企业网络资产及安全事件分析与可视化∗

飞向月球

张发  
 西安交通大学，学生/网络空间安全，研二  
西安交通大学  
 中国-西安  
 724183028@qq.com

查志超  
西安交通大学，学生/软件工程 ，研二  
 西安交通大学  
中国-西安  
  [810240960 @qq.com](mailto:email@email.com)

董博  
网络教育学院，高级工程师/大数据  
 西安交通大学  
中国-西安  
 dong.bo@mail.xjtu.edu.cn

团队简介

张发：网络空间安全专业研二学生,发表论文1篇,其中EI检索1篇,拥有专利5项(4项已公开),软件著作权5项。

查志超：软件工程专业研二学生，发表专利1篇，软件著作权1项。

杨育婷：人工智能专业博二学生，曾获2016年全国大学生数学建模竞赛陕西省一等奖, 2017年韩国第三届‘全球创新节’创新人才奖,2017年全国大学生电子设计竞赛全国二等奖, 2018年中国移动数字家庭创客马拉松全国二等奖，五届中国“互联网+”大学生创新创业大赛“青年红色筑梦之旅”赛道陕西省金奖暨冠军，国家级银奖, ICCV VisDrone2019 无人机视频单目标跟踪赛道季军, 荣获2019“中国好设计”奖项。

吴映潮：计算机专业研二学生，发表论文1篇,其中EI检索1篇。

高宇达：计算机专业研二学生，拥有专利5项.在此处插入段落文字。 在此处插入段落文字。 在此处插入段落文字。 在此处插入段落文字。 在此处插入段落文字。 在此处插入段落文字。

摘要

在本文中，我们展示资产分析及通信模式系统。该系统基于公司真实网络日志进行挖掘，得到企业内部的网络资产种类并实现不同资产的分类，展示资产间通信模式，并支持通信模式筛选，对流量进行统计和活跃机器数分析，对于复赛阶段，需要分析资产出现的异常，于是，我们对公司网络日志数据进行挖掘和分析，从周期访问突变、流量访问异常等角度找出存在的异常通信模式，并实现对异常的可视化及异常的分析。

关键词

网络资产分类，资产通信，异常分析，可视化

1 简介1

网络是信息化建设的基础，网络的安全和稳定直接影响着社会生活的运转。然而，随着今互联网的发展，潜藏的威胁也越来越多，数据安全存在未知隐患。如果网络安全性的得不到保证，信息丢失或数据泄露将会给个人、企业、国家带来不少的损失，影响个人生活，企业生存，国家发展。为了降低网络风险的影响，我们需要对网络日志进行挖掘和分析，并对网络资产进行监控并可视化。

因此，在初赛阶段，我们设计的资产分析系统实现了对企业内部的网络资产发现和分类，同时可视化资产之间的通信模式，并对流量及机器状态进行统计展示，在复赛阶段，我们设计了异常分析可视化系统。

1 初赛设计方案2

首先，我们对该公司网络日志进行统计分析，可以发现该公司的protocol一共是55种，destination\_port使用一共是\*种。其中http协议及80端口使用的最多，是该公司使用最频繁的通信模式。其次，我们根据网络日志source\_ip和destination\_ip得到网络关系结构，并将uplink\_length和downlink\_length作为关系图的边属性并进行特征处理，对节点和边进行网络表征。然后，我们利用node2vec[1]提取网络的关系，得到该网络结构的向量化特征。由于提取到的向量是高维的，因此使用TSNE [2]算法进行降维处理，发现数据是可分隔的。最后，我们使用聚类算法对资产进行发现，使用损失函数作为分类最优发现的评价指标，可知资产发现结果分为9类。我们使用kmeans 算法进行分类，并预测分类数据，得到最终的资产分类标签。

1 复赛设计方案3

首先，我们对db数据进行分析。我们对每一天的24小时的数据库操作进行统计，并用热力图进行可视化展示，热力图的颜色则表示访问量的频率，实现了纵向同比和横向环比。为了更清晰展示访问频次及周期性命令，我们设计交互，点击方块则可以展示当天具体时刻的访问频次折线图，为了更清晰的发现异常，我们也设置了对比图及缩放功能。

其次，我们对flow数据进行分析。我们先对uri，host，useragent等字段进行分词并提取特征，使用了三种不同的异常检测算法预测，isolation Forest[3],OneClassSVM[4], Local Outlier Factor[5],鼠标悬浮可以展示异常点的具体信息。

然后，我们对login数据进行分析。我们认为入侵会包含异常的登陆，如认证失败或者匿名登陆等。我们对异常登陆的信息进行可视化，同时将异常登陆的具体信息用平行坐标系进行展示。

