观安 | SICSI CYBER RESEARCH INSTITUTE 赛博研究院

人工智能赋能网络安全白皮书

ARTIFICIAL INTELLIGENCE

2021

AI Empowers Cybersecurity Whitepaper 2021

COPYRIGHT STATEMENT

版权声明

本报告版权属于出品方所有，并受法律保护。转载、摘编或利用其他方式使用报告文字或者观点的，应注明来源。违反上述声明者，本单位将追究其相关法律责任。

出品方

上海观安信息技术股份有限公司

上海赛博网络安全产业创新研究院

研究专家

张照龙	上海观安信息技术有限公司创始人，首席专家
惠志斌	上海赛博网络安全产业创新研究院首席研究员
胡绍勇	上海观安信息技术股份有限公司首席技术执行官
李　宁	上海赛博网络安全产业创新研究院高级研究员
夏玉明	上海观安信息技术股份有限公司研究院院长
魏国富	上海观安信息技术股份有限公司研究院副院长
石英村	上海赛博网络安全产业创新研究院高级研究员
刘　胜	上海观安信息技术股份有限公司研究院研究员

FOREWORD

前 言

当前，随着数字化全面融入人们的生产生活，网络安全成为数字化时代最大的挑战之一，并且随着人类社会迈入智能时代，网络安全挑战将越来越大，使得各个行业对网络安全的需求急剧增大。人工智能作为引领未来的战略性新兴技术，正在对社会进步、经济发展和人类生活产生深远影响。得益于人工智能技术领域的持续创新，人工智能技术的飞速发展不仅将变革制造、金融、医疗、商业等各个行业，也将成为网络安全产业创新和发展的强大助推器。本报告将分析人工智能如何为网络安全领域带来变革，并分十个重点领域详细阐述人工智能赋能网络安全的模式与实践，例如数据安全、身份与访问安全、物联网安全、移动安全等领域。此外，本报告还分析了人工智能自身的局限性，以及在网络安全领域应用时存在的亟需客服的障碍和问题。最后，本报告分析了人工智能赋能网络安全的未来发展趋势。

目录 CONTENTS

第一章
人工智能技术发展新阶段

人工智能（AI）作为引领未来的战略性新兴技术，正在对社会进步、经济发展和人类生活产生深远影响。当前，人工智能技术开始广泛应用于各个领域，展现出巨大的发展潜力。

近几年，人工智能相关技术持续创新和迭代，呈现出人工智能发展的新阶段。OpenAI提出的包含1750亿参数的自然语言处理模型GPT-3、Google设计的聊天机器人Meena、Facebook开源的94亿参数聊天机器人BlenderBot、DeepMind推出的蛋白质结构预测算法AlphaFold2、以及Facebook研发的超强PyTorch目标检测库Detectron2等，都是2020年以来人工智能领域影响较大的研究成果。这些技术成果证明人工智能技术发展继续保持高速推进。同时，新兴的人工智能技术领域正在蓬勃发展，将在未来几年展现出巨大的应用潜力：

● 低代码/无代码工具蓬勃发展

低代码/无代码平台能在缺乏深入编程知识的情况下构建整个生产级人工智能驱动的应用程序。去年低代码/无代码工具异军突起，应用领域包括从构建应用程序到面向企业的垂直人工智能解决方案。有数据显示，低代码/无代码工具将成为科技巨头们的下一个战斗前线，预计到2025年其市场总规模将达到455亿美元。

● 机器学习运维（MLOps）兴起

MLOps是数据科学领域一个相对较新的概念，类似于DevOps。简单来说，MLOps是机器学习（ML）领域的DevOps。MLOps为数据科学家、ML工程师提供服务，使他们转向协同工作，提高工作效率。它拥有一套完整的行为策略方式，用来解决ML和AI在运行周期内遇到的各种问题。在增长最快的GitHub项目Top-20中有5个是机器学习运维工具。这表明整个AI行业正在从“如何开发模型”转向“如何运维模型”。

高级预训练语言模型实现突破

近年来，预训练语言模型(如问答、命名实体识别、自然语言推理、文本分类等)在许多自然语言处理任务中发挥着重要作用。这些预训练语言模型非常强大，并彻底改变了语言的翻译、理解以及总结等，但这些模型非常昂贵，而且训练非常耗时。好消息是，高级预训练模型可以催生出新一代高效且极易构建的AI服务。GPT-3就是其中的翘楚。

模型可解释性将推动AI更加广泛的应用

工程师和科学家开始理解模型做出特定决策的理由，以及模型可以安全运行的范围。他们能通过试验来解释模型在不同场景中的运行方式，并借助可视化来理解模型不正常运行时的内在工作机制。随着研究者设计出更多可解释性方法，且越来越多的软件供应商将这些方法加入自己的工具，业内工作者会更愿意将 AI 创新纳入自己的工作流。

得益于人工智能技术领域的持续创新，人工智能技术的飞速发展不仅将变革制造、金融、医疗、商业等各个行业，也将成为网络安全产业创新和发展的强大助推器。

第二章
网络安全新需求与挑战

网络安全：数字化时代的最大挑战之一

当前，全球进入数字化时代，人工智能、5G、云计算、物联网、大数据等信息技术快速发展并广泛应用，万物链接将成为可能，工业、服务业、农业等传统产业领域加速数字化转型，工业互联网、在线金融、在线教育、在线医疗等在线新经济迅速发展，数字化全面融入人们的生产生活。以此同时，快速发展的数字化进程导致网络安全成为数字化时代最大的挑战之一，并且随着人类社会迈入智能时代，这个挑战将越来越大，使得各个行业对网络安全的需求急剧增大。

网络安全面临的挑战

1、网络风险呈指数级增长

根据世界经济论坛发布的《2021年全球风险报告》，网络安全风险是整个世界在今年及以后将面临的一项重大风险。报告指出，快速的数字化成倍地放大了企业所面临的网络安全风险，让网络变得更复杂，并且安全性更低。根据Cybersecurity Ventures的预测，到2021年，企业将每11秒遭受一次勒索软件攻击，而2019年仅为每14秒。这使全球勒索软件损失成本预计将达到200亿美元，远远高于2015年的3.25亿美元。在全球网络犯罪蓬勃发展的背景下，企业和组织也将继续加大在安全工具和技术方面的投入。

12.4%

Gartner最新发布的一份报告预测表明，全球在网络安全和风险管理服务方面的支出预计将增长12.4%，在2021年达到1504亿美元。2020年，安全和风险管理方面的支出增长仅为 6.4%。

2、网络安全威胁趋向智能

随着网络信息技术全面普及以及数据价值的持续增长，网络安全威胁持续升级，且呈现出智能化、隐匿性、规模化的特点，网络安全防御面临更大的挑战。网络攻击者不断创新，推出新的攻击工具和技术，以突破传统的网络安全防御。其中采用人工智能的网络威胁手段已经被广泛应用于网络犯罪，包括漏洞自动挖掘、恶意软件智能生成、智能化网络攻击等，网络攻击方式的智能化升级打破了攻防两端的平衡。网络安全攻防不对称要求网络安全防御方采取更加智能化的手段予以应对。

3、超负荷的网络安全团队

网络安全形势的日益严峻，给本就已经疲惫不堪的安全从业人员带来了更大的压力。赛门铁克发布的一项调查称，法、英、德三个国家中近一半（48%）的网络安全行业领导者们认为他们的安全团队在与黑客、网络罪犯们的对抗中处于严重落后状态。网络安全从业者们表示，他们所在的团队缺乏必要的技术能力来应对其所面对的安全威胁，超过三分之一的人则表示他们的团队根本无法在现在庞大的工作量下正常应对大规模安全事件，约有四分之三的人表示自己低估了处理网络安全事件所需要花费的精力和时间，也有相同比例的人表示多数情况下自己根本没时间去正确的评估某一事件的威胁风险。此外，网络安全系统产生的大量数据是任何人类团队都无法筛选和分析的。

4、网络安全人才缺口

美国国际战略研究中心（CSIS）2019年1月发布的《网络安全劳动力缺口》报告显示，CSIS针对八个国家的IT决策者的一项调查发现，82%的雇主表示企业缺乏网络安全人才和技能，71%的人认为网络安全人才缺口会为组织带来直接且可度量的损失。根据美国网络安全教育计划（NICE）资助的CyberSeek项目的调查结果：截至2019年1月，美国面临近31.4万名网络安全专业人员的短缺，而该国受雇的网络安全劳动力仅为71.6万人。根据《2020 (ISC)^{2}网络安全人力研究报告》，虽然2020年全球网络安全人力短缺情况有相应缓解，从去年报告的407万短缺人数降至312万，网络安全行业在全球范围内也经历了大幅增长，但数据显示，目前该领域的就业人数在美国需要增长约41%，在全球需要增长89%，才能填补人才缺口，而人才缺口仍是专业人士最关心的问题。

5、现有网络安全工具的局限性

当前，边界控制工具主要依赖于签名和规则，终端安全也是如此，并且这些类型的工具只能检测以前已经识别到的攻击，而对未知威胁无效。日志工具和SIEM需要人工来确保在整个组织中收集数据的一致性，并与安全团队对威胁的预测相匹配，这需要安全团队能够想象所有可能出错的事件。所谓的“行为分析”工具依赖于基于规则的范式，即配置某些岗位或设备的“应该”行为，然后寻找这些行为中的偏差，这种方法无法适应现代企业的复杂性和规模。因此，由于现代业务的复杂性和攻击方法的持续创新，传统安全工具已经体现出不适用和局限性，包括：传统安全工具需要掌握之前所有的攻击；需要完全理解企业业务和特定于业务的规则；需要猜测未来所有的攻击和软件漏洞；需要能够将上述所有见解转化为有效的规则或签名。最重要的是，传统安全工具需要组织被攻击后才能提供解决方案，而在这个充斥着不可预测、快速攻击的时代，这种做法显然严重不足。到2022年，全球网络安全支出预计将达到1700亿美元，所有企业都希望网络安全行业创造更好更具弹性的方法。

第三章
人工智能为网络安全带来的变革

当今，将AI应用于网络安全的组织正在看到变革的到来，并获得显著的效益。通过将AI和先进分析技术应用于大量的内外部安全数据，更加智能的网络安全技术可以产生预测性和可用的见解，帮助组织做出更好的网络安全决策，保护组织免受网络威胁。同时，通过以只有机器才能提供的速度和准确性监视网络环境，AI可以帮助组织更快地检测和响应威胁。最重要的是，利用AI可以帮助组织感知当今不断发生的、越来越复杂的网络攻击态势。虽然网络安全领域目前严重依赖人力投入，但我们正逐渐看到技术在某些特定任务上比人类做得更好。

AI可以检测和预测新兴的未知威胁

利用AI技术可补充现有的网络安全控制和应用程序，以检测渐进的、新兴的和未知的网络威胁。AI通过收集、关联和分析广泛的安全数据来加强威胁检测能力，并通过利用威胁情报、漏洞信息、设备事件日志和上下文数据来确定威胁模式，可以使企业能够检测到高级持续性威胁，并识别现有安全防御可能无法检测到的入侵指标，从而实现前瞻性和预测性的安全洞察。

AI使组织能够更快地对威胁做出响应

快速威胁响应对于确保组织免受网络攻击至关重要。利用AI，检测威胁和入侵所需的总时间最多可减少12%。AI还可将修复漏洞或执行补丁以应对攻击的时间减少12%。此外，AI通过不断扫描已知或未知异常，可使网络威胁停留时间（威胁者未被发现的时间）下降11%。

AI大幅降低检测和应对网络威胁的成本

将 AI 用于网络安全使组织能够理解威胁模式，以识别新的网络威胁。这大大减少了需要用来识别事件、调查事件和响应威胁的时间和精力。近三分之二的企业高管表示，AI 降低了检测和应对网络威胁的成本。大多数组织的成本减少幅度在 1% - 15% 之间 (平均为 12%)。此外，三分之二的组织表示，AI 提高了网络安全工具的投资回报率。例如，以西门子公司为例，西门子网络防御中心 (CDC) 使用 AWS 构建了一个支持 AI、高速、全自动和高度可扩展的平台，每秒可评估 60000 个潜在的关键威胁，基于该平台，他们能够利用一个不到 12 人的团队管理网络风险，且不会对系统性能产生影响。

AI显著提高安全分析师的工作效率

五分之三的企业高管一致认为，将AI用于网络安全领域提高了安全分析师的准确性和效率。以往，网络安全分析师需要花费大量时间查看数据日志或事件时间表。有了AI的帮助，安全分析师可以花更多的高质量时间分析AI算法识别到的威胁事件。网络安全领域的人才是稀缺的，而AI可以帮助缩小人才差距。

AI以高重复性减少人为错误

当前，没有任何人工操作能做到每一次都完美再现，尤其是在当今这种动态多变的环境中更是如此。AI技术能以高度一致和可重复的方式执行威胁检测和响应任务，减少人工干预和人为错误。这可以使网络环境的安全、管理和审计变得更容易，以满足监管法规和其他外部要求。

AI可降低专业技能要求

基于ML的安全技术可以通过关联新威胁与已知威胁，来发现人类无法及时发现的网络攻击。除此之外，这种类型的安全措施还可以提供建议的操作，以限制进一步的损害并防范未来的攻击。因此，在网络安全工作中采用AI可以让安全团队以有效、实用的方式管理更多的网络威胁，同时，借助AI，组织只需要规模较小的团队便可满足人员配备需求，而且安全专家在不需要多年经验和培训的情况下就可以获得高水平的表现。

第四章
人工智能赋能网络安全的重点领域

一、网络安全

• 恶意域名攻击风险

在网络世界中，每天都会出现新的网络犯罪组织和威胁行为者，设计逃避检测的攻击工具。这些难以捉摸的技术混淆了攻击行为，使恶意行动无法被识别。其中一种工具是通过不断改变域名，来欺骗安全人员，使用容易丢弃的域名有助于网络犯罪分子避免被安全解决方案发现。当前，这类简单的行为已经成为攻击的常态，使安全防御者的工作变得更加困难。

域名在网络攻击中通常被用于：

恶意软件

如果攻击者的目标是通过下载恶意软件来感染终端，那么域名可以被用作旁路下载、与C2服务器的通信、下载恶意的有效载荷以及向特定服务器泄露数据。

钓鱼行为

冒充合法来源和域名的电子邮件可被用于商业电子邮件攻击（BEC）诈骗。这种方法通过试图愚弄用户向Office 365等提供其云计算凭证的凭证采集攻击，可能会将用户指向一个特定的欺骗性域名。

垃圾邮件

有些垃圾邮件属于边缘性的网络钓鱼，主要是为了获取信用卡等信息，让用户点击链接到恶意网站，这会对用户及其组织构成威胁。

2019年，每天有超过27.5万个域名被注册。如此多域名的持续创建给IT组织带来了进一步的挑战。IT专家无法手动跟进哪些域名是合法的，哪些是恶意的。

• AI如何赋能恶意域名检测

AI赋能恶意域名威胁防御的模式包括：

将AI/ML应用于恶意域名检测

目前，很少有恶意攻击不涉及互联网上某种形式的通信。恶意软件需要使用C2服务器在网络内横向移动，商业电子邮件攻击往往从一个看起来像可靠企业的欺骗性域名开始，而无恶意软件的网络钓鱼攻击仍然链接到充满恶意软件的网站。在所有这些情况下，都会使用某种域名。因此，利用AI/ML作为一种手段，可以根据以前确定的恶意域名的已知特征来识别可疑的域名。对组织来说，利用AI可以实现在一个域名投入使用并被用来执行恶意行动之前就将其阻止。

使用域名特征来发现可疑的域名

域名不仅仅是一个标识符和一个顶级域名（.com，.net，等）。事实上，在DNS取证数据中，存在很多支持性的细节，ML可以利用这些细节来理解一个特定域名的意图和潜在用途。作为训练过程的一部分，ML可以处理大量的域名特征，以最终发现恶意的域名，包括：名称构成或熵、域名年龄、域名长度、域名服务器细节等。

创建威胁图谱

威胁图谱由一组有监督机器学习分类器组成，可用于预测一个域名当前是否可能是恶意的，以及在未来某个时候是否可能有恶意的意图，或者在本质上是否是非恶意的。由于一个配置文件很可能只在很短的时间内有效，因此，为了有效地发现恶意域名，需要定期重建模型，利用新的数据来了解威胁者是如何改变他们在域名方面的策略的。

赋予风险评级

ML的理想输出是一个域名风险评级。如前所述，每天有超过275000个域名被创建，而安全工具的任务是指出哪些域名是出于恶意被创建的。上述特征可以用来帮助ML算法了解一个域名的构成，并预测一个域名是否可疑。

• 企业案例

观安：基于文本和行为分析的DGA域名检测技术

【场景描述】

近年来，恶意软件的数量和复杂度持续增长，催生了大量黑色产业链和网络犯罪行为。为了维持持续的经济效益或其他目的，攻击者对肉鸡的管理是僵尸网络控制的重要问题。对肉鸡进行有效的管理，不仅有利于各种攻击类型的发起，更可以延长攻击被发现的时间，并且实现攻击者真实身份的隐藏。现代恶意软件一般通过使用DGA算法与C2服务器建立通信，从而达到上述目的。

