基于信息论的入侵检测最佳响应方案

田有亮[1,2,3]，吴雨龙[1,2]，李秋贤[1,2]

（1. 贵州大学计算机科学与技术学院，贵州 贵阳 550025；2. 贵州省公共大数据重点实验室，贵州 贵阳 550025；
3. 贵州大学密码学与数据安全研究所，贵州 贵阳 550025）

摘 要：入侵检测系统经常不可避免地出现误警、漏警错误而导致系统的重大安全隐患，然而当前未能找到一种行之有效的解决方案。针对该问题，提出一种基于信息论的入侵检测最佳响应模型。首先，将入侵检测过程中的入侵者和入侵检测系统抽象成随机变量，并根据对抗结果构建了入侵者和入侵检测系统的攻防模型。其次，根据攻防模型设计入侵检测系统的防守信道，将入侵检测系统的正确检测转换成防守信道成功传输 1 bit 信息问题。最后，通过分析防守信道的信道容量来衡量系统防守能力，其防守信道的最大互信息量就是入侵检测系统的防守极限能力，其对应的策略分布就是系统的防守能力最佳响应策略。实验结果表明，所提方案能够有效地降低系统误警和漏警所造成损失。

关键词：入侵检测系统；平均互信息量；信道容量；检测率；响应方案

中图分类号：TP309

文献标识码：A

doi: 10.11959/j.issn.1000−436x.2020111

Optimum response scheme of intrusion detection based on information theory

TIAN Youliang[1,2,3], WU Yulong[1,2], LI Qiuxian[1,2]

1. College of Computer Science and Technology, Guizhou University, Guiyang 550025, China
2. Guizhou Provincial Key Laboratory of Public Big Data, Guiyang 550025, China
3. Institute of Cryptography & Data Security, Guizhou University, Guiyang 550025, China

Abstract: Intrusion detection system (IDS) often inevitably presents major security risks caused by FPs and FNs. However, at present, an effective solution has not been found. In order to solve this problem, an optimal response model of intrusion detection based on information theory was proposed. Firstly, the intruder and IDS in the process of intrusion detection were abstracted into random variables, and the attack and defense model of intruder and IDS was constructed according to the results of the confrontation. Secondly, the defense channel of IDS was designed according to the attack and defense model, then the correct detection of IDS was transformed into the problem of successful transmission of 1 bit information in defensive channel. Finally, the defensive capability of the system was measured by analyzing the channel capacity of the defensive channel, the maximum mutual information of the defensive channel was the defensive limit capability of the IDS, and the corresponding strategy distribution was the optimal response strategy of the defensive capability of the system. The experimental results show that the scheme can effectively reduce the loss caused by FPs and FNs.

Key words: intrusion detection system, average mutual information, channel capacity, detection rate, response scheme

---

收稿日期：2020–03–24；**修回日期**：2020–04–14

基金项目：国家自然科学基金资助项目（No.U1836205, No.61662009, No.61772008）；贵州省教育厅科技拔尖人才支持基金资助项目（No.[2016]060）；贵州省科技重大专项计划基金资助项目（No.20183001）；贵州省科技计划基金资助项目（No.[2017]5788）；教育部−中国移动科研基金研发基金资助项目（No.MCM20170401）；贵州省联合基金资助项目（No.LKT201216, No.LH20147476）

Foundation Items: The National Natural Science Foundation of China (No.U1836205, No.61662009, No.61772008), The Guizhou Provincial Department of Education Science and Technology Top Talent Support Project (No.[2016]060), The Science and Technology Major Support Program of Guizhou Province (No.20183001), The Guizhou Provincial Science and Technology Plan Project (No.[2017]5788), The Ministry of Education-China Mobile Research Fund Project (No.MCM20170401), The Joint Science and Technology Foundation of Guizhou Province (No.LKT201216, No.LH20147476)

1 引言

随着互联网的飞速发展，网络安全面临的威胁也呈现爆发式增长。面对更多样、复杂且数量更庞大的安全威胁，传统的安全预防手段（如防火墙、加密、身份认证等技术）已经不足以保证网络系统的安全性[1]。为弥补这些预防手段的不足，入侵检测系统（IDS, intrusion detection system）被引入当前的网路安全技术体系结构中，并且发挥着越来越重要的作用，但是 IDS 的检测结果并不是完全准确的。受当前研究水平和应用环境等因素的制约，IDS 总是不可避免地出现两类错误：误警错误、漏警错误。误警错误将正常用户检测为入侵者，可能会造成 IDS 产生大量无用且超过管理员处理能力的错误警报，从而造成资源浪费；漏警错误将入侵者检测为正常用户，使管理员无法预防真正的攻击，从而使网络安全受到威胁。

误警错误和漏警错误严重影响了 IDS 的安全性和准确性。虽然已经存在相关研究能够在一定程度上降低 IDS 的误警率和漏警率，但是目前还不存在一种方法能够完全消除 IDS 的误警错误和漏警错误，因此有必要考虑在 IDS 不可避免发生错误的情况下，如何对警报进行响应以降低其对系统安全性带来的影响。目前常用的做法是系统管理员对发生警报的事件/用户及未发生警报的事件/用户按照一定概率进行检查[2]。关于如何设置检查的执行概率，Rhee 等[3]提出借助一种关于误警率和检测率的双向图，利用 ROC（receiver operating characteristic）曲线来权衡调整系统管理员的 2 种检查的执行概率。Cavusoglu 等[4]提出，通过刻画系统的开销，以使系统开销最小为目的来设置 2 种检查的执行概率。从整体上看，当前方案主要存在以下问题：1) 盲目地提高 2 种检测的执行概率，会增大系统因误警而产生的开销；2) 方案中存在难以实现的前提假设。

针对现有方案中前提假设难以实现，系统开销增大等不足，本文基于信息论提出了一种入侵检测最佳响应方案。本文的主要贡献如下。1) 本文将入侵检测过程中入侵者和 IDS 抽象为随机变量，考虑双方采取不同行动的概率，进一步确定入侵者和 IDS 的随机变量的概率分布，并结合双方不同行动下的对抗结果，建立了攻防模型。2) 根据攻防模型和盲自评思想设计结果随机变量，并以 IDS 的随机变量作为输入，结果随机变量作为输出，建立了 IDS 的防守信道，将 IDS 的正确检测问题转换为防守信道成功传输 1 bit 信息问题。3) 结合信息论中通信信道的特性及信道容量的定义，通过分析防守信道的信道容量来衡量系统的防守能力。当防守信道的互信息量等于信道容量时，系统的防守能力达到极限，此时的策略分布就是系统的防守能力最优响应策略。4) 在分析系统开销的约束条件下，提出了降低两类错误造成的安全隐患和损失的入侵检测最佳响应方案。

