基于 D-S 证据理论的 SPIT 检测方案①

林　荣[1,2]，李鸿彬[1,2]，王　宁[2]

[1](中国科学院研究生院，北京 100039)
[2](中国科学院沈阳计算技术研究所，沈阳 110171)

摘　要：针对基于 SIP 协议的 SPIT 攻击（Spam over Internet Telephony,SPIT），提出一种基于统计的 SPIT 检测方法。该方法提取用户多个行为属性和采用 D-S 理论将多个属性进行融合，实现对多种攻击方式的检测。同时，对域进行分类有区别地考虑域发动攻击的可能性和用户的合法性，提高了检测的准确性。实验及分析表明上述方法具有较好的准确性，能够针对 SPIT 进行有效的检测。

关键词：SPIT(Spam over Internet Telephony)；域分类；多属性；融合；D-S 证据理论

SPIT Detection Method Based on D-S Evidence Theory

LIN Rong[1], LI Hong-Bin[1], WANG Ning[2]

[1](Graduate University, Chinese Academy of Sciences, Beijing 100039, China)
[2](Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110171, China)

Abstract: To resolve the security problem of SIP, a detection method based on statistics is proposed. This method fuses multiple attributes of user's behavior into modules based on D-S evidence theory, and by such fusion it achieves the complementation of modules and the detection of multiple attacks. It also differentially considers the possibility of attack and legality of user by classifying domains,which improves accuracy of detection. Experiment and analysis results showed the high accuracy of the proposed method, by which the SPIT can be detected efficiently.

Keywords: SPIT(Spam over Internet Telephony); classified domain; multiple attributes; fusion; D-S evidence theory

SIP(Session Initiation Protocol)[1]协议是 VOIP 应用的主要信令协议，用于建立、修改和终止多媒体会话。SIP 协议在开始设计时并没有考虑安全问题，使它存在安全隐患，其中一种就是 SPIT(Spam over Internet Telephony,SPIT)[2] 。SPIT，定义为大量未经许可的会话初始尝试，这些会话包括语音、视频等，一旦用户接听电话，攻击者便会通过实时媒体流传播信息。针对此问题，本文提出了一种基于 D-S 证据理论的 SPIT 检测方案，通过对域进行分类，融合多个属性实现对多种攻击类型的正确检测。

1　研究现状

目前的 SPIT 检测技术多是借鉴反垃圾邮件的研究成果[2]，包括内容过滤、基于列表的检测、基于许可的通信、信誉系统、地址混淆、限制使用地址、图灵测试、谜语计算、风险支付等。由于垃圾语音具有实时性和直接性的特点，并且以音频作为承载手段，使得这些简单方法在使用上存在一定的局限性。

Dongwook Shin 等人[3]提出了一个基于用户当前呼叫模式的称为”Progressive Multi Gray-Leveling（PMG）”的算法，该算法实现快速有效的检测但当存在标识盗用时该算法无效。Ram Dantu 等人[4]建立一个多阶段闭环反馈的 SPIT 检测系统，但该方案需要用户反馈且需要维护一个全局的信誉系统。Bertrand Mathieu 等人[5]提出一种以网络层实体为基础的检测方法。该方案快速有效但技术的可扩展性和部署是个问题。Hannes T 等人[6]提出了一种基于 SAML 的 SPIT 防护方案，其不足之处是为网络引入了新的实体。Adrian Madhosingh[7]通过对进入的呼叫进行分类进而有所区别对待通信过程的方式来检测 SPIT。但这个方

① 收稿时间:2010-09-09;收到修改稿时间:2010-11-20

案对 SIP 消息进行修改，其通用性还有待研究。Vijay A.Balasubramaniyan 等人[8]提出一种基于呼叫时长建立全局信誉证书和社会网络的机制，但证书的携带可能需要修改 SIP 消息。图灵测试是用来区分呼叫是由人还是机器发动的一种方法，J.Quittek 等人[9]通过检查呼叫的安静时长和回答时长来检测和阻塞 SPIT 攻击。但图灵测试无法检测由人发动的攻击而且它比较耗费时间。

因此，本文在全面、深入地研究了 SPIT 检测和防范技术的基础上，提出了一种基于 D-S 证据理论融合多个属性的 SPIT 检测方法。该方法通过对用户通信模式进行统计分析，在不对用户造成干扰的情况下，简单、高效地进行检测。通过多个模块的检测，涵盖多个攻击类型，包括检测由人和机器发动的攻击、地址盗用等。

