DOI: 10.3969/j.issn.1009-6868.2018.02.007
网络出版地址：http://kns.cnki.net/kcms/detail/34.1228.TN.20180322.1340.006.html

mMTC网络中基于空口流量的入侵检测

Air-Interface Traffic-Load Based Intrusion Detection over mMTC Networks

卢楠/LU Nan
杜清河/DU Qinghe
任品毅/REN Pinyi
（西安交通大学，陕西 西安 710049）
(Xi'an Jiaotong University, Xi'an 710049, China)

中图分类号：TN929.5 文献标志码：A 文章编号：1009-6868 (2018) 02-0030-008

摘要： 提出基于空口负载特征学习的入侵检测体系与方法。基站通过分析海量机器类通信（mMTC）节点随机接入过程中的空口信号，可以智能化学习接入负载特征。在此基础上，结合常态流量负载统计信息，设计了入侵攻击检测的框架与实时检测方案。分析与仿真结果表明所提方法可以较准确地跟踪接入负载变化。与基准方案相比，可获得较高的检测概率和较短的检测时间。方案不依赖于高层安全协议，可基于底层信号实现快速入侵检测，为未来的物联网（IoT）安全防护提供了新型思路与参考方案。

关键词： 入侵检测；MTC网络；随机接入；最大似然检测

Abstract: In this paper, an air-interface traffic-load based intrusion detection approach is proposed. The base station can intelligently learn the traffic-load features by analyzing the air-interference signal in the massive machine type communications (mMTC) nodes' random access procedure. With the help of the statistic information under the normal case, the framework of intrusion and attack detection for massive machine type communications (MTC) networks is established and a real time detection scheme is designed. The performance analysis and simulation results demonstrate that our scheme can well track the arrival process with high accuracy, and outperform the baseline schemes in terms of the detection probability and the detection time. Our low layer signal based approach can make an agile intrusion detection and does not depend on security protocol applied on the high layer, which provides novel thinking and a reference scheme for the security enhancement in future Internet of things (IoT).

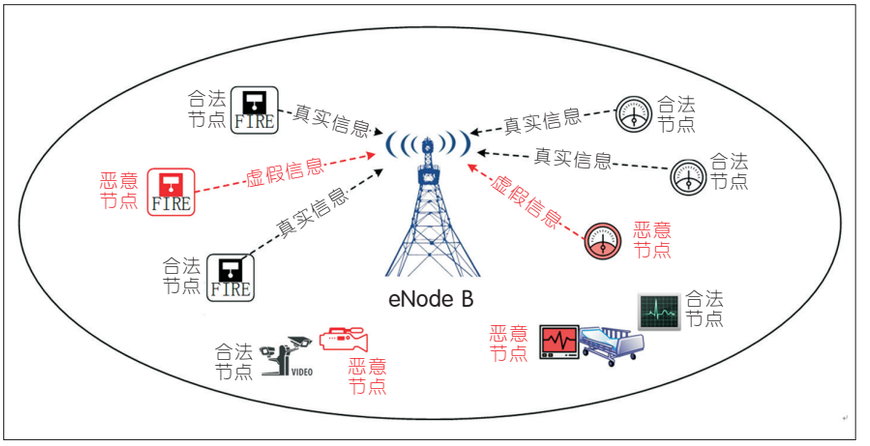
Keywords: intrusion detection; MTC networks; random access; maximum likelihood detection

机器类通信（MTC）是指不需要或只需要很少的人工干预下机器之间的通信。海量机器类通信（mMTC）面向物联网低成本节点泛在信息交换，成为5G通信系统的主要场景之一。目前，已经有50亿机器通信终端连接无线网络，到2020年这个数字预计会达到500亿[1]。机器通信网络有众多应用，如：自动驾驶、智慧医疗、智能测量、家居管理、智慧城市[2]。