2 研究现状

传统的安全措施主要采用预防技术，如防火墙、加密、身份认证[5]等。然而，这些传统的预防技术都存在一定限制。例如，防火墙一般对内部发起连接的攻击（如木马等）束手无策；身份认证则可能因为用户的密码强度较弱而被攻破。大量研究及已经发现的安全问题表明，单独使用预防技术无法保证系统的安全[6]。为弥补传统安全措施的不足，Anderson[7]在 1980 年提出了入侵检测。他把入侵行为分为内部攻击、外部入侵、越权操作/误操作，并提出了一种追踪审计的方法来监视检测入侵行为。自 Denning[8]提出入侵检测模型以来，大量的研究致力于寻找一种准确且有效的入侵检测方法，希望性能良好的 IDS 具有较高检测率和较低误警率[9]。

入侵检测根据采用技术的不同，分为 2 种类型[10]：特征检测和异常检测。特征检测[11]将已知的攻击特征存入数据库中，当一个事件行为与数据库中的一个或多个攻击特征匹配时发出警报。一般而言，特征检测的准确率是比较高的，但是新的攻击手段层出不穷，对于未知的攻击，特征检测会出现大量的漏警错误。异常检测[12]利用系统和网络的特性对正常的网络行为进行建模，再运用这些模型监督网络，并将任何与行为模型存在偏差的行为视为攻击。但是基于异常检测的 IDS 在部署前，需要大量的数据进行学习和训练，而获取适用的数据集是十分困难的，系统进行学习和训练的计算代价也是十分高昂的。同时，如果基于异常检测的 IDS 没有获得足够的数据进行学习和训练，则会在运行中产生大量的误警错误。

IDS 采用特殊的分析技术来检测攻击行为，识别其攻击源头并对系统管理员发出警报[13]。IDS 对

于网络嗅探、拒绝服务攻击（DoS, denial of service）、注入攻击等传统的攻击手段都具有良好的效果，但是高级持续性威胁（APT, advanced persistent threat）则效果不佳。由于 APT 具有隐蔽性强、潜伏期长、持续性长等特点，使其难以检测和抵御，从而对政府和企业的系统及数据安全造成严重威胁[14]。目前，针对 APT 的主要防御解决方案大致分为四类[15-16]。1) 恶意代码检测：根据行为异常特征的边界，以沙盒模式来运行程序，根据程序行为特征来判断其合法性。2) 主机应用保护：设置白名单，只允许白名单上的应用运行。3) 网络入侵检测：通过建立 IDS，检测网络流量，检测 APT 命令和控制通道的特征，及时发现和预警 APT 攻击。4) 大数据分析检测：通过对 APT 攻击产生的海量数据进行收集、处理、监控，及时准确地发现 APT 攻击。为应对 APT 带来的安全威胁，许多学者和相关安全机构的研究围绕以上 4 个方面展开。Rubio 等[17]指出，由于 APT 暴露的攻击特征是多样的，必须在不同的级别组合多个安全解决方案。从这个意义上说，IDS 是安全防御的第一道防线。例如，Luh 等[18]提出一种高级入侵检测及解释系统（AIDIS, advanced intrusion detection and interpretation system），通过一组基线过程图来计算核心事件的偏离程度，从而识别用户行为的异常，并通过一组图匹配偏差模板及随机森林和线性支持向量机的方法来智能分辨该异常是否是潜在的 APT 异常。Moon 等[19]提出了一种基于决策树的 IDS（DTB-IDS, decision tree based on IDS），该系统利用对行为信息的分析来检测入侵系统后不断变化的 APT 攻击。Vries 等[20]提出了一种高级 IDS，该系统使用来自智能数据分析领域的特征和异常检测方法，通过分析多个网络位置的网络流量和客户端数据来检测 APT 攻击。此外，国内的互联网企业及网络安全公司通过对 IDS 进行优化扩充或者对 IDS/IPS 等各种安全技术手段进行集成，并推出了多种 APT 检测系统。

许多研究工作通过使用统计方法、控制图表、机器学习、时间序列等提供多种解决方案来处理误警错误。例如，Pietraszek[21]提出了一种自适应学习警报分类器（ALAC, adaptive learner for alert classification），根据自适应的学习规则来建立和完善用于识别正确和错误警报的规则。但是，该方法需要一个庞大的训练数据集，并且训练开销十分高昂。而对于 IDS 的漏警问题，相关研究却比较少。由于识别 IDS 漏警是十分困难的，为解决这一问题，Hachmi 等[22]提出了一种多目标优化方法，用于从多个 IDS 生成的警报中识别漏警和误警错误，但是混合使用不同的 IDS 可能会造成一些冲突。

由前述可见，尽管目前已经有许多研究试图减少 IDS 的漏警错误和误警错误，并且在不同方面取得了一定成效，但是这两类错误仍然无法完全消除。在这种情况下，有必要考虑如何对 IDS 的警报进行响应，从而降低两类错误对 IDS 准确性和系统安全性的影响。Zonouz 等[23]提出了一种利用博弈论和马尔可夫决策树来寻找长期收益最优策略的方式来对入侵进行响应。Cuppens 等[24]则提出一种将警报映射到攻击上下文的方法，用来制定响应网络威胁的策略。吴姚睿等[25]提出了一种通过关系图建立攻击群模型的方法来最大程度地降低响应成本。总体上看，目前国内外入侵检测响应方案的研究容易忽略当前 IDS 的高漏警和误警问题，从而影响响应策略的准确制定。另外，相关研究缺少定量的数学模型，从而影响方案的应用。针对这一问题，本文参考了 Tian 等[5,26]利用信息论在委托计算方面对于双方攻防能力进行刻画的方法，以及在公钥密码方面对密码抗攻击能力进行衡量的理论，在权衡系统自身开销的前提下，结合信息论提出了一种基于信息论的入侵检测最佳响应方案。