2 D-S证据理论

D-S 证据理论是一种模糊推理理论，由 Dempster 于 1967 年提出[10]，并由其学生 Shafer 于 1976 年对其进行了发展[11]。D-S 证据理论可以看作是有限域上对经典概率推理理论的一般化扩展，具有直接表达“不确定”和“不知道”的能力，是带有主观概率的贝叶斯理论的一般化。

D-S 证据理论建立在非空有限域 Θ 上，Θ 称为辨识框架(frame of discernment,简称 FOD)，表示有限个系统状态{ $\theta_1, \theta_2, \cdots, \theta_n$}，而系统状态假设 H_i 为 Θ 的一个子集，即 Θ 的幂集 P(Θ)的一个元素。作为 D-S 证据理论最底层的概念，首先需要定义对某个证据支持一个系统状态的概率函数，称为信度分配函数(basic probability assignment，简称 BPA)。

定义 1. 信度分配函数定义为从Ψ的幂集到[0,1]区间的映射 $m: p(\theta) \to [0,1], m(\varnothing) = 0, \sum_{A \in P(\Theta)} m(A) = 1$。

D-S 证据理论中还提出了对多个证据的组合规则，即 Dempster 规则。

定义 2. Dempster 规则形式化定义如下：

设 m_1 和 m_2 为两个证据的信度分配函数,则对这两个证据的组合得出组合证据的信度分配函数为：$m_1(A) \oplus m_2(A) = K^{-1} \sum_{B \cap C = A} m_1(B) m_2(C)$ when A≠?。其中 K 是归一化因子，$K = \sum_{B \cap C \neq \phi} m_1(B) m_2(C)$。对 n 个证据进行组合的 Dempster 一般化规则为：

$m_{1\ldots n}(A) = K_n^{-1} \sum_{\cap A_i = A} m_1(A_1) m_2(A_2) \ldots m_n(A_n)$ when A ≠ Ø。

其中 $K_n = \sum_{\cap_i A_i \neq \phi} m_1(A_1) m_2(A_2) \ldots m_n(A_n)$。

3 整体构架

我们基于 D-S 证据理论的思想，提出了一种融合多个属性的 SPIT 检测方法。通过对通话模式进行分析和统计，提取正常用户与攻击者区分度较高的几个属性。当进行检测时，根据当前用户的呼叫模式给出攻击的可能性，并将其与域的分类进行结合得出呼叫的信度，之后通过基于 D-S 证据理论的融合算法对信度各自进行融合得到多个综合信度，最后由综合信度得出当前呼叫是否为攻击的判断。

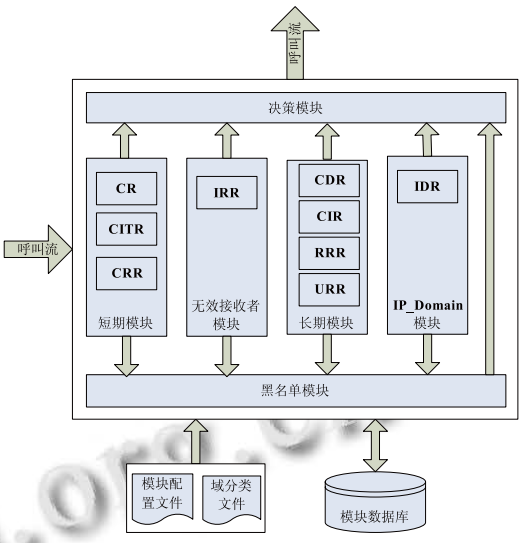

图 1 SPIT 检测系统框架

具体判断流程如图 2 所示，当一个呼叫进入系统时，首先由黑名单模块进行判断：如果呼叫处于阻塞阶段，则直接拒绝。否则，根据当前呼叫的 IP 和 URI 查找有关记录给出评价，并将其交由余下模块进行评判。余下模块各自进行评价并将各自综合信度交由决策模块进行统一处理。决策模块处理的原则是：如果各个模块都判断呼叫是正常，则将此呼叫交给用户；否则只要其中一个模块认为呼叫是攻击就直接拒绝并根据它的不信任值考虑将其加入黑名单；其它则表示综合多个模块无法得出最终结论，则将交由管理员处理。

