基于拟态防御的VPN流量劫持防御技术

高振，陈福才，王亚文，何威振

引用本文

高振，陈福才，王亚文，何威振. 基于拟态防御的VPN流量劫持防御技术[J]. 计算机科学，2023，50(11): 340-347.

GAO Zhen, CHEN Fucai, WANG Yawen, HE Weizhen. VPN Traffic Hijacking Defense Technology Based on Mimic Defense [J]. Computer Science, 2023, 50(11): 340-347.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.221000091

基于拟态防御的 VPN 流量劫持防御技术

高　振　　陈福才　　王亚文　　何威振
解放军战略支援部队信息工程大学　郑州 450001
(15048601214@163.com)

摘　要　VPN 技术能够有效保障通信流量的保密性和完整性,但是近年来出现的名为 blind in/on-path 的流量劫持攻击利用 VPN 协议规则,通过将伪造报文注入加密隧道的方式来实施攻击,严重威胁了 VPN 技术的安全性。针对此类威胁,提出了基于拟态防御的 VPN 流量劫持防御技术,并设计了拟态 VPN 架构(Mimic VPN,M-VPN)。该架构由选调器和包含多个异构的 VPN 加解密节点的节点池组成。首先选调器根据节点的可信度动态地选取若干加解密节点,来并行处理加密流量;然后对各加解密节点的处理结果进行综合裁决;最后将裁决结果作为响应报文以及更新可信度的依据。通过对来自不同节点的同一响应进行裁决,有效阻止了攻击者注入伪造报文。实验仿真结果表明,相比传统的 VPN 架构,M-VPN 可以降低 blind in/on-path 攻击成功率约 12 个数量级。

关键词:VPN;流量劫持攻击;blind in/on-path 攻击;拟态防御;M-VPN
中图法分类号　TP309.5

VPN Traffic Hijacking Defense Technology Based on Mimic Defense

GAO Zhen,CHEN Fucai,WANG Yawen and HE Weizhen
People's Liberation Army Strategic Support Force Information Engineering University,Zhengzhou 450001,China

Abstract　VPN technology can effectively guarantee the confidentiality and integrity of communication traffic. However,the traffic hijacking attack named blind in/on-path emerged in recent years,uses VPN protocol rules to implement attacks by injecting forged messages into encrypted tunnels,which seriously threatens the security of VPN technology. Aiming at such threats,this paper proposes a VPN traffic hijacking prevention technology based on pseudo defense,and designs a pseudo VPN architecture (Mimic VPN,M-VPN). The architecture consists of a tuner and a node pool containing multiple heterogeneous VPN encryption and decryption nodes. Firstly,the tuner dynamically selects several encryption and decryption nodes to process the encryption traffic in parallel according to the node's credibility. Then the processing results of each encryption and decryption node are comprehensively judged. The decision result will be used as the basis for the response message and the updated credibility. By judging the same response from different nodes,the attacker is effectively prevented from injecting forged packets. TExperimental simulation shows that compared with the traditional VPN architecture,M-VPN can reduce the success rate of blind in/on-path attacks by about 12 orders of magnitude.

Keywords　VPN,Traffic hijacking attack,blind in/on-path attack,Mimic Defense,M-VPN

1　引言

网络流量承载着通信双方重要的基本信息以及通信数据,保护其安全性的重要意义不言而喻。但是在所有流量可能路过的节点中,往往潜伏着劫持者利用各种手段实施流量劫持攻击。其中网络劫持攻击由于隐蔽性高、更新速度快、影响范围广,已成为网络安全中常见的安全威胁。

网络劫持攻击包括 DNS 劫持[1]、重定向攻击、HTTP 注入等攻击类型。根据攻击者实施攻击时位于通信路径的位置,网络劫持攻击可分为两类[2]。1)In/on-path 攻击,攻击者位于通信路径上,攻击者控制网络基础设施来注入、拦截或延迟报文,能很容易地推断是否存在活动连接、计数报文并干扰活动连接。2)Blind off-path 攻击,攻击者位于通信路径外。Blind off-path 攻击者不能获得传输中的报文,从而无法确定具体的通信信息,但是该类攻击者可以通过侧信道的手段来推断 TCP/IP 连接是否存在[3]、计数端点之间的报文[4],甚至干扰数据流并注入数据[5-6]。

针对以上两类攻击,现有防御方案如下。1)在网络协议

到稿日期:2022-10-12　返修日期:2023-02-08
基金项目:国家重点研发计划(2021YFB1006200,2021YFB1006201);国家自然科学基金(62072467,62002383)
This work was supported by the National Key Research and Development Program of China(2021YFB1006200,2021YFB1006201) and National Natural Science Foundation of China(62072467,62002383).
通信作者:陈福才(cfc@ndsc.com.cn)

中添加秘密随机值，来保护数据流免受 blind off-path 攻击。例如，TCP 客户端随机选择的端口号、初始序列号；DNS 客户端利用随机的 transaction ID(TXID)来防止 blind off-path 攻击者的欺骗响应。2)通过在客户端和服务端之间使用加密协议来缓解 in/on-path 攻击，如 TLS[7]、基于 HTTPS/TLS 的 DNS 协议[8-10]和虚拟专用网络(Virtual Private Network，VPN)。TLS 和基于 HTTPS/TLS 的 DNS 协议通过使用会话层加密来防止恶意攻击者进行流量劫持，但是攻击者仍可以通过伪造证书[11-12]来破坏 HTTPS/TLS 提供的安全性保障。VPN 是另一种加密通信技术，通常被用来提供更安全的网络通信。VPN 采用的隧道加密、密钥管理等技术增加了 in/on-path 攻击者的攻击难度。攻击者在客户端与 VPN 服务器之间只能观察到被加密的数据报文，不能获得原报文信息，暴力解密的难度大、耗时长。