最后，我们对tcpflow数据进行分析。数据泄露通常会涉及到拖库等操作，这些操作将会涉及到流量异常。

1 创新点4

1. 我们将网络表征和node2vec算法结合，并使用TSNE降维算法和聚类的思想，可以实现企业内部网络资产的发现和分类。

2. 利用统计的思想，我们可以得到该公司的通信模式及端口使用信息，对该公司的活跃工作机器进行分析

3.热力图及交互对比图可以很好的实现同比和环比及具体信息的展示，直观明了

4. 我们提取特征并结合三种不同的异常检测算法对异常点进行预测，提高预测准确性。

5. 基于提出的方法，我们提出了一套新颖的视觉分析系统，用于展示信息挖掘的结果。

2 系统结构

2.1 初赛系统结构

我们描述我们的系统如图1所示，它由三大部分组成。第一个组件是对源ip和目标ip进行关系进行的可视化，边信息包括上行带宽，下行带宽及使用的端口号，同时记录访问节点的downlink\_length流量及其uplink\_length流量大小。

第二个组件是对资产发现的结果进行展示，包括分类种类，不同的颜色代表不同的种类，同时支持按照指定的时间段展示。最后一个组件则是指定通信模式筛选符合的条件的源ip到目的ip，并展示他们的关系，也对24小时活跃机器结果展示。

IP关系：从网络日志中我们可以看到，每条记录都有源ip目标ip，每一条关系都包括协议和端口，上行流量和下行流量的信息，我们可以通过展示节点之间的关系，对固定资产的流量进行观测，并得到与其相关的其它连接节点。第一个组件展示固定时间每10分钟日志信息，包括源ip到目标ip的访问信息。同时统计目标节点的下行流量和上行流量。另外，我们可以使用第一个组件的搜索功能，在第二个组件中进行高亮显示，可以直观的展示该节点的存在的关系。

资产发现和分类：为了实现资产的发现和分类，需要利用网络日志关系及边属性信息进行网络表征，由于网络结构无法直接进行划分，需要对表征进行向量化处理，我们采用node2vec算法处理。我们利用TSNE算法将数据进行降维，利用聚类损失选取最优分类结果，划分结果如图所示。我们可以得到资产发现一共分为9类。接着我们利用KMEANs算法对资产进行分类，结果图1的第二个组件所示。由于数据量较大，我们展示类目0和类目1的结果图。我们可以通过第二个组件输入部分查询不同时间段每10分钟的网络关系结果。另外我们提供图的交互功能，鼠标悬浮可以看到边