DGA（Domain Generate Algorithm，域名生成算法）是一种利用随机字符来生成C&C域名，从而逃避域名黑名单检测的技术手段，常见于僵尸网络中。该方式作为备用或者主要的与C2服务器进行通信的手段，可以构造更加鲁棒的僵尸网络，做到对感染肉鸡的持续性控制。其原理是客户端通过DGA算法生成大量备选域名，并且进行查询，攻击者与恶意软件运行同一套DGA算法，生成相同的备选域名列表，当需要发动攻击的时候，选择其中少量进行注册，便可以建立通信，并且可以对注册的域名应用速变IP技术，快速变换IP，从而域名和IP都可以进行快速变化。

很显然，在这种方式下，传统基于黑名单的防护手段无法起作用，一方面，黑名单的更新速度远远赶不上DGA域名的生成速度，另一方面，防御者必须阻断所有的DGA域名才能阻断C2通信，因此，DGA域名的使用使得攻击容易，防守困难。

DGA域名自曝光以来，其检测工作就在持续进行，在不同场景、不同时期，检测方法也呈现出一定区别，现有方法大致可以分为以下两种：一是基于黑名单，二是基于文本、行为分析。

基于文本分析的代表工作包括通过分析DGA域名与正常域名之间字符分布的差异，对IP产生的域名进行批量分类；基于行为分析的代表工作包括对同一主机产生的NXDomain进行聚类和分类，可以发现感染主机，进一步发现C2域名。但两种方法均显示出一定弊端，基于黑名单形式复杂且难以精准定位；基于文本及行为分析难以做到实时和团伙关联，识别能力较弱。

上海观安信息研发的网络流量安全分析系统是基于半监督学习算法的DGA域名检测技术，通过对所使用算法进行特征维度的搭建、特征优化选择、模型搭建、聚合等先进技术，通过科学的数据挖掘、智能分析、大数据算法的运用、实时/离线分析处理、机器学习和数据可视化等技术手段实现信息安全威胁流量检测能力的建立。

【技术方案】

1、检测流程

本方案结合文本检测和行为检测，并结合情报分析，可以实现告警的聚合度高，且准确率和查全率高的效果。其中文本分析利用 LSTM + Attention的NLP分析技术，行为分析利用非监督学习异常检测算法，情报分析部分利用自研的快速查找图算法。

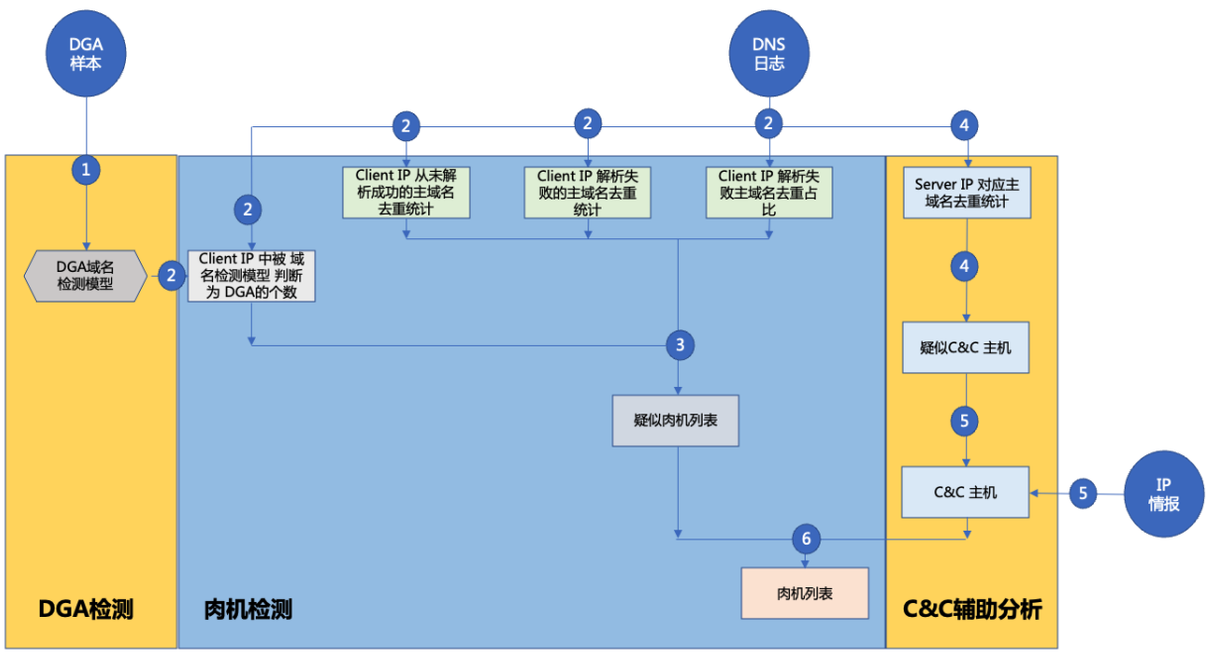


图1 DGA检测流程图

本方案包含以下几个模块：

- DNS 数据解析模块（PCAP、日志、网卡）
- DGA域名文本检测模块
- DGA域名行为检测模块
- 情报模块

其中涉及机器学习部分的有DGA域名文本检测的BiLSTM + Attention（双向长短记忆模型+注意力算法），DGA域名行为检测的异常检测算法Isolation Forest（孤立森林）和情报模块的快速查找图算法。他们的技术如下：

- **BiLSTM + Attention**

这是一个监督学习的二分类问题，其流程如图所示：

1、通过黑白样本训练生成DGA检测器；

2、给新域名打分并判断其是否为DGA 域名。

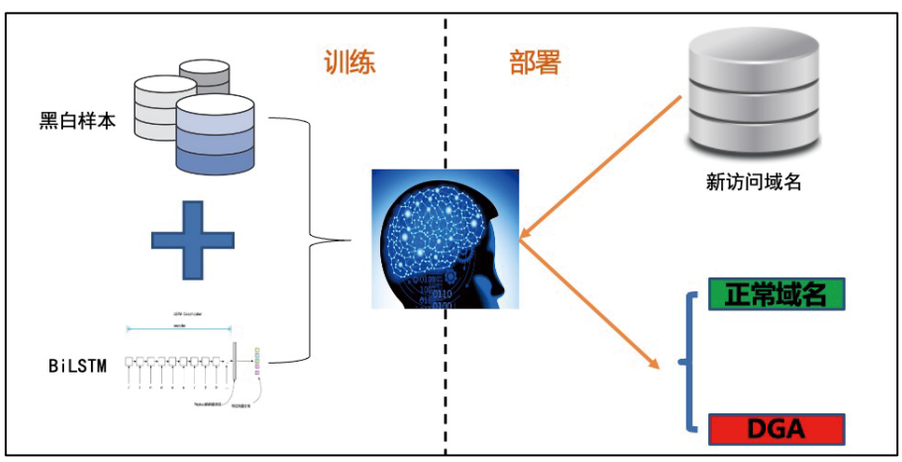


图2 DGA文本检测流程图

- **异常检测算法Isolation Forest**

对源IP抽取包含NXdomain去重统计，NXdomain占比，疑似异常访问等特征，利用孤立森林算法找到异常得分最高超过阈值的源IP并告警。

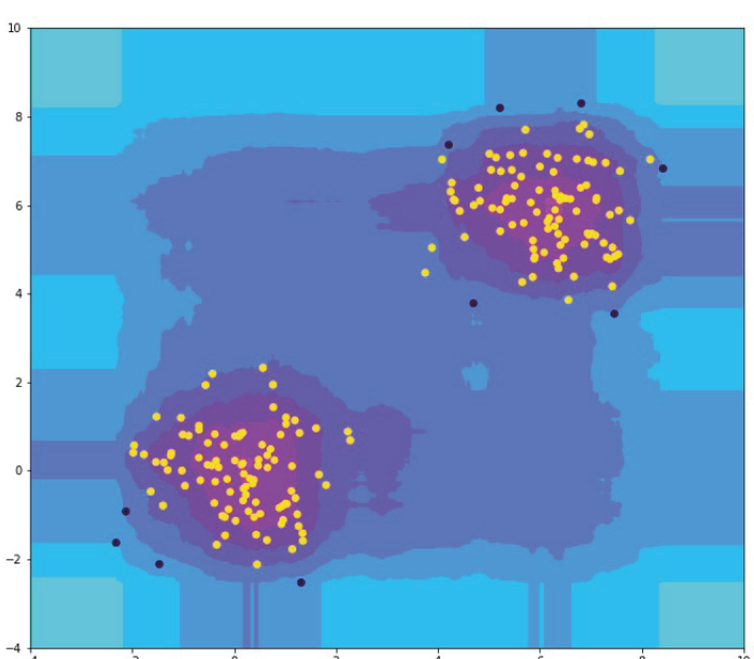


图3 孤立森林检测示意图

- **快速查找图算法**

在 DGA 检测情报部分的查找不容忽视，他可以帮你扩展你的成果。通过获取的 DGA 域名和恶意 IP 做扩展查找，可能会找到其它相似的 DGA 域名和恶意 IP。本快速查找图算法定义了节点 domain、IP 和无向边 (domain-IP)，并定义了图中边的长度。并利用 dijkstra 找到图最相关的信息。

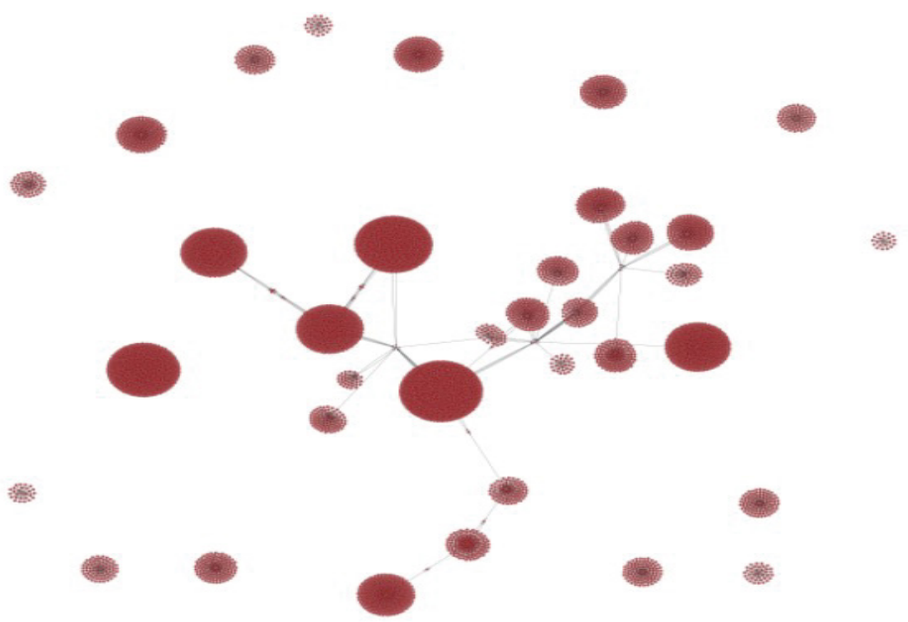


图4 快速查找图算法示意图

2、主要技术

- **流量采集技术**

深度数据包检测 (Deep packet inspection，缩写为 DPI) 是一种特殊的网络技术，它通过用“旁路接入”的方式接入到网络，会对网络中的每个数据包进行检查，一般网络设备只会查看以太网头部、IP 头部，而不会分析 TCP/UDP 里面的内容，这种被称为浅数据包检测；与之对应的 DPI 会检查 TCP/UDP 里面的内容，解析出 MAC 地址、IP 地址、端口和应用层的协议内容。

- **大数据分析技术**

部署了 HDFS、HIVE、HBASE、solar、Spark 等基础大数据组件，实现基于大数据平台的统一管控、统一开发、统一调度和运行能力。提供了基于分布式架构的 ETL 工具实现对海量数据的批处理。构建了人工智能机器学习模型库，能够对海量数据进行深度挖掘，预置了朴素贝叶斯、决策树、逻辑回归、Kmeans、关联规则、时间序列、最小哈希、线性回归、主成分分析、协同过滤、相似特征分析、全连通图形分析、邻接点分

类、影响力传播分析等人工智能算法，对未知的威胁能够及时发现，对于安全风险能够进行提前预测，提前预警，及时采取应对措施预防安全风险对系统产生的影响。

- **机器学习技术**

观安网络流量安全分析系统的文件检测中内置机器学习的检测模块，该模块中包含已经生成的机器学习模型库，通过更新机器学习模型库，可以及时把机器学习训练研究成果及时更新升级到系统中去。

机器学习模型库是离线对大量样本的静态信息和动态行为信息做特征工程，通过选择、组合、修改机器学习算法，进行反复训练并且交叉验证后生成的。模型库里会有许多经过不同算法和特征提取，训练后生成的不同的模型数据，机器学习模块会智能选择和综合不同模型的计算结果，最终给出威胁概率。

- **隐蔽信道检测技术**

观安网络流量安全分析系统能够针对隐蔽信道的通信行为进行分析与检测，能够在大流量环境下对隐蔽信道进行检测，且能够通过自定义规则快速构建新的检测模型。

【应用效果】

该技术运用于恶意域名检测以及隐蔽通信隧道识别等多种检测方式，有效帮助客户排查网络中的木马扩散、肉鸡、挖矿、勒索、被恶意监视等威胁，遏抑 DDoS 攻击，为 APT 攻击跟踪挖掘提供多维线索，充分保障客户网络安全。

二、终端安全

• 终端安全现状与EDR

目前，终端安全人员面临两大困境：一是缺乏调查终端网络安全事件以确定根本原因和攻击链的能力，二是缺乏监测终端状态以获得终端实时或接近实时的清单的能力。即使对于拥有大量 IT 人员的组织，终端安全堆栈的复杂性也使他们的工作效率非常低下。大部分 IT 团队被迫花时间处理大量的虚假警报，转移了对更重要的企业目标的注意力。事实上，大约有 1/3 的安全分析师的时间被花在处理那些在不知情的情况下已经被处理过的警报上，这对整体效率是一个巨大的消耗。随着“自带设备”策略在企业中的普及，并广泛采用一系列 SaaS 解决方案，终端已经成为网络攻击的薄弱环节，使其成为企业网络安全防护的首要关切。

当前，终端安全人员常常难以监控终端的安全性、快速调查事件并与 IT 运行协作采取补救措施。鉴于全球网络安全技能短缺，终端安全解决方案必须提高威胁预防效能，简化调查过程，并为网络安全和 IT 员工提供自动补救功能。

终端检测和响应工具 (EDR) 的发展很大程度上是由于行业已经认识到，基于签名的方法的有效性是非常有限的。为了应对这些威胁，企业越来越多地采用提供高级终端保护的终端检测和响应 (EDR) 解决方案。除了保护终端免受威胁外，有效的 EDR 解决方案至少必须具备以下能力：

调查攻击和警报数据

企业不仅需要阻止当前威胁对终端的影响，还需要阻止未来攻击发生的可能性。当威胁被挫败时，安全专业人员必须能够捕获关键数据并进行根本原因分析，以了解攻击的起源以及攻击者试图攻击终端的方式。

执行有针对性的威胁搜索

有一些恶意活动是不容易被识别的。当一台计算机开始表现得不正常，或者确定一个终端可能有被破坏的风险时，安全工具箱要为其提供做出明确判断所需的可见性。

动态威胁检测

目前，安全分析人员可以跨终端执行搜索以识别可疑行为，并通过人工调查确定是否存在威胁。虽然这个过程有巨大的价值，但它不能在整个企业中扩展。要根除隐藏在终端上的威胁，必须使用自动化的威胁检测方法。

提供响应能力

当检测到攻击时，安全工具需要在锁定终端的同时隔离可疑的攻击。根据攻击情况，可能还需要将系统恢复到以前的状态。

当前，大多数 EDR 产品包含一个基于规则的引擎，根据配置设置进行定制。其优势是能够提供无限的保护选项，但这也是它们的弱点，因为它们需要持续的人工输入来处理不断变化的攻击模式。但是，随着 IT 人员的减少，这种要求不可能得到满足。而且，即使有足够的工作人员，IT 专业人员也无法应对巧妙伪装的网络攻击速度。为了及时有效地应对日益增长的威胁，自动化是关键。

• AI如何赋能终端安全

AI 在不需要增加 IT 资源的情况下，为组织提供了网络攻击对抗能力的大幅提升。AI 通过数据分析可以获得事件风险结果，并利用历史数据创建分析模型，用于预测和分析未来的威胁。基于 AI 的方法可以通过本地实现，也可以通过基于云的工具远程实现。本地应用提供更快的响应能力，比基于云的应用使用更少的资源。而基于云的应用包含更广泛的数据集，但响应时间会减慢。

这两种基于 AI 的方案都可以取代传统的防病毒解决方案，并消除对签名的需求。利用基于 AI 的方案可以节约时间和资源，并解决当前 IT 人员短缺的问题。但机器学习可能会产生假阳性警报，使 IT 工作面临更多挑战。因此，企业需要寻求一种平衡，在确保检测恶意活动的同时最大限度地减少错误警报。

EDR 可以将机器学习用于功能性和非功能性目的。功能性目的包括能够跨大型复杂系统分析行为，识别并防止对终端执行的攻击。非功能性应用包括人工交互，例如分析人员输入时发生的交互。在这种情况下，机器学习可以减少调查错误，同时加速输入的采用。