安全性保障是大规模机器网络的重要任务之一。然而，大规模机器通信网络往往要求节点具有低成本特性，这一要求也削弱了节点的安全防护能力。安全问题有可能阻碍机器通信的发展甚至危害机器通信的各种应用。所以，机器通信中的安全性问题目前已经吸引了越来越多的研究[2-3]。大规模机器通信网络中的安全问题可以分为以下几类：物理攻击、配置攻击、协议攻击。机器通信网络中的安全保障机制也可以在网络协议栈的各层进行配置，例如：鉴权机制、加密算法、安全路由协议等。在mMTC网络中，入侵攻击是一种典型的安全隐患。如图1所示，在某些情况下，机器通信网络可能被一些恶意终端入侵，这些恶意终端接入后可以像正常的终端一样传递信息。恶意终端不会以瘫痪网络为目的进行攻击，而是在隐藏自己的同时发送虚假的感知信息，以达到其他恶意目的。目前，通信领域已经开发了一些手段来防止恶意终端入侵网络。其中，第3代合作伙伴计划（3GPP）推进了先进包交换鉴权密钥协议的标准化[4]。此外，也有大量基于鉴权和密钥协议来保障网络安全性的研究[5-6]。

上述大部分机制在网络上层增强安全性，这需要一定的信号开销和复杂的信任管理机制。所以，它们难以有效适应面向5G的大规模低成本机器通信网络的发展。作为上层安全机制的补充，我们通过研究发现可通过观测空口流量负载变化进行入

收稿日期：2018-01-13
网络出版日期：2018-03-22
基金项目：国家自然科学基金（61431011、61671371）；陕西省重点研发计划重点项目（2017ZDXM-GY-012）；中央高校基本科研业务费专项资金



▲图1 系统模型

侵检测。根据这一思路，我们提出了一种基于空口流量负载学习的入侵检测方法。本方法利用入侵攻击和正常状态网络流量的差异性来判断是否有入侵发生。已经有文献利用空口流量来研究小区网络中流量特征分类问题以提高服务可见性[7]，或者设计自适应算法改变随机接入的有关参数以均衡网络负载[8]；而尚未有利用空口流量进行安全入侵攻击检测的手段。

为了提升大规模机器通信网络的安全性，我们给出了mMTC网络的入侵模型并设计了入侵检测方法。本方法不同于传统上层安全保障机制，而是利用了空口流量负载特征进行入侵攻击发现。本文方法包含2个部分：首先，基于机器类型终端随机接入过程中碰撞与成功的状态进行流量负载估计；在此基础上，结合流量负载规律建立了入侵检测框架及实时检测方法。本方法可以作为现有安全协议的补充。同时，该方案利用底层信息，不会造成新的信令开销，并能有效地减小攻击发现时间，有望为未来物联网（IoT）安全入侵防护提供理论基础及参考方案。

1 mMTC系统模型

我们考虑如下的mMTC系统，该系统由一个基站和在其覆盖范围内的大量MTC终端组成。正常情况下，所有的MTC终端都是注册过的合法用户，其数量用 N 表示；但是在某些时候，一定数量的恶意MTC终端会在未经许可的情况下进入该网络，并且成功通过了鉴权机制。这些恶意终端会发送错误信息来扰乱IoT系统的常规运转，或者占用系统的时频资源，比如物理上行/下行共享信道（PUSCH/PDSCH），从而进一步达到不法的恶意目的。为了更好地隐蔽自己，恶意终端不会采取强烈或者易被察觉的攻击，比如：拒绝服务（DoS）攻击。当恶意终端的数量相对较小时，mMTC网络不会被明显影响到；反之，虚假或错误信息传播形成规模，造成极大危害。我们定义可以接受的最大恶意终端的数量为 N_1，如果恶意终端的数量超过 N_1，则认为mMTC网络发生了入侵行为。我们用 H_0 和 H_1 分别代表假设：入侵未发生和发生。mMTC网络中的合法MTC终端与恶意MTC终端数量的和定义为 N_0。那么，我们的检测问题可以描述为：

$$\begin{cases} H_0: N_0 \leqslant N + N_1 \\ H_1: N_0 > N + N_1 \end{cases} \tag{1}$$

