VulnData papers

1. Vulnerability databases

- NVD
- OSVDB(closed) / VulnDB(commercial)
- WhiteSource
- Advisories (project)
- Issue tracker (project)
- Vulncode-db (relevant files and patches)
- Exploit-DB
- SecurityFocus
- Mail List

2. Cleaning the NVD: Comprehensive Quality Assessment, Improvements, and Analyses [arXiv, 2020]

- 目的：改进NVD漏洞数据的准确性并分析漏洞特征
- 方法：
    - 数据
        - NVD中2018年之前的107.2k个漏洞
    - 发布日期：统计链接中数量较多的域，获取链接的（最早）发布日期
    - 厂商和软件名称：粗粒度自动匹配后人工分析
    - CVSS：2016之后使用CVSS v3.x，使用DNN模型将CVSS v2.0推导至3.0
    - 漏洞类型：使用不同方法（机器学习、深度学习等）根据漏洞描述分类，但可靠性还不够

3. Autosploit: A Fully Automated Framework for Evaluating the Exploitability of Security Vulnerabilities [arXiv, 2020]

- 目的：提出评估漏洞可利用性的自动化框架
- 方法：
    - 前提：给出有漏洞的环境和对应的exploit
    - 环境条件：
        - Access control
        - Connectivity
        - Services
        - Safeguards
        - Packages

- 框架：

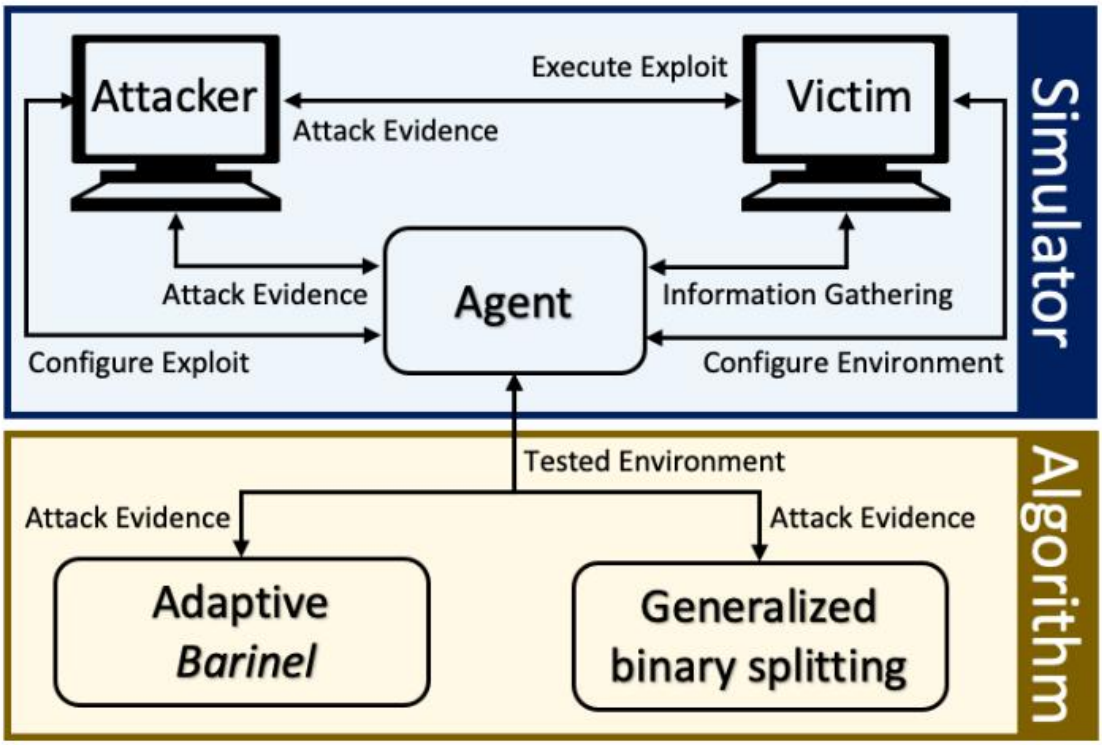


Figure 2: The architecture of Autosploit consists of two main components: a *simulator* and an underlying *algorithm*. The simulator automates the process of changing the system configuration, exploiting the system, and collecting evidence regarding the success of the exploitation. The algorithm is responsible for efficiently operating the simulator. Specifically, the algorithm determines the configuration of the environment to be tested by the simulator.

