\frac{UCN_{t_1}(net)\big/N_{t_1}(net)}{UCN_{t_2}(net)\big/N_{t_2}(net)}\right|\begin{matrix}(t_1=r,t_2=b)\vee\\(t_1=b,t_2=r)\end{matrix}\right\} \quad (9)$$

$$SCI(net)=\max\left\{\left.\frac{SCN_{t_1}(net)\big/N_{t_1}(net)}{SCN_{t_2}(net)\big/N_{t_2}(net)}\right|\begin{matrix}(t_1=r,t_2=b)\vee\\(t_1=b,t_2=r)\end{matrix}\right\} \quad (10)$$

$$RCI(net)=\max\left\{\left.\frac{RCN_{t_1}(net)\big/N_{t_1}(net)}{RCN_{t_2}(net)\big/N_{t_2}(net)}\right|\begin{matrix}(t_1=r,t_2=b)\vee\\(t_1=b,t_2=r)\end{matrix}\right\} \quad (11)$$

$$OCI(net)=\max\left\{\left.\frac{OCN_{t_1}(net)\big/N_{t_1}(net)}{OCN_{t_2}(net)\big/N_{t_2}(net)}\right|\begin{matrix}(t_1=r,t_2=b)\vee\\(t_1=b,t_2=r)\end{matrix}\right\} \quad (12)$$

其中，$ICN_t(net)$为 t 时期网络对象 net 的 ICMP 协议成分指标(ICN, ICMP composition indicator)，其计算公式如式(13)所示；$TCN_t(net)$为 t 时期网络对象 net 的 TCP 协议成分指标(TCN, TCP composition indicator)，其计算公式如式(14)所示；$UCN_t(net)$为 t 时期网络对象 net 的 UDP 协议成分指标(UCN, UDP composition indicator)，其计算公式如式(15)所示；$SCN_t(net)$为 t 时期网络对象 net 的 TCP_SYN 成分指标(SCN, TCP_SYN composition indicator)，其计算公式如式(16)所示；$RCN_t(net)$为 t 时期网络对象 net 的 TCP_RST 成分指标(RCN, TCP_RST composition

indicator)，其计算公式如式(17)所示；$OCN_t(net)$为 t 时期网络对象 net 的端口成分指标(OCN, pOrt composition indicator)，其计算公式如式(18)所示。

$$ICN_t(net)=\sum_{i\in subnet\subseteq net}ICN_t(i)=\sum_{i\in subnet\subseteq net}\frac{IM_t(i)}{AP_t(i)}(t=r\vee b) \quad (13)$$

$$TCN_t(net)=\sum_{i\in subnet\subseteq net}TCN_t(i)=\sum_{i\in subnet\subseteq net}\frac{TC_t(i)}{AP_t(i)}(t=r\vee b) \quad (14)$$

$$UCN_t(net)=\sum_{i\in subnet\subseteq net}UCN_t(i)=\sum_{i\in subnet\subseteq net}\frac{UD_t(i)}{AP_t(i)}(t=r\vee b) \quad (15)$$

$$SCN_t(net)=\sum_{i\in subnet\subseteq net}SCN_t(i)=\sum_{i\in subnet\subseteq net}\frac{TS_t(i)}{TC_t(i)}(t=r\vee b) \quad (16)$$

$$RCN_t(net)=\sum_{i\in subnet\subseteq net}RCN_t(i)=\sum_{i\in subnet\subseteq net}\frac{TR_t(i)}{TC_t(i)}(t=r\vee b) \quad (17)$$

$$OCN_t(net)=\sum_{i\in subnet\subseteq net}OCN_t(i)=\sum_{i\in subnet\subseteq net}\max\left\{\frac{ON_t^j(i)}{AP_t(i)}\middle| j=1,\cdots,m\right\}(t=r\vee b) \quad (18)$$