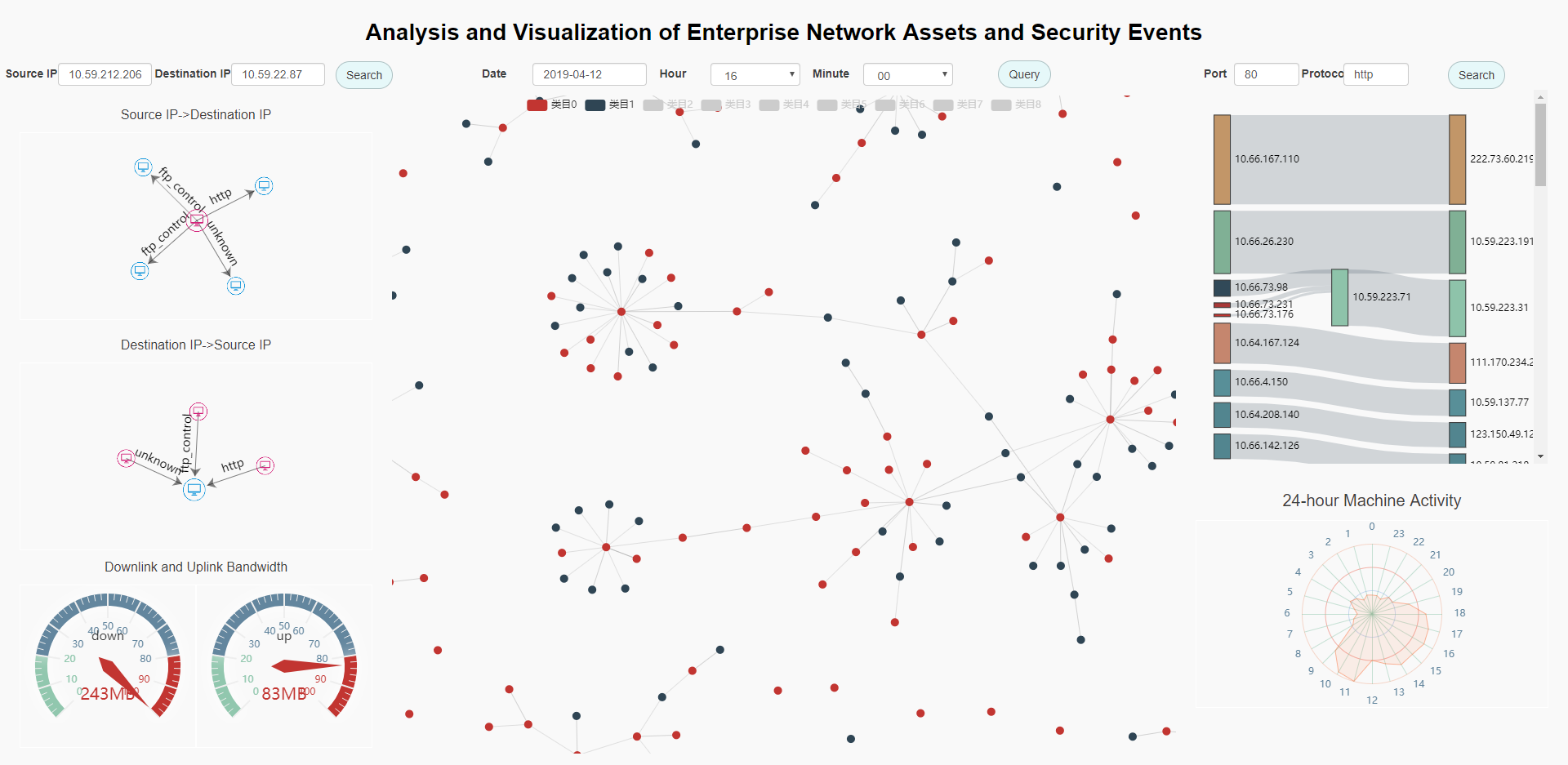


图1

上的访问信息，点击图中的节点，可以在左侧的展示框中更新当前节点的通信模式及其带宽信息。

通信模式：我们可以根据通信端口和通信协议进行条件搜索，可以找到特定通信模式下的通信数据，由于数据量较大，我们展示前100条关系数据，源ip和目标ip中间的宽度代表流量大小，点击图中的关系，可以更新第一个组件的信息，并展示该条关系的相关访问节点。

2.1 复赛系统结构

我们的系统结构如图2所示，它由以下几个部分组成

A. 周期行为异常可视化

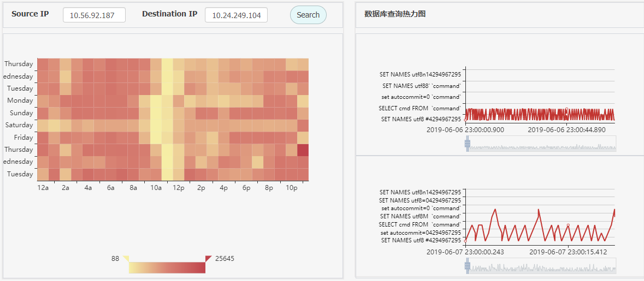
我们认为周期行为异常是可疑行为，因此我们对可疑ip的周期行为进行搜索查询，并用热力图进行可视化展示。如图1，横轴表示一天中是24小时，纵轴表示访问的日期。热力图的颜色深浅则表示访问量大小，我们认为该图的红色块为行为异常。为了实现同比和环比，我们设计了交互功能点击热力图中的方块，可以对访问细节进行展示，如图2， 

图2

B. http访问异常。

我们对主机访问行为进行分析，提取了flow表中method，uri，host等信息作为特征，然后，我们使用异常检测算法对我们提取的特征进行预测，并使用散点图进行展示，红色的点则表示异常的访问点，我们使用三种不同的算法实现异常检测，并提供交互功能选取不同的算法，鼠标悬浮则对异常的信息进行展示，如图3.我们可以看到访问的uri经过base64进行了编码，通过post方法进行脚本注入。

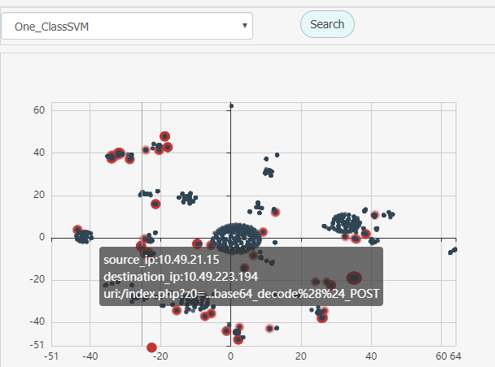


图3

C. 访问协议或流量突变异常

我们认为访问协议突变或者流量突变是嫌疑操作，因此，我们对不同时间的使用协议量进行可视化，如图4，我们认为6月5日13时出现了协议访问异常。我们也对不同时间的上行流量和下行流量进行可视化，如图5，在6月13日