  - 模拟器：
    - Victim：Docker容器
    - Attacker：选择exploit；配置payload；执行exploit；运行post-exploitation tools
    - Agent：配置Attacker；收集漏洞所需初始信息并配置Victim；与算法模块交互
  - 算法模块：
    - 使用二分法找到必要的环境条件
    - Adaptive Barinel（fault localization technique）

4. Mind your own business: A longitudinal study of threats and vulnerabilities in enterprises [NDSS, 2019]

- 目的：分析企业中威胁和漏洞状况
- 方法和结果：

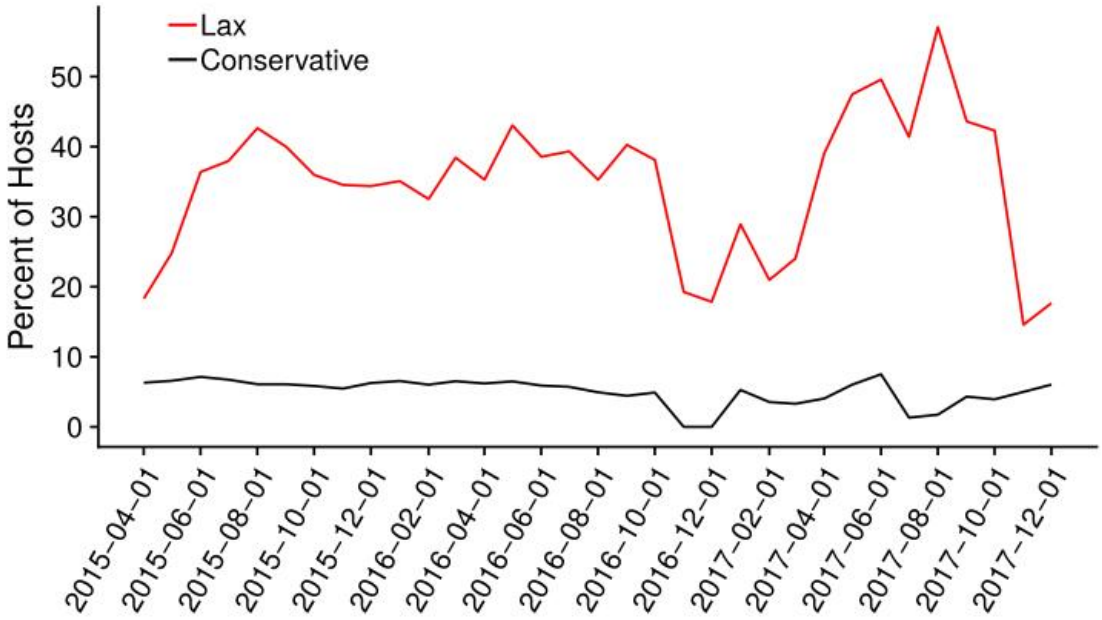
- 数据：
  - File reputation logs: 文件声誉评分，由文件特征、动态行为、流行程度、下载源和签名信息计算得出。用于区分恶意文件。
  - File appearance logs: 安装在企业主机中的可执行文件，是File reputation logs的子集。
  - Enterprise classification: 描述企业特征
  - VirusTotal: 反病毒引擎检测结果
  - NVD: 漏洞影响的软件和版本
  - Internet scans: 发现server上的应用和版本
  - IP and domiain blacklists
  - Enterprise-to-IP mapping: 映射企业和IP

Industry	Ent.	Hosts	IPs	Emp.	CC
Banks	1.1K	16.6M	7.6M	5.5M	85
IT Services	1.0K	7.5M	3,500M	3.0M	52
Healthcare Providers	1.1K	6.5M	2.9M	2.3M	46
Professional Services	875	3.8M	374.1M	1.4M	39
Commercial Services	1.2K	3.2M	366.5M	2.0M	49
Insurance	597	3.2M	1.4M	1.2M	52
Capital Markets	851	2.0M	4.1M	596K	55
Software	832	2.0M	803.8M	497K	43
Electronic Equipment	1.0K	1.7M	304.1M	1.7M	45
Machinery	1.4K	1.5M	13.3M	1.6M	49
Specialty Retail	601	1.5M	17.9M	1.6M	51
Construction & Engineering	1.3K	1.1M	471.9K	1.3M	52
Media	971	1.5M	96.6M	1.3M	44
Chemicals	850	1.0M	1.3M	909K	54
Food Products	846	872K	594.0K	1.6M	61
Financial Services	602	827K	749.1K	317K	47
Hotels Restaurants & Leisure	567	752K	1.2M	2.8M	46
Trading Companies	718	714K	12.4M	542K	40
Internet Software & Services	567	572K	407.8M	207K	34
Metals & Mining	874	506K	1.9M	1.8M	56