• 企业案例

BlackBerry：BlackBerry® Optics——AI赋能的EDR解决方案

【场景描述】

数据显示，超过 80% 的网络攻击发生在终端，企业的笔记本电脑、工作站、服务器和移动设备都处于危险之中。多年来，终端安全产品的主要威胁防护都是基于签名的，这些签名是在企业受到攻击影响和损害已经造成之后创建的。假设所有的攻击都被发现过，那么使用签名是有意义的。然而，恶意软件每天甚至每小时都在变异，因此基于签名的预防工具已经过时，并产生了对更强大的基于预防的终端安全工具的需求。

终端检测和响应 (EDR) 技术可以对安全事件进行更快的响应和补救。然而，攻击者一直在努力开发战术、技术和程序 (TTP)，以击败基于传统规则的 EDR 技术，导致它们的效率随着时间的推移而降低。展望未来，依赖规则的 EDR 产品将无法跟上新威胁的步伐。

BlackBerry Optics 解决方案为 EDR 提供了一种预防优先的方法，提供了自动机器学习威胁检测模块，旨在发现那些几乎不可能通过静态行为规则发现的威胁。

预防优先的安全防护思路可以显著减少安全堆栈生成的警报数量，减少与无休止的警报调查相关的工作负担。通过 BlackBerry® Protect 解决方案防止恶意软件、恶意脚本、流氓应用程序和无文件攻击对企业造成的伤害，BlackBerry® Optics 提供了 AI 驱动的终端检测和响应（EDR）功能，以保持数据和业务安全。

【技术方案】

BlackBerry® Optics解决方案具体功能分析：

- **预防攻击**

保护终端不受攻击者攻击的最佳方法是在攻击开始之前识别到并阻止攻击。BlackBerry® Optics 的恶意软件识别能力的核心是一个革命性的机器学习研究平台，而不是签名和沙箱。这种创新技术以机器速度实时使恶意软件、勒索软件、病毒、机器人和零日攻击变得毫无用处。在几毫秒的时间内，BlackBerry® Optics 的预防模型对每个文件的数百万个特征进行分析和分类，将它们分解到原子级别，以分辨一个对象是好是坏，并防止恶意软件在终端上执行。

- **无文件恶意软件**

由于攻击者意识到合法的管理工具和内存可以很容易地用于破坏系统而不向磁盘写入任何文件，因此无文件攻击正在增加。许多安全产品没有能力防止这一类型的攻击，但 BlackBerry® Optics 解决方案的内存漏洞预防、脚本管理和无文件威胁检测模块可在这些攻击有机会影响业务之前就将其阻止。当攻击者试图提升权限，进行进程注入，或通过其他方式不当使用终端的内存时，该解决方案可以立即检测到并阻止它。

- **恶意脚本**

出于多种原因，脚本是许多攻击者最喜欢的选择。首先，对于新手攻击者来说，恶意脚本随手可得。此外，安全产品通常难以检测脚本，因为脚本有许多合法用途。借助 BlackBerry® Optics 解决方案，组织可以获得内置的脚本管理，这意味着安全专业人员可以完全控制脚本在其环境中运行的时间和位置，从而减少攻击者利用此攻击媒介对业务造成损害的机会，同时仍允许其合法使用。

- **恶意电子邮件附件**

网络钓鱼攻击是攻击者获得端点访问权的最有效方式之一。员工在不知情的情况下打开恶意附件，认为它们是合法的，使攻击者能够进行任何数量的恶意行动。利用 BlackBerry® Optics 解决方案，恶意附件会被自动识别和阻止。例如，如果一个文件包含一个被认为是有风险的 VBA 宏，它将被阻止执行。

- **外部设备**

USB 设备在大多数组织中随处可见。这些设备使员工能够快速有效地与他人共享文件。然而，如果这些设备装载了恶意软件或被用于传输业务之外的敏感数据，它们可能会对环境造成重大破坏。为了应对这种风险，BlackBerry® Optics 解决方案具有内置的设备使用策略执行功能。这种能力使管理员能够控制哪些设备可以在他们的环境中使用。这种终极控制限制了攻击者通过外部设备成功执行攻击或窃取数据的机会。

- **基于角色的访问控制**

RBAC（基于角色的访问控制）可提高操作效率，增强遵从性，提高管理员对业务的可见性，降低成本，并将数据泄露的风险降至最低。BlackBerry® Optics 允许管理员自定义角色和权限，轻松添加新员工，快速地将访问权限限制为员工工作所需的访问权限。

- **调查攻击和警报数据**

阻止威胁对终端的影响对于确保敏感数据的安全至关重要。更进一步说，当威胁被挫败时，如果关键数据能被捕获，那么安全专家就可以看到攻击者是如何试图破坏终端的。BlackBerry® Optics 解决方案提供了这种能力，不仅适用于被阻止的攻击，而且适用于可能在终端上发现的任何潜在威胁。只需简单点击，就可以生成导致威胁的活动时间线，即所谓的焦点视图。此外，可以从受影响的终端远程收集数据，以进一步了解企图攻击或可疑的活动。

- **执行有针对性的威胁搜索发现隐藏的威胁**

一些恶意活动不容易被识别到。当一台计算机开始表现得不正常，或者确定一个终端可能有被破坏的风险时，安全工具箱需要为其提供做出明确判断所需的可见性。BlackBerry® Optics 解决方案提供对存储在终端上的证据相关数据的即时访问。在发现可疑活动的瞬间，可以针对正在调查的确切威胁进行搜索。

- **利用破坏性指标来寻找威胁**

威胁狩猎可以被描述为形成一个假设，然后使用 IOC 或其他术语进行一系列搜索和调查，以证明或反驳该假设的行为。获得正确的数据是有效执行这一技能的基本核心。BlackBerry® Optics 解决方案可以实现有针对性的威胁猎杀，并提供对当前和历史终端数据的访问。与其他工具不同的是，该解决方案只存储与取证相关的数据，这意味着安全团队不必花时间去筛选堆积如山的不相关信息来寻找威胁。

- **上下文驱动的威胁检测（静态、机器学习和自定义规则）**

BlackBerry® Optics 的能力来自于其独特而高效的威胁检测和响应能力。不像其他的 EDR 产品依赖于基于云的分析来发现威胁和做出响应，BlackBerry® Optics 将所有检测和响应决定推到终端，消除响应延迟。上下文分析引擎使安全分析师可以从各种默认检测规则中选择，包括映射到 MITRE ATT&CK 框架的规则包，或者创建满足特定业务需求的自定义规则。分析师还可以选择将机器学习威胁检测规则部署到终端，以发现那些很难用静态规则发现的威胁。

- **采取应对措施**

BlackBerry® Optics 为企业提供完全集成的自动事件响应能力。每条规则都可以配置一个响应方案，以便在规则被触发时自动启动一组离散的响应任务。响应方案驱动的响应能力帮助企业消除停留时间，确保在整个环境中对威胁做出快速和一致的响应，而不论安全人员的技能水平如何。如果检测到攻击，BlackBerry® Optics 可以自动发起响应，无需人工干预。如果需要进一步响应，则可以隔离相关项目并锁定终端，从而禁用其与任何其他终端通信的能力。该工具可以从受影响的终端收集数据，以进一步了解事件的背景。

【方案优势】

	传统 EDR	BlackBerry® Optics	优　势
安全方法	提供检测和响应	提供持续的威胁和事件预防	基于预防的方法减少了需要采取行动 / 分析的事件总数
所需技能	需要高级安全分析师技能	专为各种技能和经验水平的安全分析师打造	所有人都可以使用的解决方案，扩大了可以管理该解决方案的可能人才库
收集的数据	将所有终端活动持续地流向云端，或将其发送到专用硬件上	仅在本地收集和存储与安全相关的数据	在本地仅收集与安全相关的活动数据可显著减少责任并提高合规性
数据存储	不断地将数据流传到云端或在本地硬件上进行汇总	在每个终端本地存储数据	在本地存储数据可显著减少责任、提高合规性并优化性能和可扩展性
威胁检测技术	需要编写并不断扩充个人行为规则，以保持从云端运行的覆盖水平	将行为规则与 ML 威胁检测模块相结合，以提供更大且不断增加的覆盖范围，并在终端本地运行	消除了必须由安全专家创建和维护的多达数千条规则的需要
威胁搜寻	需要丰富的专业知识来配置和执行多种搜索功能	提供易于配置的搜索条件和优化的终端响应数据收集	提高发现难以发现的威胁的能力
根本原因分析	梳理收集的数据以确定活动威胁进入环境的位置，以确定如何阻止持续的破坏	使用阻止威胁时收集的数据来了解恶意行为者选择的攻击向量	自动化方法缩短了分析完成的时间

【应用效果】

BlackBerry® Optics 已在医疗、制造等多个领域广泛应用。例如某大型跨国制造企业通过在其企业内部 45000 个终端上运行 BlackBerry®PROTECT 和 BlackBerry® Optics，获得了更高级别的威胁防护和更快的威胁响应，提高了对员工下载的可见性和控制，使该企业节省了 840 万美元的费用，同时减少了 190 万美元的安全漏洞成本，并节约了员工 95% 的攻击响应时间。

三、身份与访问安全

• 身份与访问管理（IAM）的重要性

在当今全球化和高度互联的商业环境中，业务变得更有生产力和效率。但另一方面，企业成为数据泄露或其他网络威胁受害者的可能性也在不断增大。对于许多企业而言，决定谁应该访问哪些信息是一项艰巨的任务，而将这个问题搁置不管则会使系统变得非常脆弱。这就是为什么不应低估一个成熟的 IAM 策略的重要性。分析公司的研究报告称，超过 70% 的组织没有认真对待 IAM。这意味着这些组织遭受数据泄露的风险是采用 IAM 策略的组织的两倍。研究报告还表明，IAM 方法越智能，安全风险就越小。

如上所述，对于许多组织而言，IAM 是其网络安全工具集中的关键武器。它作为一个很有效的解决方案，可以减轻数据泄露以及管理远程工作和自带设备 (BYOD) 带来的额外风险。IAM 涉及跟踪 IT 环境中每个人和资产的行为和行动，特别是系统管理员和关键任务资产。IAM 使个人能够以适当的理由在正确的时间访问正确的资源，这需要大量的系统集成，以便所有平台都具有正确执行策略所需的态势感知。如果部署得当，IAM 可以显著提高可见性和安全性。

然而，IAM 中存在的一个典型问题是，用户需要根据其在组织中的角色被授予访问权限，但员工很少只拥有单一的角色，它们可能需要一次特殊的访问，或者每个拥有相同角色的人可能需要略微不同的访问类型。这导致了非常复杂的情况，通常需要许多部门之间开展协作。因此，完善的管理涉及到一个组织的各部门员工。这可能会由于存在大量的技术数据、困难的决策过程以及与他们的日常工作缺乏相关性，而导致人们遭受所谓的“安全疲劳”。但管理不善的 IAM 基础设施将给企业带来灾难性的后果。

• AI如何赋能 IAM

人工智能和机器学习技术能为实现IAM的有效管理带来帮助，可以使企业从过于技术化的访问管理方法发展为企业内部各个层面都能理解的访问管理形式。为了实施基于风险的IAM方法，企业将需要由机器学习驱动的高级身份分析。整个行业的最佳实践已经证明，基于ML的身份分析能够显著改善IAM架构。

- **先进的分析技术**

  将分析与AI相结合，可以提供更多的上下文洞察力，使技术和非技术员工可以更有效率地工作。AI技术提供了学习新的见解和自动化流程的方法，能够大幅加快现有的IAM合规控制。他们可以检测异常情况和潜在威胁，而不需要安全专家。这为员工提供了做出正确决定所需的信息。

- **精准的访问控制**

从生物识别密码出发，不难想象，AI可以通过视觉和听觉以额外的安全性来识别用户。通过视觉和听觉的线索，机器将能够理解并确认一个人是否是他们所声称的人，而不是根据预先定义的证书进行检查。它还可以学习何时授予访问权限，并采取相应的行动。

在用户的访问权限内，AI系统也可以实时监测任何不寻常或不合理的行为。他们可以检测到用户是否试图访问他们通常不会访问的系统的某个部分，或下载比通常更多的文件。还可以通过观察用户的键盘和鼠标移动的节奏，以确定不规则或不寻常的模式。这些安全策略使企业能够安全地开展业务。

- **自动化和灵活性**

AI 能够监控用户操作的细微细节，因此可以针对低风险访问情况自动进行身份验证，从而减轻IT部门IAM管理的部分负担。在授予网络访问权之前考虑这些细节，使IAM具有情境性和颗粒性，并能控制由不适当的配置或取消配置引起的潜在问题。由 AI 驱动的系统能够根据需求和情况将适当的 IAM 策略应用于任何访问请求，这样 IT 部门就不必浪费时间为每个用例找出 " 最小特权 " 的基本要素，或解决特权变动的问题。

· 企业案例

Acceptto：eGuardian——无密码连续认证平台

【场景描述】

在身份持续受到攻击的世界中，用户的凭据有被盗用的风险。由于目前的安全解决方案无法满足企业需求，系统漏洞和威胁成本急剧上升，包括：双因素身份认证在传输过程中很容易被拦截；当前的多因素身份认证 (MFA) 安全解决方案缺乏上下文，并且依赖的属性太少；虽然指纹、人脸或视网膜识别等生物识别技术看起来很安全，但他们也很容易被欺骗。

在网络安全方面，企业必须考虑三个主要因素：脆弱性、业务影响和成本。当使用不安全的二进制身份验证方法时，脆弱性会增加；而当使用更多的工具导致访问变得不方便时，对业务的影响就会增加。成本则会伴随着脆弱性和对业务的影响而累积。同时，控制的有效性会随着时间的推移而降低。因此，将身份认证视为连续过程的技术对于防御威胁行为者日益先进的策略至关重要。

Acceptto's Passwordless Continuous Authentication™对用户行为进行建模和分析，以推断访问尝试在认证前、认证中和授权后是否合法，同时还能减少对业务的影响和成本。Acceptto 是同类产品中第一个提供持续身份认证的产品，市场上没有其他解决方案可以在授权后验证用户的身份。Acceptto 的 AIML 方法利用上下文和行为，在每个应用环境中创建一个丰富的用户档案，取消了对密码的需求。

【技术方案】

Acceptto 采用现代数据科学方法进行无密码身份认证和异常检测，利用以流为核心的现代系统设计和架构，将认证视为一个连续的过程。简言之，传统的身份认证类似于公司门口的守卫，而 Acceptto NGA（下一代认证）则是遍布公司所有走廊的安全摄像头，昼夜不停地进行记录。

Acceptto NGA 执行行为验证，通过进行异常检测和共性分析，检测出使用被盗凭证的冒名顶替者。与数据流相结合，该系统能够保持实时更新，绕过主流控制所带来的效能下降。通过 AI 模型对数据流进行预测，能够在低延迟下检测冒名顶替者。

Acceptto NGA 可实现的能力包括：

- **动态风险得分**

根据 AI/ML 算法计算的风险分数，计算动态保障水平（LoA）。该方法可为每笔交易自动找到最佳策略，以最大限度地提高安全性，同时利用机器学习和人工智能分析，最大限度地减少对用户体验的影响。这在不牺牲安全认证的前提下提供了更顺畅的用户体验。

- **无密码认证**

Acceptto 的 It'sMe 移动应用程序通过智能 MFA 认证所有用户，以随时随地授权访问应用程序，防止黑客窃取身份并访问账户和数据。

- **数字DNA**

设备和浏览器指纹是一个信息集合，用于唯一识别试图访问的远程设备。对于 DBFP 支持的网站上的每一次离散访问或用户验证尝试，Acceptto 收集各种参数的信息，包括用户代理、浏览器类型和屏幕分辨率。这些数据点形成一个整体，以描述每个特定的访问者，使企业能够确保智能和安全的认证，并通过使用多因素认证来消除潜在的威胁者。

- **平台会话管理 (PSM)**

Acceptto 的平台会话管理 (PSM) 功能 提供对企业网络特权会话的 360°洞察力，获得对所有远程特权会话的前所未有的可见性，同时仍然保持管理员完全控制。