3　基础知识

本文在建模及分析过程中运用到的信息论及相关背景知识如下。

定义 1　盲对抗[27]。参与攻防对抗的双方，在每次攻防回合结束后，双方都会对自身在本回合的结果产生一个评价，由于这个评价是只有自己知道的，因此双方都不必作假。以随机变量 X、Y 分别代表攻防双方的自评结果。当攻击方认为他在一次攻防对抗中取得成功时，记为 $X=1$；当认为失败时，记为 $X=0$。同理，当防守方认为他在一次攻防对抗中取得成功的，记为 $Y=1$；当认为失败时，记为 $Y=0$。

定义 2　信道。信道是通信系统的组成部分，它的输出信号依赖于输入信号，其依赖关系由转移概率矩阵 $\boldsymbol{P}(y|x)$ 决定。

定义 3　互信息量。互信息量可以被视为一个随机变量因已知另一个随机变量而减少的不确定性。$p(x)$、$p(y)$分别表示随机变量 X、Y 的概率分布函数，$p(x,y)$表示随机变量 X 和 Y 的联合概率分布函数，则随机变量 X

和 Y 的互信息量 $I(X,Y)=\sum_{x\in X}\sum_{y\in Y}p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$。

定义 4 信道容量。信道容量反映了信道所能传输的最大信息量。对于输入信号为 X、输出信号为 Y 的信道，其信道容量为 $C=\max_{p(x)} I(X;Y)$。

引理 1 红客守卫能力极限定理[31]。设由随机变量(Y,Z)组成的防守信道 F 的信道容量是 C，如果防守方想真正地抵御 k 次入侵，那么一定有某种技巧，使其能够在 $\frac{k}{C}$ 次防守中，以任意接近于 1 的概率达到目的；如果经过 n 次防守，防守方获得了 S 次真正成功的防守，那么一定有 $S\leqslant nC$。

4 入侵检测最佳响应方案

入侵检测最佳响应方案是结合信息论思想，参考盲对抗方法并依据各种事件发生的概率因素，将攻防双方的对抗结果定义为按照一定概率分布的随机变量，并通过攻防随机变量来建立防守信道。然后，利用防守信道的信道容量来刻画系统当前的防守能力，在权衡系统开销的前提下，调整对 IDS 警报执行两类检测错误的概率来调整防守信道的信道容量，从而寻找最优的响应方案。

入侵检测最佳响应模型包括形式化定义、攻防模型构建、防守信道构建、防守期望开销和响应策略分析五部分。

4.1 形式化定义

本文所提的基于信息论的入侵检测最佳响应方案是一个八元组$(\rho_1,\rho_2,X,Y,F,C,\mathrm{cost}_1,\mathrm{cost}_2)$，具体含义如下。

1) ρ_1 表示系统对报警事件及其用户采取审查的比例，$\rho_1\in[0,1]$。

2) ρ_2 表示系统对未报警事件及其用户采取抽查措施的比例，$\rho_2\in[0,1]$。

3) 随机变量 X 表示攻击者在一次攻防回合对自身的评价。$X=1$ 表示攻击者自评为成功；$X=0$ 表示攻击者自评为失败。

4) 随机变量 Y 表示防守方在一次攻防回合对自身的评价。$Y=1$ 表示防守方自评为成功；$Y=0$ 表示防守方自评为失败。

5) 防守信道 F 表示根据双方自评结果建立的信道模型。

6) 防守信道容量 C 为防守信道的信道容量，用于刻画系统的防守能力。

7) cost_1 为系统管理员对一次事件执行检查的期望开销。

8) cost_2 为系统管理员对一次事件不作为的期望开销。

4.2 攻防模型构建

在一次攻防对抗过程中，对于攻击者而言有 2 种行动策略，即发动攻击（记为 A_1）和不发动攻击（记为 A_2）。对于防守方而言也有 2 种行动策略，即执行检查（记为 D_1）和不执行检查（记为 D_2）。将攻防双方的行动策略进行两两组合，可以得到攻防双方的所有策略组合，如矩阵 $\boldsymbol{S}$ 所示。

$$\boldsymbol{S}=\begin{bmatrix}(A_1,D_1) & (A_1,D_2)\\ (A_2,D_1) & (A_2,D_2)\end{bmatrix}$$

防守方在一次对抗回合中是否执行检查需要参考 IDS 的检测结果，由于 IDS 不是完全可靠的，因此防守方对于 IDS 发出警报和不发出警报的回合都会按一定概率执行检查。根据攻防双方的所有行动策略和 IDS 是否报警，可以组成 8 种不同的事件，如表 1 所示。

当 IDS 在其应用环境中运行足够长的时间后，其保护的系统遭受入侵的平均概率是可以统计的。令 P_C 表示 IDS 正确检测出一次入侵的概率，那么 $1-P_C$ 代表 IDS 的漏警率。P_F 表示 IDS 将一次正常事件检测为入侵的概率，ϕ 表示系统在一定时间内遭受攻击的平均概率，得到以下事件的发生概率，具体如下。

$$P_1=P(\text{攻击者入侵},\text{IDS报警})=\phi P_{\mathrm{C}}$$

$$P_2=P(\text{攻击者入侵},\text{IDS未报警})=\phi(1-P_{\mathrm{C}})$$

$$P_3=P(\text{攻击者未入侵},\text{IDS报警})=(1-\phi)P_F$$

$$P_4=P(\text{攻击者未入侵},\text{IDS未报警})=(1-\phi)(1-P_F)$$

不同的事件发生，将导致 IDS 和入侵者之间的攻防对抗产生不同的结果。结合定义 1，从信息论的角度看，IDS 和入侵者依据其盲自评可以被视为按照一定概率分布的随机变量。根据前文所述，系统管理员会按照概率 ρ_1 手动地检查 IDS 发出警报的日志文件，并跟踪审查相关用户，从而确认或者排除入侵。与此同时，为避免遗漏，系统管理员会按照概率 ρ_2 检查未报警的日志文件和相关用户。在本文假设系统管理员执行检查时，能够正确地识别并处理事件。那么系统管理员对一个事件是否执行

