



一种基于同源行为分析的 APT 异常发现策略

俞艺涵, 付钰, 吴晓平, 李洪成

(海军工程大学信息安全系, 湖北 武汉 430033)

摘要: APT(advanced persistent threat)攻击的日益频繁对 APT 攻击行为的检测提出了更高的要求,对同源行为进行分析是尽早发现 APT 攻击行为的一种有效方法。针对数据量过大造成数据对比认证效率低下的难题,提出了借助数据标签技术,建立历史同源行为数据库,并将数据库存储到云端;依托 Hadoop 平台和 MapReduce 聚合计算能力,基于伪随机置换技术完成网络全流量并行检测,通过与数据库中的数据标签进行对比验证,来判断是否有 APT 攻击行为。测试结果表明,该方法可尽早从网络中发现 APT 异常行为,提高全数据流检测的效率。

关键词: APT 防御;同源策略;实时检测;数据标签;伪随机置换

中图分类号: TP309.7

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2016012

A discovery strategy for APT anomaly based on homologous behavior analysis

YU Yihan, FU Yu, WU Xiaoping, LI Hongcheng

Department of Information Security, Naval University of Engineering, Wuhan 430033, China

Abstract: As APT (advanced persistent threat) attacks are increasingly frequently, higher requirements for the detection of APT attacks were proposed. It was an effective method to early discover the attack behavior of APT based on homologous behavior analysis. Aiming at the problem of low efficiency of data authentication caused by excessive data, the historical behavior database with data label technology was established and the database was stored in the cloud. Relying on the Hadoop platform and the aggregate computing ability of MapReduce and the pseudorandom permutation technique, the whole traffic parallel detection of the network was realized. In order to determine whether there was a APT attack behavior, the detection of APT attacks was implemented by comparing the data labels in the database. Test results show that the proposed method can detect the abnormal behavior of APT from the network as soon as possible and improve the efficiency of the whole data flow detection.

Key words: APT defense, homologous strategy, real-time detection, data label, pseudorandom permutation

收稿日期:2015-07-06;修回日期:2015-11-02

基金项目: 国家自然科学基金资助项目(No.61100042);湖北省自然科学基金资助项目(No.2015CFC867);信息保障技术国防重点实验室基金资助项目(No.KJ-13-111)

Foundation Items: The National Natural Science Foundation of China(No.61100042), The Natrual Science Foudation of Hubei (No.2015CFC867), Project of National Defense Key Laboratory of Information Security Technology (No.KJ-13-111)

1 引言

近年来,高级持续性威胁(advanced persistent threat, APT)越来越引起人们的广泛关注,APT具有应用技术手段高超、潜伏时间长、目的明确等特点,其引发的安全事件往往是对关键要害部门造成极大安全威胁的社会性事件^[1]。由于APT攻击者在进行攻击前往往进行长时间的准备,且进行攻击时的时间跨度可长达数月甚至数年,这使得如何准确预测APT的攻击时间成为难点^[2,3]。

目前,已有大量的学者对APT攻击的原理及特点进行了相关研究,也有学者提出了应对APT攻击的安全架构和防御策略。杜跃进等人^[4]提出了一种高低位协同监测的方法,剖析了APT的外延和内涵,将其抽象为数学模型并建立了威胁模型,从APT攻击的源头、途径和终端3个层面监测和发现APT异常行为;李凤海等人^[5]提出了站在全局角度,利用大数据全方位地对APT攻击进行统筹预测与防御的方案;Cole E^[6]提出了应对APT几个关键性阶段针对性防御的策略。但是,目前对于如何尽可能早地发现APT攻击行为还没有具体可实施的方案。