其中，$IM_t(i)$表示 t 时期节点 i 的 ICMP 分组数指标，$AP_t(i)$表示 t 时期节点 i 的总分组数指标，$TC_t(i)$表示 t 时期节点 i 的 TCP 分组数指标，$UD_t(i)$表示 t 时期节点 i 的 UDP 分组数指标，$TS_t(i)$表示 t 时期节点 i 的 TCP_SYN(标记比特只为 SYN 的 TCP 分组)分组数指标，$TR_t(i)$表示 t 时期节点 i 的 TCP_RST(标记位只为 RST 的 TCP 分组)分组数指标。$ON_t^j(i)$ 表示将 2^{16} 个端口号分成 m 段后属于第 j 段端口内的分组数指标，例如当 m=3 时，0~1 024 为第一段，1 025~9 999 为第 2 段，10 000~65 535 为第 3 段。

3.3 分布特征指数

定义 7 分布特征指数(DCI, distribution characteristics index)，是指用于反映网络数据在统计分布特征上的变化程度的一类指数，如 IP 分布指数等。其计算公式为

$$DCI(net)=ALI(net)+\max\{SII(net),DII(net)\}+\max\{SPI(net),DPI(net)\} \quad (19)$$

其中，$ALI(net)$表示网络对象 net 的流平均分组长分布指数（ALI, flow average packet length distribution index），$SII(net)$表示网络对象 net 的源 IP 分布指数（SII, source IP distribution index），$DII(net)$表示网络对象 net 的目的 IP 分布指数（DII, destination IP distribution index），$SPI(net)$表示网络对象 net 的源端口分布指数（SPI, source port distribution index），$DPI(net)$表示网络对象 net 的目的端口分布指数（DPI, destination port distribution index），其计算公式分别如式(20)~式(24)所示。

$$ALI(net)=\max\left\{\frac{ALN_{t_1}(net)\big/N_{t_1}(net)}{ALN_{t_2}(net)\big/N_{t_2}(net)}\middle|\begin{matrix}(t_1=r,t_2=b)\vee\\(t_1=b,t_2=r)\end{matrix}\right\} \quad (20)$$

$$SII(net)=\max\left\{\frac{SIN_{t_1}(net)\big/N_{t_1}(net)}{SIN_{t_2}(net)\big/N_{t_2}(net)}\middle|\begin{matrix}(t_1=r,t_2=b)\vee\\(t_1=b,t_2=r)\end{matrix}\right\} \quad (21)$$

$$DII(net)=\max\left\{\frac{DIN_{t_1}(net)\big/N_{t_1}(net)}{DIN_{t_2}(net)\big/N_{t_2}(net)}\middle|\begin{matrix}(t_1=r,t_2=b)\vee\\(t_1=b,t_2=r)\end{matrix}\right\} \quad (22)$$

$$SPI(net)=\max\left\{\frac{SPN_{t_1}(net)\big/N_{t_1}(net)}{SPN_{t_2}(net)\big/N_{t_2}(net)}\middle|\begin{matrix}(t_1=r,t_2=b)\vee\\(t_1=b,t_2=r)\end{matrix}\right\} \quad (23)$$

$$DPI(net)=\max\left\{\frac{DPN_{t_1}(net)\big/N_{t_1}(net)}{DPN_{t_2}(net)\big/N_{t_2}(net)}\middle|\begin{matrix}(t_1=r,t_2=b)\vee\\(t_1=b,t_2=r)\end{matrix}\right\} \quad (24)$$

其中，$ALN_t(net)$为 t 时期网络对象 net 的流平均分组长分布指标(ALN, flow average packet length distribution indicator)，其计算公式如式(25)所示；$SIN_t(net)$为 t 时期网络对象 net 的源 IP 分布指标(SIN, source ip distribution indicator)，其计算公式如式(26)所示；$DIN_t(net)$为 t 时期网络对象 net 的目的 IP 分布指标(DIN, destination IP distribution indicator)，其计算公式如式(27)所示；$SPN_t(net)$为 t 时期网络对象 net 的源端口分布指标(SPN, source port distribution indicator)，其计算公式如式(28)所示；$DPN_t(net)$为 t 时期网络对象 net 的目的端口分布指标(DPN, destination port distribution indica-

tor)，其计算公式如式(29)所示。

$$ALN_t(net)=\sum_{i\in subnet\subseteq net}ALN_t(i)$$
$$=\sum_{i\in subnet\subseteq net}\left(\begin{array}{l}\ln\sum_{j=1}^{n}AL_{jt}(i)-\\ \dfrac{\sum_{j=1}^{n}AL_{jt}(i)\times\ln AL_{jt}(i)}{\sum_{j=1}^{n}AL_{jt}(i)}\end{array}\right)(t=r\vee b)\quad(25)$$