表 1　　**针对不同事件攻防双方自评取值及防守开销**

事件描述	执行检查	事件概率	X	Y	防守期望开销
攻击者入侵，IDS 报警	是	$P_1\rho_1$	0	1	$P_1\rho_1 c+(1-\varphi)dP_1\rho_1$
	否	$P_1(1-\rho_1)$	1	0	$P_1(1-\rho_1)d$
攻击者入侵，IDS 未报警	是	$P_2\rho_2$	0	1	$P_2\rho_2 c+(1-\varphi)dP_2\rho_2$
	否	$P_2(1-\rho_2)$	1	0	$P_2(1-\rho_2)d$
攻击者未入侵，IDS 报警	是	$P_3\rho_1$	0	0	$P_3\rho_1 c$
	否	$P_3(1-\rho_1)$	0	1	0
攻击者未入侵，IDS 未报警	是	$P_4\rho_2$	0	0	$P_4\rho_2 c$
	否	$P_4(1-\rho_2)$	0	1	0

检查响应，将直接影响攻防双方对当次对抗的自评成败。其中在攻击者未入侵的情况下，系统管理员检查了相应事件，则代表其误警率对维护系统安全产生了不良影响，从而防守方自评为失败，即 $Y=0$。对于攻击者而言，其目的一般是获取额外资源，在未发动攻击的情况下，则代表其放弃了本次攻防对抗，则攻击者自评为 $X=0$。针对不同事件，攻防自评随机变量 X、Y 的取值及系统的防守开销如表 1 所示。其中，c 表示系统管理员对一次事件进行检查的开销，d 表示系统管理员未对一次攻击执行响应的损失，φ 表示防守方对一次攻击执行了响应后能减少损失的比例。显然存在 $c \leqslant \varphi d$，否则系统不会为任何事件花费代价请系统管理员进行检查。

根据表 1 可得，攻防自评随机变量 X、Y 的概率分布如下。

$$P(Y=1)=P_1\rho_1+P_2\rho_2+P_3(1-\rho_1)+P_4(1-\rho_2)$$

$$P(Y=0)=P_1(1-\rho_1)+P_2(1-\rho_2)+P_3\rho_1+P_4\rho_2$$

$$P(X=1)=P_1(1-\rho_1)+P_2(1-\rho_2)$$

$$P(X=0)=P_1\rho_1+P_2\rho_2+P_3+P_4$$

4.3　防守信道构建

由于防守方的目的是抵御攻击方的入侵以保证系统安全，那么是否真正防守成功，不能由防守方的盲自评决定，而应该由对手的盲自评决定。也就是说，当攻击者盲自评为失败时，防守方才真正的取得了一次防守成功。此时本文定义第三个随机变量 $Z=(X+Y)\bmod 2$，当 Z=0 时，表示攻防双方同时认为自己在本次对抗中取得了成功或者失败的事件，即双方自评相同的事件。当 Z=1 时，表示当攻击方认为自己在本次对抗中成功且防守方认为自己失败的事件，与攻击方认为自己在本次对抗中失败且防守方认为自己成功的事件的并集，即双方自评不同的事件。其概率分布为

$$P(Z=1)=P(X=1,Y=0)+P(X=0,Y=1)=$$
$$P(X=1)P(Y=0)+P(X=0)P(Y=1)$$
$$P(Z=0)=P(X=1,Y=1)+P(X=0,Y=0)=$$
$$P(X=0)P(Y=0)+P(X=1)P(Y=1)$$

在信息论当中，任意 2 个随机变量都可以组成一个通信信道。据此，以随机变量 Y 作为输入、随机变量 Z 作为输出，构成一个防守信道 F。那么防守方一次真正防守成功的定义为

{防守方某次真正成功}={防守本次盲自评成功∩攻击方盲自评失败}∪{防守方盲自评失败∩攻击方盲自评失败}={Y=1, X=0}∪{Y=0, X=0}= {Y=1, Z=1}∪{Y=0, Z=0}={1 bit 信息成功地从防守信道的发送端传输到输出端}

当防守信道建立起来后，防守方和攻击方在入侵中的攻防对抗问题就转换为在防守信道上传输信息的问题。当信道传输信息的能力越强时，防守方的防守能力就越强（防守信道的信道容量就代表着防守方的防守能力）。与此类似，如果以随机变量 X 作为输入，随机变量 Z 作为输出，就构成了一个攻击信道 G。那么对于攻击方而言，一次成功的入侵同样可以被定义为 1 bit 信息从攻击信道 G 的发送端传输到输出端，而攻击信道的信道容量则代表着攻击方的攻击能力。由于本文的主要工作是研究并寻找防守方的响应策略，因此不对攻击信道做进一步展开讨论。

4.4　防守期望开销

当 IDS 在其工作场景运行了足够长的时间后，其潜在的攻防环境是已知的。也就是说，系统遭到

攻击的可能性、管理员执行检查的开销、遭到攻击所承受的损失等都是可以预估的。虽然这些因素的确切值不太容易获取，但是仍然可以从系统以往的运行日志文件和系统管理员的经验中得到。系统管理员对不同事件采用不同响应时的防守期望开销如表 1 所示。根据表 1，可以得到对一次事件执行检查的期望开销 cost_1 和对一次事件不作为的期望开销 cost_2，计算方式如下。

$$\begin{aligned}\text{cost}_1 &= P_1\rho_1 c + (1-\varphi)dP_1\rho_1 + P_2\rho_2 c + \\ &\quad (1-\varphi)dP_2\rho_2 + P_3\rho_1 c + P_4\rho_2 c\end{aligned}$$

$$\text{cost}_2 = P_1(1-\rho_1)d + P_2(1-\rho_2)d$$

显然，对一次事件执行检查的期望开销不应大于对该事件不作为的期望开销，否则系统管理员在任何情况下都不会执行检查。即

$$\begin{aligned}&\text{cost}_1 \leqslant \text{cost}_2 \\ &P_1\rho_1 c + P_2\rho_2 c + 2P_1\rho_1 d + 2P_2\rho_2 d + P_3\rho_1 c + \\ &P_4\rho_2 c - P_1\rho_1\varphi d - P_2\rho_2\varphi d - P_1 d - P_2 d \leqslant 0 \qquad (1)\end{aligned}$$

式(1)中只有 ρ_1和ρ_2 是未知的，因此可以解出 ρ_1和ρ_2 的取值范围。

4.5 响应策略分析

4.3 节已经建立了防守信道，根据第 3 节的定义 4 和引理 1，只要得到防守信道的信道容量，那么防守方的防守能力就确定了。下面将继续计算防守信道 F 的信道容量。首先，随机变量 (Y,Z) 的联合概率分布为

$$\begin{aligned}P(Y=0,Z=0) &= P(Y=0,X=0) \\ P(Y=0,Z=1) &= P(Y=0,X=1) \\ P(Y=1,Z=0) &= P(Y=1,X=1) \\ P(Y=1,Z=1) &= P(Y=1,X=0)\end{aligned}$$