APT攻击的一个显著特点是其在一个长时间跨度内进行持续性进攻,可能是几个月甚至几年。在这样一个长时间跨度的攻击中,如何能尽可能早地发现攻击者的行为是APT攻击的防御方面面临的难题。而在攻击实施的最初阶段,攻击者往往寻找被攻击方的一个薄弱点进行攻击(单点突破阶段)。在此过程中,攻击者会利用技术手段或社会工程学手段先控制被攻击网络中的一台主机、一个路由器或者一个用户权限作为跳板来为下一步攻击行为做好准备。在这一时段内,同一来源的主机、设备或者操作者的行为必将出现反常,如访问非常用IP地址、权限扩大、频繁登录等^[7,8]。本文旨在抓住APT攻击的这一特点,对同源行为进行分析,从而尽早发现APT攻击。

在一个较长的可信时间段内,以网络中各台主机为对象,收集其在该时段内的所有单位时间窗内的网络数据流信息,包括操作者身份、使用权限、源IP地址、目的IP地址、源端口信息和目的端口信息,生成历史同源行为文件 M_i ,对其进行相应处理后建立历史同源行为数据库。进行检测时,收集实时同源行为数据与历史同源行为数据做对比验证,判断是否有异常同源行为的发生,为进一步判断是否存在APT攻击行为提供依据。

以APT攻击的单点突破阶段为突破点,创新性地提出一种基于同源行为分析的APT异常发现策略,该策略依托Hadoop架构以及MapReduce技术,运用基于散列函数的数据标签技术和伪随机置换并行验证技术,对防御网络中来自同一主机的网络行为进行学习并建立云数据库,对网络数据流进行实时监控,通过与数据库中的数据标签进行对比来判断是否有APT攻击行为。本文的创新点在于:利用APT攻击在单点突破阶段的特征,提出同源行为的概念,并通过分析同源行为对APT攻击进行检测;提出基于散列函数的数据标签技术来对网络流量数据进行处理,从单个数据对比验证的角度降低数据对比验证时的开销;提出基于伪随机变化的数据标签对并行验证技术,对实时监测时的历史行为数据库进行抽样,从全流量检测的角度进一步降低数据对比验证时的开销。

2 基于散列函数的数据标签技术

如果简单地将一定时间段内一台主机的不同历史同源行为数据进行存储,实时同源行为数据通过遍历数据库进行对比验证,这样不仅不能满足数据库的安全要求,当历史同源行为数据的大小超过某个值以后,进行验证的开销和耗时将随实时同源行为数据大小的增加而呈指数倍增长。最直接的数据对比验证方法是将原始数据与现有数据进行直接对比验证,虽然这种方法能够完全准确地比较出数据是否相同,但是在APT防御中将实时同源行为文件与海量历史同源行为文件进行对比验证,显然开销太大且太过低效,为此本文提出一种基于散列函数的数据标签技术。

数据标签^[9]是由历史同源行为文件进行相应运算后生成的、唯一代表历史同源行为的一种信息。在验证时,无需访问原始历史同源行为文件,也不需要网络传输文件,只需少量的数据传输即可进行实时同源行为与历史同源行为的对比验证。

数据标签生成模块将历史同源行为文件 M_i 以时间窗序列分割成 N 个子文件,对每个子文件进行运算生成数据标签,计算式为:

$$T_{im} = (h(W_i) \times g^m)^d \bmod n \quad (1)$$

其中, $W_i = v \parallel i$, i 为历史同源行为子文件 m_i 中的索引 i , v 为一个安全素数,在文件处理的过程中随机生成,将 v 和 i 连接之后相当于对索引文件进行了加密处理,可以保证 W_i 不同且不可预测,满足了安全性要求。 $h(x)$ 为散列函数,



在这里使用 MD5 消息摘要算法,经散列转换后可得到一个 128 位的二进制数并将其转换成大数,参与随后的运算。 g 为一个安全素数,并满足 $g \bmod n=1$, m 为文件块转化的十进制大数, d 是 RSA 算法中生成的密钥, n 是 RSA 算法中生成的大素数模底。