目前，5G mMTC网络仍然处于标准化的初始阶段，具体协议尚未确定。因此，本文中我们暂时遵循长期演进（LTE）网络规范[10]。LTE网络中的机器类型网络规范的核心特征包括Beta分布到达模型、访问类别限制（ACB）机制，及4次握手接入协议。

1.1 流量模型

在网络中有 N 个注册过的合法MTC终端。文献[9]给出了两种流量模型：模型1可以视为MTC终端在一段时间内均匀地接入网络，比如非同步模式；模型2可以视为大量MTC终端以高度同步的模式接入网络，比如一次断电后的接入。考虑到网络流量的突发性，我们采用文献[9]的模型2来描述本网络中的合法MTC终端的到达过程。具体说来，MTC终端在 t 时刻发送接入请求的数量满足概率密度函数 $g(t)$，其中 $g(t)$ 服从Beta分布，如下：

$$g(t) = \frac{t^{\alpha-1}(T-t)^{\beta-1}}{T_A^{\alpha+\beta-1} Beta(\alpha,\beta)} \quad \alpha>0, \beta>0, 0 \leqslant t \leqslant T_A \tag{2}$$

其中，T_A 是时间长度，$Beta(\alpha,\beta)$ 是Beta函数[10]，对 $g(t)$ 在时间上积分可以求出在第 i 次接入中的到达终端数 $A_{[i]}$，下标“$[i]$”表示第 i 个时隙。在我们的模型中，恶意MTC终端以一种最隐蔽的方式存在于网络中。也就是说，它们有和合法终端同样的到达过程、时间起点和接入过程。

1.2 接入控制

在本系统模型中，时间划分为时隙，每个时隙由下标“$[i]$”索引，且MTC终端遵循LTE网络中的ACB机制[11]。在每个时隙开始前，eNodeB广播ACB因子 p。在每一个时隙，每个准备接入的MTC终端生成一个0和1之间的随机数 q。如果 q 小于 p，该终端则通过ACB过程，进入基于竞争的随机接入。否则，该终端退避一段时间，时间长度为随机变量 T_1，由公式（3）给出：

$$T_1 = (a_0 + b_0 \times rand) \times T_0 \tag{3}$$

其中，$rand$ 表示在区间 $[0,1]$ 中产生的均匀随机数，a_0 和 b_0 是正实数，T_0

是退避时间参数。在 T_1 S后，被退避的终端重新开始ACB过程。我们定义在第 i 次接入机会时的准备接入终端数为 $D_{[i]}$，它是在该时刻新到达的终端数、被退避至该时刻的终端数和在上一次接入中被碰撞的终端数之和。另外，在第 i 次接入机会时，通过ACB过程的终端数为 $M_{[i]}$。