防守信道 F 由随机变量 Y 和 Z 构成，它的二阶转移概率为

$$A=[A(y,z)]=[\Pr(z\mid y)],\ y,z=0,1$$

$$A(0,0)=\Pr(Z=0\mid Y=0)=\frac{P(X=0,Y=0)}{P(Y=0)}=P(X=0)$$

$$A(0,1)=\Pr(Z=1\mid Y=0)=\frac{P(X=1,Y=0)}{P(Y=0)}=P(X=1)$$

$$A(1,0)=\Pr(Z=0\mid Y=1)=\frac{P(X=1,Y=1)}{P(Y=1)}=P(X=1)$$

$$A(1,1)=\Pr(Z=1\mid Y=1)=\frac{P(X=0,Y=1)}{P(Y=1)}=P(X=0)$$

那么，随机变量 Y 和 Z 的互信息量为

$$\begin{aligned}I(Y,Z) &= \sum_y\sum_z p(y,z)\text{lb}\frac{p(y,z)}{p(y)p(z)} = P(X=0,Y=0)\cdot \\ &\text{lb}\frac{P(X=0)}{P(Z=0)} + P(X=1,Y=0)\text{lb}\frac{P(X=1)}{P(Z=1)} + \\ &P(X=1,Y=1)\text{lb}\frac{P(X=1)}{P(Z=0)} + P(X=0,Y=1)\text{lb}\frac{P(X=0)}{P(Z=1)}\end{aligned} \qquad (2)$$

根据前文所述，决定随机变量 X 和 Y 概率分布的参数中，大部分是稳定的，可以将其视为常量，只有 2 个参数（ρ_1和 ρ_2）是变化的。于是，以 Y 为输入、Z 为输出的防守信道 F 的信道容量为 $\max\limits_{0\leqslant\rho_1,\rho_2\leqslant 1}[I(Y,Z)]$。也就是说，防守信道的信道容量 C 是参数 ρ_1和 ρ_2 的函数。结合引理 1，当 ρ_1和 ρ_2 取值越大时，防守信道的互信息量越大，防守能力越强，当 $\rho_1=\rho_2=1$ 时，防守信道的互信息量和防守能力达到最大值。此时防守方对每一个事件都执行了检查，在这种情况下攻击方的所有入侵都会被发现，所以系统的安全性是最高的。然而这种响应策略显然是不可取的，首先防守方的能力显然不可能是无限的，使其无法对网络系统中的海量事件进行逐一检查；其次防守方即使能够对事件执行逐条检查，也会使系统的防御开销变得非常大，这会严重影响网络的可用性和生命周期。因此，对于 ρ_1和 ρ_2 的设置应该限定在系统可以接受的防守开销范围内。由 4.4 节可知，对于任何一个可取的响应策略 ρ_1和 ρ_2 而言，系统在当前策略下的系统开销不应大于系统不执行任何操作的开销。当系统选择了某一个响应策略，使系统执行检查的开销与系统不执行任何操作的开销相等时，系统不会因为执行检查而获得额外收益，但是由于系统采取了很高的比例对未报警事件和报警事件进行检查，因此系统的安全性是在系统开销约束范围内的最佳点；当系统选择了某一个响应策略，使系统执行检查的开销最小时，系统因执行检查而获得最大收益，但是由于此时系统没有花费代价去尽可能地执行检查，系统的安全性较弱。因此在实际对抗当中，应该在系统开销允许的范围内，根据系统对于安全性和系统收益的重要程度来调整对报警事件及其用户的审查比例 ρ_1，对未报警事件及其用户采取抽查措施的比例 ρ_2，从而调整信道容量 C，使自身的防守能力在当前情况下达到最大。

5　模型理论分析

通过本文方案，可以利用信道通信来形式化攻防过程，也可以利用信道容量来刻画攻防双方的对抗能力。在此基础上，得到如下理论结果。

定理 1　在入侵检测过程中，防守方进行一次成功的防守，等价于在防守信道 F 中 1 bit 信息成功地从发送端传输到接收端。

证明　由于防守方是否防守成功，不能由其主观的盲自评决定，因此防守方是否取得一次防守成功，应该由攻击方的盲自评决定。当攻击方自评失败，即 $X=0$ 时，防守方才取得一次防守成功。那么在入侵检测过程中防守方进行一次成功的防守的情况为

{防守方取得一次防守成功}={防守本次盲自评成功∩攻击方盲自评失败}∪{防守方盲自评失败∩攻击方盲自评失败}={Y=1, X=0}∪{Y=0, X=0}={Y=1, Z=1}∪{Y=0, Z=0}={1 bit 信息成功地从防守信道 F 发送端传输到输出端}

若 1 bit 信息成功地从防守信道 F 的发送端传输到输出端，则信道的两端取值相等，即 $Y=Z$。满足此条件的情况只有 2 种，即 $Y=Z=1$，$Y=Z=0$。由于 $Z=(X+Y)\bmod 2$，在以上条件下，可得 $X=0$，此时攻击者自评失败，相当于防守方取得一次防守成功。

令随机变量 X 为输入，随机变量 Z 为输出，构建信道容量为 R 的攻击信道 G，那么适用于防守方的定义，定理同样适用于攻击方。证毕。

定理 2　入侵检测中，当攻击信道容量大于防守信道容量，即 $R>C$ 时，入侵者的攻击能力大于 IDS 的防守能力，此时的对抗中入侵者占优势，其更容易成功入侵系统；当攻击信道容量小于防守信道容量，即 $R<C$ 时，入侵者的攻击能力小于 IDS 的防守能力，此时的对抗中 IDS 占优势，其更容易成功抵抗入侵。

证明　根据定理 1，在防守信道（攻击信道）上每成功传输 1 bit 信息，相当于 IDS（入侵者）取得了一次防守（攻击）成功；而根据定义 2、定义 4、引理 1 及香农定理可知，信道容量是在信息率一定时，信道所能传输的最大比特数。那么从整体上来看，防守信道（攻击信道）的信道容量就是其防守能力（攻击能力）的最大值，显而易见，在对抗过程中能力强的那一方更容易取得成功。证毕。