但是 2021 年发表于 30th USENIX Security Symposium 的研究[2]介绍了一种针对 VPN 协议的新型劫持攻击手段——blind in/on-path 攻击。区别于上述两种攻击，该攻击要求攻击者位于 VPN 加密隧道上；并且由于攻击所需的报文字段是不可见的，如端口号和序列号，因此攻击者是盲目的。图 1 给出了 blind in/on-path 攻击较前两种攻击类型在通信路径中的不同位置。攻击者通过伪装成中间人，在进行加解密的 VPN 服务器处注入伪造数据报文，从而达到劫持流量的目的。文献[2]中的实验结果表明，针对 VPN 服务器端的 DNS 劫持攻击，成功率最高可达 75.3%。由于攻击者采用明文注入-密文嗅探的方式猜测秘密随机值，致使目前针对网络劫持攻击的两类防御方法无法奏效，且目前还没有研究人员进行相关的防御研究，也无有效的防御策略被提出。

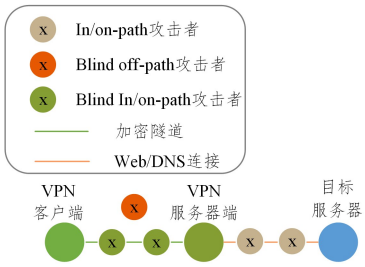


图 1　3 种不同类型攻击者在通信路径中的不同位置

Fig. 1　Different locations of three different types of attackers in communication path

针对这种新型的攻击手段，本文结合拟态防御的动态异构冗余[13-14](Dynamic Heterogeneo-us Redundancy，DHR)的思想，提出了一种拟态 VPN 架构(Mimic VPN，M-VPN)来防御此类攻击。本文的具体贡献如下：

1)针对 blind in/on-path 攻击的特点，设计了区别于传统 VPN 的动态异构冗余架构 M-VPN。M-VPN 架构由加解密池和选调器组成，加解密池由若干 VPN 加解密节点组成。选调器通过对加解密节点返回的响应进行裁决，能够有效发现并过滤伪造报文，提高加密隧道末端的 VPN 服务器的安全性。

2)通过建模对文献[2]中攻击者 DNS 劫持攻击的成功率进行了仿真实验，原实验攻击的成功率[2]与仿真实验结果的平方误差为 6.38×10^{-5}。并在白盒测试的条件下对 M-VPN 进行了仿真实验，仿真实验结果表明，M-VPN 可以降低 blind In/On-path 攻击成功率约 12 个数量级。

2　威胁模型

VPN 主要采用 4 项技术来保证安全，分别是隧道技术(Tunneling)、加解密技术(Encryption & Decryption)、密钥管理技术(Key Management)、使用者与设备身份认证技术(Authentication)。对 VPN 隧道路径上的 blind in/on-path 劫持者来说，攻击所需的报文字段是不可见的，例如端口号和序列号，因此攻击者是盲目的。但是现有的 VPN 技术不能隐藏数据包数量、大小和时间等粗粒度属性，攻击者仍可以利用这些信息推断出攻击所需的信息[15-16]。

blind in/on-path 分为对客户端的攻击和对服务端的攻击两种类型。客户端攻击要求劫持者与客户端在同一物理网络中，即在链路层相邻。而服务器端攻击只要求劫持者是 VPN 客户端到 VPN 服务端路径上的路由。这两种类型的攻击有两个相同的基本特点：1)能够将伪造报文路由到加密隧道；2)即使无法解密 VPN 隧道上的密文，也能通过加密流量的时间、大小等粗粒度信息推断出劫持攻击所需的字段信息。

对于劫持者来说，实施服务器端攻击较客户端攻击有两个优势：1)反向路径过滤、Martian 或 bogon 过滤对服务器端攻击无效；2)VPN 类型、配置、实现以及操作系统的多样性不会影响服务器端攻击。blind in/on-path VPN 服务器端攻击的攻击依赖和具体攻击细节如表 1 所列。

表 1　blind in/on-path 攻击的两种类型

Table 1　Two types of blind in/on-path attacks

攻击类型	攻击假设和依赖	攻击阶段	攻击过程	攻击结果
VPN 服务器端攻击(DNS 劫持)	1. 攻击者位于隧道经过的路由 2. 服务器的响应报文不参与竞争 3. 攻击者已知攻击域名	Phase1：推断活动连接的临时端口	发送不同端口报文到 VPN 服务器，如果端口猜中则能在隧道中观察到对应的加密报文	Phase1：推断出了 DNS 响应的临时端口
		Phase2：推断有效的 TXID	攻击者发送所有可能的伪造报文，在响应超时前碰撞出有效的响应报文	Phase2：将伪造的 DNS 响应返回给了客户端