图2给出了入侵检测方案的框图，图2的上半部分是接入控制和随机接入过程的示意图，下半部分是入侵检测过程的示意图，图中各符号省略了下角标。

1.3 随机接入过程

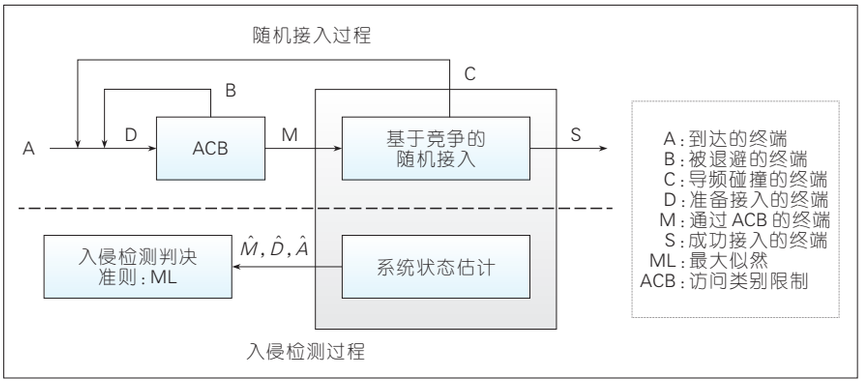
所有通过ACB过程的终端都需经过随机接入过程来传输它们的数据。实际中有2种接入模式：适用于高优先级终端的非竞争模式和适用于普通终端的时隙化竞争模式。在本文中，我们考虑采用竞争模式，其适用于存在大量普通终端的一般物联网。LTE网络下的竞争接入模式包含4个阶段[12]：第1阶段，每个终端从所有可选导频信号（Preamble）中随机选择一个，并在当前时隙通过物理随机接入信道发送该导频（在网络中，假设一共有 K 个可用导频信号，它们之间两两正交，典型的持续时间为1 ms）；第2阶段，eNodeB对每个被选择导频进行回应，发送随机接入响应消息（RAR），每个RAR包含对应于某一导频的资源块分配命令；第3阶段，每个终端根据自己在第1阶段中发送的导频检索RAR中信息，并在得到的物理上行共享信道上传输连接请求信息；第4阶段，eNodeB向数据包被成功解码的终端发送竞争解决方案消息。

对于在第1阶段选择相同导频信号的终端，其传输可能发生碰撞；但由于发射的信号相同，基站通常也可能正确接收。然而，即使在第1阶段不发生碰撞，在第3阶段用户发送连接请求数据的时候，传输的数据包在同一资源块并且不同用户的信号不同。此时碰撞不可避免，不能被eNodeB成功解码。这些导频碰撞的终端将在下一次随机接入机会时从ACB过程开始它们的接入过程。对于每一个终端，随机接入机会每 T_m 秒出现一次，通常 T_m 为0.005。

在本文中，为了便于分析并更好地关注如何入侵检测，我们采用简化的握手传输模型，即如果用户选择同一导频，那么则假设会发生碰撞。发生碰撞的用户在下一个接入机会开始时仍可进行接入竞争。这里需要指出：恶意终端因为需要尽可能伪装成合法节点而仅散播虚假或错误信息，所以它们也会遵循协议的随机接入与退避策略，从而避免基站很容易发现它们的非常规接入行为。

2 基于空口流量的入侵检测方案

我们的目标是估计网络中MTC终端的到达过程，并由此判断是否出现了异常流量，也就是MTC网络是否被入侵。如图2所示，本入侵检测方案包括两部分：系统状态估计即到达过程的估计和实时的入侵检测判决。下面分别描述这两部分。



▲图2 入侵检测方案框图

2.1 到达过程的估计算法

我们假设在每次随机接入过程中，eNodeB知道空闲导频、只被一个MTC终端占用的导频和被多个MTC终端选择的导频的数量。这3类导频的数量分别定义为A，B，C。假设基站掌握这3类导频数量的合理性在于：对于仅有只被一个MTC终端占用的导频，基站可以正确检测到并统计数量；对于被多个MTC终端选择同一导频而发生碰撞的情况，基站可以检测到较强的信号能量但不能正确译码请求数据包，从而可以区分这一类导频并统计数量；对于空闲导频，基站将仅观测到很低的能量，进而也可区分这类导频并统计数量。我们将导频状态向量 S 定义为 $(A,\mathcal{B},C)$。M 的最大后验（MAP）估计可以由公式（4）得到：

$$\hat{M}=\arg\max_{0\leqslant m\leqslant N}\left\{\Pr\left(M=m\middle|A=a,\mathcal{B}=b,C=c\right)\right\} \quad (4)$$

其中，$a,b,c=0,1,\cdots,K$。在用户独立选择的条件下，最大后验估计退化为最大似然（ML）估计。M 的最大似然估计由公式（5）给出：

$$\hat{M}=\arg\max_{0\leqslant m\leqslant N}\left\{\Pr\left(A=a,\mathcal{B}=b,C=c\middle|M=m\right)\right\} \quad (5)$$

