Computer Science and Application 计算机科学与应用, 2021, 11(4), 832-839
Published Online April 2021 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2021.114085

Hans汉斯

网络安全数据采集关键技术研究

张海霞[1*]，乔赞瑞[2]，潘　啸[2]，黄克振[1]，连一峰[1]

[1]可信计算与信息保障实验室，中国科学院软件研究所，北京
[2]北京市公安局网络安全保卫总队，北京
Email: *zhanghx@tca.iscas.ac.cn

收稿日期：2021年3月12日；录用日期：2021年4月7日；发布日期：2021年4月14日

摘　要

多源异构的网络安全数据是开展安全保护工作的基础。针对网络资产、安全漏洞、网络流量、软件代码等安全数据，需要通过不同类型的关键技术进行综合采集。本文提出了从数据来源维度对网络安全数据采集技术进行分类的方法，详细阐述了流量检测、行为分析和网络探测等典型的关键技术方法，分析相关技术的特点和关联性，为建立网络安全大数据平台，开展数据汇聚、治理、分析和挖掘提供参考。

关键词

网络安全，数据采集，模式匹配，行为分析，人工智能

Research on Key Technologies of Data Collection for Cyberspace Security

Haixia Zhang[1*], Zanrui Qiao[2], Xiao Pan[2], Kezhen Huang[1], Yifeng Lian[1]

[1]Trusted Computing and Information Assurance Laboratory, Institute of Software, Chinese Academy of Sciences, Beijing
[2]Network Security Corps of Beijing Municipal Public Security Bureau, Beijing
Email: *zhanghx@tca.iscas.ac.cn

Received: Mar. 12th, 2021; accepted: Apr. 7th, 2021; published: Apr. 14th, 2021

Abstract

Multisource heterogeneous network security data is the foundation of security protection. In view

*通讯作者。

文章引用: 张海霞, 乔赞瑞, 潘啸, 黄克振, 连一峰. 网络安全数据采集关键技术研究[J]. 计算机科学与应用, 2021, 11(4): 832-839. DOI: 10.12677/csa.2021.114085

of network assets, security vulnerabilities, network traffic, software code and other security data, it is necessary to conduct comprehensive collection through different types of technologies. This paper proposes a method to classify network security data collection technologies from the dimension of data source. It describes in detail the typical key technologies such as traffic detection, behavior analysis and network detection, analyzes the characteristics and correlation of relevant technologies, which contributes to establish the big data platform for data aggregation, governance and knowledge mining.

Keywords

Cyberspace Security, Data Collection, Pattern Matching, Behavior Analysis, Artificial Intelligence

1. 技术背景

当前，网络空间安全形势日益严峻，网络空间成为国家间竞争、对抗和博弈的新战场，跨空间跨领域安全威胁不断加剧。伊朗、委内瑞拉、乌克兰、南非等国家都曾遭受过针对电力、水利、金融、交通管控等设施的网络攻击，造成大规模设施瘫痪，从而引发民众恐慌和社会动荡。另一方面，商业领域的各类大数据平台、云服务平台、无线接入网络等基础设施也逐渐成为网络攻击的焦点，利用安全漏洞进行攻击并植入后门程序，从而实施数据窃取、非法远控或恶意勒索，也成为境内外黑客组织和网络犯罪团伙的惯用伎俩。

《中华人民共和国网络安全法》规定：保护关键信息基础设施免受攻击、侵入、干扰和破坏，依法惩治网络违法犯罪活动，维护网络空间安全和秩序。因此，有必要依据国家网络安全保护的法律和政策要求，通过部署一系列技术手段、管理措施和服务机制，实现网络安全主动防御和纵深防御。