其中，$AL_{jt}(i)(j=1\sim n)$表示在 t 时期节点 i 的具有相同平均分组长的流数指标，n 为不同平均分组长的个数。

$$SIN_t(net)=\sum_{i\in subnet\subseteq net}SIN_t(i)$$
$$=\sum_{i\in subnet\subseteq net}\left(\begin{array}{l}\ln\sum_{j=1}^{n}SI_{jt}(i)-\\ \dfrac{\sum_{j=1}^{n}SI_{jt}(i)\times\ln SI_{jt}(i)}{\sum_{j=1}^{n}SI_{jt}(i)}\end{array}\right)(t=r\vee b)\quad(26)$$

其中，$SI_{jt}(i)(j=1\sim n)$表示在 t 时期节点 i 的具有相同源 IP 地址的流数指标，n 为不同源 IP 地址的个数。

$$DIN_t(net)=\sum_{i\in subnet\subseteq net}DIN_t(i)$$
$$=\sum_{i\in subnet\subseteq net}\left(\begin{array}{l}\ln\sum_{j=1}^{n}DI_{jt}(i)-\\ \dfrac{\sum_{j=1}^{n}DI_{jt}(i)\times\ln DI_{jt}(i)}{\sum_{j=1}^{n}DI_{jt}(i)}\end{array}\right)(t=r\vee b)\quad(27)$$

其中，$DI_{jt}(i)(j=1\sim n)$表示在 t 时期节点 i 的具有相同目的 IP 地址的流数指标，n 为不同目的 IP 地址的个数。

$$SPN_t(net)=\sum_{i\in subnet\subseteq net}SPN_t(i)$$
$$=\sum_{i\in subnet\subseteq net}\left(\begin{array}{l}\ln\sum_{j=1}^{n}SP_{jt}(i)-\\ \dfrac{\sum_{j=1}^{n}SP_{jt}(i)\times\ln SP_{jt}(i)}{\sum_{j=1}^{n}SP_{jt}(i)}\end{array}\right)(t=r\vee b)\quad(28)$$

其中，$SP_{jt}(i)(j=1\sim n)$表示在 t 时期节点 i 的具有相同源端口的流数指标，n 为不同源端口的个数。

$$DPN_t(net)=\sum_{i\in subnet\subseteq net}DPN_t(i)$$
$$=\sum_{i\in subnet\subseteq net}\left(\begin{array}{l}\ln\sum_{j=1}^{n}DP_{jt}(i)-\\ \dfrac{\sum_{j=1}^{n}DP_{jt}(i)\times\ln DP_{jt}(i)}{\sum_{j=1}^{n}DP_{jt}(i)}\end{array}\right)(t=r\vee b)\quad(29)$$

其中，$DP_{jt}(i)(j=1\sim n)$表示在 t 时期节点 i 的具有相同目的端口的流数指标，n 为不同目的端口的个数。本论文采用已公开的信息熵[13]的方法计算所述分布指标。

3.4　模式特征指数

定义 8　模式特征指数(PCI, pattern characteristics index)是指用于反映网络数据在行为模式特征上的变化程度的一类指数，如同源 IP 同目的 IP 模式指数等。其计算公式为

$$PCI(net)=STPI(net)+SDPI(net)+DDPI(net)\quad(30)$$

其中，$STPI(net)$为网络对象 net 的同源 IP 同目的端口模式指数（STPI, SIP-DPORT pattern index），$SDPI(net)$为网络对象 net 的同源 IP 同目的 IP 模式指数（SDPI, SIP-DIP pattern index），$DDPI(net)$为网络对象 net 的不同源 IP 同目的 IP 模式指数（DDPI, DSIP-DIP pattern index），其计算公式分别如式(31)~式(33)所示。

$$STPI(net)=\max\left\{\frac{STPN_{t_1}(net)\big/N_{t_1}(net)}{STPN_{t_2}(net)\big/N_{t_2}(net)}\middle|\begin{array}{l}(t_1=r,t_2=b)\vee\\(t_1=b,t_2=r)\end{array}\right\}\quad(31)$$