生成的数据标签传输给云数据库进行保存,原始历史同源行为文件则可以不作保留。在这里需要解释的是,考虑到云数据库也作为需防御保护的一部分,也可能遭受 APT 攻击,为满足其安全性,采用散列函数的数据标签技术而不是应用简单的数据标签技术。

3 基于伪随机置换的数据标签对并行验证

在实时检测过程中,可根据实际防护需求将系统部署在网络的主干链路中,针对全网进行检测,也可将系统部署在分支网络中,特定地针对某一台主机进行检测。在对网络中的全数据流进行检测的过程中,可简单地通过抓取分组技术获取分组头五元组信息作为数据源,从而能使检测数据的获取相对简单,但可靠性有所降低;也可以把通过深层协议解析技术获取的协议类型、使用权限等同源信息作为数据源,从而使检测数据的可靠性更高,但网络开销更大。为实现对网络全数据流的检测,解决检测中的开销问题,本文从硬件架构和检测策略两个方面进行解决。

在硬件架构方面,将数据库存储在云端,采用 Hadoop 架构,并利用 MapReduce 技术对数据标签的对比验证请求进行处理,合理分配分布式计算资源,使大规模网络数据环境下的实时检测成为可能。实时检测时获得的数据将通过系统直接上传至云端,数据的处理与对比验证在云端完成,占用的是检测系统的计算资源而不是网络传输资源。同时,利用 Hadoop 架构可随时增加 DataNode 的数量,系统具有很强的易扩展性。

在策略方面,传统的全流量检测通常是对数据源进行抽样来降低开销,这样做将使漏报率大大增加。提出了一种全流量数据源采集而抽样数据库对比的思想,在验证实时行为数据标签时,采用基于伪随机置换的数据标签对验证方法,在降低漏报率的同时大大降低了对比认证时的网络开销。伪随机置换(pseudo-random permutation, PRP)是序列密码中产生伪随机数的一种方法。由于一台主机的历史同源行为数据量很大,且有相同行为的数据同时存在云数据库中,只是因为序列索引

使得数据标签有所不同。同时,对于一台主机而言,由网络行为学分析可知^[10],在绝大多数时间段内,其网络行为是稳定且“一致”的。由此,在进行实时同源行为数据标签与历史同源行为数据标签的对比时,并不需要遍历 100% 的历史同源行为样本。

同源行为并行对比验证过程如图 1 所示。

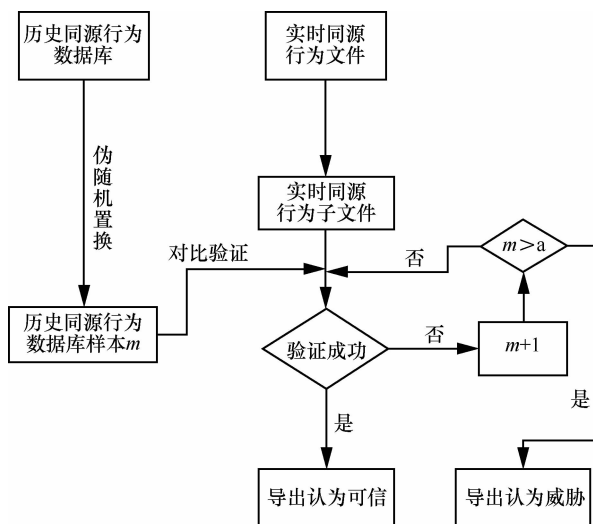


图 1 同源行为并行对比验证过程

具体步骤如下。

步骤 1 由相应主机对应的历史同源行为文件分割成的子文件数 N , 确定需对比的子文件数据标签数 C 。 C 是一个小于 N 的数, 其值越接近 N 则可认为数据标签对比的完整性越高, 相应的计算开销也会越大。 C 的具体取值是由 N 、检测对比需求和计算开销综合决定的。