网络安全领域经历了几十年的发展过程，已经形成了多个较为成熟的关键技术门类，典型的如加解密技术、身份认证技术、访问控制技术、安全审计技术等。防火墙、入侵检测系统、漏洞扫描系统等商业化的网络安全产品也已经在各行各业广泛应用。传统的网络安全技术和产品受制于技术发展的局限性，多注重单点防护，数据来源单一，通常由单一设备完成所有的数据采集和后续的处理操作。以防火墙技术为例，通过采集流经防火墙设备的网络数据包，将其与预置的网络过滤规则进行匹配，例如禁止对内部网络除 HTTP 服务和邮件服务以外的其他应用服务，或仅允许来自特定 IP 地址段的数据包进入内部网络进行访问。随着网络信息技术的快速发展，系统规模的不断扩大，系统范围的逐步延伸，网络空间安全也进入了大数据时代。大量零日漏洞、恶意代码变种，以及分布式、协作式攻击方式的出现，对网络安全保护工作提出了新的挑战。为了有效应对网络空间威胁态势，所需采集的网络安全数据已不仅限于传统的网络流量和系统日志，而是需要及时掌握网络资产、威胁和脆弱性等各类安全要素信息，通过大数据分析挖掘，实现安全事件的监测发现和处置应对。因此，如何对各类多源异构的网络安全数据进行全面高效的采集汇聚、信息提取和数据融合，以支撑后端的数据分析，是当前网络安全保护工作的技术难点。

本文重点对网络安全数据采集的各类关键技术进行研究，在此基础上分析相关技术研究方向的发展

趋势。

2. 技术分类

网络安全数据多种多样，相关的采集技术也各有侧重。由于网络安全问题涉及网络架构、协议、流量、服务器、终端、操作系统、应用系统、数据库、人员、管理机制、安全措施等多方面因素，目前业界并没有形成统一的技术分类方法。我们首先分析网络安全数据类型，以此作为对数据采集技术进行分类的参考依据：

- 网络基础信息：包括网络拓扑结构、接入方式、带宽、协议类型、分区分域情况等；
- 网络资产信息：包括网络中各类网络设备、计算设备、安全设备的资产信息，如路由器、防火墙、服务器的 IP 地址、域名、品牌型号、操作系统类型和版本、数据库系统和版本、配置策略、应用服务等；
- 网络流量：包括通过分光、交换机镜像口或其他辅助设备提取的网络传输流量，例如原始的数据包或经过协议还原后的应用协议记录；
- 日志数据：包括设备中操作系统、数据库、应用系统、安全系统所产生并记录的各类安全相关日志，例如系统登录日志、数据库访问日志、Web 访问日志等；
- 样本文件：从网络流量、电子邮件协议、文件传输协议中提取的文档或软件代码文件，其中可能包含对目标系统进行植入、远控、数据窃取的恶意代码，样本文件可用于从中分析提取恶意代码的特征、操作行为和相关联的攻击方信息；
- 安全知识数据：网络安全漏洞库、病毒库、资产库等基础知识数据，通常由第三方服务机构提供，用于进行数据的关联比对；
- 威胁情报：由第三方提供的网络安全威胁情报，例如公开发布的漏洞隐患数据、恶意 IP 地址/域名信息、恶意样本信息等；
- 人员信息：系统管理人员、运维人员和用户的相关信息，例如人员登录系统的用户名、访问权限等；
- 管理信息：系统的安全保障策略、访问控制策略(规则)、认证方式、权限管理策略、安全审计策略、安全管理制度等；
- 其他参考信息：例如从相关媒体、网站、论坛上获取的安全相关信息，如突发的勒索病毒、突发的网络攻击等。

上述数据的类型、内容、格式各不相同，针对每类数据需要有针对性的技术对其进行采集并汇聚到网络安全大数据平台中。我们将典型的网络安全数据采集技术分为三类，包括流量检测技术、行为分析技术和网络探测技术。表 1 给出了三类技术各自适用的数据类型、技术原理、商业化产品、技术发展现状和发展趋势的概览。后续章节我们将针对三类典型网络安全数据采集技术进行详细阐述。

3. 流量检测技术

流量检测技术的数据来源是网络流量，目标是从中发现网络攻击行为(包括扫描、渗透、远控、病毒传播等)，提取攻击活动线索或证据，作为开展安全保护工作的前提。流量检测技术是目前网络安全领域中应用最为广泛的技术，防火墙、入侵检测系统(IDS)、入侵防御系统(IPS)、网络审计系统等均会使用流量检测技术。近年来，由于加密流量的逐渐增多，以及各类新型病毒、新型攻击方式的快速涌现，传统的特征匹配已越来越无法满足检测要求，各类基于统计分析、数据挖掘、人工智能的异常检测方法不断出现，推动着流量检测技术的更新换代。典型的流量检测技术包括协议还原与载荷提取、深度流检测、深度包检测，以及基于人工智能的检测技术。