以下将从属性选取和量化、基于 D-S 证据理论的融合方案以及域分类三个方面对本文的检测方案进行

详细论述。

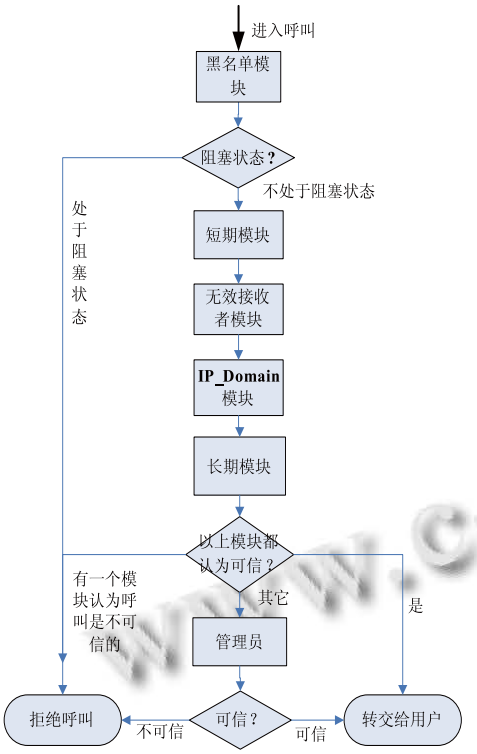

图 2 检测主流程

3.1 属性选取和量化

通过分析正常用户和攻击者的通话模式，提取区分度较高的属性作为检测的标准，以下是该方案选取的属性以及正常用户与攻击者在这些属性上的区别：

1） Call_Rate，Call_Interval_Time_Rate：由于利益驱使，大部分情况下攻击者会以一种贪心的方式进行攻击，此时与正常用户区别最明显的就是呼叫率和呼叫间隔时长。

2） Invalid_Receiver_Rate：攻击者事先收集许多用户地址信息，这些地址信息有些是有效的，有些是无效的，所以通过查看收到的错误信息数目可以判断是攻击者的可能性。

3） Call_Rejection_Rate：一般来说正常用户呼叫拒绝率比较低，而攻击者的呼叫拒绝率比正常用户高。

4） Call_Interation_Rate，Repeated_receiver_Rate，Unknown_receiver_Rate：正常用户一般会拨打和接收呼叫且会对其中一些号码呼叫多次，同时接收方会回电给它。但是，攻击者几乎不会接收到呼叫或很少接收到呼叫，它的呼叫一般不重复同时它经常拨打从未回电的用户。

5）Call_Duration_Rate:正常用户的很多呼叫都会持续较长时间，而攻击者的呼叫时长一般很短。

6）IP_Domain_Rate：通过计算在某种方式上标识呼叫方的多个标识与其它消息的标识的距离来进行检测[12]，该属性的目的在于阻塞试图通过盗用隐藏它们的标识的可疑源。

属性的量化由信度分配函数完成。信度分配函数的作用是将结果转化为适于 D-S 理论的形式。根据 D-S 证据理论，取辨识框架为{N,S}，N 为正常，S 为攻击，有 $N \cap S = \varnothing$，定义信度分配函数 $m: P(\{N,S\}) \to [0,1]$，$m(\varnothing)=0, m(N)+m(S)+m(\{N,S\})=1$。其中 m(N)表示支持正常行为的信度，m(S)表示支持攻击行为的信度，而 $m(\{N,S\})=1-m(N)-m(S)$ 表示不能确定属于正常行为或攻击行为的信度，即未知的信度[12]。在该方案中，信度分配函数的设计原则是：根据呼叫方域的类别给出呼叫的基本信任度、基本不信任度和基本未知度，之后各个属性分别通过评分系统给出是攻击的可能性。然后由融合函数给出最终信任度、不信任度和未知度。

