关于选择人工智能方法提高网络安全的建议

Roumen Trifonov†
保加利亚索菲亚技术大学
r_trifonov@tu-sofia.bg

保加利亚索非亚
Slavcho Manolov 技术大学 s_manolov@tu-sofia.bg

格鲁吉亚索非亚
理工大学，保加利亚 gtsochev@tu-sofia.bg

保加利亚索非亚
帕夫洛娃技术大学
raicheva@tu-sofia.bg

摘要

在教育部科学研究基金项目的实施过程中，在对网络安全威胁进行多因素分析的基础上，将军事理论引入网络安全程序，对网络防御的阶段和可使用人工智能方法解决的任务类别进行了新的划分。

基于理论模型和具体原型的实验，作者提出了适合网络防御各个阶段的人工智能方法选择的建议

CCS 的概念

• 正式的安全模型•安全需求

关键字

网络防御，网络威胁情报，人工智能方法

ACM 参考格式:

Roumen Trifonov, Slavcho Manolov, Georgi Tsochev 和 Galya Pavlova。2020.关于选择人工智能方法提高网络安全的建议。*《计算机系统技术学报》，2020 年 6 月 19-20 日，保加利亚，鲁泽，5 页。https://doi.org/10.1145/3407982.3407987*

1 介绍

索菲亚工业大学计算机系统与技术学院于 2013 年开展了人工智能方法在网络安全领域的应用研究。自 2017 年以来，这些研究获得了科技部研究基金的支持。

Acm isbn 978-1-4503-7768-3/20/06
https://doi.org/10.1145/3407982.3407986

在项目网站[1]以及其中提到的 14 篇出版物和向权威国际会议发表的报告中，详细描述了所构建的模型、实验装置和实验结果。本文旨在向专家社区宣布这些结果的摘要，并准备了一组建议。　由于会议形式的限制，它们是以相当零碎的形式给出的。

在该项目的工作开始时，其承包商计划试验有效使用理论上选择的人工智能方法来检测攻击，防止入侵，以及信息系统网络防御中的更典型行动。自然，研究开始于网络威胁的评估和分类[2]，并越来越多地投入到网络安全军事概念的理论和实践中，如“网络威胁情报”;[3]和“杀戮链”;(图 1)[4]。

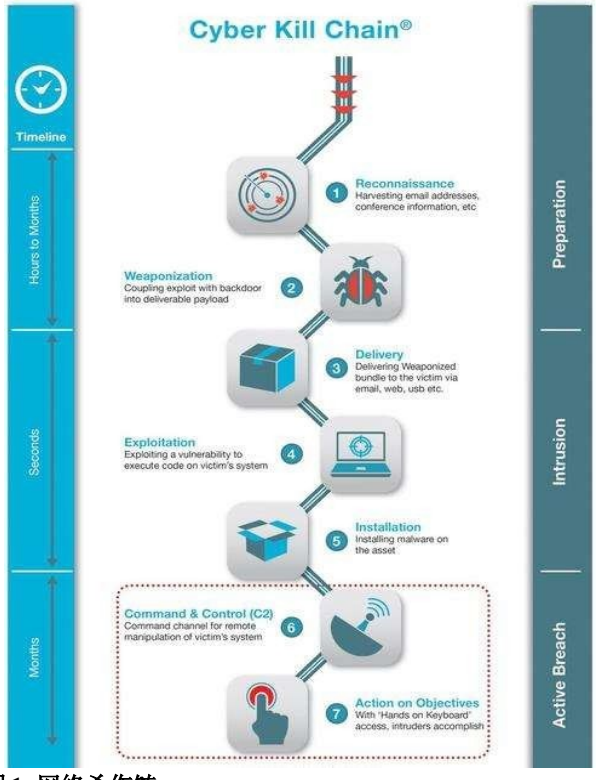

图 1: 网络杀伤链

但在研究过程中，项目组得出了三个基本结论(在参考资料中)，迄今为止还没有明确的形式:

A. 网络防御(取决于所使用的环境、方法和工具)可分为三个主要部分:

A1。立即排斥攻击-就上述网络安全专家社区采用的军事概念而言，这可以被称为所谓的战术网络威胁情报。监控计算机系统或网络中发生的事件，并分析可能违反或即将威胁违反计算机安全政策、可接受使用政策或标准安全措施的事件的迹象的过程;

A2。预测潜在对手的行动——同样，这可以被称为所谓的作战网络情报，其主要目标是通过以下方式降低组织关键资产和活动的安全风险:定义作战环境，描述作战环境的影响，识别潜在对手并识别这些潜在对手的意图和行动。网络作战情报以积极防御理论为基础

A3。排除网络攻击造成的后果

-这可称为所谓的事件处理。在全球范围内，计算机安全事故管理的概念已被广泛接受和实施。它有助于系统地应对事件(即遵循一致的处理方法)，并帮助员工尽量减少信息丢失或被盗以及与事件有关的服务中断。此概念的另一个优点是能够使用在事件处理期间接收到的信息为将来的事件做好准备，并为系统和数据提供更好的保护。作为一项正式程序，事故处理受多项国际标准规管，例如: 国际电联 E.409 和 H.1500 建议书、美国国家标准化与技术研究院(NIST)的 ISO 18044 和 SR 800-61。

B. 国际上在网络安全领域使用人工智能方法的理论和实践的发展表明，这些方法适用于上述网络防御的所有阶段。同时，文献(以及在一定程度上在本项目范围内进行的研究和实验)清楚地表明，没有一种通用的方法在所有(甚至大多数)可能的应用中都同样有效。

选择对某一类任务有效的人工智能方法的可能性(和必要性)(在项目各自的活动报告中介绍了这些方法的某种分类)要求，首先，根据文献来源和具体的研究和实验，制定选择有效方法的标准。

C. 除了上述关于应用特定人工智能方法解决特定网络防御任务的动机的一般结论外，该团队还在项目中得出了另一个基本结论:

事件是由于所谓的“威胁分类”中的一种类型的威胁造成的。ENISA[5]。

A2。**决策辅助——在很**大一部分网络防御中，人工智能系统并不直接触发行动，而是用于支持网络安全相关决策(包括基础设施、安全政策等的基本决策，以及短期行动的一次性决策)。

B. 任务的性质——这种分类不像上面那样明显。这是项目参与者的原创贡献——分析了大量使用人工智能方法解决的任务，作者得出结论，它们可以分为两大类:

B1。**异常检测**-检测通信网络或其对象的异常行为，某些样本的突变等。实践表明，这类任务主要在战术网络情报中解决;

简而言之，它可以表述为:通过某些标准选择的最有效的人工智能方法可以被认为是基本的。在大多数情况下，可以通过另一种适当的方法加以补充来提高其效力。基本方法和互补方法的集合可以称为混合人工智能方法。

如上所述，这种对网络防御阶段的分类以及为每个阶段选择具体人工智能方法的方法是作者的原创贡献。

2 .通过不同人工智能方法解决的任务评估

许多专家认为，人工智能领域的基础研究和应用研究的重要领域之一，首先是其一般理论，就是所谓的“任务导向方法”。以研发为主。这种方法的基本思想是，将应用人工智能的主体所涉及的整个活动过程，适当地描述、建模和设计为解决各种任务的过程系统。这种发展的有效性最终取决于什么任务以及以什么顺序解决这些任务。因此，描述任务的定性和定量特征，以及解决任务的手段和方法，对于创建有效的人工智能系统非常重要。

这种“任务导向的方法”对人工智能系统的研究主要表现在确定所解决问题的类型，制定解决问题的抽象方法，以及“以任务为导向”地分析人与人工智能系统之间的各种类型的关系。此外，这种方法的目标之一是构建能够解决一种或另一种类型任务的结构，另一个目标可能是确定与一种或另一种人工智能系统交互的用户训练的内容。

按照这种思路，为了创建解决特定类型问题的系统，建立决策过程的收敛性是至关重要的。

根据两个主要标准对这些任务进行分类后，可以对网络防御领域的任务进行具体评估:

A. 运用这些方法所采取的行动的性质:

A1。**自动操作**——顾名思义，这是一种人工智能系统的操作，没有直接的人类参与，自然地，与性能要求有关，即，它首先适用于响应时间是网络防御中关键因素的情况。这决定了它在战术网络情报(或攻击的即时反映)中的应用。

在某种程度上，自动化操作也适用于事件处理，但只是部分地适用(例如，当基于形式化的事件报告(所谓的故障单或其他标准化数据结构)时，系统对事件进行唯一的分类

B2。**解决分类问题——一**般来说，这涉及到确定对象(情况、模型等)与公认的分类方案(或分类法)的一个组件之间的关系，或者确定对象(情况、模型等)。)通过其特征与某些模式的特征进行比较。我们认为这类任务将主要在事件处理中解决，并且很有可能在作战网络情报中解决。

C. 在区分任务类型时(分别是选择有效的人工智能方法)，根据威胁的结构和威胁的代理，可以在每个部分中确定特定的应用领域。

建立了网络防御各个阶段的标准和有效方法的具体选择

需要强调的是，根据批准的“科学描述”;和“工作计划”;该项目的活动完全集中在战术网络情报或直接检测和行动，以反映攻击，防止未经授权访问系统资源，阻止恶意软件等。

为了执行这些行动，项目小组开发了原型和实验室产品，进行了模拟真实攻击的实验，分析了结果，并采取了旨在改进原型工作的行动。

因此，使用一种或其他人工智能方法的具体建议仅适用于网络防御的这一要素(根据我们的分类)。

在项目实施过程中制定的其他两个网络防御阶段的描述和初步考虑超出了当前项目的范围，仅用于概述问题。

3.1. 战术网络情报

传统上，网络攻击的检测和预防过程基于两个基本原则:异常检测和滥用检测，尽管它们的特征没有显著差异。入侵检测机制通常检测网络中的以下可疑行为:

(a) 试图使用防火墙阻止的服务;

b) 意外的请求，尤指来自未知地址的请求;

c) 意外的加密消息;

d) 来自未知服务器和设备的过度活跃流量;

(e) 与以前的网络操作相比发生了重大变化;

(f) 试图利用已知的错误或漏洞;

(g) 试图从意外地址访问未知用户;

(八)不当或者可疑地使用行政职能的;

(i)用户日常活动发生重大变化等。

在网络流量中检测用户的可疑活动是机器学习中最广泛的应用。现代系统在检测大数据流中的异常事件和解决标准分析问题方面越来越成功。

攻击检测的类型取决于所使用威胁的性质: 通知已知，已知未知;和“未知未知”;[6]。全球专家已经制定了评估检测效力和应对水平的标准。在假阳性和假阴性结果之间取得适当的平衡也是至关重要的。假阳性(所谓的假警报)的破坏性不亚于假阴性。

项目团队采用以下方法来分析和比较人工智能的不同方法:将一个基本标准(或一组基本标准)与附加标准相结合。对于战术网络情报，选择的主要标准如下:最大生产力(即检测效率与性能水平相结合)和最小错误警报百分比。

我们还接受了其他标准，包括:在不同环境中使用的灵活性;通用的方法;分析数据包内容以排除丢失数据包所需的处理速度[7]。

在上述多元分析的基础上，作者选择了可能适合这些应用的人工智能方法。所获得的数据表明，在管理设备进行入侵检测和防御时，使用自学习多智能体系统网络发现异常行为和未知性质的攻击比其他方法具有更高的效率和生产力。此外，该方法还具有最小的误报报警值。(图 2、3)。移动药剂的使用可能会被进一步推荐，尽管它们的效果还没有像固定药剂那样得到充分的研究。移动代理不能直接改进攻击检测技术，但可以重构所应用的技术，从而提高效率。让代理访问数据仓库并监视结果是一种理想的替代方案，它适合移动代理最小化计算的能力。除了减少网络负载外，专门的代理还可以