服务器端攻击主要依赖大多数 VPN 服务器执行网络地址转换(NAT)。RFC 2663[17]中明确定义了 VPN 的规范，其中指定的 NAT 是基于五元组(协议、源地址、目的地址、源端口、目的端口)进行工作的。大多数 VPN 在隧道端点进行 NAT，以提供跨外部网络域的安全 VPN 传输。解密后的请求报文通过 NAT 转发到公网上，并在内核中留下对应的

conntrack 条目。如果响应报文字段与其相匹配，则将响应报文 NAT 到加密的 VPN 隧道中，如果不匹配则丢弃。虽然存在不使用 NAT 的 VPN 的实现，例如使用 SOCKS 代理的 Outline13，但其在本质上也遵循着与 NAT 同样的原则：具有正确字段的数据包会通过隧道传输，而具有不正确字段的数据包则不会。这意味着位于加密隧道的攻击者可以根据此原则向 VPN 隧道注入伪造报文，以推断正确的报文字段。通过这种方式，攻击者不仅可以推断出报文中的秘密随机值，也绕过了 VPN 的加密保护，使得伪造报文被成功返回给客户端。因此，目前存在的两种防御措施，即协议中添加的秘密随机值以及加密，都无法有效阻止 blind in/on-path 攻击。

针对该攻击所披露的 CVEs，已经有一些操作系统或者 VPN 供应商提供了某种补丁来减少客户端攻击，这些补丁大部分依赖于过滤伪造报文。但是，对于 VPN 服务器端的攻击，目前还没有研究人员提出有效的防御策略，防御 blind in/on-path 服务端攻击主要面临以下挑战：1）VPN 服务器无法区分伪造报文与合法报文。伪造报文与合法报文到达相同的端口，且在报文头部信息上无法进行区分。2）无法阻止攻击者通过流量分析获取粗粒度信息。尽管 VPN 采取的加密技术使得攻击者很难直接获取报文中的具体信息，但是利用报文大小和时间等粗粒度信息，攻击者仍能推断出所需要的信息。

3 基于拟态防御的 M-VPN 架构

针对 blind in/on-path 攻击中的 VPN 服务器端的攻击，本文提出了 M-VPN 架构，通过动态异构冗余机制增加了 VPN 服务器整体的随机性；并通过投票机制赋予了 VPN 服务器能及时发现注入隧道中的伪造报文的能力，保证了 VPN 隧道末端的可靠传输。

3.1 M-VPN 架构

如图 2 所示，M-VPN 架构由加解密池和选调器组成，加解密池由若干 VPN 加解密节点组成。

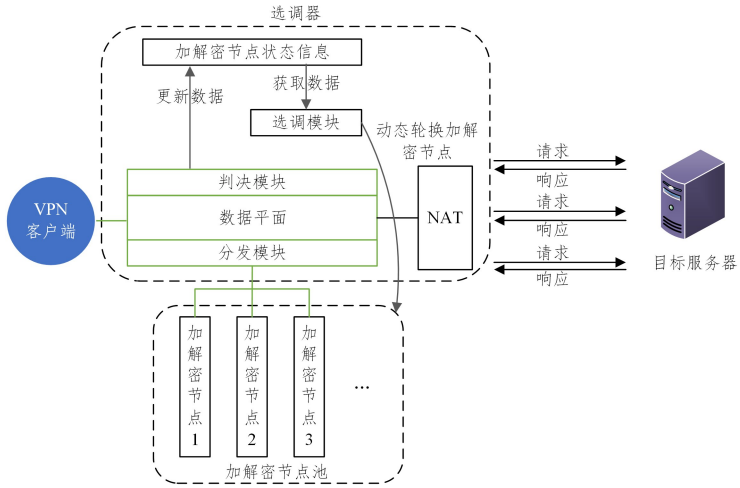


图 2 M-VPN 逻辑框架图

Fig. 2 Diagram of M-VPN logical frame

选调器由数据平面和控制平面组成，其中前者完成数据分流和裁决的功能，后者负责节点信息更新和选调节点。选调模块根据加解密节点的状态信息来动态地选取若干节点作为活跃集，并向选调器数据平面下发分流指令。裁决模块对收到的各个执行体的响应进行综合裁决，并将裁决结果作为选择响应报文的依据以及判定执行体是否异常的依据，并根据判定结果更新加解密节点的状态信息。

3.2 M-VPN 的通信流程

M-VPN 服务器的通信流程如图 3 所示。具体的服务步骤如下：

1）用户通过加密隧道发送 DNS 请求；

2）选调器动态选取 *nos* 个服务器，并为其分发加密流量；

3）对加解密节点进行解密，将解密后的 *nos* 个请求发回选调器；

4）选调器经过 NAT 将 *nos* 个请求发给一个或多个 DNS 服务器；

5）DNS 服务器对 *nos* 个请求进行响应；

6）选调器将响应 NAT 给选定的加解密节点进行加密；

7）选调器根据各个加解密节点返回的响应进行处理：

（1）若满足 $RES1=RES2=,\cdots,=RES_{nos}$，即所有响应一致，则将该响应返回给用户；

（2）若不满足 $RES1=RES2=,\cdots,=RES_{nos}$，即响应不一致，则对各个结果进行统计，依据统计结果更新加解密节点状态信息；

8）将最后的结果返回给用户。

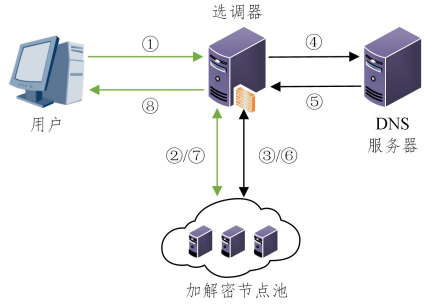


图 3 VPN 服务器的服务流程图

Fig. 3 Flow chart of VPN server service

3.3　裁决策略

裁决模块是 M-VPN 实现感知攻击的关键模块。裁决模块对各个加解密节点发过来的响应密文进行裁决，其逻辑流程如图 4 所示。正常情况下，当同一 DNS 请求的响应报文到达不同节点进行加密后，产生的响应密文负载内容应该是相同的，通过裁决模块将一致的响应返回给客户端。但是当攻击者试图对 DNS 响应报文进行 blind in/on-path 攻击时，由于攻击者伪造的报文不能同时通过 NAT 到达各个加解密节点，伪造报文不能通过大数裁决被丢弃，攻击者在隧道上不能嗅探到伪造报文相对应的密文。因此，在 M-VPN 框架下，攻击者在 phase1 阶段通过注入负载大小不同的密文来进行端口猜测的方法是行不通的。