$$SDPI(net)=\max\left\{\frac{SDPN_{t_1}(net)\big/N_{t_1}(net)}{SDPN_{t_2}(net)\big/N_{t_2}(net)}\middle|\begin{array}{l}(t_1=r,t_2=b)\vee\\(t_1=b,t_2=r)\end{array}\right\}\quad(32)$$

$$DDPI(net)=\max\left\{\frac{DDPN_{t_1}(net)\big/N_{t_1}(net)}{DDPN_{t_2}(net)\big/N_{t_2}(net)}\middle|\begin{array}{l}(t_1=r,t_2=b)\vee\\(t_1=b,t_2=r)\end{array}\right\}\quad(33)$$

其中，$STPN_t(net)$为 t 时期网络对象 net 的同源 IP 同目的端口模式指标(STPN, SIP-DPORT pattern indicator)，其计算公式如式(34)所示；$SDPN_t(net)$为 t 时期网络对象 net 的同源 IP 同目的 IP 模式指标(SDPN, SIP-DIP pattern indicator)，其计算公式如式(35)所示；$DDPN_t(net)$为 t 时期网络对象 net 的不同源 IP 同目的 IP 模式指标(DDPN, DSIP-DIP pattern indicator)，其计算公式如式(36)所示。

$$STPN_t(net) = \sum_{i \in subnet \subseteq net} STPN_t(i) = \sum_{i \in subnet \subseteq net} \sum_{j=1}^{p} STP_{jt}(i)(t = r \vee b) \quad (34)$$

其中，具有相同源 IP 地址和相同目的端口的流属于同一个模式，$STP_{jt}(i)$表示在 t 时期节点 i 流数为第 j 多的模式的流数指标。该模式一般用于度量水平扫描行为。

$$SDPN_t(net) = \sum_{i \in subnet \subseteq net} SDPN_t(i) = \sum_{i \in subnet \subseteq net} \sum_{j=1}^{p} SDP_{jt}(i)(t = r \vee b) \quad (35)$$

其中，具有相同源 IP 地址和相同目的 IP 地址的流属于同一个模式，$SDP_{jt}(i)$表示在 t 时期节点 i 流数为第 j 多的模式的流数指标。该模式一般用于度量垂直扫描行为。

$$DDPN_t(net) = \sum_{i \in subnet \subseteq net} DDPN_t(i) = \sum_{i \in subnet \subseteq net} \sum_{j=1}^{p} DDP_{jt}(i)(t = r \vee b) \quad (36)$$

其中，具有不同源 IP 地址和相同目的 IP 地址的流属于同一个模式，$DDP_{jt}(i)$表示在 t 时期节点 i 流数为第 j 多的模式的流数指标。该模式一般用于度量分布式拒绝服务攻击行为。

3.5 异常性指数的定量计算

基于上文所述，给出网络异常性指数的计算公式，如式(37)所示，为方便阐述，将该直推式定量计算方法称为 QCDP 法。

$$NBI(net) = QCI(net) + CCI(net) + DCI(net) + PCI(net) \quad (37)$$

通过对网络异常性指数的定义及其定量计算方法的分析，可以看出：1）如果选取网络正常行为的一个具有代表性的时期作为基期，那么指数的大小可以反映出网络异常性的变化程度，即能够刻画出网络在异常性上的安全态势；2）以基期的网络行为作为统一的评价标准，异常性指数能够度量和评价不同安全事件对网络异常性的安全影响；3）异常性指数能够从宏观上反映网络的异常性随时间的演化和发展趋势。

4 实验分析

4.1 实验设计

为了进一步验证异常性指数及其定量计算方法的有效性和优势，分别针对正常行为、突发群体性(crowd)访问、UDP 洪泛分布式拒绝服务(DDoS, distributed denial of service)攻击、同步洪泛分布式拒绝服务攻击(SYN Flood DDoS)攻击、蠕虫传播、垂直扫描等 6 种典型情景设计了 7 个实验。

软硬件环境：采用曙光服务器，4 个 CPU (Dual-Core AMD Opteron，2 211 MHz，64 bit)，2 GB 内存，CentOS Linux 5.2 64 bit 操作系统。