步骤 2 在对比验证时, 由 $\pi: \{0,1\}^k \times \{0,1\}^{ln} \rightarrow \{0,1\}^{ln}$ 随机生成一个验证密钥 K , 最后生成 i_j , 其中 i_j 与 m_j 中的 i 有确定的对应关系, 且 $1 \leq j \leq C$ 。由此, 随机选择出历史同源行为子文件数据标签中的一部分与实时同源行为进行对比验证, 如图 2 所示。

步骤 3 对一定时间段内的不同时间窗中的实时数据并行进行验证, 即无需依循时序限制, 生成随机验证样本后将 i 值赋予各时间窗的实时数据生成验证标签, 采用并行计算, 同时进行多个标签对的对比。

步骤 4 一旦发现有与历史数据标签吻合的实时数据则马上滤除并认为其可信。

步骤 5 将通过一轮随机验证样本而没有发现吻合验证的实时数据导入下一轮。如果在一个合理的时间段内都不能找到吻合样本标签, 则可判断其为异常行为, 将其

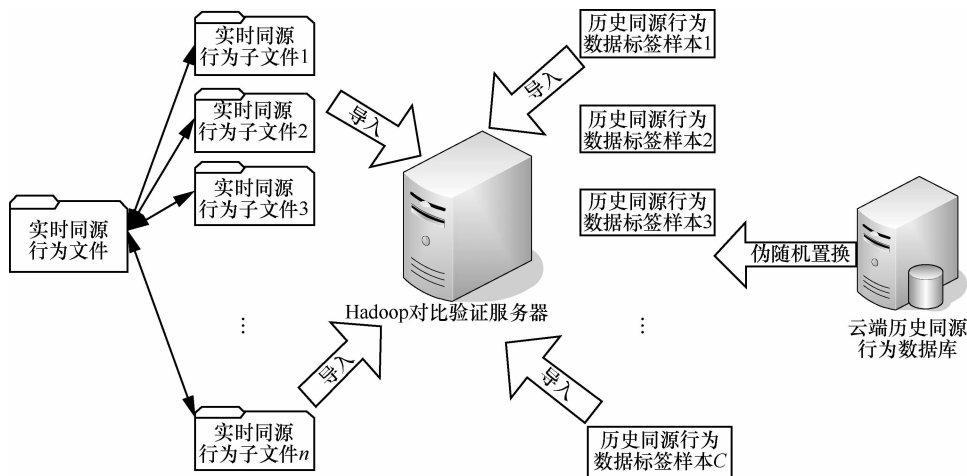


图2 实时同源行为对比验证

标记并上报。

另一方面,在检测过程中,由于是对全流量进行检测,对比验证结果会有一定的时延,并不能做到绝对实时性,但这样的时延对整个单点突破阶段及时间跨度长达数十个月的整个APT攻击来说可忽略不计。因此,可认为这样的同源行为异常发现是有效的。

4 实验测试

利用5台IBM X系列机架式服务器搭建Hadoop平台,10台PC终端和一个交换机组成测试环境,其网络拓扑如图3所示。准备不同大小的历史同源行为数据分组和实时同源行为数据分组,分别对本文提出策略的效率和可靠性进行测试。

4.1 效率测试

在不同大小历史同源行为数据库下,对不同大小的实时同源行为数据按照本文提出的策略和一般对比验证策略进行对比测试,测试结果见表1。

由测试结果可以看出,当检测的历史同源行为样本与实时同源行为数据都较小时,一般的对比验证方法在效率上高于基于本文策略的对比验证方法。然而,当历史同源行为样本与实时同源行为数据的大小增加时,一般的对比验证方法的耗时呈线性增长,而基于本文策略的对比验证方法的耗时的增长并不明显且远小于一般的对比验证方法,符合在大规模网络环境下应用的要求。