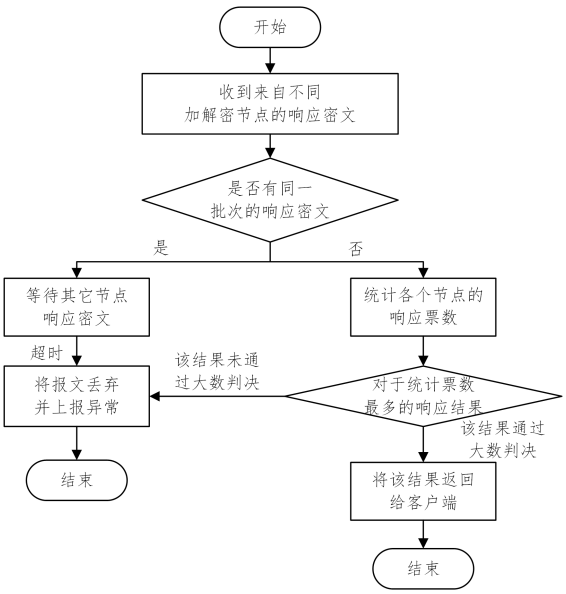


图 4　裁决策略逻辑图

Fig. 4　Diagram of decision strategy logic

3.4　动态调度

选调策略主要是依据各个加解密节点的状态信息来决定，先给出加解密节点的状态信息的相关参数和定义。

3.4.1　参数及定义

定义 1(可信度(reliability))　设加解密节点集合为 A，$\forall d\in A$，d 的可信度计算式如下：

$$d_{\text{relia}}=\begin{cases}1-\mathrm{e}^{-\frac{z\cdot x}{1-x}}, & 0\leqslant x<1\\ 1, & x=1\end{cases}\tag{1}$$

x 的计算式如下：

$$x=1-\frac{n_{\text{att}}}{\sum\limits_{n\in A}n_{\text{att}}},\ \sum_{n\in A}n_{\text{att}}\leqslant th_{\max}\tag{2}$$

其中，n_{att} 表示某一节点，$th_{\max}$ 表示最大阈值。

Wang 等[18]利用生物种群竞争的生存灭亡规律，分析了服务器遭受的攻击次数与可信度的关系。由于各个加解密节点提供相同的功能，因此可认为加解密节点池是一个生物种群，攻击者的攻击可视为捕食行为。由式(2)可以看出，每个个体的生存能力 x 都会受到攻击者捕食和种内竞争的影响：1)节点遭受的攻击次数越多，其生存力就越小；2)其他节点遭受攻击的次数越多，该节点的生存力就越大；3)但是整个种群能够承担的攻击并不是无限的，存在一个最大阈值 $th_{\max}$。综合式(1)、式(2)可以得出可信度和被攻击次数的关系，如图 5 所示(假设 $th_{\max}=100$)。

参数 z 的取值反映了选调器对该节点判决结果的敏感程度，即当该节点判决结果出现异常时，z 值越小，该节点的可信度下降得越快，越容易触发选调条件，轮换掉该节点。当然 z 值不宜设置得过小，否则可能导致系统长期处于轮换状态，使得整个系统的开销变大，影响正常的通信效率。

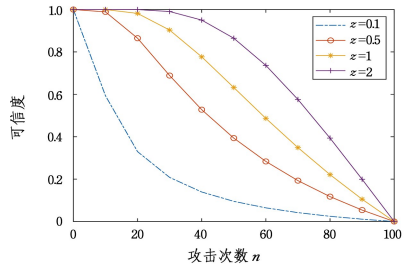


图 5　可信度与攻击次数的关系图

Fig. 5　Relationship between credibility and the number of attacks

3.4.2　选调策略

首先设置一个可信度的基本阈值 d_{basic}，如果节点的可信度小于该阈值，则将该节点加入黑名单，选调器将在节点池中重新选择一个执行体轮换不可信节点，不可信节点将不再执行加解密操作。将轮换下的节点进行初始化后从黑名单中移除。设不在黑名单中的服务器个数为 N，则具体的选调步骤为：

1)若 $N<3$，则随机选择一个节点，并定期轮换该节点；

2)若 $N\geqslant 3$，选调器生成随机奇数 nos($3\leqslant nos\leqslant n$)，在节点池中选取 nos 个节点作为活跃集。

对于选调器何时轮换与可信度之间的关系，根据式(2)可以看出，选调器进行轮换的动作是由基本阈值 d_{basic}、最大阈值 $th_{\max}$ 和敏感参数 z 共同决定的。在基本阈值 d_{basic} 和最大阈值 $th_{\max}$ 不变的情况下，发生轮换时有以下两种情况：