实验数据：不失一般性，本实验采用 NetFlow[14] 流数据，选取某运营商某省 3 个出口路由器节点上某年 5 月 15 日至 21 日一周的 NetFlow 流数据。

参数设置：不失一般性，选定其中 5 月 15 日的每一小时为基期 b，随意选定 5 月 17 日的相应时间段为报告期 r。为使实验更具说服力，实验 2~6 均选取 b 为基期，以 5 月 17 日的网络流数据为背景流量。此外，$N_t(net)=3(t=r \vee b)$，模式特征指数中 p=20。

SYN Flood DDoS 攻击流量构造方法：在插入时段内每遇到 1 条 TCP 流则添加 5 条攻击流，其中攻击流的源 IP 地址随机产生，目的 IP 地址为此期间出现过的若干真实目的 IP 地址，分组数为 1，字节数为 100，标志比特为 SYN，其他信息与背景流数据一致。

UDP Flood DDoS 攻击流量构造方法：在插入时段内每遇到 1 条 UDP 流则添加 1 条攻击流，其中攻击流的源 IP 地址随机产生，目的 IP 地址为此期间出现过的若干真实目的 IP 地址，字节数为 1 500 与该 UDP 流的分组数的乘积，其他信息与背景流数据一致。

Crowd 访问流量构造方法：在插入时段内每 5 条流添加 10 条访问流，其中访问流的源 IP 地址为该环境下出现过的真实 IP 地址，目的 IP 地址为此期间出现过的若干真实目的 IP 地址，分组数为 1，字节数为 120，其他信息与背景流数据一致。

蠕虫扫描流量构造方法：在插入时段内每遇到 1 条 TCP 流则添加 1 条扫描流，其中扫描流的源 IP 地址为此期间出现过的若干真实源 IP 地址，目的 IP 地址随机产生，分组数为 1，字节数为 100，标志位为 SYN，目的端口为 80，其他信息与背景流数据一致。

垂直扫描流量构造方法：在插入时段内每遇到 1 条 TCP 流则添加以该流的目的地址为被扫描目的地址，目的端口为常用 185 个端口的 185 条扫描流，其中扫描流的源 IP 地址为此期间出现过的若干真实源 IP 地址，分组数为 1，字节数为 100，标志位 SYN，源端口是 2 000 到 65 535 之间随机产生，其他信息与背景流数据一致。

上述攻击构造方法均符合安全实践中获知的典型攻击原理。

4.2　传统流量法

实验 1　基于传统流量法的异常性指数对各种安全事件的反应情况。

选取 5 月 17 日 24 h 的 NetFlow 流数据为背景流量，在 7 时针对 3 个目的 IP 融入构造的随机伪造源地址的 SYN Flood DDoS 攻击流量；在 10 时针对 3 个目的 IP 融入构造的随机伪造源地址的 UDP DDoS 攻击流量，在 13 时针对 3 个目的 IP 融入构造的 Crowd 访问流量，在 16 时针对 3 个源 IP 融入构造的蠕虫扫描流量，在 19 时针对 3 个源 IP 融入构造的垂直扫描流量。实验结果如图 1 所示。

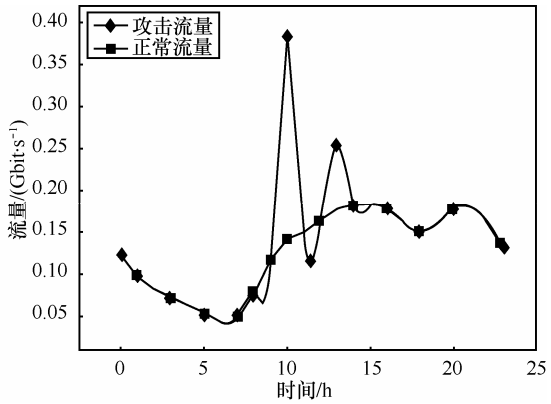


图 1　基于流量法的异常性指数

通过图 1 可以看出，传统流量法对 UDP 分布式拒绝服务攻击的反应强烈，对群体性访问情况的反应较强烈，而对同步洪泛攻击、蠕虫扫描、垂直扫描等安全事件的反应几乎看不出来。