10点，出现了上行流量和下行流量过大的异常，因此该时间段内操作可以认为是异常的。

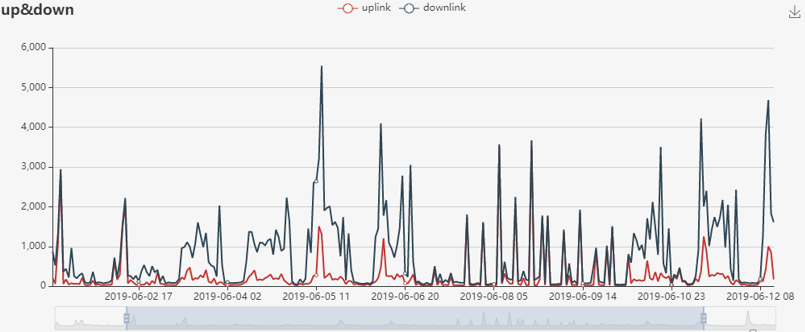


图4

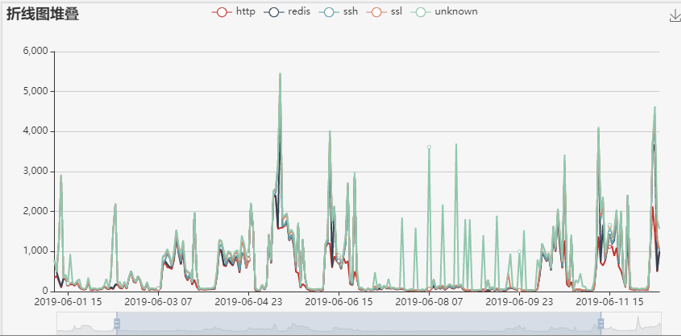


图5

D. 登陆异常

在攻击过程中可能需要多次注入访问才能获得权限，但是在访问过程中可能出现失败的登陆，图6则利用平行坐标轴对出现这些的异常行为的ip访问情况进行了可视化展示，线条颜色的深浅代表了该条操作的频次

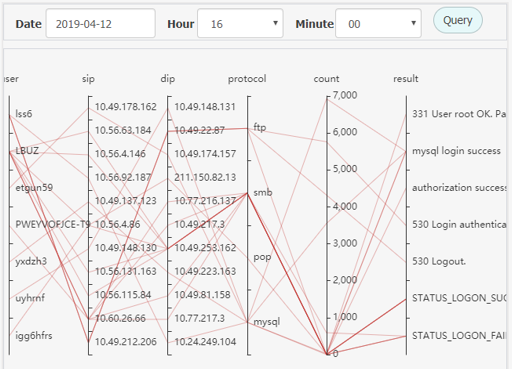


图6

**致谢**

比赛时光中，我尤为感谢一起奋战的小伙伴，感谢和我一起并肩作战的哥们，感谢给我提供动力的女朋友，感谢悉心指导的老师。因为你们，我们才能一步一个脚印，最终走到了比赛的终点。在这次比赛中，我意识到团队的重要性，合作的必要性和认真耐心给我们带来的飞跃性成长，没有团队的小伙伴们，是不可能获得这样的结果，感谢努力的我们，感激奋斗中的我们。

**参考**

1. A. Grover and J. Leskovec, “node2vec: Scalable feature learning for networks,” arXiv preprint arXiv:1607.00653, 2016.
2. L. v. d. Maaten and G. Hinton, “Visualizing data using t-sne,” Journal of machine learning research, vol. 9, no. Nov, pp. 2579–2605, 2008.
3. F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation forest,” in 2008 Eighth IEEE International Conference on Data Mining. IEEE, 2008, pp. 413–
4. K.-L. Li, H.-K. Huang, S.-F. Tian, and W. Xu, “Improving one-class svm for anomaly detection,” in Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 03EX693), vol. 5. IEEE, 2003, pp. 3077–3081.
5. M. Amer and M. Goldstein, “Nearest-neighbor and clustering based anomaly detection algorithms for rapidminer,” in Proc. of the 3rd Rapid-Miner Community Meeting and Conference (RCOMM 2012), 2012, pp.

会议名称：ACM伍德斯托克会议

会议简称：WOODSTOCK'18

会议地点：美国德克萨斯州埃尔帕索

ISBN：978-1-4503-0000-0 / 18/06

年份：2018

时间：6月

版权年份：2018年

版权声明：维护权利

DOI：10.1145 / 1234567890

RRH：F。Surname等。

价格：$ 15.00