1)当活跃集内某一节点 $d_{\text{relia}}\geqslant d_{\text{basic}}$ 时，说明此时的参数设置对攻击者的攻击较为敏感。

2)当攻击的总数达到最大阈值 $th_{\max}$ 时，则在活跃集中轮换可信度最低的节点，这说明此时的参数设置对攻击者的攻击较为迟钝。为保证参数设置的准确性，通过算法 1 动态计算敏感参数 z 的取值，其中 $rate$ 是值动态变化率。

通过动态更新系数 z，每次触发轮换的条件动态变换，使得选调具有了动态性，增加了攻击者的攻击难度。

算法 1　动态计算敏感参数 z 的值

输入：(rate，num_{att})

输出：(z)

1. if $\text{d}_{\text{relia}}\geqslant\text{d}_{\text{basic}}$
2. $z=\text{rate}\cdot z$;
3. alter_nodes();%轮换该节点
4. else if $\text{d}_{\text{relia}}\leqslant\text{d}_{\text{basic}}$ && $\text{num}_{\text{att}}\geqslant\text{th}_{\max}$
5. $z=(-\text{rate})\cdot z$;
6. alter_nodes();
7. end if

3.5 M-VPN 架构的安全性与可行性分析

M-VPN 的动态冗余性增加了攻击者推断报文安全字段的难度。更重要的是，M-VPN 很大程度上阻断了攻击者企图通过探查报文推测字段信息的途径。也就是说，劫持者必须同时对半数以上的加解密节点实施成功的劫持攻击，才会使得攻击者的探查报文或者伪造响应出现在加密隧道中并返回给客户端。M-VPN 通过冗余以及投票的方式获得了对攻击的感知能力，并根据裁决结果及时动态地调度加解密节点，使攻击者的攻击成本增加。

拟态动态冗余模型有以下前提假设：1）系统的输入和输出一一对应；2）异构执行体对同一输入有同一输出。因此对 M-VPN 应用存在的问题进行以下说明。某些互联网公司为实现负载均衡，可能存在多个 IP 对应一个域名的情况。例如，针对某个域名的查询可能会返回 IP^1，IP^2，IP^3 或 IP^1，IP^3，IP^2 等，这种情况可以将这多个 IP 视作一个集合，从而实现域名到 IP 地址集的一一对应。

4 理论分析

本章从理论上对比分析了 M-VPN 和典型防御方案的效果，对比指标为一个有效时间内 blind in/on-path 攻击成功的概率。所需要的变量如表 2 所列。

表 2 变量名称及其含义

Table 2 Variable names and their meanings

变量名称	变量含义
P	攻击者单个伪造报文“猜中”的概率
P_{success}	攻击者攻击成功的概率
TO	DNS 超时时间
NoA	OT 时间内，攻击者发送的伪造报文数量
NoS	攻击者每秒发送伪造报文的个数
$NoTX$	TXID 数
NoP	端口数
nos	选取加解密节点的个数
$rate_{\text{loss}}$	攻击者误判率

blind in/on-path 攻击者首先通过流量分析等方式来分析加密的 VPN 数据包是否为攻击者所寻找的某一域名的 DNS 请求。并且为了避免与正确的响应竞争，攻击者对 DNS 服务器进行按 IP 地址的拒绝服务攻击，使其停止对 VPN 服务器的响应，因此攻击者必须在 DNS 响应超时前将有效的伪造报文送达客户端。攻击者能伪造的报文数量为：

$$NoA = NoS \times TO \tag{3}$$

攻击者针对同一域名的 DNS 查询请求，发送不同的端口号或 TXID 的伪造报文进行碰撞，符合生日攻击[19]的条件。根据生日攻击算式，单个报文“猜中”的概率 P 和攻击成功的概率 $P_{\text{birth\_att}}$ 的关系式如下：

$$P_{\text{birth\_att}} = 1 - \prod_{k=1}^{NoA}\left(1 - \frac{P}{1 - P \cdot (k-1)}\right) \tag{4}$$

在 phase1 阶段，攻击者计算加密隧道中加密报文的负载大小，从而匹配相对应的端口。由于加密隧道中往往存在其他通信报文，从而影响了攻击者判断的准确率。假设攻击者的误判率为 $rate_{\text{loss}}$，则单个报文“猜中”的概率 P、误判率 $rate_{\text{loss}}$ 和攻击成功的概率 P_{success} 的关系式如下：

$$P_{\text{success}} = 1 - \prod_{k=1}^{NoA}\left(1 - \frac{P}{rate_{\text{loss}} - P \cdot (k-1)}\right) \tag{5}$$

本文将求证下面各种情况下的概率。

1）基本模型：源端口和 TXID 是随机的。该模型下攻击者在 phase1 阶段需要猜中端口，并在 phase2 阶段猜中 TXID。因为攻击者对临时端口进行遍历，所以端口随机化不论发生在客户端发送 DNS 请求时，还是在 VPN 服务器转发 DNS 请求时，采取的防御策略对 blind in/on-path 攻击都是无效的。假设攻击者在 phase1 阶段尝试的报文数为 H_1，在 phase2 阶段尝试的报文数为 H_2，则：

$$\begin{aligned} P_{\text{basic}} &= P(H_1 + H_2 \leqslant NoA) \\ &= P(H_1 = h_1, P_1, rate_{loss\_1}) \cdot \\ &\quad P(H_2 = h_2, P_2, rate_{loss\_2}) \end{aligned} \tag{6}$$