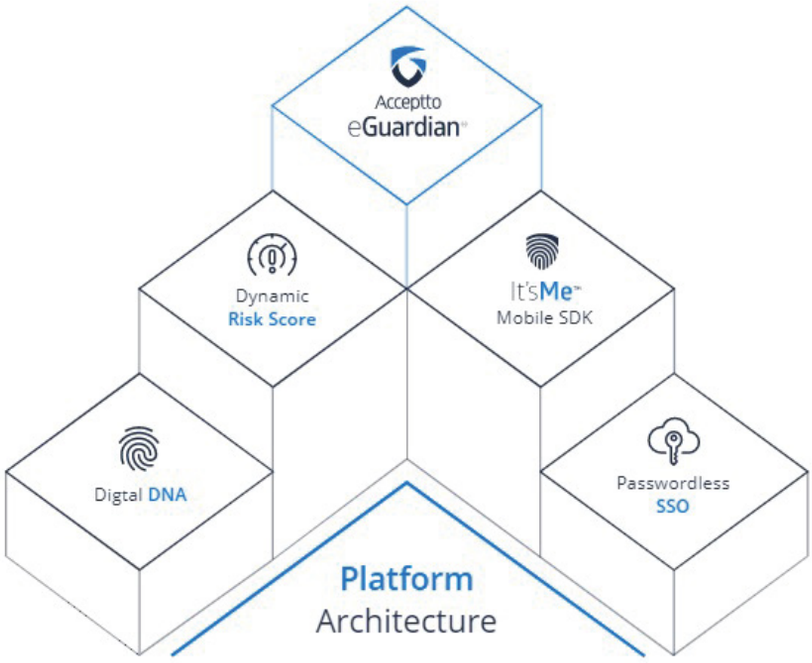


图5 eGuardian平台架构

【应用效果】

Acceptto 在全球拥有 2000 万用户，每天保护 300 万次认证，阻止了 97% 的账户接管攻击。企业通过采用 Acceptto 的无密码解决方案，减少了与用户管理相关的 TCO，满足了员工的灵活性需求，提高了终端用户的工作效率，并在不影响终端用户的情况下提高了公司的安全性，并大大降低了安全成本。利用 Acceptto 身份验证风险引擎，企业减少了 98% 的用户帐户接管 (ATO) 欺诈损失。

四、应用安全

• 应用层安全风险

随着我们的生活越来越数字化，企业的软件资产组合在各个垂直领域都在增加。当今，由于代码和架构复杂性的增加，为所有这些软件应用程序建立和维护安全性变得非常困难。因此，我们看到源自应用层的漏洞和攻击数量大幅增加。数据显示，目前高于 75% 的网络攻击发生在应用层。

根据 OWASP Top 10，最普遍的应用层安全风险包括窃取敏感数据、安全错误配置和跨站点脚本攻击。通过利用这些工具，黑客能够破坏敏感数据、获得对系统数据或功能的未经授权的访问、窃取凭证、向受害者发送恶意软件等等。

同时，大多数传统的安全解决方案仍然主要在网络层提供保护，侧重于基于签名的入侵防御和检测，对应用层发生的攻击几乎没有影响。这种差距要求提高对软件应用程序安全性的认识，并需要使用机器学习等新技术来弥补传统解决方案的不足。

• AI如何赋能应用安全

当前，网络攻击变得愈发自动化和智能。牛津大学最近牵头进行的一项联合研究显示，黑客正越来越多地使用人工智能技术取代人工，并引入全新的威胁场景，导致现有威胁的扩大。这显著提高了攻击的有效性和效率，并且归因变得越来越困难。攻击者现在拥有造成更多伤害的工具。这引入了将机器学习应用于应用安全的强烈需求。

机器学习在应用安全中的另一个强大应用是防御零日漏洞。防范零日攻击至关重要，因为它们很少被立即注意到。发现和解决这些漏洞通常需要几个月的时间，与此同时，会导致大量敏感数据的泄露。机器学习不仅可以基于规则识别恶意行为，还可以通过识别异常数据迁移并帮助发现异常值，从而提供一种防范零日攻击的方法。

AI的具体用途

异常检测

异常检测首先需要定义正常行为，然后将所有其他行为识别为异常，从而识别潜在威胁（类似于白名单）。

恶意行为检测

与异常检测相反，恶意行为是通过对带标签数据进行训练来识别的。通常，所有未被归类为恶意的流量都被允许通过（类似于黑名单）。

数据探索

数据探索用于识别数据的特征，通常使用可视探索，它可以作为异常检测或恶意行为检测的基础，也可以通过增加传入请求的“可读性”直接协助安全分析师。

风险评分

风险评分可用于评估某个用户的行为是恶意的概率，这可以通过分配绝对风险评分或根据用户是不良行为者的概率对用户进行分类来完成。

• 企业案例

NeuraLegion：NexPloit——AI赋能的新一代动态应用安全测试

【场景描述】

当前，绝大多数的网络攻击发生在应用层。应用层的漏洞越来越多，极易成为攻击者的目标，主要原因如下：

- **开发人员在编写代码时总会犯错误，不管他们的经验如何，也不管他们对安全性的了解程度如何。**
- **对快速交付软件的需求导致安全测试的缺乏。**安全测试通常非常耗时、消耗资源，并且在大多数情况下是不切实际且通常无法克服的任务，这导致大量的数据测试被跳过，企业被迫发布相对安全、但并非完全安全的应用程序，以满足紧迫的业务时间要求。
- **新技术（云计算、无服务器计算等）仍在成熟和不断发展。**由于很少有人拥有必要的技术专长和正确使用这些新技术的经验，持续保持跟进这些技术的最新进展变得越来越困难。这导致即使在使用最佳实践的同时，错误仍然经常发生。

为了克服这些挑战，组织必须利用多种工具和方法来加强其应用安全。理想情况下，如果我们能够塑造完美的开发人员，他们将有能力来预测软件将遇到的每一个可能的场景，包括来自其他人编写的通信或集成软件的任何场景，并编写代码以有效覆盖并针对这些情况提供安全保障。不幸的是，这是不切实际的。最重要的因素是跨开发和安全的“规模”问题，特别是人类无法学习和掌握大量不同的技术和方法，从而有效地生成 100% 安全的应用程序。

然而，由于需要测试的数据量不断增长和加速，而且其复杂性不断增大，使用简单的依赖于手动编码的自动化是不可能的。因此，需要一种开发自动化的“自动化方式”，这就是 AI。新的 AI 解决方案正在成为自动化发展的重要新步骤，能够应对常规的、基于启发式的自动化无法解决的挑战。

【技术方案】

NexPloit 通过使用机器学习算法来理解复杂的软件架构和 API，然后利用这些知识来检测软件逻辑流和单点故障（例如 API 本身）中的安全漏洞。使用最先进的进化策略，NexPloit 可以生成复杂且不断变化的恶意场景，执行真正的新攻击，而不仅仅是通过手动更新已知漏洞的数据库。

NexPloit 的强化学习特性使其能够利用每次攻击的影响，从每个目标中学习并适应每个目标。通过只将真正的影响视为真正的漏洞，误报被完全消除，无需安全专家重新检查和过滤每个扫描报告。此外，每个新漏洞都会自动存储在自己的知识库中，从而提高未来扫描的效率。

图6 检测界面

【方案亮点】

- **零日漏洞和业务逻辑流漏洞检测**

NexPloit 的 AI 引擎除了能在组织的应用程序中查找所有 OWASP Top 10 Plus 技术漏洞之外，还可以自动检测未知的零日漏洞和业务逻辑流漏洞，减少冗长且昂贵的手动测试，通过无误报报告和修复指南节省组织的时间和金钱。

- **以DevOps的速度通过安全测试实现自动化和扩展**

NexPloit 能无缝集成到 SDLC 中。作为唯一自动检测零日漏洞的解决方案，NexPloit 的无误报报告是实时生成的，具有精确的代码检测，使组织的 DevOps 能够达到最高安全标准，而不会降低开发速度或敏捷性。

- **易于使用**

NexPloit 是一个基于云的解决方案，不需要昂贵的集成或由安全人员进行复杂的配置。只需登录 > 上传 > 扫描，或使用开放 API 实现完全自动化。

- **无误报**

作为唯一真正的无误报解决方案，NexPloit 只报告经过验证的可利用漏洞，消除了重新检查和过滤每个扫描报告的需要。

- **自动化渗透测试**

NexPloit 可使用应用程序层的自动化渗透测试减少对冗长、昂贵和侵入性的手动测试的需要，并可根据需要验证组织的网络安全状况。

- **最高合规标准**

NexPloit 能提供全面的测试以及识别漏洞的即时报告，简化组织的合规性验证过程。NexPloit 支持 ISO27001、PCI DSS、HIPAA、NIST 800 系列等相关标准和法规的应用安全测试。

【应用效果】

企业反馈，NexPloit 能以快速、高效和简单的流程为他们提供全面的合规报告，并无需进行任何复杂的集成。使用 NexPloit，企业能在更短的时间内获得最佳结果，并节省了手动合规服务的大笔费用。企业在使用 NexPloit 执行的安全测试中，能达到最高的应用程序覆盖率。此外，NeuraLegion 在提供支持和调整产品以满足企业的特定需求方面非常迅速。同时，NexPloit 使用的简单性和完全的自动化使企业能够真正开始将应用安全测试向开发靠拢。

五、数据安全

• 数据安全风险加剧

在过去的十年中，数据为我们最有价值的技术和进步提供了动力。但是如今，保护消费者数据隐私成为一个重大挑战。当前，数据的增长速度比以往任何时候都要快。数据显示，每秒有超过 1.7 兆字节的新数据产生。企业不仅要保护客户的个人信息，而且要保护敏感的个人信息。然而缺乏足够的安全措施继续使组织面临数据泄露的风险。

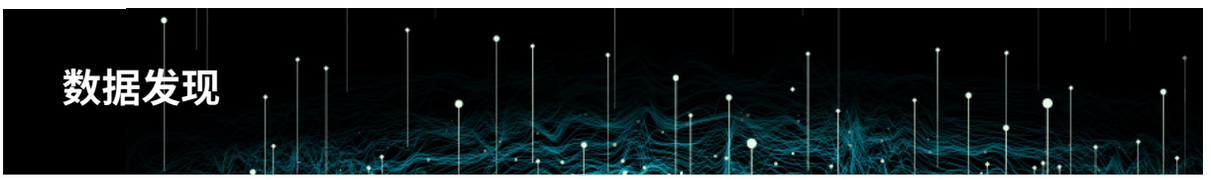
近年来，数据泄露事件频频发生。根据华盛顿州总检察长办公室发布的年度《数据泄露报告》，影响华盛顿州居民的数据泄露在 2020 年几乎翻了一番。同时，数据泄露事件也给企业组织造成了高额的财产损失。IBM Security 发布的《2020 年数据泄露成本报告》显示，数据泄露事件给企业造成的平均成本为 386 万美元。如果组织遭到攻击造成数据泄露，他们将面临来自监管机构的严厉处罚。例如，根据 GDPR，在欧盟境内运营或使用客户数据的公司如果因缺乏安全控制而遭受大规模数据泄露，可能面临高达总收入 4% 或 2000 万欧元的罚款。因此企业组织必须对许多关键安全技术进行投资，例如数据归档、备份和冗余基础架构，以确保其数据受到保护并且可以恢复。

根据 Verizon 发布的《2021 年数据泄露调查报告》，网络钓鱼、勒索软件和 Web 应用攻击是 2021 年数据泄露的三大主要原因。其中，Web 应用攻击导致了去年 39% 的数据泄露事件。Verizon 发现，61% 的数据泄露与凭证数据有关，人为疏忽依然是安全的最大威胁。同时，IBM 发布的报告也显示，凭证被盗或受攻击以及云配置错误是导致恶意数据泄露事件的最常见原因，占比近 40%。

• AI如何赋能数据安全防护

AI 在数据安全领域有多种用途，同时可以显著降低企业数据泄露成本。根据 IBM Security 发布的《2020 年数据泄露成本报告》，智能技术可将数据泄露成本降低一半。通过完全部署安全自动化技术，企业响应数据泄露所需的时间大幅缩短，这是降低数据泄露成本的一个关键因素。该报告显示，人工智能、机器学习、分析和其他形式的安全自动化技术使得完全部署了安全自动化技术的公司对数据泄露的响应速度比尚未部署安全自动化的公司要快 27% 以上，后者平均需要多出 74 天才能识别并遏制数据泄露。

AI在数据安全领域的具体应用包括：



数据发现是所有类型数据管理的基础——从网络安全到数据隐私再到数据治理。为应对当今的数据挑战，企业需要一个多管齐下的策略来准确识别组织中所有类型的敏感、关键和个人数据。通过应用机器学习和数据关联技术，组织可以更准确地识别所有类型的敏感数据，并能够理解数据上下文和关系（而不是孤立地看待数据点）。首先，通过正则表达式（RegEx）寻找和识别敏感数据。这种技术通过模式匹配，可以知道需要识别的敏感数据的确切格式，例如银行账户、电子邮件地址、身份证号码等。第二步是利用机器学习技术和基于上下文的分类器发现传统上更难定义的个人信息（PI），例如出生日期、投票趋势、名字、姓氏、居住地等。通过增加相关性，不仅能够发现暗数据，而且能够揭示敏感数据之间的关系。

当前，越来越多的公司需要对基于上下文的敏感数据进行分类。此外，在数据泄露预防、数据库活动监控和数据访问治理产品中占主导地位的基于正则表达式的分类器，往往只能在有限的数据源上操作，如关系型数据库或内部非结构化文件。这些都需要一种新的分类方法，能够识别所有现代数据存储中的上下文敏感数据，包括非结构化、结构化、半结构化、大数据、云和企业应用（如 SAP）。利用机器学习和上下文智能技术，可以为隐私保护提供先进的数据分类、归类、编目和关联能力。其次，利用机器学习还可以基于相关性和关联性匹配不同类别的关联数据，这种基于关联性的分类对隐私保护至关重要。此外，利用机器学习还可以整合元数据分析，为数据及其使用情况提供更丰富的视角。这种元数据输入可用于更好、更自动地对数据进行编目，以便通过搜索更容易地发现数据，并衡量敏感度风险。

AI 和 ML 可以加强传统的数据泄露预防(DLP)解决方案，以大大降低数据泄漏的风险。当前，企业经常需要安全专家来处理 DLP 解决方案中无法处理的更复杂的安全决策。将 AI/ML 添加到 DLP 解决方案中，可以通过将决策自动化，创造更高的效率。此外，AI/ML 驱动的 DLP 解决方案可以根据使用或行为模式自动阻止或关闭特定的高风险用户，防止数据泄露。数据模式也可以跨地域、部门或流程进行实时扫描，使公司能够识别薄弱环节。再者，AI 和 ML 可以同时筛选和分析大量的数据，比任何人类或传统的 DLP 解决方案能更快、更准确地检测威胁事件。

• 企业案例

观安：基于文本识别的敏感数据发现方法及系统

【场景描述】

数据是企业运营的支撑基础，也是企业信息系统的核心部分，一旦数据相关的管理及应用系统出现问题，将严重影响企业形象和发展，因此数据安全问题一直是企业备受关注的主题。目前实际应用中数据保护方案主要有数据隔离、权限设定、数据脱敏等。对数据的保护方案中，敏感数据的保护尤为重要，敏感数据保护方案的核心部分就是从海量的数据中挑选出敏感数据，完成对敏感数据的精准识别。

目前敏感数据的识别主要依赖于字典匹配方法和人工识别的方法。字典匹配方式主要是数据进行逐一匹配，这个需要先能提供字典，适用范围有限，而人工识别方法主要依赖于风险评估敏感数据识别精准度低且会造成分类结果干扰。

针对以上问题，观安发挥创新、研发优势，设计基于文本识别的敏感数据发现方法。该方法基于多种机器学习算法模型的敏感数据发现系统，能够从样本中自动提取特征并进行分析检测，有效提升敏感数据识别的精准度，并对未知的敏感数据有一定的检测识别能力，既减少了识别规则定义方面的人工投入，又可以通过持续使用新样本进行迭代训练的方式实现模型检测能力的自动提升。

【技术方案】

架构设计

图7 敏感数据发现系统架构设计

敏感数据发现系统主要包含三个场景，分别为**数据梳理场景、分类分级场景、合规分析场景**。

1. 数据梳理场景

通过对客户的数据资产进行梳理、我们可以摸清客户的数据资产中敏感数据的分布情况，这为后面保护这些敏感数据提供数据支撑，也为之后对这些数据进行分类分级做好充分的准备。

2. 分类分级场景

为了更好的将数据保护起来，一些行业发布了分类分级指引，敏感数据发现系统可以根据这些分类分级的指引将客户数据分类分级，对于一些没有发布分类分级指引的行业，也可以通过自定义分类分级，将客户数据分类分级，这样可以将这些数据管控起来，达到保护的效果。

3. 合规分析场景

由于一些业务要求，客户的一些敏感数据会经过处理后外发出去，合规分析场景就是对这些数据进行分析处理，看是否满足合规要求，如果不满足合规要求，则需要对数据进行整改，这样可以防止敏感数据泄露。

· 技术方案优势

基于文本识别的敏感数据发现方法是敏感数据发现系统的核心技术。

在深度学习方法流行之前，对于文本类型数据的检测主要是以手动提取特征为主，通过设定滑动窗口提取分类，最后汇总为文本区域。本方案利用深度学习方法实现文本的标注，并结合相关算法训练分类模型，能够有效提高敏感数据识别的精度与速度，其具体实施过程如下。