Table 1. Overview of data collection technology
表 1. 数据采集数据总览表

	流量检测技术	行为分析技术	网络探测技术
适用数据	网络流量、网络基础信息、网络资产信息	日志数据、人员信息、样本文件	网络基础信息、资产信息、威胁情报、安全知识
技术原理	基于流量镜像和协议还原，检测流量中包含的攻击行为或异常活动	对网络会话、用户、样本等对象的行为进行建模和异常检测	主动探测网络中包含的资产、拓扑、脆弱性、威胁情报等安全相关信息
商业产品	入侵检测系统、防火墙	沙箱、网络审计系统	漏洞扫描器、Web 扫描器
发展现状	从早期的数据包过滤和数据流检测技术，发展到目前的深度包检测(DPI)和加密流量检测技术	从基于统计的异常流量检测和异常会话检测，发展到当前针对样本文件的 APT 攻击检测，以及利用人工智能实现用户行为分析(UEBA)	从简单的主机探测、端口扫描、漏洞扫描，发展到当前针对互联网空间的资产测绘和网络测绘
发展趋势	利用人工智能技术实现针对未知威胁行为和零日漏洞利用行为的检测	构建网络安全知识图谱，形成网络安全目标画像，实现智能推理	实现跨地理空间和网络空间的网络空间地理全域测绘

3.1. 协议还原与载荷提取技术

流量检测通过获取网络数据包，对其传输协议和传输内容进行分析，以判断其中是否包含攻击行为或异常行为。网络数据的传输必须遵循一定的通信协议，在基于 TCP/IP 协议的互联网通信中，常用的网络协议有位于传输层的 TCP、UDP，以及位于应用层的 HTTP、HTTPS、SMTP、POP、SSH、FTP、DNS 等。这些协议都有统一的字段格式和交互流程。

为了从网络流量中对通信内容进行全面分析，必须对原始流量包进行协议解析，根据其协议类型将其还原为不同种类的应用访问记录，其中包含通过应用协议传输的载荷内容，例如访问网页的 URL 请求和网站返回内容、用户登录 SSH 服务器执行的操作命令和返回结果等。

目前，由于各类网络设备安全能力的提升，传统的利用网络协议漏洞实施的攻击行为(例如 Flood 攻击)已经逐渐失去效果。威胁较大的主要是各类针对操作系统和应用服务的攻击活动，例如口令爆破利用服务器口令的复杂性缺陷，注入攻击利用 Web 服务程序在输入校验方面的漏洞，缓冲区溢出则利用应用程序在编码实现方面的疏漏。因此，必须对网络流量中的载荷内容进行完整的提取，获取通信双方的交互内容，才能作为判断是否存在攻击活动的依据。协议还原和载荷提取技术相对成熟，主要的技术差距体现在对于各类应用协议的覆盖能力以及协议还原的性能方面。

3.2. 深度流检测技术

“爱因斯坦”计划是一个网络安全自动监测项目，由美国国土安全部(DHS)下属的美国计算机应急响应小组(US-CERT)开发，用于监测针对政府网络的入侵攻击行为，保护政府网络系统安全。

“爱因斯坦”计划分为三个阶段，第一个阶段始于 2003 年，采用了基于深度流检测(DFI)的分析技术，对网络异常流量进行检测和趋势判断。深度流检测技术不提取应用载荷内容，只针对网络流传输协议字段进行分析，包括 IP 地址、端口、协议类型、标志位、时间戳、流长度、流持续时间等。深度流检测技术仅能发现网络流层面的异常和利用协议漏洞实施的攻击活动，监测发现能力十分有限。

3.3. 深度包检测技术

由于深度流检测技术无法发现应用协议内的攻击活动，因此于 2009 年启动的“爱因斯坦 2”计划全面应用了深度包检测(DPI) [1]技术，基于典型的特征匹配方式，通过提取网络数据包的载荷内容，与预置

的已知攻击行为特征进行比对，一旦比对成功就发出安全告警。深度包检测技术也是目前入侵检测系统商业产品中普遍使用的技术。

深度包检测技术对数据包全部内容进行匹配，能够发现包括协议攻击和应用层攻击在内的各类攻击行为。其检测能力主要体现在预置行为特征库的覆盖度和准确度方面，既要保证能够检测出所有已知的攻击行为，又不会被数据包中的无关内容触发从而引发误报，因此需要对特征库进行实时更新。