在实际应用环境下,随着网络中主机用户的增多,Hadoop平台中的分布式计算节点的数量可以相应地增

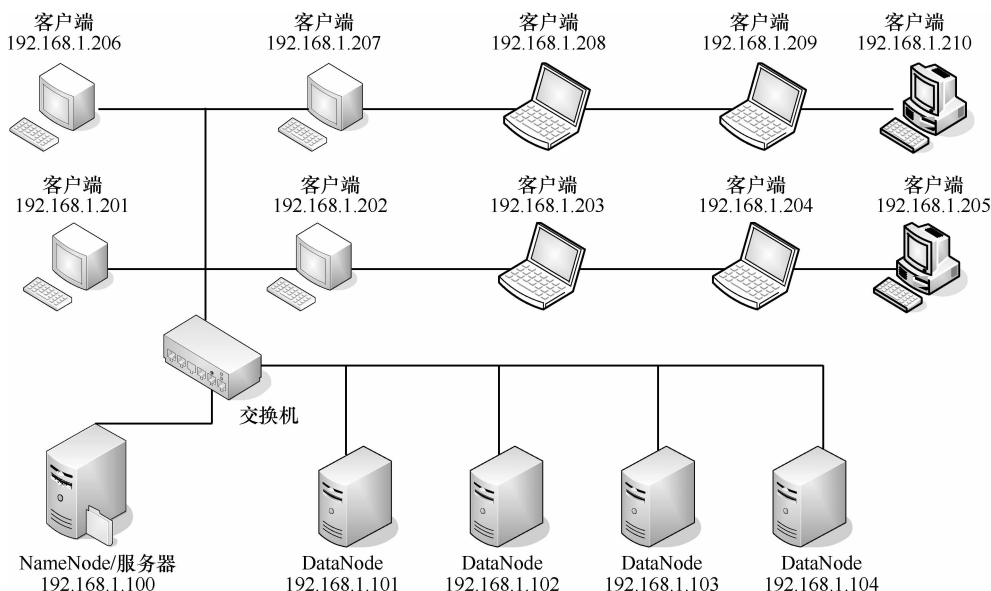


图3 测试平台网络拓扑



表 1 效率测试结果

历史同源行为 文件大小/GB	实时同源行为 文件大小/MB	基于本文策略对 比验证耗时/ms	一般对比验 证耗时/ms
1	0.1	122	4
	0.5	198	20
	1	278	45
	4	465	180
	16	745	645
4	0.1	319	17
	0.5	416	78
	1	678	204
	4	1 452	746
	16	1 671	2 559
16	0.1	1 467	72
	0.5	1 612	323
	1	2 788	852
	4	3 678	2 265
	16	4 121	9 743
64	0.1	2 678	345
	0.5	3 832	1 546
	1	5 897	3 567
	4	7 653	9 078
	16	13 234	45 342
512	0.1	3 456	3 423
	0.5	4 563	14 356
	1	6 784	26 789
	4	11 481	78 906
	16	18 145	401 456

多,由于是并行分布式计算,单位节点下处理数据的开销趋于一致,所以可以认为效率测试的结果具有普遍性。

4.2 可靠性测试

在历史同源行为文件大小为 512 GB 的情况下,对不同大小且存在威胁的实时同源行为文件分别进行 100 次测试,统计其误报率和漏报率,测试结果见表 2。

由测试结果可以看出,基于本文策略的对比验证方法的准确性极高,可以说几乎不会出现漏报情况,而误报率则会随实时同源行为文件大小增加相应出现,合理控制实时检测时实时同源文件的分块大小将有效解决这一问题。