6　仿真实验

6.1　实验参数

由于本文更加关注于对 IDS 警报的响应策略，因此并没有对 IDS 进行建模而将其视作一个黑盒。在实验中，借鉴文献[28]对 IDS(Snort)进行配置，并使用 UNSW-NB15 数据集[29]作为输入。其中，UNSW-NB15 是澳大利亚网络安全中心用于入侵检测评估的数据集，该数据集包含了现代正常的数据流量和 8 种不同的攻击类型的数据流量。UNSW-NB15 数据集包含了 2 540 044 条数据样本，共 100 GB。由于原始数据集过于庞大且存在许多冗余[30]，本文采用部分数据集 UNSW-NB15 testing set，它是原始数据集经过处理后不包含任何冗余，专门用来进行入侵检测测试的数据集。表 2 总结了该数据集的分布情况。

表 2　UNSW-NB15 testing set 数据分布情况

类别	统计数/条
正常事件	37 000
Fuzzers 攻击	677
分析攻击	583
拒绝服务攻击	11 132
Exploits 漏洞攻击	6 062
泛型攻击	18 871
侦察攻击	3 496
Shellcode 攻击	378
蠕虫攻击	44
合计	82 332

根据本文的实验配置，IDS 检测率保持在 51.5%左右，误检率保持在 30.7%左右。可以看到，Snort 在当前配置下对于该数据集的表现不是很好，但这并不违背本文不改进 IDS 精度而是调整响应策略来提高系统安全性并平衡开销的初衷。为方便分析和研究，本文将系统未对一次攻击进行响应的损失设置为 50，对一次攻击进行了响应则可以挽回 80%的损失，管理员执行一次响应需要花费的开销为 20。相关参数如表 3 所示。

表 3　　相关参数

参数描述	参数值
IDS 的正确检测率 P_C	51.5%
IDS 的漏检率 $1-P_C$	48.5%
IDS 的误警报率 P_F	30.7%
系统遭受攻击的平均概率 ϕ	55.1%
管理员对一次事件进行响应的开销 c	20
防守方未对一次攻击执行响应的损失 d	50
响应攻击所减少损失的比例 φ	80%

6.2 结果分析

如前文所述，对事件执行响应的期望开销不应大于不对事件执行响应的期望开销，将表 3 中的响应参数代入式(1)，得到 ρ_1和ρ_2取值范围如图 1 所示，其中 z 轴表示对事件执行检查响应的期望开销减去对事件不作为的期望开销的差值。在最极端的情况下，系统采用很大的概率令系统管理员进行检查，这将导致系统用于检查事件的开销大大增加，使此时对事件执行检查的期望开销和对事件不作为的期望开销相等。这种极端情况的 ρ_1和ρ_2取值如图 2 所示。在这种情况下，系统并不会因对事件执行检查而获得比对事件不作为更高的收益，但是由于系统对事情采取很高的概率执行检查，此时的系统防守能力是更强的，因此这种极端取值是被允许的。

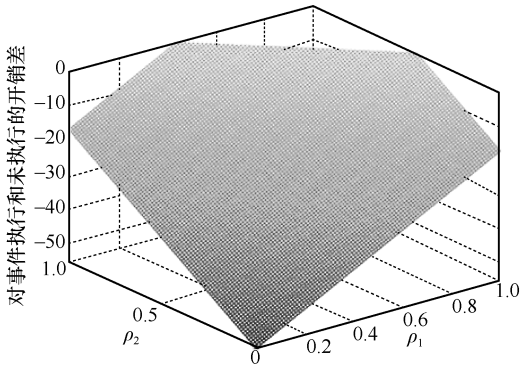

图 1 ρ_1和ρ_2取值范围

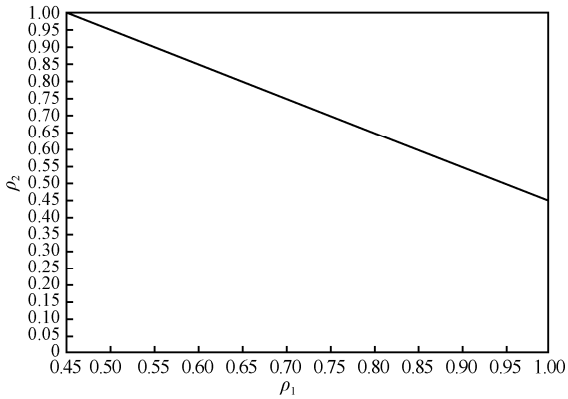

图 2 极端情况 ρ_1和ρ_2的取值

随后将 ρ_1和ρ_2代入式(2)中，寻找在允许范围内的防守信道的信道容量最大值，即是允许范围内的防守能力最大点，其结果如图 3 所示。由图 3 可知，当 $\rho_1 = 0.72$、$\rho_2 = 0.64$时，随机变量 Y 和 Z 的互信息量达到最大，也就是防守信道的信息传输率达到最大并等于当前防守信道的信道容量。此时，系统的防守能力达到最强。但是值得注意的是，此时 ρ_1和ρ_2取值的点是落在图 2 上的，也就是说，此时对某一事件进行处理的期望开销和不处理的期望开销是一样的。系统在不同响应策略下的防守能力（防守信道的信道容量）及系统开销如图 4 所示。由于响应策略是对于报警和不报警事件按一定概率抽查，为避免偶然性，所有的开销都是在对应策略下重复运行 10 次的标准差。