3.4. 智能检测技术

无论是深度流检测还是深度包检测，其采用的检测模型都是基于模式匹配或统计分析。从当前攻击活动的发展趋势来看，越来越多的攻击活动通过加密流量实施(例如网站普遍开始采用 HTTPS 协议替代原先的 HTTP 协议)，同时，攻击者为了躲避检测系统，更多地倾向于利用 0 day 漏洞渗透高价值目标，或利用各类变种、变形的木马病毒实现入侵和远控，而基于模式匹配和简单统计分析的检测方法无法应对此类攻击态势的转变，更多的高危攻击活动无法实时发现，导致关键基础设施或重要信息系统遭受恶意入侵而引发重大损失。

近年来，以神经网络为代表的人工智能技术在经济社会各个领域得到了广泛应用。由于具备自学习和迭代更新的特点，人工智能技术能够从数据样本中学习获取隐含的知识，突破了传统机器学习技术的局限性，甚至在某些特定领域已经能够超越人脑的分析能力。

人工智能技术同样适用于网络安全领域的流量检测工作[2]，从早期的数理统计、数据挖掘到现在的深度学习、增强学习，基于人工智能的流量检测已经具备了一定程度上的实用性。通过将攻击检测问题转化为基于流量数据的行为推理问题，攻击活动由于在本质上与正常网络活动存在的显著差异性，即便在具体行为方式上出现较大变化，仍然有可能利用神经网络的泛化学习能力，判断出新型攻击活动或病毒变种。人工智能技术的进步，为流量检测工作创造了新的价值，突破了以往必须依靠更新特征库才能实现检测能力提升的限制，摆脱了攻击检测工作长期存在的被动局面。

当前，基于人工智能的流量检测技术发展势头迅猛，CISCO 公司提出了加密流量分析(ETA)技术，针对加密流量 TLS 建立有监督机器学习模型，利用连接的初始数据包、数据包长度和时间顺序，以及数据包有效载荷的字节分布等参数，通过机器学习得到包含攻击行为的加密流量特征，从而实现对加密流量的攻击检测。华为、深信服、360、知道创宇等国内安全厂商也陆续提出基于人工智能的相关技术用于攻击检测，在一定程度上解决了加密流量检测的难题。

4. 行为分析技术

行为分析的概念是相对于传统检测技术中使用的特征匹配而提出的。由于特征匹配只进行数据流的简单比对，从语义角度而言并没有理解攻击行为，也无法形成针对攻击活动的完整画像，已无法满足当前应对网络安全严峻形势的需要。因此，研究人员提出行为分析的概念，尝试对攻击活动及其人员、样本、环境等关联因素进行综合分析，以便更为准确、全面地把握网络安全态势[3]。典型的行为分析技术包括软件源码分析、软件行为分析以及用户实体行为分析技术。

4.1. 软件源码分析技术

大量应用软件由于设计和实现时的疏漏而存在漏洞隐患。源码分析技术在可以获取软件源代码的前提下，利用源码扫描、污点分析、代码插桩等技术，分析软件程序的行为特点和安全缺陷，识别软件开发过程中可能存在的缓冲区溢出、输入校验错误、内存泄漏等隐患，作为网络安全工作的重要数据来源，用于判断安全事件的真实性和有效性，或进行资产关联以实现安全预警。

4.2. 软件行为分析技术

软件行为分析技术是构造虚拟的动态运行环境，观察软件在虚拟环境中运行时的行为特征，包括系统中断、系统调用、文件操作、注册表操作、开放端口、网络回连等，从而判断软件是否存在可疑的操作行为，作为检测木马病毒和 APT 样本的依据。典型的软件行为分析技术产品称为“沙箱”，业界主流的沙箱产品可以模拟 Windows、Linux、Unix 等典型操作系统、常用的数据库管理系统和应用服务，部分产品可以直接实现硬件 CPU 指令集的模拟，从而提供更为真实的模拟环境，以便全面收集软件的运行特征。近年来，研究人员开始将卷积神经网络等深度学习算法应用到软件行为分析中，通过自动训练神经网络获得软件代码的行为画像从而检测恶意软件[4]。