其中，P_1 和 P_2 分别表示攻击者在不同阶段发送单个报文猜中的概率，即：

$$P_1 = \frac{1}{NoP}; P_2 = \frac{1}{TXID} \tag{7}$$

2）M-VPN 模型：攻击者若要攻击成功，需要同时成功攻击至少 $l=(nos+1)/2$ 个节点。假设攻击者成功时对每个节点尝试注入的伪造报文数量为 $k_1, k_2, \cdots, k_l$ 且它们相互独立，那么攻击成功的概率为：

$$\begin{aligned} P_{\text{D-VPN}} &= P(K_1 + K_2 +, \cdots, + K_l \leqslant \frac{NoA}{l}) \\ &= \sum_{k_1=1}^{\frac{NoA}{l}} P(K_1 = k_1) \cdot \\ &\quad P\left(K_2 + K_3 +, \cdots, K_l \leqslant \frac{NoA - l \cdot k_1}{l}\right) \\ &= \cdots \\ &= \sum_{k_1=1}^{\frac{NoA}{l}} \sum_{k_2=1}^{\frac{NoA - l \cdot k_1}{l}} \cdots \sum_{k_1=1}^{\frac{NoA - l \cdot (k_1 + k_2 + \cdots + k_{l-1})}{l}} P(K_1 = k_1) \cdot \\ &\quad P(K_2 = k_2) \cdot, \cdots, \cdot P(K_l = k_l) \end{aligned} \tag{8}$$

其中：

$$P(K_m = k_m) = \frac{P}{1 - P \cdot (k_m - 1)} \cdot \prod_{i=1}^{k_m - 1}\left(1 - \frac{P}{1 - P \cdot (i-1)}\right) \tag{9}$$

上述计算是假设攻击者已知 M-VPN 内部结构的白盒测试条件。1）攻击者知道选调器选出的作为活跃集的节点，并针对该活跃集的节点进行劫持攻击。2）针对某一节点攻击成功时，攻击者能瞬间感知并立即转向下一个节点。而实际上在拟态 DHR 模型下，攻击者对 M-VPN 具体的构造和内部节点的选择是不知情的，而这样的假设是攻击者成功概率的上限。

5 仿真实验

5.1 原实验攻击成功率验证

下文针对理论分析中得到的概率进行仿真验证。根据文献[2]中 blind in/on-path 攻击中已知的攻击参数，本小节通过基本模型公式对文献[2]中攻击者 DNS 劫持攻击的成功率进行对比验证。本文中基本参数的取值如表 4 所列。

5.1.1 报文分配与攻击成功率的关系

为了保证 blind in/on-path 攻击的成功率，攻击者需要将

TO 时间内产生的伪造报文数目 NoA 合理分配给两个攻击阶段。假设攻击者在 phase1 阶段消耗的报文数量为 NoA_{ph1}，则攻击成功率与 NoA_{ph1} 之间的关系如图 6 所示。

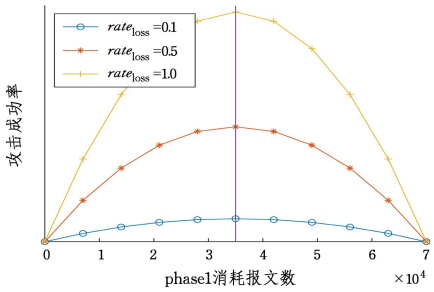


图 6　phase1 在不同误判率下消耗的报文数与攻击成功率的关系

Fig. 6　Relationship between the number of packets consumed by phase1 and attack success rate at different false positive rates

在不同条件下进行测试后发现，当攻击者将报文平分给两个攻击阶段时攻击成功率最高。下面的实验默认攻击者以这种方式进行攻击。

5.1.2　不同误判率下的误差

由于加密隧道中通常不只有加密后的伪造报文，一些正常通信的加密报文可能会给攻击者造成误判，设误判率为 $rate_{loss}$。由于攻击者只在 phase1 攻击阶段通过嗅探加密负载大小来确认临时端口，因此 phase1 存在误判，设 phase1 阶段的误判率为 $rate_{loss\_1}$。而 phase2 攻击阶段不需要嗅探报文确认是否注入成功，因此 phase2 阶段的误判率 $rate_{loss\_2}$ 为 0。探索误报率 $rate_{loss\_1}$ 的不同取值，对仿真实验结果与原实验结果之间的误差的影响，结果如图 7 所示。

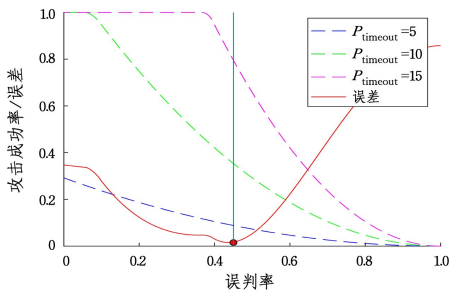


图 7　不同误判率下的误差

Fig. 7　Errors at different false positive rates

当误判率 $rate_{loss\_1}$ 为 0.45 时，误差最小，说明当误判率为 0.45 时，最接近原攻击实验的实验结果，其仿真结果与原实验结果的对比如表 3 所列。当攻击者单位时间伪造的报文数量为 12180 时，误差达到最小值，为 6.38×10^{-5}。结果证明，仿真实验可以有效地计算出在现实实验环境下攻击者的攻击成功率。