4.3　QCDP 法

实验 2　正常行为情况下网络异常性指数的计算。

实验结果如图 2 中实线所示，正常情况下网络异常性指数呈直线状态，在 10 左右，说明正常情况下网络呈现高度自相似性。

实验 3　Crowd 访问情况下网络异常性指数的计算。

为了与实验 2 进行对比，以 17 日（报告期）的流数据为背景流量，在此流量中融入构造的 Crowd 访问流量。不失一般性，选取了 17 日的 7:00、9:00、11:00 这 3 个小时段作为插入段，分别针对 2 个、1 个和 3 个目的 IP。

实验结果如图 2 中虚线所示，通过与正常行为情况下的曲线对比可以看出：Crowd 访问发生时段网络异常性指数较网络正常行为情况增加 3 倍左右，约 40。

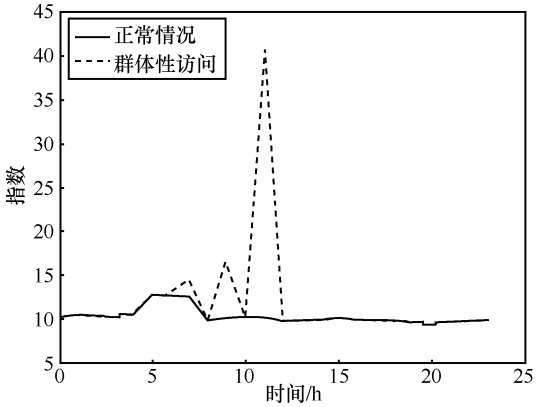


图 2　群体性访问下的网络异常性指数

实验 4　SYN Flood DDoS 攻击情况下网络异常性指数的计算。

同理，为与其他实验进行对比，仍以 17 日（报告期）的流数据为背景流量，在此流量中融入构造的随机伪造源地址的 SYN Flood DDoS 攻击流量，并将融合后的流数据作为报告期的流数据。不失一般性，选取了 17 日的 7:00、9:00、11:00 这 3 个小时段作为插入时段，分别针对 2 个、1 个和 3 个目的 IP。

实验结果如图 3 所示，通过对比可以看出，SYN Flood 攻击发生时段网络异常性指数较网络正常行为情况增加 6 倍以上，在 70 以上。

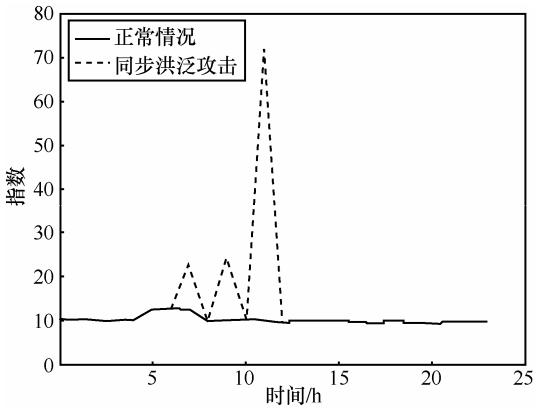


图 3　SYN Flood DDoS 下的网络异常性指数

实验 5　UDP Flood DDoS 攻击情况下网络异常性指数的计算。

同理，在背景流量中融入构造的随机伪造源地址的 UDP DDoS 攻击流量，并将融合后的流数据作为报告期的流数据。不失一般性，选取了 17 日的 7:00、

9:00、11:00 这 3 个小时段作为插入时段，分别针对 2 个、1 个和 3 个目的 IP。

实验结果如图 4 所示，通过对比可以看出，UDP DDoS 攻击发生时段网络异常性指数较网络正常行为情况增加 2 倍左右，约 30。

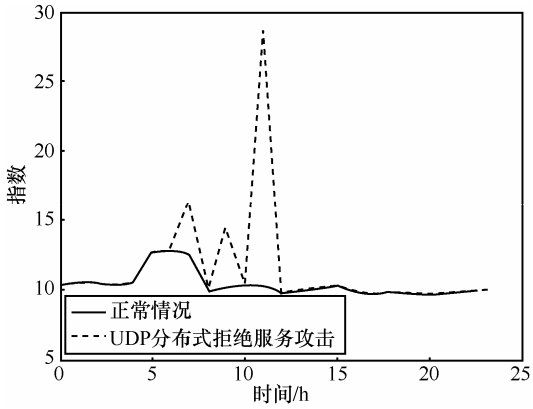


图 4　UDP Flood DDoS 下的网络异常性指数

实验 6　蠕虫传播情况下网络异常性指数的计算。

同理，在背景流量中融入构造的蠕虫扫描流量，并将融合后的流数据作为报告期的流数据。不失一般性，选取了 17 日的 7:00 2 个源 IP、9:00 1 个源 IP、11:00 这 3 个源 IP 进行蠕虫扫描。