表 2 可靠性测试结果

实时同源行为文件大小/MB	误报率	漏报率
0.1	0	0
0.5	0	0
1	1%	0
4	1%	0
16	3%	0

总体而言,可认为系统安全可靠。

5 结束语

本文提出的基于同源行为分析的 APT 异常行为检测策略,针对的是 APT 攻击的单个突破阶段,尽可能早地发现 APT 攻击行为。依托 Hadoop 平台与 MapReduce 的分布式计算能力,运用数据标签技术和伪随机置换并行标签对比技术,做到了全流量检测,达到了预期效果。然而,APT 攻击由于其攻击手段的复杂性、攻击技术的高超性、攻击时间的持续性等特点,对其的防御不可能仅仅依赖单一的技术手段与方案,往往需要在多个层面运用多种技术手段来综合应对 APT 攻击,而本文提出的策略可作为应对 APT 攻击的第一步,为后续研究起到抛砖引玉的作用。

参考文献:

- [1] CHEN P, DESMET L, HUYGENS C. A study on advanced persistent threats [C]//Proceedings of the 15th International Conference Conference on Communications and Multimedia Security, September 25-26, 2014, Aveiro, Portugal. Berlin: Springer Press, 2014:56-73.
- [2] NIKOS V, DIMITRI G. The big four-what we did wrong in advanced persistent threat detection [C]//Proceedings of International Conference on Availability, Reliability and Security, September 2-6, 2013, Washington DC, USA. New Jersey: IEEE Press, 2013: 248-254.
- [3] YANG G M Z, TIAN Z H, DUAN W L. The prevent of advanced persistent threat [J]. Journal of Chemical and Pharmaceutical Research, 2015, 6(1): 572-576.
- [4] 杜跃进, 翟立东, 李跃, 等. 一种应对 APT 攻击的安全架构: 异常发现[J]. 计算机研究与发展, 2014, 7(7): 1633-1645.
DU Y J, ZHAI L D, LI Y, et al. Security architecture to deal with APT attacks: abnormal discovery [J]. Journal of Computer Research and Development, 2014, 7(7): 1633-1645.
- [5] 李凤海, 李爽, 张佰龙, 等. 高等级安全网络抗 APT 攻击方案研究[J]. 信息安全, 2014(9): 109-114
LI F H, LI S, ZHANG B L, et al. An anti-APT scheme research for high-security network[J]. Netinfo Security, 2014(9): 109-114.
- [6] COLE E. Advanced Persistent Threat: Understanding the Danger and How to Protect Your Organization[M]. Boca Raton: CRC Press, 2012:1-280.
- [7] 郑黎明, 邹鹏, 贾焰, 等. 网络流量异常检测中分类器的提取与训练方法研究[J]. 计算机学报, 2012(4): 719-729.
ZHENG L M, ZOU P, JIA Y, et al. How to extract and train the classifier in traffic anomaly detection system [J]. Chinese Journal of Computers, 2012(4): 719-729.

- [8] 许婷. 一种有效防范 APT 攻击的网络安全架构[J]. 信息安全与通信保密, 2013(6): 65-67.

XU T. A hierarchical-centralized network security architecture effectively preventing APT attacks [J]. China Information Security, 2012(4): 65-67.

- [9] SEJONG O, SEOG P.Task-role-based access control model[J]. Information System,2003(28): 533-562.

- [10] 张鸽. 基于网络行为分析的跨站攻防技术的研究 [D]. 郑州: 解放军信息工程大学, 2012:46.

ZHANG G. Analysis of offensive and defensive techniques cross station based on network behavior [D]. Zhengzhou: PLA Information Engineering University, 2012:46.

[作者简介]



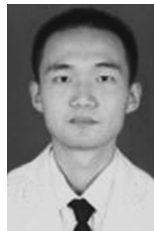
俞艺涵(1992-),男,海军工程大学信息安全系硕士生,主要研究方向为信息系统安全。



付钰(1982-),女,博士,海军工程大学信息安全系副教授,主要研究方向为信息安全风险评估。



吴晓平(1961-),男,博士,海军工程大学信息安全系教授,主要研究方向为系统分析与决策。



李洪成(1991-),男,海军工程大学信息安全系博士生,主要研究方向为大数据安全与风险评估。