3.2 基于 D-S 证据理论的融合方案

由于 SPIT 存在多种攻击模式，使用单个属性值很难同时将多个攻击行为与正常行为正确区分开，所以如果通过计算单一属性进行检测很难保证检测的准确性。事实上，正常行为很难在几个属性同时呈现较异常的取值。因此，我们考虑基于 D-S 证据理论，通过同时考虑多个属性来提高检测的准确性。本方案的五个模块是：黑名单模块，短期模块，无效接收者模块，IP_Domain 模块和长期模块。黑名单模块由固定部分和动态部分组成。动态部分维护用户被判定是攻击者的次数和用户阻塞的时间这两个表。短期模块包括 Call_Rate,Call_Rejection_Rate,Call_Interval_Time_Rate。长期模块由 Call_Interation_Rate, Call_Duration_Rate, Repeated_receiver_Rate ，Unknown_receiver_Rate 这几个属性构成。无效接收者模块由 Invalid_Receiver_Rate 属性构成。

IP_Domain 模块由 IP_Domain_Rate 属性构成。、

各个模块分别根据各自的信度分配函数和 D-S 证据理论得出综合信度评价，然后根据信任度、不信任度、未知度这 3 个值的最大者给出各自关于当前呼叫是正常、攻击或不能确定的综合评判，并与其它模块

合作得出最后的决定。

3.3 域分类

对域进行分类的思想来自邮件系统，域分类体现了域发动攻击的可能性和用户的合法性。可以根据域的用户名管理政策、认证和授权机制等对域进行分类，例如：域是否对呼叫的用户的身份进行验证；在域内得到用户名的难易程度，申请用户名时需要什么特殊身份或验证措施；域是否有相应措施惩罚发送垃圾信息的用户；域是否有相应的措施来检测外发的呼叫是否是攻击。我们根据域的类别给予域一个基本的可信度，例如一个域是公司内部的，只有员工才享有，由于员工使用这个用户名进行非法操作的可能性极小，所以该域用户的信任度比较高。相反，如果一个域对用户的申请无所限制，且不记录用户的有关信息，那么该域用户的信任度就较小。

4 实验与分析

根据文献[12,13]的仿真方法：由呼叫的开始时间和结束时间、SIP URI、IP 地址、标识符 GOOD/BAD 这几个标识符标识一个呼叫；根据泊松分布产生正常呼叫到达时间；为呼叫随机选择相关的域（IP，SIP URI）；攻击的间隔时间和呼叫时长根据不同场景的需要由不同的分布来产生；进行 SIP 盗用时，从整个攻击者的地址空间中随机选取呼叫的用户名和域名；进行 IP 盗用时，从为每个攻击者预留的 IP 地址中随机选取 IP 地址。

下面仿真四个攻击场景对各个模块的性能以及总体性能进行分析，并通过与 PMG 的比较来查看方案的性能。由于无效接收者模块与实际收集的用户标识有关，呼叫拒绝率与实际用户有关，所以仿真中并不包含这两个模块。情况 1 指没有域分类和黑名单的系统，情况 2 指情况 1 加域分类，情况 3 指情况 2 加上黑名单。Smodule，Lmodule，Imodule，Bmodule，System 分别是指短期模块，长期模块，IP_Domain 模块,黑名单模块和整个系统。FA（False Alarm）是指误报率，LA(Leaking Alarm)是指漏报率。以下数据都是指百分比。

1) 朴素攻击：

指攻击者以贪心的方式进行攻击，此时没有盗用 SIP 或 IP 地址。

表 1 朴素攻击时系统性能

	情况1		情况2		情况3	
	FA	LA	FA	LA	FA	LA
Smodule	0	0.29	0	0.28	0	33.33
Lmodule	0	0.32	0	0.26	0	33.33
Imodule	0	100	0	100	0	100
Bmodule					0	0.12
System	0	0.29	0	0.26	0	0.06

2) 朴素攻击，SIP 盗用和 IP 盗用：

表 2 朴素攻击、SIP 盗用和 IP 盗用时系统性能

	情况1		情况2		情况3	
	FA	LA	FA	LA	FA	LA
Smodule	0	79.56	0	75.94	0	70.41
Lmodule	0	100	0	100	0	100
Imodule	0	2.18	0	1.84	0	23.01
Bmodule					0	5.88
System	0	1.54	0	1.26	0	1.11

3) 软攻击：

指攻击者为了避开以呼叫率为基础进行检测的方法而采取的攻击方式，这时攻击者是以一种懒惰的方式进行攻击。

表 3 软攻击时系统性能

	情况1		情况2		情况3	
	FA	LA	FA	LA	FA	LA
Smodule	0	100	0	99.80	0	90.00
Lmodule	0	1.38	0	1.13	0	40.00
Imodule	0	100	0	100	0	100
Bmodule					0	0.88
System	0	1.38	0	1.01	0	0.38