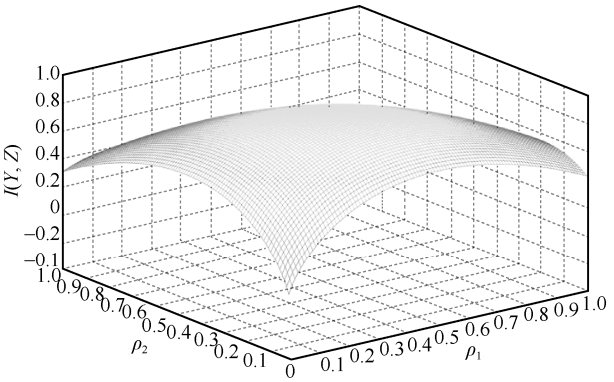

图 3 防守信道的信道容量

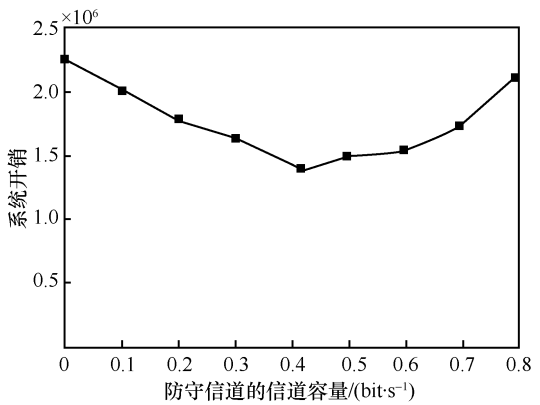

图 4 不同防守能力下系统开销

从图 4 中可以看到，当防守信道的信道容量为 0 时，此时不对任何事件做出响应，系统因遭受入侵而造成大量损失，此时的系统开销是最大的。随着系统调整响应策略，开始恢复对警报和未报警事件的抽查，防守信道容量持续增大，系统防守能力逐渐增大，系统开销逐渐降低，当防守信道容量达到 0.421 时，系统开销降到最低，此时 $\rho_1 = 0.58$、$\rho_2 = 0.46$。而当系统的防守能力继续增大时，系统的开销也开始持续上升，当系统达到允许的最大防守能力，即防守信道容量达到 0.793 时，系统开销已经与不进行响应时的情况十分接近，此时 $\rho_1 = 0.72$、$\rho_2 = 0.64$。这是由于系统为了获得更高

的安全性，花费了更多的代价对事件进行了响应。针对不同的应用环境，相关系统管理员可以根据自身应用需求对系统的安全性和开销进行考虑和权衡，在系统开销最小和防守能力最大的响应策略中选择最符合自身利益的响应方案。

下面，将本文方案与文献[3]方案、文献[4]方案、完全信任 IDS 方案等进行对比。在此，本文定义评价指标响应准确率为

$$\text{响应准确率} = \frac{\text{正确响应次数}}{\text{所有响应次数}}$$

各方案在运行过程中的响应准确率如图 5 所示，表 4 从各个方案关于响应准确率和系统开销方面进行了比较。文献[3]方案需要依赖于 IDS 对每个事件的总体预期风险进行评估，因此当 IDS 的准确率不足够高时，方案的准确率和系统开销的表现都不是十分良好。文献[4]方案由于忽视了提高执行检查的概率而造成误警成本的增加，因此其方案没有达到预期的最优效果。完全信任 IDS 方案则只对 IDS 报警的事件进行响应，对未报警的事件则信任，此方案受到 IDS 准确率的严重制约，使系统开销极大时，响应准确率却最差。本文方案是根据 IDS 运行过程中各种相关安全参数在严格的数学模型和信息论方法下得到的基于信息论的最优响应方案。实验结果表明，本文方案能够在合理的系统开销前提下，获得更高的响应准确率。

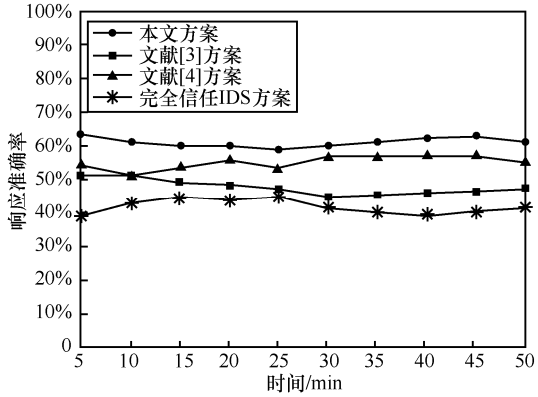

图 5　各方案在运行过程中的响应准确率

表 4　各方案关于响应准确率和系统开销的对比

方案	响应准确率	系统开销/bit
本文方案	61.2%	2 170 691
文献[3]方案	47.6%	2 402 100
文献[4]方案	55.3%	1 981 301
完全信任 IDS 方案	41.5%	2 031 310

本文方案是将攻防双方抽象成攻防随机变量，并利用 IDS 的检测率、漏警率、误警率、系统在一段时间内遭受攻击的概率等参数来计算双方不同行动策略下，产生不同结果的概率分布，并以此建立防守信道，在系统开销的约束条件下寻找防守信道的最大信道容量，而在此情况下的响应策略就是本文所需要的基于信息论的最优响应方案。根据实验结果显示，本文方案相较于传统的方法[22]而言，拥有严格定量的数学模型，能够更准确地帮助安全人员或者自动入侵响应模块[31]对入侵进行响应而非对部分入侵选择忽略，同时也能帮助系统以更小的代价获得更高的安全性。所提策略既可以帮助系统管理员应对和处理相关的网络事件，也可以应用于各种自动入侵响应系统或模块。对于初次上线的系统而言，可以通过业内认可且较全面的公开数据集对其 IDS 进行测试，同时可以结合正在运营的同类系统的日志文件作为参考来获取相关参数。对于正在运营的系统，则可以直接利用历史日志文件对其 IDS 性能和运行环境进行分析，从而获取所需的相关参数。

7　结束语

本文基于信息论，研究了 IDS 漏警和误警错误的响应问题，并提出了基于信息论的入侵检测响应方案，详细分析了 IDS 运行中各种事件发生的概率和采取不同响应的结果，并以此建立攻防对抗模型；结合信息论建立防守信道模型，将攻防对抗问题转换为信道上信息传输的问题；最后通过防守信道的信道容量来刻画系统防守能力，在研究 IDS 防守能力的同时，讨论了系统采取不同行动策略时的开销，并且给出了 IDS 响应策略的系统开销最小点和防守能力最优点。

由于 IDS 或入侵者都能调整自身的行动策略，因此在尚未稳定的应用场景中，双方为使对抗结果更符合自身目标会不断调整自身策略，直到攻防对抗过程达到平衡。显然，这一攻防过程是一种非合作博弈，而当博弈达到均衡时，攻防对抗达到平衡状态。关于这一博弈的讨论，将是下一步的研究工作。由于 APT 具有隐蔽性强、潜伏期长、持续性长等特点，因此其成为当前入侵检测技术面临的严峻挑战。利用信息论对 APT 攻击的攻击行为进行抽象提取，并进一步研究其攻击能力，同样是下一步的研究工作。

参考文献：

[1] WU S X, BANZHAF W W. The use of computational intelligence in intrusion detection systems: a review[J]. Applied Soft Computing, 2010, 10(1):1-35.