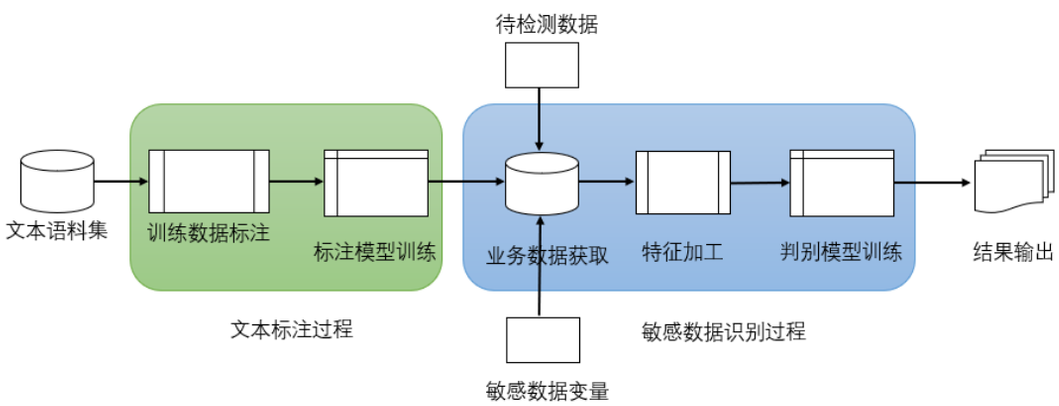


图8 敏感数据识别实施过程

1. 文本标注过程

- 对获取的大量语料集进行关键词标注，生成训练样本；
- 结合 Bi-LSTM 以及 CRF 构建文本标注模型，获得每个词的类别。

2. 敏感数据识别过程

- 获取业务数据及待检测数据集，以及敏感数据变量集合，针对业务数据进行特征加工；
- 识别方法采用 catboost 算法，建立判别模型。利用加工好的业务数据进行训练，最终可应用于识别待测数据集中的敏感字段及其敏感类型。

【应用效果】

本案例可应用于国内大量重要客户的敏感数据保护，包括运营商、金融、医疗、教育、能源、电力、国家机关等。这些客户由于大量的业务系统产生了庞大的敏感数据，要将这些敏感数据很好的管理起来极其困难。借助基于文本识别的敏感数据发现方法，可以很好的识别分散在各处的敏感数据，帮助客户将这些数据很好的管理起来，并方便做相应的安全策略。

六、物联网安全

• 物联网安全挑战

物联网 (IoT) 可以广义地理解为配备传感器和 IP 地址以通过互联网进行通信的计算设备的全球网络。物联网设备的用途非常广泛，超出了任何行业安全解决方案的范围，使得实现物联网设备的安全性及其具有挑战性。物联网由种类繁多的设备组成，每个设备都有各自的操作系统和多种安全漏洞。这种异构性使得物联网网络难以用单一的防御源来覆盖。此外，由于物联网设备的设计成本低，且它们通常是非常低功率、高能效的设备，因此没有内置或仅有极少的安全框架，通常只配备一个简单的密码，这使得物联网设备极易受到黑客攻击。物联网由数以千计，甚至数以百万计的设备组成，通过互联网向其提供数据，这使整个网络安全的实施具有非常高的操作复杂性。即使是为了满足最低要求，安全人员也需要确保所有操作系统、网络应用程序的定期更新，衡量安全风险，检测潜在目标等。这就是安全专家寻求利用人工智能对抗物联网安全威胁的原因。

• AI如何赋能物联网安全

AI可以帮助物联网设备在很短的时间内解析难以想象的数据量。对于拥有大量物联网设备和传感器的组织而言，AI和物联网的结合可以为组织提供更大的可见性和控制力，换言之，AI可以使企业将通过物联网收集到的数据转化为更有价值的信息。这对保护设备和网络免受未经授权的访问和渗透尤为重要。

为物联网构建安全框架首先需要识别网络上的所有设备。对于拥有数百万传感器和设备的大型网络来说，这是一项艰巨的任务。然而，利用AI，发现过程会变得更加容易，并可以提供关于设备性质的全面、详细的信息。有效的网络安全是通过识别和监控网络中每个节点实现的，AI的这种识别和资产管理能力使其可以极大地赋能物联网安全。

其次，AI还可以通过数据分析赋能物联网安全。AI在对庞大的物联网网络进行持续监控、检测异常方面，将比人类更有效。但同时，AI也会导致许多误报的情况，因为任何异常都可能被认为是一个潜在的攻击。这可以通过使用ML训练AI识别攻击模式来解决，并通过其他异常行为减少误报。

机器学习在识别潜在威胁、发现网络中的漏洞和识别系统性 IoT 漏洞（例如 IoT 设备上缺少密码保护或密码保护较弱）以及解决网络配置来构建安全防御方面非常有用。ML 基于海量网络安全数据集和 IoT 设备配置文件，在对抗 DDoS 攻击和改善物联网网络的整体安全状况方面非常有效。机器学习的数据还可以

帮助物联网开发人员创建更安全的设备。通过及早识别漏洞，开发人员会在可能的情况下发送安全补丁，或创建新版本的设备以更好地保护用户。由于大多数现有的物联网设备缺乏有效的加密和安全框架，机器学习可以非常有效地在网络级别提供适应性强且灵活的物联网安全措施。

• 企业案例

Extreme Networks：ExtremeAI™ Security 利用AI来识别和修复针对物联网设备的高级威胁

【场景描述】

当前，多云、移动性和企业中大量涌入的物联网设备扩大了攻击面，使得企业必须在网络深处部署高级别安全措施，而不仅仅是在外围部署，才能实现有效的安全性。终端和网络流量的这种爆炸式增长带来了极大的复杂性，使网络管理员和安全团队难以通过传统解决方案获得对混乱局面的可视性。由于物联网设备从价值百万的智能 MRI 机器到价值五美元的传感器不等，单靠设备层面的安全是无法保证终端安全的。这给企业安全团队带来了很大的工作负荷。

Extreme Networks 利用创新的机器学习技术，将强大的流量分析和可见性功能嵌入到 ExtremeAI™ Security 解决方案中，可以帮助企业识别和修复物联网中的网络威胁。

【技术方案】

ExtremeAI™ Security 提供对恶意流量的深入可见性和检测，并实时监控物联网设备的行为异常，照亮企业网络，使攻击者无处藏身。通过对可疑设备和流量的全自动修复，ExtremeAI™ Security 能够确保威胁在没有人工干预的情况下得到控制，防止它们在网络中移动。

ExtremeAI™ Security 的主要功能包括：

行为监控和基线	利用机器学习理解物联网设备的典型行为，实现大规模可扩展的行为异常检测，并在终端以异常方式运行时自动触发警报。

无监督学习

零接触、零配置的方法使ExtremeAI™ Security易于实施。先进的机器学习算法从多个角度自动学习物联网设备行为，而不需要在大数据湖上进行监督训练，并使用零足迹建模技术对设备进行建模（每个设备需要<5K的存储）。

洞察力和精准分析

通过利用ExtremeAnalytics™（端到端分析应用程序），用户可以深入了解恶意流量的横向移动以及对关键网络服务产生的影响。通过分析平台，用户可以按严重程度、类别、高风险终端和地理位置查看威胁。

多厂商互操作性和集成

ExtremeAI™ Security 与所有领先的威胁情报源合作，并与 Extreme Workflow Composer 紧密集成，实现了威胁自动缓解和修复。自动标记功能与各种普遍使用的 IT 工具（如 Slack、Jira 和 ServiceNow）集成，整个解决方案与许多安全工具可以互操作。

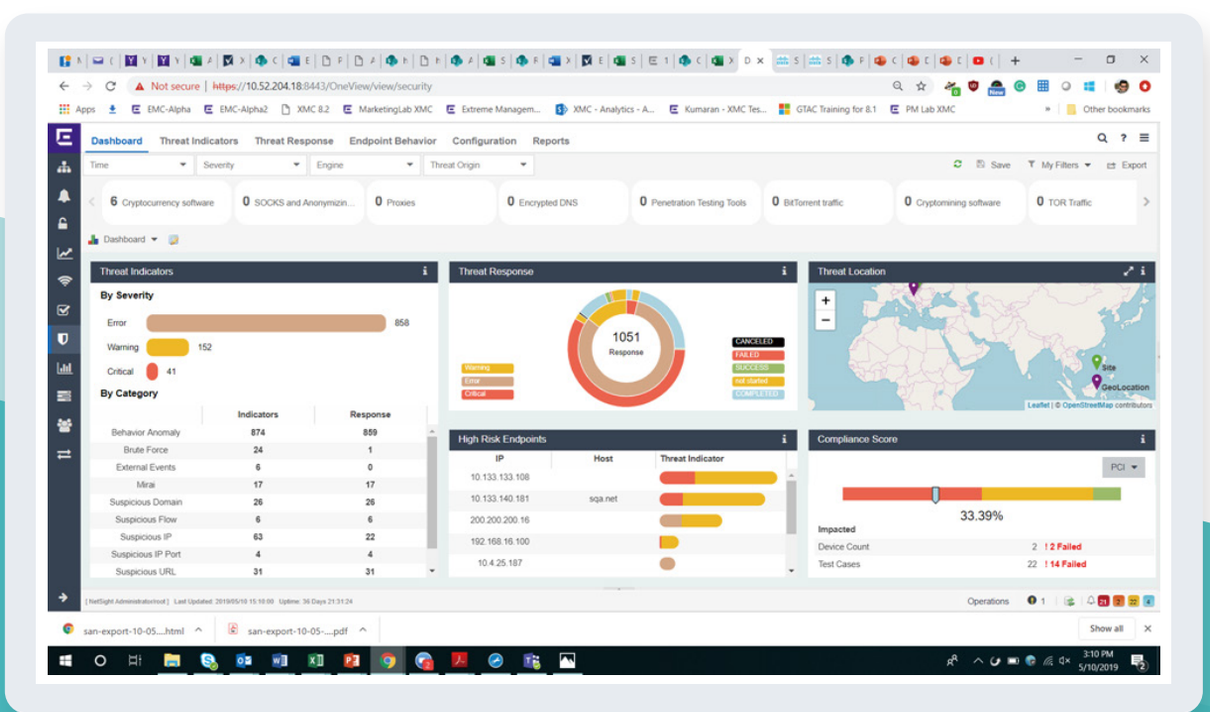


图9 ExtremeAI™ Security application安全分析界面

【应用实例】

• 物联网异常行为检测

安全分析的一个新兴领域是异常行为检测，ExtremeAI™ Security 利用 ML 来学习物联网终端的预期行为，并在终端的行为超出常规时触发警报。如下图所示，Extreme 提供了一个直观的三维可视图，显示被追踪的每个物联网终端的当前行为。每个彩色点对应一个物联网终端。当一个物联网终端首次上线时，ML 算法开始建立它的行为信息。如果两个终端经常有类似的行为，它们对应的点就会显示在彼此的附近。通常情况下，在网络上表现出类似活动的同一类别的设备（如温度传感器、闭路电视监控摄像机或工业自动化设备），将倾向于“聚集”成小群。

该系统不断收集信息，一旦收集到足够的数据（通常在一周左右），终端行为模型就会被更新，他们相应的 3D 点会相应地移动。例如，一旦识别出威胁指标，系统可以进行数据包捕获以收集更多信息，生成 JIRA 标签以通知安全操作员，并自动隔离网络终端以限制对网络其他部分的损害。

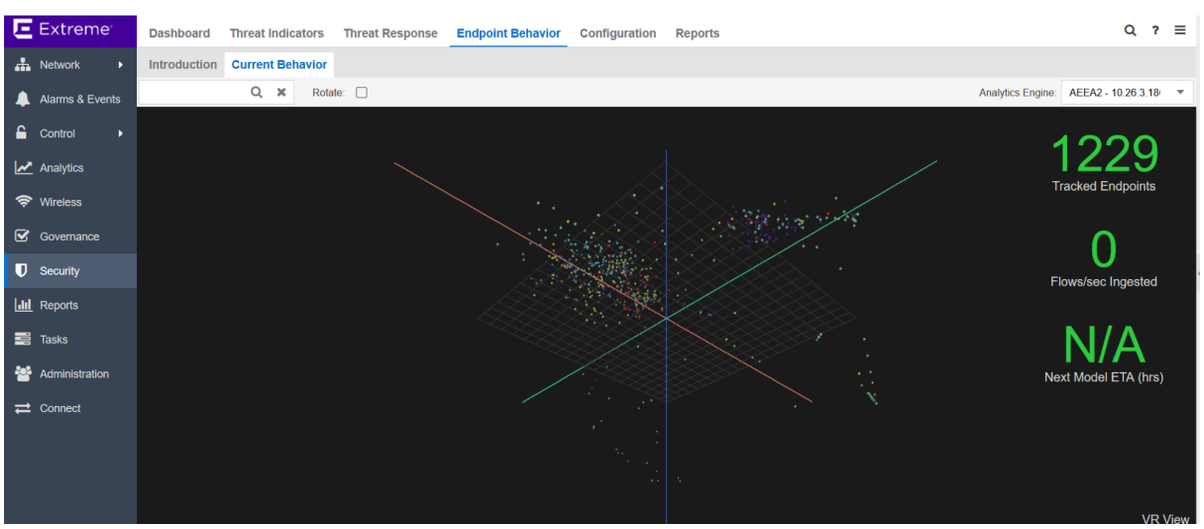


图10 ExtremeAI™ Security application网络终端安全试图

七、移动安全

• 移动安全风险加剧

企业移动性是当今组织的新规范。在企业采用这种新的生产力工具为企业带来好处的同时，针对这些设备的攻击比任何其他目标的增长都要快。在今天的组织中，访问企业数据的终端中 60% 为移动终端，但其中大多数都没有任何安全保护。随着攻击者越来越具有创造力，利用移动安全性和可见性的不足，移动威胁和攻击的数量和影响都在增加。事实上，移动威胁的数量正以每年 100% 以上的速度增长，到 2019 年将占所有威胁的 30% 以上。

据《Zimperium 企业移动安全态势报告 2019》数据显示：

- 24% 的企业移动终端面临设备威胁。
- 68% 的恶意配置文件被认为是“高风险”的，这意味着他们有更高的访问权限，可能导致数据泄露。
- 超过一半的企业移动终端连接了危险的网络。
- 网络（Network）攻击占本报告所涵盖的所有攻击的 92%。
- 19% 的企业移动终端遭遇过网络攻击。
- 13% 的企业安卓设备检测到了恶意应用。
- 85% 的 iOS 应用和 21% 的安卓应用的隐私评分不及格。
- 71% 的 iOS 应用和 68% 的安卓应用没有通过安全等级。

当前安全工具的痛点：

1. 基于云的威胁检测会延迟威胁响应

基于云的威胁检测解决方案为移动威胁防御提供了一种方法，但它们也有一些缺点。首先，这些解决方案可能会延迟对设备威胁的响应时间，因为它们首先必须扫描设备，然后将数据从设备上移出，并将其发送到云端。然后，云端代理必须将潜在的安全入侵行为通知 UEM 解决方案，并向设备上的 UEM 代理发送响应指令 -- 从设备扫描到威胁响应，这是一个漫长的五步过程。

作为这个过程的一部分，敏感数据（比如用户的位置）被传输到云端。这种数据传输可能容易受到中间人攻击，从而阻止在设备上执行威胁响应。其次，通过此过程提取敏感数据还可能违反许多数据保护法规，例如 GDPR。此外，目前基于云的解决方案主要检测已知的威胁。由于它们不使用机器学习算法，因此它们不太擅长发现未知的零日威胁。现代的网络攻击者已经非常擅长改变威胁的某些部分以避免被发现，所以移动威胁防护解决方案必须能够在这些异常发生时立即检测到威胁。

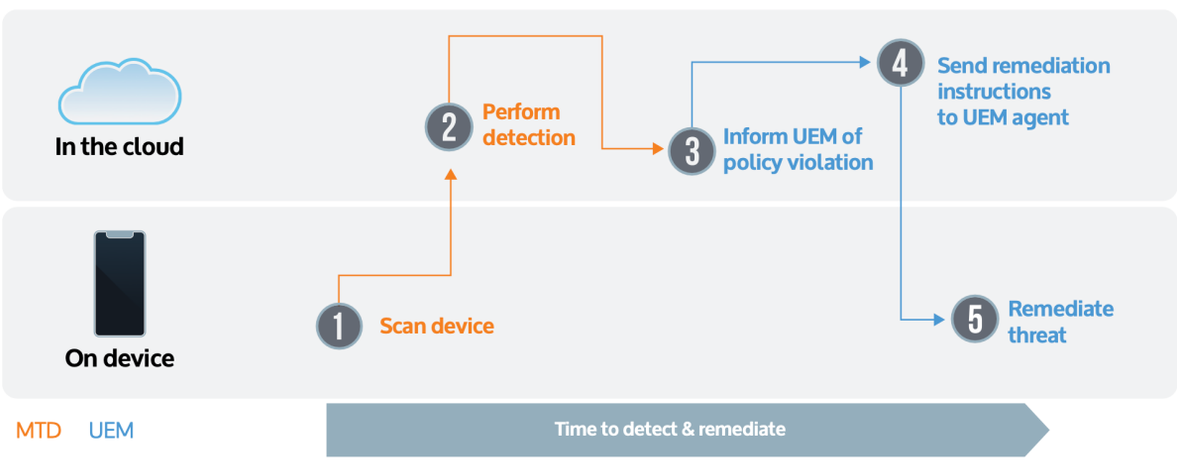


图11 基于云的威胁检测和响应过程

2. 传统的反病毒工具采用被动方式

为了应对威胁，传统的安全方法大多使用签名和沙箱。这些技术首先需要识别和获得攻击的样本。然后，安全专家必须在沙盒中运行该攻击以创建一个签名。这远远不能满足阻止以机器速度移动的高度适应性移动威胁所需的反应时间。移动操作系统的攻击签名很难开发，因为安全解决方案不能像防病毒解决方案在传统终端上那样访问相同级别的数据，如详细的文件注册表。此外，移动设备上的沙盒限制了对用户数据的访问。因此，更新的签名文件必须交付给设备（如果解决方案提供设备上的检测）或云（基于云的查询）。这需要耗费很长的时间，无法达到有效的安全效能。