4.3. 用户及实体行为分析技术

用户及实体行为分析技术(UEBA) [5]是目前网络安全领域开展异常发现的重要分析技术，无论是网络环境的态势感知，用户上网行为管理，还是数据防泄漏，都是不可或缺的重要能力。UEBA 利用大数据相关技术，通过多源采集的人员信息，从人员的岗位、管理归属、访问权限、访问轨迹、可疑操作等方面进行综合画像，掌控人员行为规律，作为判断是否存在异常行为(包括攻击行为)或造成数据泄漏的依据。

5. 网络探测技术

流量检测技术和行为分析技术均采用被动方式获取网络安全数据，主要采取旁站式监测，缺乏与探测目标的交互式过程，数据的全面性和精准性有所欠缺。例如，通过流量检测技术可以发现攻击者尝试对网站服务器的攻击行为，但由于缺乏对目标服务器的深入了解，无法准确判断此次攻击行为是否对目标构成实际威胁，难以确定攻击活动是否奏效。网络探测技术是通过主动探测的方式，感知获取网络的拓扑结构、资产信息、系统配置、威胁情报等。通过主动探测方式获取的信息，可以作为网络流量和行为分析数据的有效补充。典型的网络探测技术包括漏洞探测、网络测绘和威胁情报获取技术。

5.1. 漏洞探测技术

漏洞扫描是指基于漏洞数据库[6] [7]，通过扫描等手段对指定的远程或者本地计算机系统的安全脆弱性进行检测，从而发现安全漏洞的一种检测行为。漏洞扫描包括网络漏扫、主机漏扫、数据库漏扫等不同类型。通过漏洞扫描可以了解网络的安全设置和运行的应用服务，及时发现安全漏洞，评估网络的风险等级，根据扫描结果进行漏洞修补和系统加固。

漏洞扫描通过主动获取目标系统的版本、系统设置、配置参数等信息，与已知漏洞的存在环境进行比对，从而实现漏洞信息的检出。漏洞扫描通常需要结合渗透测试技术，通过模拟攻击者的行为，确认可被成功利用的安全漏洞，形成实战化的漏洞探测结果。有效的渗透测试可以发现操作系统、应用程序、数据库、网络设备等存在的可实际利用的漏洞，从而全面评估被测试网络中存在安全方面的技术风险。

5.2. 网络测绘技术

网络测绘技术是资产探测技术的升级，传统的资产探测技术主要关注资产的基础信息和安全配置信息，网络测绘技术则在此基础上扩展了针对资产地理位置和关联性的探测，通过绘制网络空间的节点和连接关系图，形成覆盖全网络空间的图谱。

网络测绘技术是当前开展大规模网络安全监测分析的基础，通过准确掌握互联网资产状况，实现网络资产画像，与漏洞隐患和威胁攻击行为进行关联分析，有效支撑跨地理空间和网络空间的安全保护工

作，提高监测发现的精准性。当前，互联网上有多个项目致力于网络测绘，例如 Shadon、Zoomeye 等。相关机构的研究人员提出了网络空间地理学的概念，从传统地理学的“人地”关系演变为更为复杂的“人地网”关系，从而为网络空间的科学认知、地理学与网络空间安全等学科建设以及国家网络安全防控和全球网络空间命运共同体的构建提供了新视角[8]。预计未来网络测绘作为网络空间地理学的核心领域，将会发展成为跨越地理学、计算机科学和网络空间安全的跨领域交叉学科。

5.3. 威胁情报获取技术

威胁情报是关于网络安全威胁、攻击活动、攻击样本、漏洞隐患、攻击资源等相关的信息，这些信息与已经发生过的攻击活动相关，也包括对未来可能的攻击活动的预测预警。目前，国内外诸多安全机构均提供威胁情报，例如 FireEye、VirusTotal、360、微步在线等，其主要内容是用于识别和检测威胁的失陷标识，如恶意样本 HASH、恶意 IP 地址、恶意域名、程序运行路径、注册表项等。

威胁情报旨在为面临威胁的资产主体提供全面的、准确的、带有预警性质的安全知识和信息。获取威胁情报可以通过商业交换或共享方式，也可通过主动获取方式得到开源威胁情报，后者利用威胁情报源提供的 API 接口或直接采用网页爬虫获取数据，将其解析为可供机读的要素信息，或利用自然语言处理模型，对网页内容进行语义识别和分类，提供给网络安全大数据平台使用[9]。