[2] ZHU J M, RAGHUNATHAN S. Evaluation model of information security technologies based on game theoretic[J]. Chinese Journal of Computers, 2009, 32(4):828-834.

[3] RHEE H, RYU Y. Evaluation of intrusion detection systems under a resource constraint[J]. ACM Transaction on Information and System Security, 2008, 11(4): 95-118.

[4] CAVUSOGLU H, RAGHUNATHAN M S. The value of intrusion detection systems in information technology security architecture[J]. Information Systems Research, 2005, 16(1):28-46.

[5] TIAN Y L, LI Q X, HU J, et al. Secure limitation analysis of public-key cryptography for smart card settings[J]. World Wide Web, 2020(23):1423-1440.

[6] SUBBA B, BISWAS S, KARMAKAR S. False alarm reduction in signature-based IDS: game theory approach[J]. Security and Communication Networks, 2016, 9(18): 4865-4881.

[7] ANDERSON J P. Computer security threat monitoring and surveillance[Z]. [S.n.:s.l.], (1980-04-15)[2020-03-24].

[8] DENNING D E. An intrusion-detection model[J]. IEEE Transactions on Software Engineering, 1987,13(2): 222-232.

[9] ATHANASIADES N, ABLER R, LEVINE J, et al. Intrusion detection testing and benchmarking methodologies[J]. IEEE Proceedings First IEEE International Workshop on Information Assurance, 2003:63-72.

[10] JIANG J C, MA H T, REN D E, et al. A survey of intrusion detection research on network security[J]. Journal of Software, 2000, 11(11): 1460-1466.

[11] PAXSON V. Bro: a system for detecting network intruders in real-time[J]. Computer Networks, 1999, 31(23-24):2435-2463.

[12] GARCÍA-TEODORO P, DÍAZ-VERDEJO P, MACIÁ-FERNÁNDEZ G, et al. Anomaly-based network intrusion detection: techniques, systems and challenges[J]. Computers & Security, 2009, 28(1-2):18-28.

[13] LIN W C, KE S W, TSAI C F. CANN: an intrusion detection system based on combining cluster centers and nearest neighbors[J]. Knowledge-Based Systems, 2015, 78:13-21.

[14] CHEN P, DESMET L, HUYGENS C. A study on advanced persistent threats[C]//15th International Conference on Communications and Multimedia Security. New York: ACM Press, 2014:63-72.

[15] FRIEDBERG I, SKOPIK F , SETTANNI G , et al. Combating advanced persistent threats[J]. Computers & Security, 2015, 48(C): 35-57.

[16] ZHANG Y, PAN X M, QING Z L, et al. APT attacks and defenses[J]. Journal of Tsinghua University (Science and Technology), 2017(11): 10-16.

[17] RUBIO J E, ALCARAZ C, ROMAN R, et al. Current cyber-defense trends in industrial control systems[J]. Computers & Security, 2019: 87.

[18] LUH R, JANICKE H, SCHRITTWIESER S. AIDIS: detecting and classifying anomalous behavior in ubiquitous kernel processes[J]. Computers & Security, 2019(84):120-147.

[19] MOON D, IM H, KIM I, et al. DTB-IDS: an intrusion detection system based on decision tree using behavior analysis for preventing APT attacks[J]. The Journal of Supercomputing, 2015(73):1-15.

[20] VRIES J D, HOOGSTRAATEN H, BERG J V D, et al. Systems for detecting advanced persistent threats: a development roadmap using intelligent data analysis[C]//International Conference on Cyber Security. Piscataway: IEEE Press, 2013.

[21] PIETRASZEK T. Using adaptive alert classification to reduce false positives in intrusion detection[C]//International Workshop on Recent Advances in Intrusion Detection—RAID 2004. Berlin: Springer, 2004: 102-124.

[22] HACHMI F, BOUJENFA K, LIMAM M. Enhancing the accuracy of intrusion detection systems by reducing the rates of false positives and false negatives through multi-objective optimization[J]. Journal of Network & Systems Management, 2019, 27(1): 93-120.

[23] ZONOUZ S A, KHURANA H, SANDERS W H, et al. RRE: a game-theoretic intrusion Response and Recovery Engine[J]. IEEE Transactions on Parallel and Distributed systems, 2013, 25(2): 395-406.

[24] CUPPENS N, CUPPENS F, VERAGRA J, et al. An ontology-based approach to react to network attacks[J]. International Journal of Information & Computer Security, 2008, 3(3/4):280-305.

[25] 吴姚睿, 刘淑芬. 基于攻击群模型的协同入侵的响应方法[J]. 电子学报, 2009, 37(11): 2416-2419.
WU Y R, LIU S F. A response method for cooperative intrusions based on the attack group model[J]. Acta Electronica Sinica, 2009, 37(11): 2416-2419.

[26] TIAN Y L, GUO J, WU Y L, et al. Towards attack and defense views of rational delegation of computation[J]. IEEE Access, 2019, PP(99):1.

[27] 杨义先, 钮心忻. 安全通讯[M]. 北京: 电子工业出版社, 2018.
YANG Y X, NIU X X. The general theory of information security[M]. Beijing: Publishing House of Electronics Industry, 2018.

[28] LIN W C, KE S W, TSAI C F. CANN: an intrusion detection system based on combining cluster centers and nearest neighbors[J]. Knowledge-Based Systems, 2015(78):13-21.

[29] MOUSTAFA N, SLAY J. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)[C]//2015 Military Communications and Information Systems Conference. Piscataway: IEEE Press, 2015: 1-6.

[30] MOUSTAFA N, SLAY J. The evaluation of network anomaly detection systems: statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set[J]. Information Security Journal A Global Perspective, 2016, 25(1-3):1-14.

[31] 彭凌西, 谢冬青, 付颖芳, 等. 基于危险理论的自动入侵响应系统模型[J]. 通信学报, 2012, 33(1): 136-144.
PENG L X, XIE D Q, FU Y F, et al. Automated intrusion response system model based on danger theory[J]. Journal on Communications, 2012, 33(1): 136-144.

[作者简介]

田有亮（1982– ），男，贵州盘县人，博士，贵州大学教授，主要研究方向为博弈论、密码学与安全协议。

吴雨龙（1995– ），男，贵州贵阳人，贵州大学硕士生，主要研究方向为密码学与网络安全。

李秋贤（1992– ），女，河南温县人，贵州大学硕士生，主要研究方向为密码学与理性密码协议。