专注于特定类别的入侵，例如协调攻击，这些攻击发生在很长一段时间内，来自不同的来源。

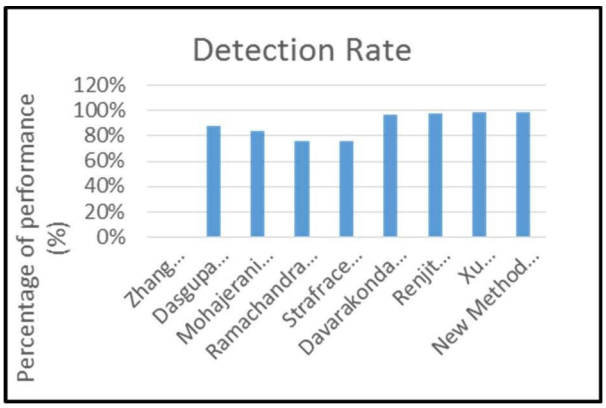

图 2: 各种方法的检出率

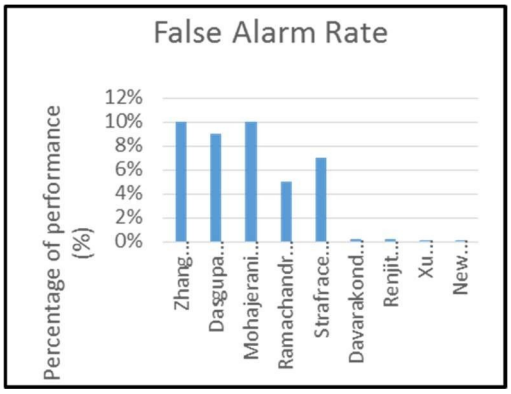

图 3: 各种方法的虚警率

在项目过程中进行的实验中，在由自学习静态智能体组成的多智能体系统的基本方法之外，补充了所谓的强化学习(Reinforcement Learning, RL)和模糊集(Fuzzy Sets, FS)。进行了一系列测试，以确定 RL 和 FS 的使用是否可以允许代理学习使用有关信息流的信息对网络中的正常和异常活动进行分类。结果表明，在异常情况下，该组合内部分析并更新受训代理的 q 值，实现其训练周期的递归迭代。

实验表明，混合模型的检测精度平均提高了 6%左右(图 4)。

3.2. 作战网络情报

我们对操作网络情报自动化的假设是，这可能是基于潜在对手的行为模型，基于他在网络中的活动和出站流量。目前，在没有这样的参考资料

国际标准(例如所谓的故障单[11])。

相反的任务似乎更困难——它涉及更改统一事件报告的属性，以便它们更充分地反映特定事件与分类方案的一个元素

研究认为，目前这种行为可以分为两种基本类型:敌对和非敌对。

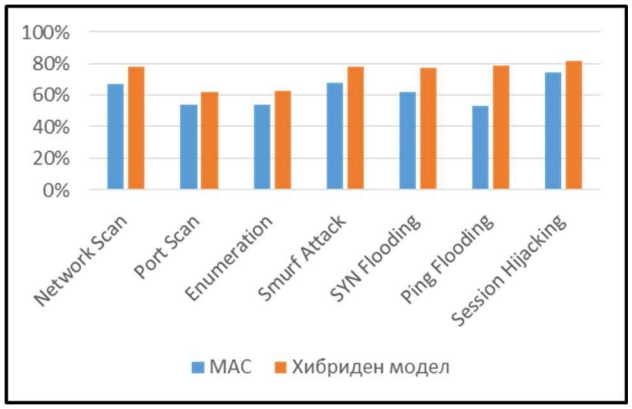

图 4: 针对各种攻击的检测精度

这假设了一个闭环场景，由四个步骤组成:

(a) 使用数据包捕获设备[8]等方法测量网络活动;

(b) 预处理;

(c) 提取被称为“特征”的特定特征;

(d) 指定与所识别的行为类型相对应的一类特征集的分类。

在文献综述的基础上，提出了回声状态网络(回声状态网络，ESN)方法作为特征选择机制。回声状态网络结构由具有 s 型非线性(通常为双曲切线)的随机耦合动态神经元库组成[9]。

通过研究在各种主题上训练分类器的可行性，我们建议将顺序特征选择(SFS)过程用于行为分类是最合适的。这可以减少数据的固有可变性，并可以在识别受试者之间的行为状态方面具有很高的准确性[10]。