TABLE II: Number of enterprises, hosts, IPv4 addresses, employees, and country codes for the top 20 industries sorted by number of hosts. The high number of IPs for IT Services is due to that industry including ISPs and hosting providers.

- 威胁格局分析
  - 分类恶意的文件和PUP（有害程序）
  - 分析主机和企业遇到的恶意软件和有害程序：41%的主机和97%的企业存在一个或以上这样的软件；电子设备企业中最多76.4%

        - 纵向分析：以月为时间间隔



(a) By hosts

        - 将企业IP和黑名单结合分析其恶意行为，最多的是垃圾邮件

    - 漏洞修复行为分析

        - 客户端程序漏洞修复时间：根据File appearancre log，分析软件漏洞版本-无漏洞版本的时间间隔
        - 服务端程序漏洞修复时间：根据Internet scans，分析漏洞修复时间

5. Analysis of operating system diversity for intrusion tolerance [Software: Practice and Experience, 2014]

- 目的：在入侵容忍系统中应用操作系统多样性所获得的安全特性是怎样的?
- 方法：

    - 数据：

        - NVD中2563个OS-level漏洞，1994-1010，11个操作系统

    - 人工将漏洞分配到内核、驱动、系统软件和应用
    - 分析不同系统中的共有漏洞
    - 分析共有漏洞随时间的变化
    - 入侵容忍系统中部署操作系统多样性的策略，减少共有漏洞
    - 结果：在一段时间间隔（几年）中，有几对OS共享了少量漏洞，甚至没有共有漏洞。

6. From Analysing Operating System Vulnerabilities to Designing Multiversion Intrusion-Tolerance Architectures [Transactions on Reliability, 2019]

- 目的：分析操作系统中的漏洞，
- 方法：
  - 数据：NVD中的漏洞数据，2012-2017，6个操作系统
  - 统计分析漏洞生命周期特征
    - Days-of-Grey-Risk
    - Forever-Day Vulnerability Statistics
    - Vulnerability Severity
    - The Most Critical Types

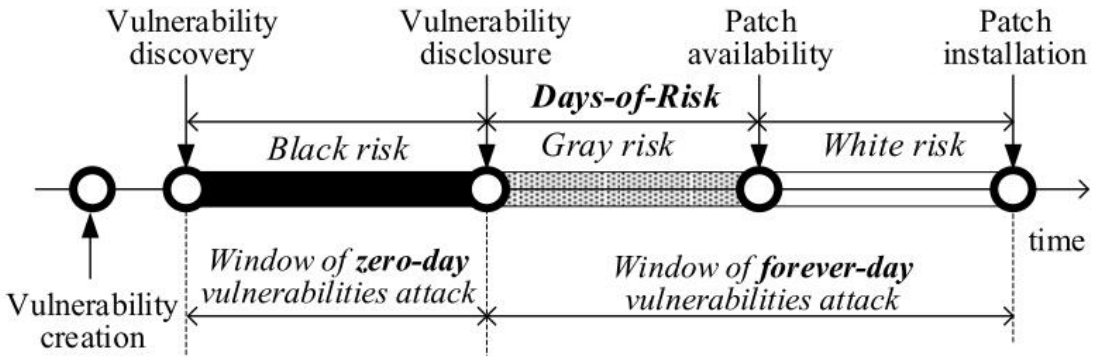


Fig. 1. Vulnerability lifecycle.