此外，签名和沙箱对于缓慢或较早的威胁来说是足够的，但对于那些旨在保持隐蔽性和持久性的恶意的、未知的和有针对性的攻击来说，是完全不够的。同时，虽然沙箱和签名分析对已知的攻击有防护效果，但不足以打击未知的移动攻击和威胁变体。

• AI如何赋能移动安全

机器学习在提高移动安全方面发挥着变革性的作用。近年来，机器学习在企业安全中的应用势头良好，因为与传统的桌面防病毒等更常见的安全方法相比，它提供了速度、准确性和规模。随着移动威胁形势的发展，安全专业人员需要能够在设备层面快速检测和响应威胁的技术。这就是基于机器学习的移动威胁防御可以提供的帮助，因为它的能力远远超出了基于云和防病毒解决方案所提供的能力。当前，机器学习已经迅速成为移动威胁防御中的一项重要技术，可以持续监控移动设备上的应用程序和用户的行为，以便识别正常和异常行为之间的区别。

当前，许多组织已经在使用机器学习来解决移动安全挑战，例如：

机器智能攻击

如今的攻击，包括未知的零日威胁，都是以机器速度执行的，因此 IT 部门必须同样迅速地做出反应。机器学习可以帮助管理员快速发现这些威胁并采取适当的行动。

设备安全

世界上有数十亿台设备，其中许多极易受到移动攻击。UEM 通过部署基于机器学习的威胁检测到世界各地数以百万计的设备上，是保护大规模移动设备免受攻击的理想方式。

移动威胁情报

有针对性的攻击通常会在设备上产生非常微妙的变化，其中大多数变化是人工分析师看不见的。有时，只有通过机器学习将数千个设备参数关联起来，才能检测到威胁。

IT管理人员短缺

当今，需要在合理的时间内处理的安全警报太多，而 IT 安全人员非常稀缺。利用机器学习可以减轻他们处理威胁的重担。

• 企业案例

MobileIron：基于机器学习的移动威胁检测解决方案

【场景描述】

攻击者通常使用三种技术来执行有针对性的移动攻击：

1. 使用未知的或变种的攻击来避免被发现。

2. 专注于入侵设备，这是保持持久性和控制设备的主要方式。

3. 使用 MITM 攻击或钓鱼技术来传递攻击设备所需的漏洞。这比把应用放在 App Store 或 Google Play 上，然后指望受害者下载这些应用要有效得多。

MobileIron 的基于机器学习的移动威胁检测解决方案可以为用户提供机器速度级别的威胁响应，并且可以检测未知威胁。

【技术方案】

组织可以在 UEM 提供的基本安全基础上，通过在移动设备上部署 MobileIron 移动威胁防御工具，来应对上述威胁。

该工具提供的能力包括：

- 通过基于机器学习的检测来增强传统的反病毒安全，以阻止未知或变种攻击。
- 同时覆盖设备、网络和应用的所有攻击向量，防止单点故障危及设备。
- 在设备上进行威胁检测，不需要基于云。
- 将云计算、机器学习训练和设备内检测进行有机结合，这是企业移动安全最有效的方法。
- 环境的选择因问题而定。对需要分析的数十亿数据点进行机器学习训练，应该在云中离线进行。但威胁检测应该发生在设备上，以防止 MITM 攻击和其他与基于云的方法有关的风险。

1. 基于云的机器学习引擎训练

为了创建高度准确的移动威胁预测器，机器学习引擎必须分析数十亿数据点，然后利用数十个位于云中的高性能计算集群来构建机器学习模型。然后，这些模型被分发到设备上。在设备没有连接互联网的情况下，也能立即检测到之前未知的威胁。

基于云的训练是如何实现的

在云端

1. **初始数据收集**
   首先，该模型从正常使用的几十个设备中收集和分析数百个属性。
2. **行为特征**
   每一个被分析的行为都被分类为“正常”或“异常”(潜在的恶意行为)。
3. **去噪和特征选择**
   每次攻击都有一个独特的“反应”或变化，这可以在系统数据中看到。其中，需要识别并消除可能导致不准确结果的数据点。对于每种攻击类型，选择相关属性并进行加权。
4. **机器学习模型训练**
   引擎通过迭代和训练，获得最佳检测精度。
5. **模型的创建**
   机器学习引擎基于每个威胁的独特的、优化的和去噪的数据点组合来创建终端威胁检测模型。

在设备上

6. **数据收集**
   机器学习的大部分过程发生在云端。数以亿计的取证数据点从世界各地数以百万计的设备中收集，并在云端进行分析。然后，设备提供的数据被用来不断重复整个学习过程，并对模型进行完善，然后部署到设备终端。
7. **终端部署**
   一旦模型交付给设备，所有实际的检测都将实时发生在设备上。与签名不同，优化的行为检测模型不需要针对任何新的威胁进行更新。这是因为每个威胁的独特的系统数据组合不会变化，并且很容易被模型检测到。

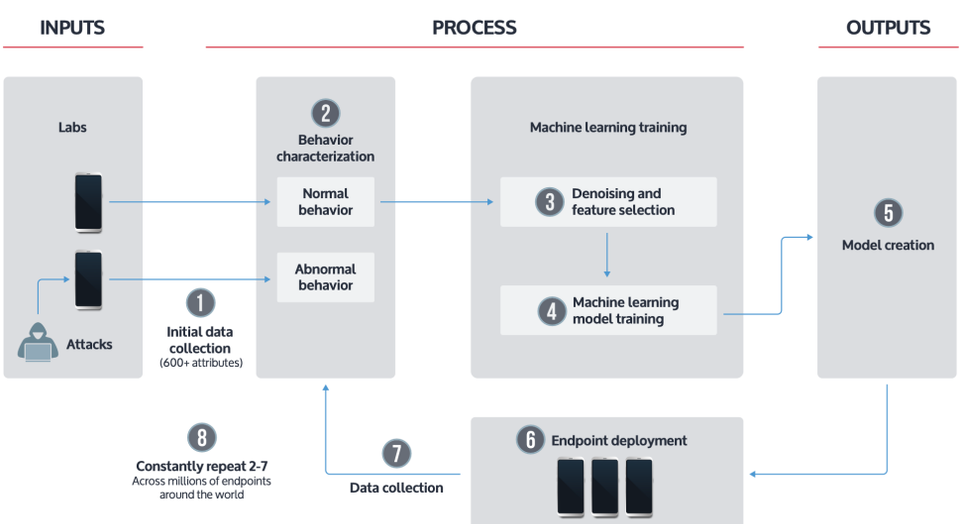


图12　基于云的机器学习引擎训练过程

【方案优势】

- **未知威胁检测**

  与传统的反病毒解决方案不同，机器学习可以检测以前未知的或零日威胁。

- **机器速度级别的检测**

  因为移动攻击是以机器的速度发生的，所以安全工具必须能够同样迅速地做出反应。基于云的工具无法与设备上检测的速度相匹配。

- **最佳的隐私保护**

  通过在设备上进行所有检测，敏感的数据不需要被上传到云端。

- **无需网络连接**

  基于设备的检测可以提供对网络攻击的即时保护，如 MITM。即使断开网络连接，也只有设备上的检测可以继续提供保护。

八、工业互联网安全

• 工业互联网安全挑战

当前，全球化带来的竞争加剧，正在推动网络 - 物理领域和更普遍的、不同信息网络的融合和协同。因此，CISO 不仅需要对 IT 安全负责，还需要对 OT 安全负责。随着 OT 和 IT 加速融合，IT 网络成为最有可能破坏 ICS 的攻击媒介。

除了 OT 和 IT 系统融合所带来的安全挑战外，随着工业物联网 (IIoT) 的兴起，OT 的范围也在不断扩大。工业物联网带来了两个挑战——更加复杂和动态的网络，以及新型、独特的技术部署。随着工业环境中的连接设备数量急剧增加，IIoT 导致网络攻击面持续扩大，数字环境的完全可见性也变得越来越复杂和难以实现，这对安全团队产生了重大影响。

首先，IIoT 引入了无数的设备类别，所有形式的网络通信都发生了广泛的变化。越来越多的智能、小型设备的出现，使计算的方向从单体平台转向高度分布的节点。其次，工业控制系统通常很难更新，且一些现有的 ICS 网络在设计、部署和运行中存在系统性的身份验证不足，这导致任何可能与互联网的连接都可以被利用。同时，打补丁是极其困难的，因为在操作环境中交付更新的内置方法不适合系统不间断和可用性的要求。在安装过程中对操作系统的安全支持也被证明不会像控制系统本身那样持久，安全团队无法将安全功能更新到使用寿命还剩几十年的设备上。

ICS 环境面临众多网络安全威胁，包括：

- 高级持续性威胁（APT），包括将 IT 恶意软件和 OT 控制系统攻击技能结合起来的以 OT 为目标的活动；
- 对企业网络入侵波及到 OT 系统；
- 内部人员破坏；
- 供应链中断和受攻击的供应商或承包商；
- 人为的错误配置；
- 分布式拒绝服务（DDoS）攻击。

随着工业互联网在我国众多行业加快部署，电力、轨交、制造、钢铁、石化等重要工业领域的关键基础设施、生产设备、控制系统将逐步联网化、数字化、智能化，一旦受到恶意攻击，将有可能造成重大经济损失，甚至威胁人身安全，给环境和国家安全带来重大安全风险。同时，工业领域关键信息基础设施在全球范围内成为黑客重点关注和攻击目标，安全防护压力空前增大。当前，安全保障能力已成为工业互联网创新发展的关键因素。

AI如何赋能工业互联网安全

基于 AI 技术，例如贝叶斯数学和无监督机器学习等，可以为 OT 环境提供可靠的安全防护。AI 可以通过学习每个网络、设备和用户的“行为模式”，来分析复杂的网络环境。通过识别异常情况，安全团队能够在恶意软件危害和内部风险出现时以及在攻击生命周期的所有阶段开展监测和调查。

AI 可提供的能力包括：

实时检测出现的威胁：AI 基于对网络环境正常模式的学习，可以通过检测预期行为的微妙变化来发现未知威胁，并向组织发出威胁警报。同时，利用 AI 技术，安全系统可以在整个部署过程中继续自适应和自学习，而不需要操作员手动进行维护。此外，利用 AI 的先进数学技术，可以使其能够突出重要的潜在威胁，而不会将它们隐藏在许多无关紧要或重复的警告之下。它可以将许多微妙的威胁指标关联起来，形成威胁的有力证据，从而减少大量的误报。

规模安全覆盖：不同于传统方法（随着设备数量、连接或网络带宽进行线性扩展），基于 AI 技术的安全系统在判断网络威胁的可能性时，可以利用增加的上下文可用环境，更有效地扩展警报。因此，安全系统可以在复杂环境中处理威胁检测，并防护拥有数百万设备的跨国企业。

• 企业案例

观安:5G工业互联网高仿真欺骗防御威胁检测技术

【场景描述】

对于工业互联网来说，数字化、网络化、智能化是其发展的主要特征，工业互联网应用需要进行大规模数据的连接和运算，如智能制造、机器人、车联网、智慧能源等应用场景。具有高带宽、覆盖广、高可靠性等特点的 5G，正是这些应用场景最有效的技术支撑。然而随着 5G 在工业互联网中的大量应用，MEC 下沉到网络边缘供用户接入应用，企业核心数据未来会直接在 MEC 上运转，因此一旦 MEC 被入侵，就意味着企业的核心资产数据被直接盗用，同时，多种不安全的接入方式和终端设备，一旦感染某个 NF 感染病毒、蠕虫（例如勒索病毒）等，将会在内网中大量传播导致病毒被植入 MEC 的管理系统中，造成网络威胁的风险在大规模终端中传播，甚至利用僵尸网络对核心网进行网络攻击等安全挑战。

针对 5G 工业互联网企业面临的诸多安全问题，上海观安信息创新性的研究出 5G 工业互联网高仿真欺骗防御威胁检测技术，通过暗设定制 MEC 应用服务陷阱，主动诱导攻击，能够动态感知对 MEC 应用服务的攻击行为，及时精准定位攻击源，并隐匿真实应用，保障 5G 工业互联网 MEC 的安全运行，改变网络攻防博弈不对称局面。同时，可延缓攻击进程，感知内网攻击情况，以极低的误报率对内网攻击行为进行预警，通过回溯攻击行为，构建行为分析模型，依托威胁情报，扩充特征库，有针对性地预防可能发生的攻击，提供更大范围的网络安全运行保护。

【技术方案】

1、技术方案概述：

通过深入研究工业互联网 5G＋MEC 业务的安全威胁近况以及趋势，最终研究制定了高精准的安全告警、低侵入式部署、相对低成本全覆盖、具有未知威胁检测能力的可行性创新产品，详细技术方案内容如下：

• 服务仿真欺骗防御

通过在边缘平台管理器、虚拟化基础设施管理器和移动边缘平台部署 5G 工业互联网高仿真蜜罐代理

形成诱捕节点并暴露仿真服务，当攻击者进入 MEC 可信任区后进行探测时，诱捕节点将记录其触碰行为上报，同时把攻击流量引入蜜网。

- **MEC核心仿真**

MEC 可以通过在更靠近用户的网络边缘，提供低时延、高带宽服务，提升客户体验与业务粘性，节省带宽资源，另一方面通过将计算能力下沉到网络边缘，借助运营商边缘节点广覆盖优势，构建“联接＋计算”的能力。5G 工业互联网高仿真蜜罐在 MEC 中仿真出移动边缘平台 MEP 仿真、移动边缘平台管理器 MEPM、虚拟化基础设施管理器等，并将仿真出的服务隐藏在真实的服务中，混淆攻击者的视野。

- **5GC网管系统仿真设计**

5GC 的网络服务及切片是基于 SDN 和 NFV，SDN 和 NFV 这样的技术引入，可以构建逻辑隔离的安全切片，用来支持不同应用场景差异化的需求。但这些技术的引入也对安全造成带来了巨大的挑战，由于它使网络边界变得十分模糊，以前依赖物理边界防护的安全机制难以得到应用。SDN 控制器，虚拟化平台及云平台等等的网管平台成为整个 5GC 的控制中心，一旦其中某个平台被攻击瘫痪，将造成整体网络及所有网络服务的瘫痪 。5G 工业互联网高仿真蜜罐针对 5GC 网络中的硬件设备厂商，部署出与该厂商相同的网管服务（如云管平台、虚拟化平台、网管平台、SDN 控制器等等），干扰攻击者的视野。

2、架构设计

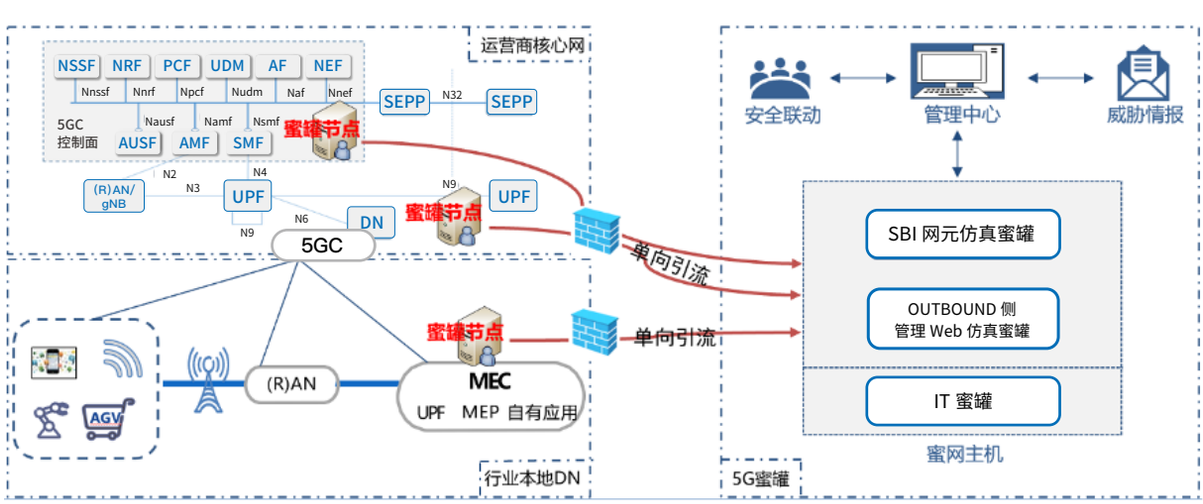


图13 5G工业互联网高仿真蜜罐架构设计

5G 工业互联网高仿真蜜罐包含三个组件，分别为诱捕节点、蜜网主机、中心管理平台三层架构。根据 5GC 及 MEC 网络的内部业务需求，分别部署在支撑管理域、控制域及转发域部分。

- **诱捕节点**

诱捕节点包含伪装成真实主机的诱捕节点和撒放在真实主机上的诱饵所组成。诱捕节点由部署在业务网络中的一台主机上的 Agent 程序所形成，伪装成网络中的一台主机，上面暴露一些虚假服务来吸引和干扰攻击者的视线。同时，在一些易被攻击真实主机上可以撒放一些引向诱捕节点的信息诱饵，一旦攻击者攻陷此主机并翻看此主机的相关信息时，就有可能看到撒放的诱饵，从而增加攻击者去触碰诱捕节点的机会。