3.3. 用于事件处理阶段

从描述尝试在事件处理中应用人工智能方法的稀缺文献来源中，项目团队得出结论，人工智能的主要功能目前可以集中在解决分类任务上，即将当前事件明确地引用到已批准的分类方案的一个要素，从而为方案的每个要素开发程序和工作流程。

显然，解决这个分类任务最重要的部分是找到所谓的“特征”。即，充分反映客观依赖于分类状态的特征。

如果我们坚持这种做法，应该注意的是，通过人工智能在事件处理中的应用必须解决正反两方面的任务。

正确的任务可以概括如下:根据统一事件报告中包含的属性找到“特征”

的关联。在这里，我们可以说这个任务还没有被项目团队认为是合适的。

对文献来源的分析使我们在未来的研究中推荐“强化学习”。-“试错”;学习范式[12]。在模型无法以足够高的置信度预测值的情况下，此事件的结果可以由人手动修正。在这样做的过程中，相关的变更事件必须被固定并用于改进模型。

4. 结论

项目最后部分制定的对政府、企业和学术界的建议并没有侧重于研究和实验人工智能的具体方法，而是坚持在网络安全阶段使用上述分类，以及在明确标准的基础上选择合适的人工智能方法的方法。

此外，他们强调需要进行一些新的研究领域，例如:

a) 探索人工智能在工业网络安全方面的应用，特别是针对欧盟的“工业 4.0”倡议;

b) 研究这些在网络防御各个阶段的应用之间的关系(例如，在作战网络情报领域的应用支持与战术网络情报相关的系统运行的能力)。

致谢

本研究是在科研项目№H07/56“使用智能方法提高网络和信息安全水平”下实现和资助的，由合同 DH 07/4 与保加利亚国家科学基金。

参考文献

[1] Roumen Trifonov, Ognian Nakov, Slavcho Manolov, Radoslav Yoshinov, Georgy Tsochev 和 Galya Pavlova。*2020* 年，以智能化手段提高网络和信息安全水平。*研究项目:国家自然科学基金资助，https://iti.tu-sofia.bg/en/national-science-funds- project-h-07-56/*

[2] Roumen。特里福诺夫，斯拉夫乔·马诺洛夫，格奥尔基·措切夫，加利亚·帕夫洛娃。2018.审查网络威胁的新方法。信息技术国际*会议论文集(Info-Tech 2018) 2018 年 9 月 20-21 日，保加利亚瓦尔纳*

[3] 2016 年 ENISA 威胁态势报告

[4] www.lockheedmartin.com/content/dam/lockheed/data/corporate/documents / lm 白皮书- intel - driven - defense .pdf

[5] ENISA 威胁分类:构建威胁信息的工具 2016 年 1 月

[6] 国家网络安全战略“保加利亚网络弹性”，索菲亚，2016 年

[7] Roumen Trifonov, Slavcho Manolov,Radoslav Yoshinov, Georgy Tsochev, Galya Pavlova。通过人工智能方法充分应对新的网络安全挑战。商业与经济应用，

《经济与经济研究》，14 页，272 - 281,E-ISSN: 2224 - 2889

[8] 埃里克。Hjelmwik 2011。被动网络安全分析与网络矿法医焦点，http://www.forensicfocus.com/passive-network-security- 分析 - 网络矿工

[9] Mantas Lukosevicius 和 Herbert Jaeger 2009 水库计算方法的递归神经网络训练，计算机科学评论，vol. 3, p.127-149。

DOI: 10.1016 / j.cosrev.2009.03.005

[10] Isabelle Guyon 和 André2003 Elisseeff; a Introduction to Variable and Feature Selection, Journal of Machine LearningResearch, vol.3，

DOI: 10.1162 / 153244303322753616

[11] 征求意见 1297“内部综合故障单系统-功能规格清单”，Merit Network 公司，1992 年 1 月

[12] Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, Anil Anthony Bharath 2017。深度强化学习概述，IEEE 信号处理杂志，*图像理解的深度学习特刊，DOI 10.1109/MSP.2017.2743240*