表 3　仿真结果与实验结果的对比

Table 3　Comparison of simulation results with experimental results

超时时间	实验结果/%	仿真结果/%
15	75.3	78.5
10	48.1	48.6
5	11.6	12.1

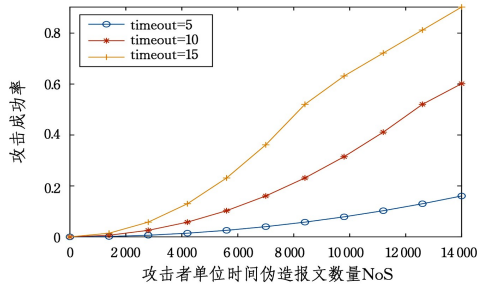
5.2　不同模型攻击成功率仿真分析

接下来对所提模型进行仿真实验，模型中参数的取值如表 4 所列，其中 nos 取 3，是 M-VPN 最简单的情况，也是该模型所能提供的安全下限。攻击成功率和攻击者每秒发送报文数量之前的关系如图 8 所示。

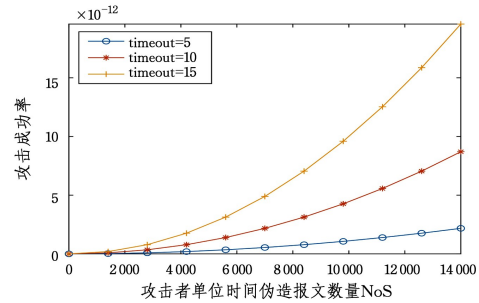
表 4　变量仿真取值

Table 4　Variable simulation values

变量名称	变量取值
Nop	64000
$NoTX$	65535
nos	3
$TO1$	5
$TO2$	10
$TO3$	15



(a)基本模型



(b)D-VPN 模型

图 8　攻击成功率与攻击者单位时间伪造报文数量的关系图

Fig. 8　Relationship between attack success rate and the number of forged packets per unit time of the attacker

表 5 列出了 $NoS=14000$，$timeout=15$ 时各个模型的攻击成功率的数量级。可以看出，对于基本模型，在超时时间足够长的情况下，攻击者可以在数量级为 10^{-1} 的概率下攻击成功。而 M-VPN 模型的数量级为 10^{-12}，可见 M-VPN 的安全性得到显著提升。

表 5　仿真结果

Table 5　Simulation results

模型	攻击成功概率数量级
基本模型	10^{-1}
M-VPN 模型	10^{-12}

5.3　时延分析

对传统 VPN 与 M-VPN 模型的时延进行对比分析，涉及的相关变量及含义如表 6 所列。

表 6 时延变量及其含义

Table 6 Delay variables and their meanings

变量名称	变量含义
T_{cv}	VPN 客户端与 VPN 服务器间的时延
T_{vs}	VPN 服务器与 DNS 服务器间的时延
T_{ct}	VPN 客户端与选调器间的时延
T_{tn}	选调器与加解密节点间的时延
T_{ts}	选调器与 DNS 服务器间的时延
T_{proc_v}	VPN 服务器处理时延(传统 VPN 服务器加解密时延)
T_{fp_v}	VPN 服务器转发时延
T_{proc_t}	选调器处理时延(选调时延、分发时延和裁决时延)
T_{pf_t}	选调器转发时延
T_{proc_n}	加解密节点处理时延
T_{fd_n}	加解密节点转发时延
$T_{\text{tradition}}$	M-VPN 相比传统 VPN 额外增加的时延

以 VPN 服务器响应正确的响应为例进行分析,传统 VPN 响应时延 $T_{\text{tradition}}$ 和 M-VPN 时延 $T_{\text{D-VPN}}$ 可分别表示为:

$$T_{\text{tradition}}=2T_{cv}+T_{proc\_v}+T_{fd\_v}+2T_{vs} \tag{10}$$

$$T_{\text{D-VPN}}=2T_{ct}+T_{proc\_t}+T_{fd\_v}+2T_{vs}+T_{proc\_n}+T_{fd\_n}+T_{fd\_t}+2T_{ts} \tag{11}$$

传播时延相比其他时延相对较小,可忽略选调器与加解密节点之间的传播时延 T_{tn}。因此 T_{addition} 可以近似为:

$$T_{\text{addition}}=T_{proc\_t}+T_{proc\_n}+T_{fd\_n}+T_{fd\_t} \tag{12}$$

由于各个加解密节点是并发执行的,加解密节点处理时延 T_{proc_n} 与传统 VPN 服务器处理时延(主要是加解密时延)相差不大;而且选调器和加解密节点接收和转发的时延可以忽略。因此 T_{addition} 可近似为:

$$T_{\text{addition}}\approx T_{proc\_t} \tag{13}$$

为进行定量分析,在原 IPsecVPN 的环境下进行了 10000 次 DNS 时延测试,统计 10000 个时延数据在不同时延区间内的个数,以此计算 IPsecVPN 下 DNS 查询时延的概率分布。得到 IPsecVPN 下 DNS 查询时延的概率分布后,再计算 M-VPN 查询时延,当 *nos* 分别为 3,5,7 时的概率分布如图 9 所示。

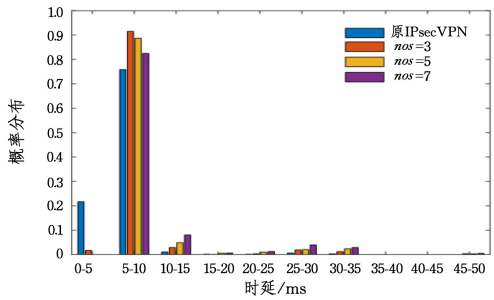


图 9 M-VPN DNS 查询时延概率分布图

Fig. 9 M-VPN DNS query delay probability distribution

接下来求解时延均值,根据概率分布图,求得在不同 *nos* 取值下的时延均值和,如表 7 所列(原 IPsecVPN 时延为 6.751 ms)。

表 7 在不同 nos 取值下的平均时延和额外增加时延

Table 7 Average delay and extra added delay for different nos values

nos/个	平均时延/ms	额外增加时延/ms
3	8.505	1.754
5	9.135	2.404
7	9.903	3.152

相比 IPsecVPN 下的 DNS 查询时延,分别增加了 25.9%,35.6%,46.6%的查询时延。在实际应用和部署中可以让每个加解密节点查询不同的 DNS 服务器以降低查询时延。