- **蜜网主机**

在多个诱捕节点能访问到的公共区域的一台主机上集中部署多个 5G 工业互联网高仿真蜜罐，形成蜜网。其中的 5G 工业互联网高仿真蜜罐支持与来自诱捕节点转移而来的攻击请求进行交互，从而与攻击者形成高交互反馈。5G 工业互联网高仿真蜜罐主机上支持部署管理进程监控并上报攻击者在 5G 工业互联网高仿真蜜罐中的行为和流量。

- **管理平台**

管理平台由事件汇总引擎、数据库和 web 管理界面组成，事件汇总引擎收集来自诱捕端和蜜网中捕获的行为，实时产生告警事件，并在数据库中持久化。同时，管理员通过 web 管理界面可以查看告警事件，并对诱捕节点，5G 工业互联网高仿真蜜罐进行配置管理。

3、技术方案优势

- **数据蜜网技术**

采用用于自然语言处理的深度机器学习模型学习现实业务系统中的真实数据，可以有效提取出真实数据中的特征，进而可以使深度学习模型标记的数据更加贴近真实数据，可以得到一套更加真实可靠的蜜罐业务数据，进而提高了蜜罐系统的威胁诱捕和威胁欺骗效果。

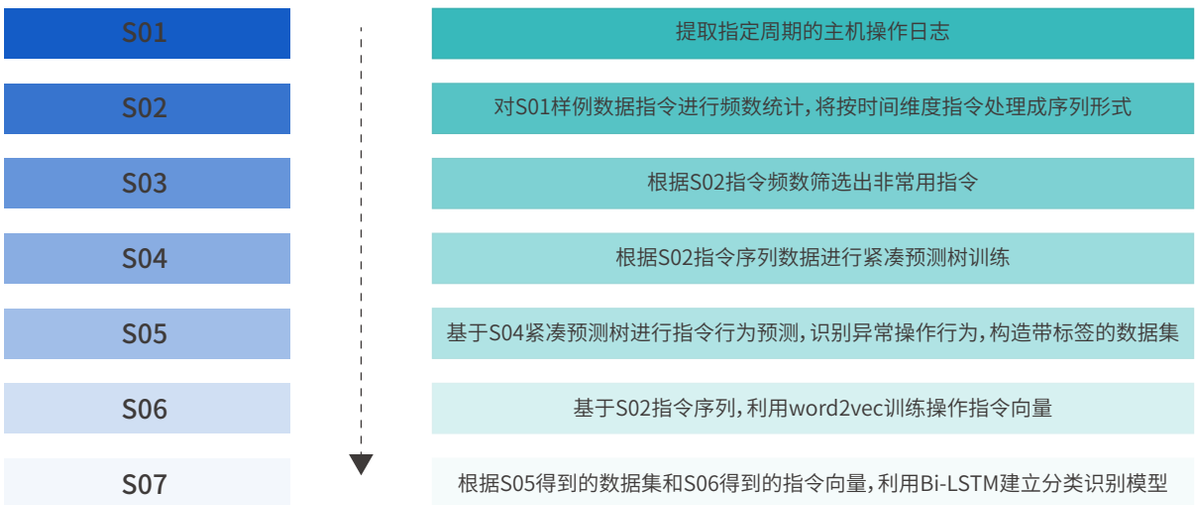


图14 样本算法流程图

- 蜜罐自应答技术

利用在互联网中对设备收集捕获请求的潜在响应，我们能够获得不同类型互联网设备的行为。但是，为了通过攻击者的检查，我们还需要学习最精确的响应，它有更高的概率成为攻击者的预期响应。我们利用机器学习机制来学习系统业务交互过程，并改进响应逻辑，以更高的机会来扩展会话以捕获黑客的攻击行为。

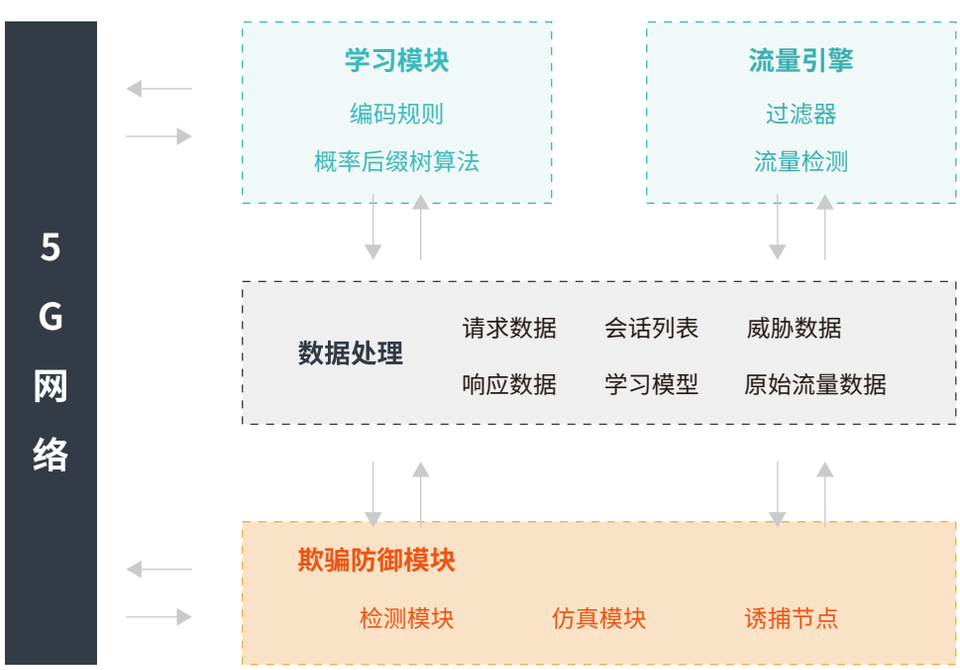


图15 自应答技术架构图

【应用效果】

本方案自 2020 年起，在智能制造等企业选取试点启动上线应用，应用期间发现多起重大网络安全事件并预警响应。基于 5G 工业互联网的高仿真欺骗防御威胁检测技术研究，通过部署适用于 5G 工业互联网边缘层仿真蜜罐，提供设陷引诱攻击、威胁情报分析、追溯攻击源头、行为留存取证、告警、统计分析等功能，实现被探测、被攻击、被威胁等攻击活动的检测与分析，助力工业互联网企业及时、准确掌握攻击者的攻击手段、技巧、战术，实现更深层次的网络安全保护。

九、业务安全

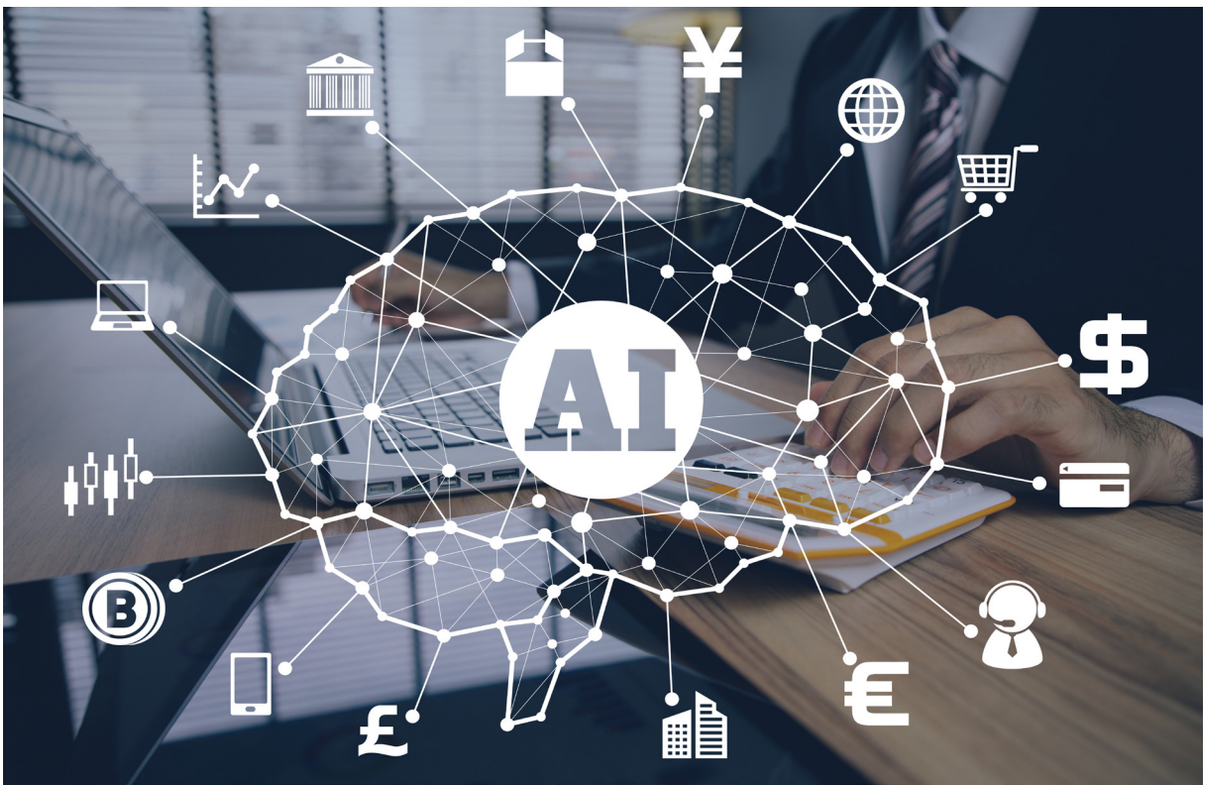
• 业务安全风险持续升级

当今，随着各行各业实现互联网化，业务安全问题无处不在，涉及游戏、电商、支付、直播、金融、内容、社交等多个领域，例如广告场景中的虚假流量，营销场景中的虚假用户裂变，电商场景中的羊毛党，直播场景中的刷榜、虚假粉丝，金融场景中的骗贷等。有获利空间的地方，就有网络黑产从业人员存在，也就带来了业务安全风险。现如今，网络上的“羊毛党”已经具备了明确的分工和周密的组织安排。既有找寻风控薄弱产品或活动的项目发起人，也有收集所有参与有奖活动的线报群，最后还有高度军事化管理的羊毛社群。一旦发现目标，立刻群起而攻将企业薅得“颗粒无收”并损失惨重。据《2018 网络黑灰产治理研究报告》数据显示，2017 年我国网络黑灰产已达近千亿元规模，全年因此造成的经济损失估算达 915 亿元。由网络黑灰产导致的新型安全风险已经成为企业开展线上业务的重大挑战。对于企业来说，企业自身的数字化转型，已经成为趋势，如何避免转型中可能遭遇的各类欺诈威胁风险，保障企业整体业务逻辑的顺畅，成为企业必须考虑的问题。

面对上述业务安全风险，传统的防御手段面临三大挑战：一是防御能力单薄，传统的黑名单、简单规则等，会很容易被绕过；二是仅仅依赖人工判断非常耗时、耗力；三是防御时效性差。

• AI如何赋能业务安全

将 AI 应用于业务欺诈预防，可以结合有监督和无监督的机器学习算法。有监督的机器学习擅长于检查过去的事件、因素和趋势。基于历史数据来训练有监督的机器学习模型，可以发现规则或预测分析无法辨别的模式。无监督机器学习善于发现异常情况、相互关系，以及新出现的因素和变量之间的有效联系。结合无监督和有监督的机器学习可以使 AI 在欺诈预防方面发挥重要作用，包括：



AI正在重新定义欺诈预防。传统的欺诈预防系统仅仅依靠规则，擅长分析过去的欺诈模式，而不提供对未来威胁的洞察力。通过将在历史数据上训练的监督学习算法与无监督学习相结合，数字企业可以对客户行为的相对风险获得更大程度的敏锐性和清晰度。

AI使得实时检测欺诈攻击成为可能。当数字企业仅仅依靠结构化学习和规则时，新的攻击很难被发现。利用基于AI的评级技术（如Omniscore）在不到一秒的时间内检测出欺诈攻击的能力是欺诈管理的未来。

利用AI可以挫败更复杂、更细微的滥用攻击。包括推荐朋友滥用、促销滥用或市场上的卖方串通。如果数字企业曾经遭受过欺诈，他们往往会采用更严格的交易审批控制标准。其带来的结果是糟糕的用户体验。利用AI来评估历史数据和异常情况，可以阻止更复杂的细微滥用攻击，并获得更好的客户体验。

AI可以为欺诈分析师提供实时的风险分数，并能更深入地了解在何处设置最佳阈值，以最大限度地提高销售额和减少欺诈损失。利用AI，欺诈分析人员可以获得对交易的360度视图。利用无监督的机器学习，以及异常检测和对实时活动的洞察，欺诈分析师可以立即验证或重新定义他们关于阈值水平的决定，从而很好地管理风险。

人工智能可以通过帮助商家轻松批准网上购物并减少误报，提升用户体验。欺诈分析师可以利用基于人工智能的评分技术，如Omniscore，减少假阳性，从而获得更好的客户体验。

• 企业案例

观安：基于知识图谱与机器学习的互联网业务风控防护系统

【场景描述】

随着互联网在线服务的快速发展，越来越多的企业将线下的业务、营销活动等迁移到互联网平台如网购平台、P2P 平台等，从而拥有大量用户参与营销活动、注册、登录、充值、购买等交易活动。很多电子商务平台在实际业务运营和推广中投入了大量资金，取得巨大的经济效益和用户体验，得到广大用户的认可和支持。同时也给互联网“黑色产业”提供了滋生的土壤，各类业务安全导致经济风险问题层出不穷，如恶意“薅羊毛”、撞库、账号盗用、养卡、非授权调用、敏感信息泄露、套取流量等业务漏洞，这些违规甚至违法行为会造成企业经济的损失和对企业形象产生负面影响。

本方案将前沿的知识图谱、机器学习等技术与互联网业务相结合，实现了具有高效率、高准确度、自主学习并自我迭代优化的业务风控及经济安全防护的大数据解决方案。

【技术方案】

1、方案流程架构

该方案通过营销推广活动中团伙行为甄别、机器注册登录行为识别、平台业务操作异常行为分析三大流程，识别高风险平台操作与交易行为，在各个阶段及时实行针对性的业务防范措施，解决“黑产”群体所引起的经济损失和企业形象影响的问题。

该方案以数据为基础，采用多维度业务算法应用，实现以实时分析、离线分析相结合的方式，进行业务风控和安全防护。下图为本解决方案的流程架构图。

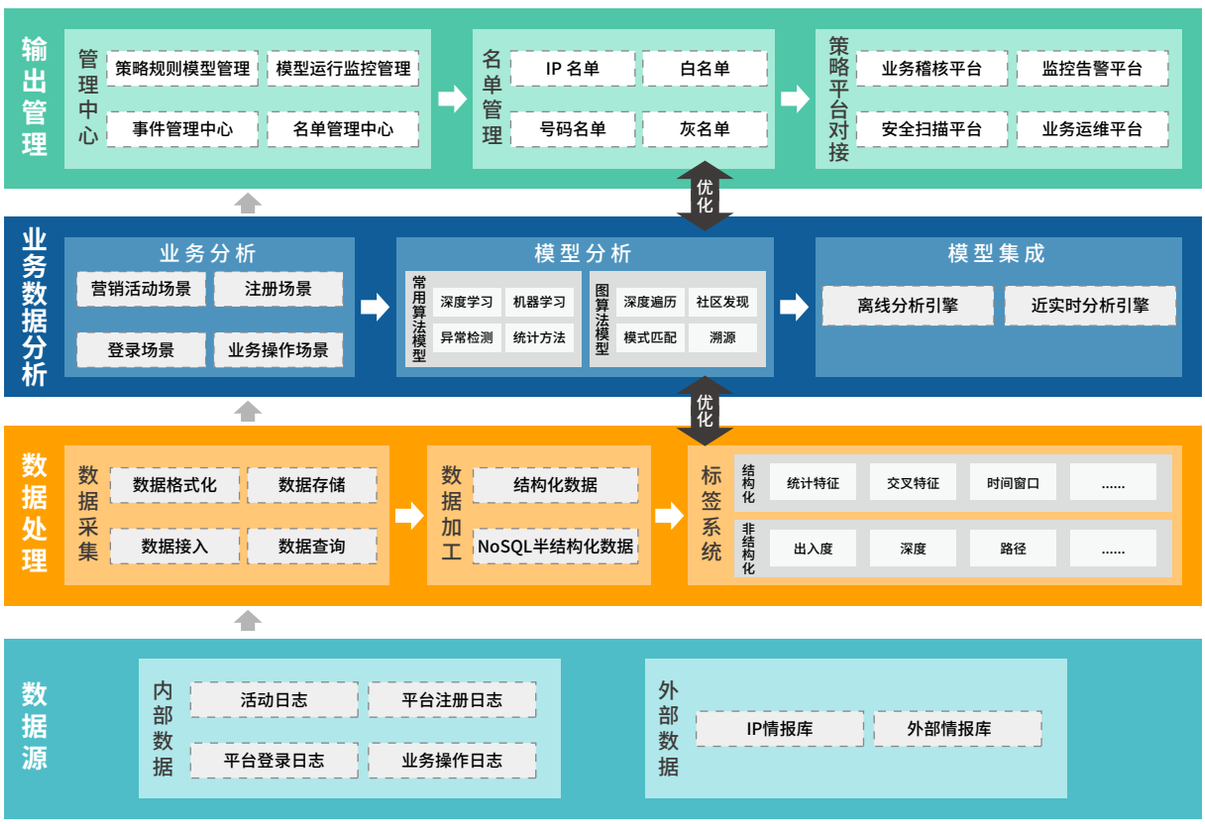


图16 基于知识图谱与机器学习的互联网业务风控防护系统架构图

2、业务分析

该方案从用户来源、用户注册、用户平台业务操作三个方面，利用机器学习、基于图数据库技术的图分析、动态行为基线等技术，对营销活动参与、平台账号注册、业务订单办理、业务接口访问等行为进行分析，挖掘出“黑产”、“恶意攻击”等风险群体，并予以相应安全策略进行风险阻断。