4) 软攻击，SIP 盗用：

表 4 软攻击、SIP 盗用时系统性能

	情况1		情况2		情况3	
	FA	LA	FA	LA	FA	LA
Smodule	0	99.90	0	99.80	0	97.90
Lmodule	0	100	0	100	0	100
Imodule	0	1.64	0	1.39	0	23.8
Bmodule					0	4.67
System	0	1.58	0	1.20	0	1.14

5) 与 PMG[3]对比：

表 5 与 PMG 性能比较

	场景1		场景2		场景3		场景4	
	FA	LA	FA	LA	FA	LA	FA	LA
System	0	0.06	0	1.11	0	0.38	0	1.14
PMG	0	0.12	0	100	0	0.42	0	100

对数据进行分析可知：场景 1 是典型的攻击方式，短期模块和长期模块这时都发挥主要作用。在场景 2 中由于地址盗用的存在，短期模块和长期模块的准确率并没有场景 1 高，但也发挥了积极的作用，IP_Domain 模块这时发挥主要作用。在场景 3 中主要角色是长期模块。在场景 4 中 IP_Domain 模块发挥主要作用。同时可以看到，各个模块在不同攻击下有效地合作，实现互补，使系统的整体性能得到保证；黑名单的加入，虽然使后续模块准确率有所下降，但由于它过滤了部分流量，到达后续模块的流量大量减少，从而使整体检测速度加快，系统整体性能提高；域分类的引入使准确率进一步提高；而 PMG 在场景 1 和场景 3 下准确率没有本文的检测方法高的原因在于它只是通过计算两个灰度值进行检测，没有考虑攻击行为会在多个方面表现异常。同时，由于 PMG 以 SIP URI 为标识使其对场景 2 和场景 4 的攻击无能为力。

参考文献

1 Rosenberg J, Schulzrinne H, Camarillo G, Johnston A, Peterson J, Sparks R, Handley M, Schooler E. SIP: session initiation protocol. RFC 3261. United States. Internet Engineering Task Force. June 2002.

2 Rosenberg J, Jennings C. The session initiation protocol(SIP) and Spam. RFC 5039. United States. Internet Engineering Task Force. January 2008.

3 Dongwook S, Jinyoung A, Choon S. Progressive multi gray-leveling: a voice spam protection algorithm. IEEE NETWORK, 2006,20(5):18－24.

4 Dantu R, Kilan P. Detecting spam in voip networks. Proc of Steps to Reducing Unwanted Traffic on the Internet Workshop(SRUTI'05). July 2005: 5－5.

5 Mathieu B, Loudier Q, Gourhant Y. SPIT mitigation by a network-level anti-SPIT entity. Proc.of the 3rd Annual VoIP SecurityWorkshop (VSW'06).ACM Press, June 2006.

6 Tschofenig H, Falk R, Peterson J, Hodges J, Sicker D, Polk J, Siemens AG. Using SAML to protect the session initiation protocol (SIP). IEEE NETWORK, 2006,20(5):14－17.

7 Madhosingh B.The design of adifferentiated session initiation protocol to control VoIP spam[Master. Thesis]. Florida: Florida State University, 2006.

8 Balasubramaniyan V, Ahamad M, Park H. Callrank: combating SPIT using call duration,social networks and global reputation. Fourth Conference on Email and Anti-Spam(CEAS2007).CA, 2007.

9 Quittek J, Niccolini S, Tartarelli S, Stiemerling M, Brunner M, Ewald T. Detecting SPIT calls by checking human communication patterns. Proc. of the IEEE International Conference on Communations (ICC). IEEE Press, Glasgow, Scotland, June 2007.

10 Dempster A. Upper and lower probabilities induced by a multivalued mapping.Ann Math Statist, 1967,38(2): 325－339.

11 Shafer G. A mathematical theory of evidence. Princeton: Princeton university, 1976:19－63.

12 Menna F, Cigno RL, NiccoliniI S. Simulation and optimization of SPITdetection framework. IEEE GLOBECOM'07. Washington DC, IEEE, 2007.

13 Menna F. Spam over internet telepnony prevention systems:design and enhancements with a simulative approach. Anno Accademico, 2006.