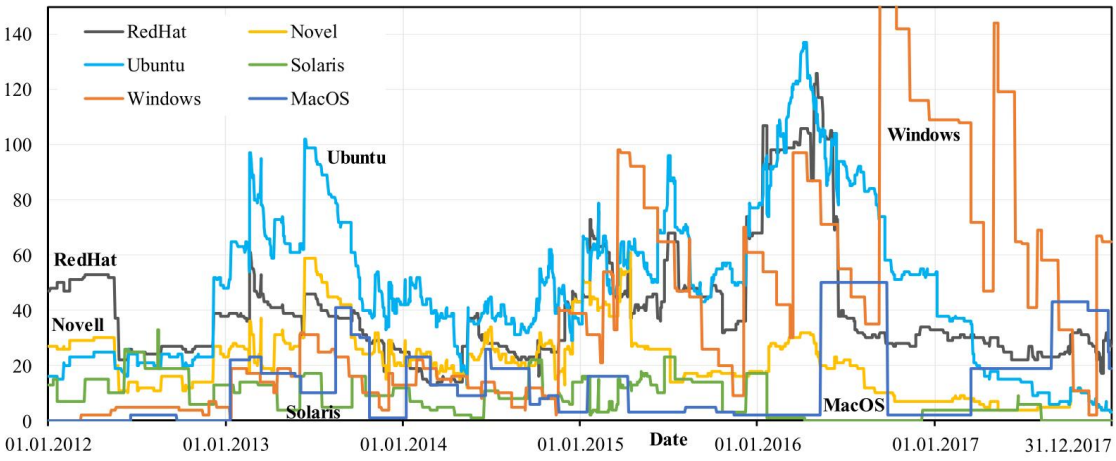


Fig. 4. Forever-day vulnerabilities.

- 分析共有的OS漏洞

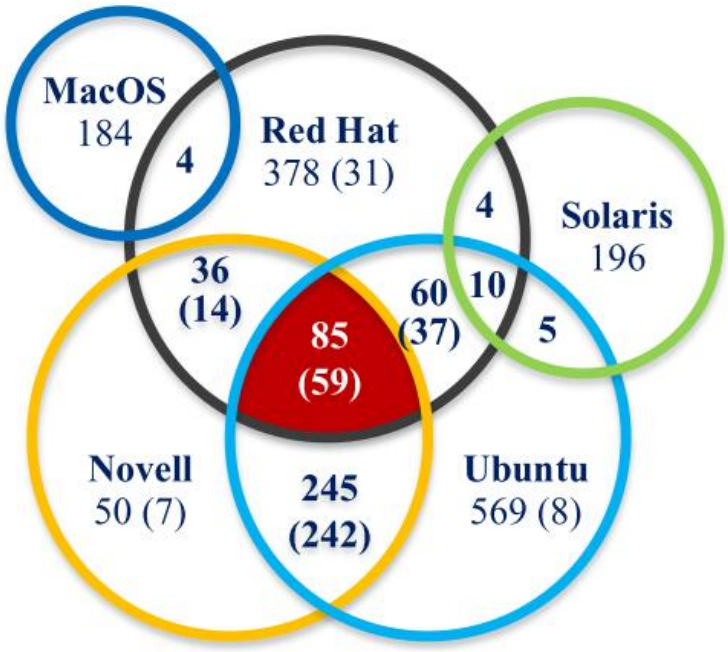


Fig. 9. Number of individual and common vulnerabilities shared by Linux (Ubuntu, Novell and Red Hat) and Unix (MacOS and Solaris) families of OSes.

- 操作系统多样性和入侵容忍架构

7. On the Effectiveness of Software Diversity: A Systematic Study on Real-World Vulnerabilities [International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment, 2009]

- 目的：分析应用软件中的漏洞、共有漏洞和多样性
- 方法：
  - 数据：
    - NVD中2007年的6340个漏洞
  - 软件分类：
    - 应用软件
    - Web脚本模块
    - 操作系统
    - 程序语言和库