6. 小结

网络安全工作涉及多源异构的各类数据，包括资产数据、漏洞数据、流量数据、行为数据等，需要综合利用各类数据采集技术，实现网络安全大数据的汇聚和关联分析，为开展网络安全监测发现和处置应对工作提供全面、精准的数据内容。本文重点针对流量检测技术、行为分析技术和网络探测技术三类典型的网络安全数据采集技术进行了阐述，分析了相关技术的特点和局限性，明确了关键技术之间的关联性。

当前，随着信息技术领域和网络空间安全领域的快速发展，针对网络安全数据采集的相关技术呈现出以下发展趋势：一是异构多源化，网络安全数据的类型和范围不断扩展，不仅限于以往所认知的漏洞、病毒、攻击事件等安全强相关数据，而是逐步覆盖到网络空间的各个方面，包括资产数据、用户数据、地理数据等；二是分析智能化，深度学习、增强学习等人工智能算法逐步应用到网络安全数据的采集过程中，成为分析研判网络用户行为、软件行为、威胁情报线索的有力武器，从而适应网络安全数据格式及内容的不断扩充和演变；三是协同共享化，网络安全数据采集工作越来越依赖于多方的协同共享，通过采集汇聚跨空间、跨地域、跨行业的网络安全数据，能够为精准掌控网络安全态势，实现网络威胁预测预警，指导开展网络安全主动防御提供支撑。

综上所述，数据采集是开展网络安全保护工作的前提环节，也是关系到保护工作实效性的关键要素，为了提高安全保护的有效性和精准性，有必要致力于对采集技术进行研究、扩展和升级，以适应网络安全攻防态势的不断演变和技术发展，持续提升网络安全保障能力。

基金项目

本论文得到“异构多源网安大数据采集汇聚和清洗融合技术研究”课题资助。

参考文献

[1] Cheng, Z.H., Beshley, M., Beshley, H., *et al*. (2020) Development of Deep Packet Inspection System for Network Traffic Analysis and Intrusion Detection. 2020 *IEEE* 15*th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering* (*TCSET*), Lviv-Slavske, 25-29 February 2020, 877-881. https://doi.org/10.1109/TCSET49122.2020.235562

[2] Li, Z.Y., Xian, M., Liu, J., *et al*. (2020) The Development Trend of Artificial Intelligence in Cyberspace Security: A Brief Survey. *Journal of Physics*: *Conference Series*, **1486**, Article ID: 022047. https://doi.org/10.1088/1742-6596/1486/2/022047

[3] Ahn, S., Paek, Y., *et al*. (2020) Hawkware: Network Intrusion Detection Based on Behavior Analysis with ANNs on an IoT Device. 2020 57*th ACM/IEEE Design Automation Conference* (*DAC*). San Francisco, 20-24 July 2020, 1-6. https://doi.org/10.1109/DAC18072.2020.9218559

[4] Huang, X., Ma, L., Yang, W.Y., *et al*. (2020) A Method for Windows Malware Detection Based on Deep Learning. *Journal of Signal Processing Systems*, **93**, 1-9.

[5] Das, S. (2019) Taking Cyber Security to the Next Level. *Dataquest*, 37, 44-45.

[6] Common Vulnerabilities and Exposures (CVE). The MITRE Corporation (2011) http://cve.mitre.org

[7] Mell, P.M. and Scarfone, K. (2010) The Common Configuration Scoring System (CCSS): Metrics for Software Security Configuration Vulnerabilities. *NIST Interagency/Internal Report*, Report No. 7502.

[8] Gao, C.D., Guo, Q.Q., Jiang, D., *et al*. (2019) The Theoretical Basis and Technical Path of Cyberspace Geography. *Journal of Geographical Sciences*, 29, 1949-1964. https://doi.org/10.1007/s11442-019-1698-7

[9] Husari, G., Al-Shaer, E., Chu, B., *et al*. (2019) Learning APT Chains from Cyber Threat Intelligence. *Proceedings of the* 6*th Annual Symposium on Hot Topics in the Science of Security*, Article No. 19, 1-2. https://doi.org/10.1145/3314058.3317728