实验结果如图 5 所示，通过对比可以看出，蠕虫扫描发生时段网络异常性指数较网络正常行为情况增加 9 倍左右，约 100。

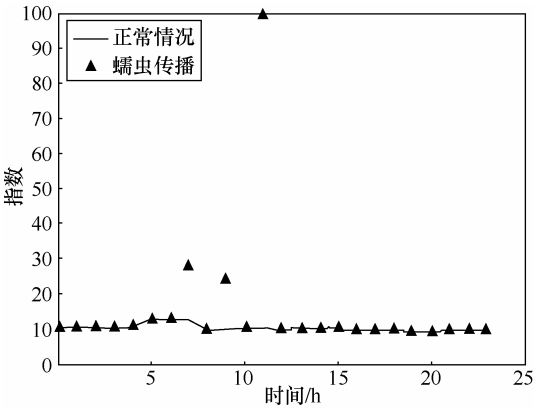


图 5　蠕虫扫描下的网络异常性指数

实验 7　垂直扫描情况下网络异常性指数的计算。

同理，在背景流量中融入构造的垂直扫描流量，并将融合后的流数据作为报告期的流数据。不失一般性，选取了 17 日的 7:00 2 个源 IP、9:00 1 个源 IP、11:00 3 个源 IP 进行垂直扫描。

实验结果如图 6 所示，通过对比可以看出，垂直扫描发生时段网络异常性指数较网络正常行为情况增加 4 倍以上，在 50 以上。

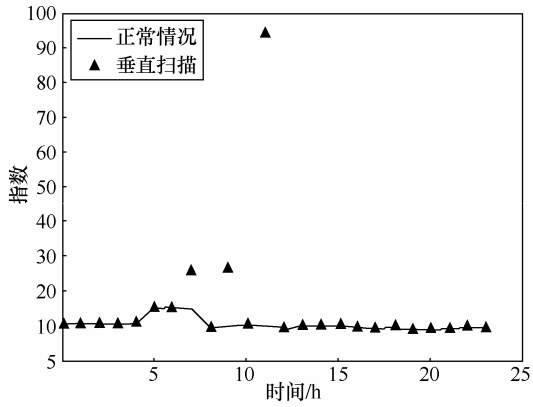


图 6　垂直扫描下的网络异常性指数

综上所述，较传统流量法，本文提出的 QCDP 法能够更有效地反映出典型网络安全事件对网络安全宏观态势产生的影响，且通过指数的对比和计算能够评价和度量不同安全事件对网络的不同影响。

5　讨论

从安全态势感知技术手段的角度，不难看出，异常性指数方法与传统的网络异常检测方法存在着相同之处：都选取适当的网络数据特征进行检测或计算以反映网络的异常性，但还存在以下本质区别：1）所反映网络安全态势的层面不同，“网络异常检测”主要面向微观安全态势，重在识别出异常事件；而“网络异常性指数”主要面向宏观安全态势，重在反映和度量态势变化；2）反映网络安全态势所采用的依据不同，“网络异常检测”以检测出的异常事件的数量间接地反映态势；而“网络异常性指数计算”以计算出的指数直接地反映态势，并不去挖掘其中是否存在异常事件或存在多少异常事件；3）关于异常的计算方法不同，“网络异常检测”一般基于阈值或数据挖掘、人工智能等各类模型方法，对是否存在异常事件进行判定，一般需要学习正常行为轮廓；而本文提出的 QCDP 法基于不同时期数据特征的相对数及其累加和的方法来计算异常性指数，并不判定是否存在异常事件，只通过客观数据表现出这些特征的异常程度，也不需要学习正常行为轮廓。