增量型推广营销活动面临“薅羊毛”团伙使用效率最高和最不容易被发现的方式绕过现有规则进行薅羊毛。由于群体关系较深，行为特征显著性较低，仅仅依靠传统薅羊毛识别方法，“黑产”群体很难被基于常规业务特征的机器学习模型和规则场景模型识别。为了避免潜在的业务风险如黑产“薅羊毛”，从业务安全角度出发，利用外部情报，结合业务数据，加工特征工程，同时利用图分析技术提取动态特征，离线训练模型，利用训练好的模型实时计算用户行为是否属于黑产，从而有效的拦截黑产的攻击。同时不定期迭代模型，从而实现模型的闭环和有效性。

【应用效果】

目前基于机器学习算法与知识图谱相结合的AI模式识别、全天候动态监控的业务安全风控的解决方案已在多家互联网平台使用，并取得了以下效果：

（1）系统产生的有效拦截量高达千万次级，其中90%以上为准确拦截，对比情报名单覆盖面更广，且灵活性和误报率更低。

（2）通过阻断养卡和对应活动参与金额统计计算，从现行运行以来，每月发现和阻断薅羊毛优惠金额100万元左右，预估一年累计挽回损失在1200万左右，将这些优惠活动普及更多普通用户群体，保证了业务活动效果。

（3）通过将发现养卡名单加入活动限制名单中，降低了参与活动中奖概率，避免经济损失。

（4）发现一批接口访问次数异常用户，通过限制其访问次数，降低整个接口服务器压力和节省服务器资源，提升用户体验。

（5）通过对重点关注人员业务操作异常告警，及时分析其业务操作可能存在的业务安全风险，避免经济损失。

（6）通过实施后与实施前的数据进行对比，业务指标有实质性进步，用户体验显著提升；体验类用户投诉占比下降5%，用户平均受理时长缩短了60%，业务数据准确度提升10%以上，减少无效业务办理数据。

十、网络内容安全

• 网络内容安全挑战

随着数字化时代的到来，终端不断升级，能提供更加便利的移动网络接入，任何智能设备都有可能成为信源或信息终端，即将迎来万物皆媒的新传播景象。信息的及时性、创新性和多态性在新技术的加持下得到增强，写作机器人、虚拟主播出现在日常新闻内容中，以 AR、VR 为代表的虚拟现实、增强现实技术成为新兴的内容业态而且新技术应用门槛和成本的降低将极大助推用户生成内容（UGC）的激增，加之云技术泛化和 5G 商用普及的步伐加快，数字内容的生成和传播触手可及，无处不在。

当今时代数字内容极大丰富，但同时内容审核能力不足矛盾凸显，导致网络内容安全面临重大挑战。

据不完全统计，过去两年里生成的数据占到了全球总数据的 90%。预计到 2022 年，全球互联网流量将达到每秒 7.2PB，未来数字内容必将继续呈指数级增长。然而，面对庞大的数字内容洪流和各国日趋严格的内容审核政策，试图依靠传统审核模式实现内容含义的准确判断并及时应对信息爆炸引发的各类问题，越发捉襟见肘。内容审核能力不足的问题将会更加凸显，集中表现为人工审核难以应对海量负面信息。

• AI如何赋能网络内容安全

作为多技术集合体，AI 在数字内容治理方面有着极大的应用前景，数据挖掘、归因分类、机器学习、自然语言处理、计算机视觉、模式识别等各种技术方向均会在数字内容治理的不同场景中发挥重要作用。具体应用领域包括：

内容审核

AI技术发展与应用，极大地提高了不良信息识别发现、审核判别、处置处罚等治理效率，有效节省人力、物力成本。在信息识别发现环节，利用AI技术可加强对不良信息的特征识别，提高筛选素材的效率和质量，为高效治理奠定基础；在审核判别环节，传统的人工审核机制需要配备大量审核人员，对于违规特征不明显、不易识别的恶意信息，此审核方式效率极低，且不同审核人员存在标准认知差异，审核输出质量偶有不同，AI技术的应用使机器审核在准确率、效率和标准化方面均能得到保障。

事实核查

虚假新闻、网络谣言和深度造假等内容造假是目前数字内容治理影响最大的问题之一。基于AI技术的数据挖掘、文本汇聚、深度学习等技术能够有效检索虚假内容的传播源头，构建各类结构数据库和标识体系，帮助核查者对海量资讯进行针对性处理，随着对合成图片、声音和视频的鉴伪技术和溯源技术的研发精进，对虚假信息的识别取得了较大突破。

舆情监测

网络主体在数字内容的生产、传播和交流中的表达情绪和立场汇聚成为网络舆情，对现实社会产生有着极大的影响力和观念塑造能力，不仅仅关系到网络安全和舆论安全，更关系到整个社情民意和国家稳定，是数字内容治理中十分重要的命题。基于 AI 构建的情感分析技术、舆论模型、态势感知、可视化呈现、应急处置等机制，能够实现对网络舆论各个主体的标准化信息采集、汇聚、分类，对特定事件和整体舆论环境的实时态势做出理解和评判，对舆情的发展趋势、关切热点、各方态度做出预测和提供处理建议，从而推动舆情治理工作更加准确和有效。

• 企业案例

观安：基于多机器学习的网站内容安全智能分析监测

【场景描述】

近些年来，Web 应用的发展，使 Web 系统发挥了越来越重要的作用，与此同时，越来越多的 Web 系统也因为存在安全隐患而频繁遭受到各种攻击，导致 Web 系统出现各类违规内容、存在敏感数据、页面被篡改、甚至成为传播木马的傀儡，最终会给更多访问者造成伤害，带来严重损失。

另一方面，运营商、各大门户网站的网页违规内容安全事件频繁发生，不仅对运营商自身的同行业竞争力和市场声誉造成了严重影响，甚至会造成不良的政治风险。因此，防范网页出现违规内容已成为安全工作的重要目标和任务。

针对以上问题，观安信息基于自研的视觉 AI 技术，通过机器学习结合深度学习，为用户提供音视频、图片、文本的内容审核能力，主要涉及涉政、色情、暴恐、违禁等审核场景，可以大大提高音视频内容审核的效率，降低人工审核漏审风险，缩减平台的人力成本。

【技术方案】

观安网站安全监控平台，是在多年安全研究基础上自主开发的基于 B/S 架构的网站集中安全监控平台，该平台以高性能服务器为硬件载体，在具备极高检测性能的同时，可负责对大规模网站 / 网站群的安全状态进行全方位实时监控，包括内容安全 (如敏感词、网页木马、暗链、网页变更等)、可用性 (网站应用状态、网站访问速度、域名劫持检测) 及网站信息 (能够提供周期性的任务调度功能，用户无需干预即可实现周期性全自动化监控 HTTP 请求方式、IP 及定位等) 等，并且，系统同时提供详细的监控报告和图形化的网站安全状态集中展示，让客户对所监控网站的安全状况一目了然。

图17 观安网站安全监控平台架构

其中，网页违规内容检测模型基于自研的视觉 AI 技术，通过机器学习结合深度学习，为用户提供音视频、图片、文本的内容审核能力，主要涉及涉政、色情、暴恐、违禁等审核场景，可以大大提高音视频内容审核的效率，降低人工审核漏审风险，缩减平台的人力成本：

1、违规内容检测模型能力介绍

违规内容监测引擎具备全面、灵活、高效等特点，相较传统视觉技术对违规内容的识别能力不足，人工智能预测准确率整体高达 85%+，且准确率随着数据量提升可持续优化，在单项审核上具备高召回、高准确的审核能力。

在涉黄监测方面：需要色情审核模型，针对特殊垂类，如健身、艺术品等正常较裸露场景，保持更高精度，针对指定评测集，保持准确率95%以上。

在暴恐监测方面：需要暴恐审核模型，可以精准识别图像中是否包含杀人流血场景、 暴恐袭击场景、恐怖分子头目照片、恐怖组织的旗帜、暴力行为等内容，针对指定评测集，保持准确率 90%以上。

在涉政监测方面：需要人脸识别模型，精准识别图像中的政治人物，最小识别人脸　80px*80px，保持高准召；利用文本和音频审核，规避财经类股评里提及政治敏感问题。针对指定评测集，保持准确率　90%以上。

在违禁监测方面：需要ASR、OCR、NLP　模型，针对语音、文本以及图像中的文字，准确识别各类违禁品，如器官变卖、毒品等，避免产品涉及违禁品风险。针对指定评测集，保持准确率 95%以上。

2、违规内容检测模型技术原理

OCR 技术是一种字符识别技术，是通过算法对输入图片进行文字提取的过程。通过 OCR 技术与违规内容信息关键词库、文本处理技术相结合，从而能够有效识别网站是否包含违规内容。

文本技术识别网站违规内容的处理过程为：

图片爬取过程

1. 搭建出针对网站图片的爬虫系统，将待检测网站的 url，输入至爬虫系统，输出对应网站的图片集合；

OCR识别过程

2. 对图片进行二值化、去噪、倾斜校正等操作；
3. 针对预处理之后的图片集合，使用版面分析算法将图片进行分段、分行操作，利用字符切割算法处理字符之间的粘连、断笔造成的字符间难以切割问题；
4. 特征提取及匹配：使用字符特征提取算法提取图像多维特征，结合字符识别算法将当前字符提取的特征向量与模板库进行分类和匹配；
5. 对已识别字符根据语义进行校正，输出成文本格式；

分词及特征匹配过程

6. 利用分词工具（如 jieba 分词器）对输出文本进行分词处理；
7. 通过 FuzzyWuzzy 工具计算分词后的字符与关键词库进行差异比对，若相似度大于 80，则判断为违规网站；

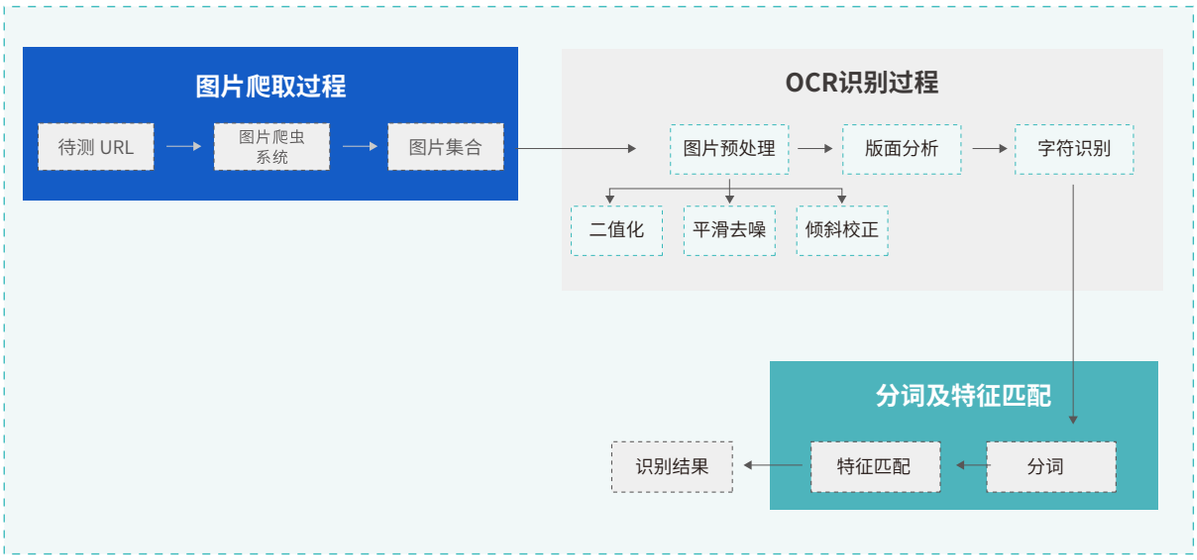


图18 文本技术识别网站违规内容的处理过程

【应用效果】

本案例可应用于国内外大量重要客户的网站的安全监测与性能监测，包括运营商、金融、医疗、教育、能源、电力等。涉及网站可用性；业务能否正常使用的业务可用性；网站是否存挂马事件或网站存在的漏洞修复时间；网站内容是否存在违规内容。借助人工智能技术可以大幅提升检测的能力，有效减少对专业安全分析人员的依赖，减少低效重复的人力劳动。

第五章
当前人工智能的局限性与需克服的障碍

当前，虽然利用AI技术可以为网络安全防御带来多方面的能力，但AI不是万能的，其也存在自身的局限性，并且在AI的应用过程中，也存在亟需客服的障碍和问题，以使AI能发挥更大的效用。

AI不能解决所有问题

AI目前并不是缓解所有网络安全挑战和解决技能短缺的灵丹妙药。它能在解决特定问题中发挥作用，但并非所有的网络安全问题都可以用AI来解决。

人类团队仍将是必不可少的

很多人认为我们可以从数据中学到一切，但这是错误的。在网络安全方面过度依赖AI会造成一种虚假的安全感，这就是为什么有些网络安全公司除了应用算法外，还聘请了网络安全专家、数据科学家和心理学家。因此，机器学习是对人类能力的补充和增强，而不是取代它们。目前，机器学习为网络安全提供了多种实际应用，但至少在短期内不太可能完全消除对人类智能的需求。相反，人类与AI之间的关系应该是共生关系。例如，首先需要网络安全专业知识来准备和分类训练数据、选择合适的算法；其次，根据算法的结果，再次需要人类专业知识，来根据机器分类或可视化的数据做出决策。

准确性问题

AI驱动的网络安全解决方案并非总是100%准确。有些时候，这些解决方案可能无法识别恶意活动与一般活动之间的区别。此外，通常情况下，这些类型的工具会创建安全事件警报，但在当前，这些警报必须由人工调查以确保准确性。因此，用于网络安全的AI解决方案必须与人类分析师一起使用，以确保安全的网络环境。

数据完整性问题

在AI模型的构建中，有监督的机器学习算法需要学习数据集进行训练，因此安全团队需要掌握许多不同的恶意代码、恶意软件和异常数据集。但目前，有些公司没有足够的资源和时间来获取所有这些准确的数据集。因此，训练数据不足在一定程序上阻碍了一些安全厂商采用AI技术。

对AI网络安全专家的需求

将AI应用于网络安全，需要更多该领域的AI和ML网络安全专家。智能网络安全技术将大大受益于能够根据需要进行维护和调整的专业技术人员。然而，当前在全球范围内，此类合格的、训练有素的专业人员数量远远满足不了需求。

第六章
人工智能赋能网络安全的未来趋势

预测算法的精度将持续提高

机器学习技术的不断创新发展将持续提高AI模型预测网络安全风险的能力，以准确识别新出现的安全风险。此外，神经形态计算的发展将有潜力增强AI在网络安全中的作用，产生更快更准确的预测结果。例如空客等公司正在与卡迪夫大学等学术机构合作，开发基于神经形态计算芯片组的更强大的恶意软件检测技术。

技术投入成本将不断下降

当前，企业需要在计算能力、内存和数据等资源上投入大量的时间和成本来构建和维护AI系统，这会增加额外的企业开支。不过，针对AI的高成本问题，产业界正在加快技术创新，以降低算力成本。数据显示，AI训练成本从2017年到2019年下降了100倍。同时，自2012年以来，AI模型在ImageNet分类中训练神经网络达到相同性能所需的计算量，每16个月减少2倍。随着自动化和计算能力变得更快、更便宜，在网络安全解决方案中部署AI技术将变得更加经济。

丰富数据源的可用性将不断增加

据IDC 预测，到2025 年全球产生的数据量将会增长到175ZB，其中超过80%-90%的数据都会是处理难度较大的非结构化数据。但随着针对非结构化数据分析技术的不断发展，丰富的外部和内部数据集的可用性将持续增加，这对AI网络安全工具的快速演进和能力迭代将产生非常积极的影响。

AI将具有更强的可解释性

随着AI/ML在对抗网络攻击检测方面的应用激增，对可解释AI (XAI)的需求也在增长。XAI对AI“黑盒子”提供了更多的洞察力，可以使采用AI工具支持网络威胁防御的利益最大化。通过XAI，安全团队可以洞察其算法背后的“原因”，从而持续改进网络安全生命周期。未来，随着XAI技术的快速发展，AI在网络安全领域将发挥更重要的作用。

观安 | SICSI CYBER RESEARCH INSTITUTE 赛博研究院

人工智能赋能网络安全白皮书

2021.07

观安信息

赛博研究院