        - 其他

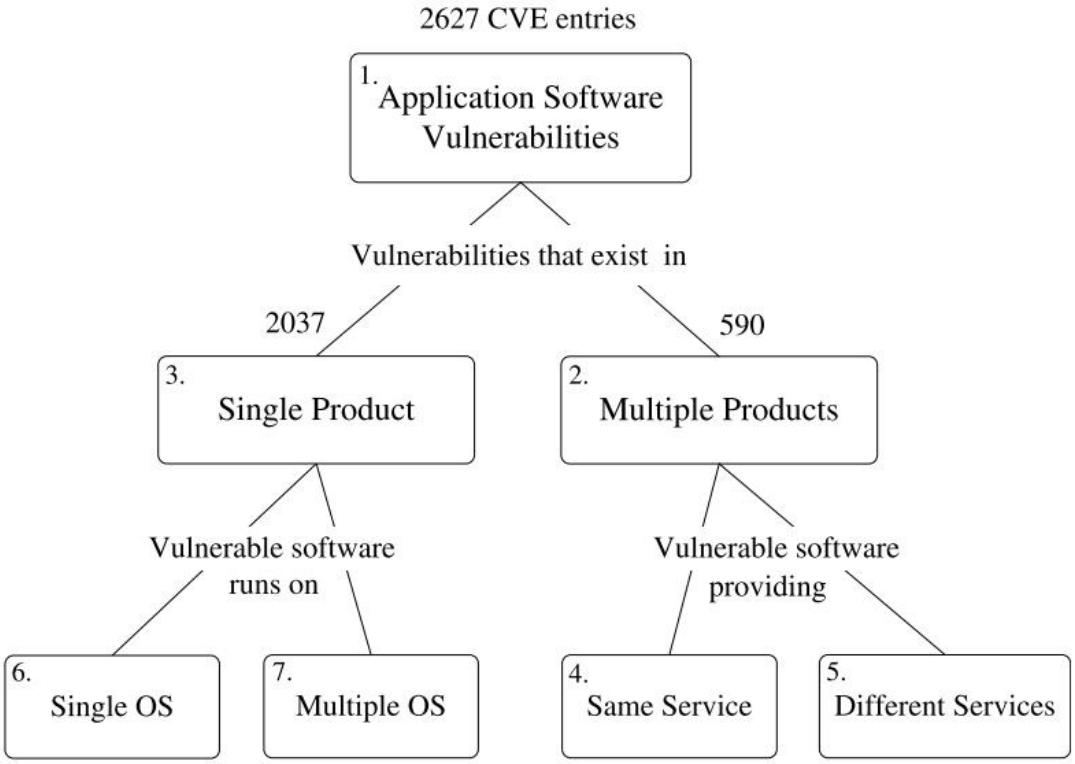


Fig. 3. Analysis on application software vulnerabilities

    - 漏洞的分布
    - 不同软件分类中的共有漏洞（case study）

8. Small World with High Risks: A Study of Security Threats in the npm Ecosystem [USENIX, 2019]

- 目的：分析javascript npm软件生态中的安全问题
- 方法和结果：
    - 数据：
        - npm仓库中的的package release，构造依赖图
        - 开发者
        - 漏洞（issue advisories）
    - Package依赖分析
        - 直接和传递依赖分析：一个package平均依赖80个其他package
        - package影响分析：popular package可能影响10万个其他包
    - 开发者分析
        - 每个开发者负责的package数量和变化
        - 可信的开发者
        - 开发者的影响（package的依赖）
    - 漏洞的演化
        - 漏洞总数、未修复漏洞的数量和密度变化

研究计划：

- 目标：刻画软件系统在任意时刻的安全状况和演化
- 方法：
    - 数据：
        - 漏洞数据收集：
            - 尽可能全面的漏洞数据，漏洞数据来源：
                - NVD
                - SecurityFocus
                - WhiteSource
                - Vulncode-db (relevant files and patches)
                - Exploit-DB
            - 漏洞属性：
                - CVSS
                - CWE
                - exploit
                - **product and version** （包含闭源软件，无法确定漏洞的引入时间，只能以漏洞库中给出的为准）
        - 漏洞数据对齐、聚合
            - 解决不同数据源的重复、缺失问题，保证准确性
            - （漏洞知识图谱，有关联则判断是否是同一实体）
        - 软件分类：
            - OS
            - Library
            - Application
            - Network firewall
        - 时刻T不同主机上的软件栈：
            - 真实数据：file log / software list
                - 文章4中，imdea（马德里研究所）和symantec使用真实数据。没有途径获取
            - 自己构造：
                - 操作系统安装后的默认软件栈
                - 需要解决软件何时升级的问题
        - 不同类软件中漏洞的分析和演化
        - 刻画任意时刻T，不同配置的系统中的漏洞和演化
        - 是否存在共有漏洞，共有漏洞的影响分析
        - 软件多样性对安全的影响