6　结束语

针对网络异常性指数的计算问题，本文基于数量特征指数、成分特征指数、分布特征指数和模式特征指数提出了一种直推式定量计算方法——QCDP 法，通过应用真实流数据的 7 个实验验证了该方法的有效性。理论分析与实验结果表明：1）与传统的基于流量的直推式方法相比，QCDP 法能够更有效地反映出典型网络安全事件对网络安全宏观态势产生的影响，且通过指数的对比和计算能够评价和度量不同安全事件对网络的不同影响；2）与归纳式方法相比，QCDP 法具有简单明了、客观性强、实时性强、计算复杂度低、实用性强等优势。

参考文献：

[1] 张永铮，云晓春. 网络运行安全指数多维属性分类模型[J]. 计算机学报, 2012, 35(8): 1666-1674.
ZHANG Y Z,YUN X C. Network operation security index classification model with multidimensional attributes[J]. Chinese Journal of Computers, 2012, 35(8): 1666-1674.

[2] ZHANG Y Z, YUN X C. Modeling of hierarchical index system for network operation security[A]. ISCTCS[C]. Beijing, China, 2013. 580-590.

[3] HE Y H, ZHANG Y Z. Network availability index and its flow-based quantitative calculation method[A]. ICCET[C]. Chengdu, China, 2010.551-555.

[4] 李洋, 方滨兴, 郭莉等. 基于直推式方法的网络异常检测方法[J]. 软件学报, 2007, 18(10):2595-2604.
LI Y, FANG B X, GUO L, *et al*. A network anomaly detection method based on transduction scheme[J]. Journal of Software, 2007,18(10): 2595-2604.

[5] NYCHIS G, SEKAR V, ANDERSEN D G, *et al*. An empirical evaluation of entropy-based traffic anomaly detection[A]. IMC[C]. Vouliagmeni, Greece, 2008. 151-156.

[6] ZHANG J, ZULKERNINE M. Anomaly based network intrusion detection with unsupervised outlier detection[A]. ICC[C]. Istanbul, Turkey, 2006. 2388-2393.

[7] BRAUCKHOFF D, DIMITROPOULOS X, WAGNER A, *et al*. Anomaly extraction in backbone networks using association rules[A]. SIGCOMM[C]. Chicago, Illinois, USA, 2009. 28-34.

[8] YAN H, FLAVEL A, GE Z, *et al*. Argus: end-to-end service anomaly detection and localization from an ISP's point of view[A]. INFOCOM[C]. Orlando, FL, USA, 2012. 2756-2760.

[9] HSIAO S W, SUN Y S, CHEN M C, *et al.* Behavior profiling for robust anomaly detection[A]. WCNIS[C]. Beijing, China, 2010. 465-471.

[10] QIN T, GUAN X, LI W, *et al*. Monitoring abnormal network traffic based on blind source separation approach[J]. Journal of Network and Computer Applications, 2011, 34(5):1732-1742.

[11] ZIVIANI A, GOMES A T A, MONSORES M L, *et al*. Network anomaly detection using nonextensive entropy[J]. IEEE Communications Letters, 2007, 11(12):1034-1036.

[12] BARFORD P, DUFFIELD N, RON A, *et al*. Network performance anomaly detection and localization[A]. INFOCOM[C]. Rio de Janeiro, Brazil, 2009. 1377-1385

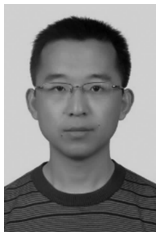
[13] SHANNON C E. A mathematical theory of communication[J]. Bell System Technical Journal, 1948,27(3):379-423.

[14] Cisco netflow[EB/OL]. http://www.cisco.com/en/US/products/ps6601/products_ios_protocol_ group_home.html, 2013.

作者简介：

张永铮（1978-），男，黑龙江哈尔滨人，博士，中国科学院副研究员、博士生导师，主要研究方向为网络安全态势感知。

周勇林[通信作者]（1974-），男，辽宁抚顺人，国家计算机网络应急技术处理协调中心高级工程师，主要研究方向为互联网安全监测、应急响应处理。E-mail: zyl@cert.org.cn。



杜飞（1982-），男，山西太原人，硕士，中国科学院工程师，主要研究方向为网络安全。