结束语 本文针对一种 VPN 的新型流量劫持攻击技术——blind in/on-path 攻击,在分析已有的防御机制的基础上,基于拟态防御技术的思想,提出了 M-VPN 架构。M-VPN 通过冗余以及大数裁决获得了对攻击的感知能力,并根据裁决结果及时动态地调度加解密节点,使攻击者的攻击成本增加,从而降低攻击成功的概率。

但是,M-VPN 也存在一定的不足,主要表现在以下方面:1)M-VPN 的选调器承担了较大的处理量,容易达到性能瓶颈;2)M-VPN 的选调策略成本和开销较大。

下一步的工作计划如下:1)利用 DPDK(Data Plane Development Kit)提供的数据平面开发套件实现收发包处理功能,进一步实现 M-VPN 架构,并优化裁决算法,进一步降低开销;2)拟态理论中选调器是相比传统 VPN 架构新增的设备,负责对执行体的流量进行转发、裁决以及动态调度。但是,目前缺少对其安全性的研究和度量,下一步计划对选调器的安全进行风险评估。

参 考 文 献

[1] HOUSER R, HAO S, LI Z, et al. A Comprehensive Measurement-based Investigation of DNS Hijacking[C]// 2021 40th International Symposium on Reliable Distributed Systems (SRDS). IEEE, 2021: 210-221.

[2] TOLLEY W J, KUJATH B, KHAN M T, et al. Blind In/On-Path Attacks and Applications to VPNs[C]// 30th USENIX Security Symposium (USENIX Security 21). 2021: 3129-3146.

[3] ALEXANDER G, ESPINOZA A M, CRANDALL J R. Detecting TCP/IP Connections via IPID Hash Collisions[J]. Proc. Priv. Enhancing Technol., 2019, 2019(4): 311-328.

[4] KNOCKEL J, CRANDALL J R. Counting packets sent between arbitrary internet hosts[C]// 4th USENIX Workshop on Free and Open Communications on the Internet (FOCI 14). 2014.

[5] FENG X, FU C, LI Q, et al. Off-path TCP exploits of the mixed IPID assignment[C]// Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security. 2020: 1323-1335.

[6] CAO Y, QIAN Z, WANG Z, et al. Off-Path TCP Exploits: Global Rate Limit Considered Dangerous[C]// 25th USENIX Security Sympo-sium (USENIX Security 16). 2016: 209-225.

[7] KOTZIAS P, RAZAGHPANAH A, AMANN J, et al. Coming of age: A lon-gitudinal study of tls deployment[C]// Proceedings of the Internet Measurement Conference 2018. 2018: 415-428.

[8] EGEVANG K, FRANCIS P. The IP network address translator (NAT)[R]. 1994.

[9] BUSHART J, ROSSOW C. Padding ain't enough: Assessing the privacy guarantees of encrypted DNS[C]// 10th USENIX Workshop on Free and Open Communications on the Internet (FOCI 20). 2020.

[10] SIBY S, JUAREZ M, DIAZ C, et al. Encrypted DNS—> Priva-

cy? A traffic analysis perspective[J]. arXiv:1906.09682,2019.

[11] RANJAN A K,KUMAR V,HUSSAIN M. Security analysis of TLS authentication[C]// International Conference on Contemporary Computing and Informatics (IC3I 2014). IEEE, 2014: 1356-1360.

[12] CHENG K,GAO M,GUO R. Analysis and research on HTTPS hi-jacking attacks[C]// 2010 Second International Conference on Net-works Security, Wireless Communications and Trusted Computing. IEEE,2010,2:223-226.

[13] WU J X. Meaning and vision of mimic computing and mimic security defense[J]. Telecommunications Science, 2014,30(7): 2-7.

[14] WU J X. Research on cyber mimic defense[J]. Journal of Cyber Security,2016,1(4):1-10.

[15] IACOVAZZI A,SARDA S,FRASSINELLI D,et al. DropWat: An invisible network flow watermark for data exfiltration traceback[J]. IEEE Transactions on Information Forensics and Security,2017,13(5):1139-1154.

[16] IACOVAZZI A,SARDA S,ELOVICI Y. Inflow:Inverse network flow watermarking for detecting hidden servers[C]// IEEE INFOCOM 2018-IEEE Conference on Computer Communications. IEEE,2018:747-755.

[17] EGEVANG K,FRANCIS P. The IP network address translator (NAT)[R]. 1994.

[18] ZHENPENG W,HONGCHAO H,GUOZHEN C. A DNS Architecture Based on Mimic Security Defense[J]. Acta Electronica Sinica,2017,45(11):2705-2714.

[19] KONG Z,JIANG X Z. DNS spoofing principle and its defense scheme[J]. Computer Engineering,2010,36(3):125-127.

GAO Zhen, born in 1997, postgraduate. His main research interests include network security and mimic defense.

CHEN Fucai, born in 1974, professor. His main research interests include network security and so on.

（责任编辑：喻藜）