在 $A+\mathcal{B}+C=K$ 的约束下，S 一共有 $(K+1)(K+2)/2$ 种可能。对于 $(A,\mathcal{B},C)$，它的状态索引定义为：

$$n=\frac{1}{2}(2K-A+3)A+\mathcal{B}+1 \quad (6)$$

我们可以假设每次接入竞争机会中用户的选择是独立的，但是顺序发生。这样，我们可以采用马尔科夫链模型来完成最大似然估计中条件概率的计算。具体的概率转移矩阵及相关计算参见我们的前期工作[13]。为了本文的完整性，我们将过程做如下简述。每次接入机会中用户逐个

选择导频的概率转移矩阵定义为 P。其中，P_{n_1,n_2} 代表 S 从 n_1 状态转移到 n_2 状态的概率。对于当前状态 n_1（其状态矢量描述记为 $(A,\mathcal{B},C)$）在添加一个导频后，转移为状态 $(A-1,\mathcal{B}+1,C)$，$(A,\mathcal{B},C)$，$(A+1,\mathcal{B},C-1)$ 的概率分别为 A/K，$\mathcal{B}/K$，C/K。因此有：

$$P_{n_1,n_2}=\begin{cases}\frac{A}{K}, & if\ n_2=\frac{1}{2}(2K-A+4)(A-1)+\mathcal{B}+2\\ \frac{\mathcal{B}}{K}, & if\ n_2=\frac{1}{2}(2K-A+3)A+\mathcal{B}+1\\ \frac{C}{K}, & if\ n_2=\frac{1}{2}(2K-A+2)(A+1)+\mathcal{B}+1\\ 0, & others\end{cases} \quad (7)$$

构造出概率转移矩阵 P 后，我们可以得到第 m 步后的概率分布向量 $\pi(m)$ 为：

$$\pi(m)=\pi(0)P^m \quad (8)$$

其中，$\pi(0)$ 是初始概率的分布向量。对于第 n 种状态，公式（5）可以写为：

$$\hat{M}=\arg\max_{0\leqslant m\leqslant N}\left\{\left[\pi(m)\right]_n\right\} \quad (9)$$

其中，下标 n 代表向量的第 n 个元素。由公式（9）可以求出 M 的最大似然估计。M 的取值范围是 $[0,N]$，但在实际系统中，遍历所有的可能性会引入较大的复杂度。因此，在本文仿真中，我们设定一个 M 的上限 $M_{\max}$，仅在 $[0,M_{\max}]$ 范围内考察负载大小。这里 $M_{\max}$ 的取值远远大于每一次竞争的节点平均数目，因此不会对方案性能产生显著影响。

上述方法完成了对于通过 ACB 的终端数的估计，下面我们需要对网络内的到达节点数进行估计。如果网络中目前有 D 个准备接入的 MTC 终端，它们首先需要进行 ACB 过程，通过 ACB 过程的 MTC 终端数为 M，M 服从二项分布，试验次数为 D，概率为 p。那么，已知 $\hat{M}$ 后，D 的最大似然估计值为：

$$\hat{D}=\frac{\hat{M}}{p} \quad (10)$$

在第 i 个时隙，准备接入的 MTC 终端由 3 部分组成：由公式（2）定义的新到达的 MTC 终端，在 $i-1$ 时隙碰撞的 MTC 终端和退避到第 i 个时隙的 MTC 终端。所以，第 i 个时隙的新到达 MTC 终端的数量可以由式（11）计算：

$$\hat{A}_{[i]}=\hat{D}_{[i]}-\bar{B}_{[i]}-\hat{C}_{[i-1]} \quad (11)$$

其中，

$$\hat{C}_{[i-1]}=\hat{M}_{[i-1]}-\mathcal{B}_{[i-1]} \quad (12)$$

在公式（12）中，$\mathcal{B}_{[i-1]}$ 是 $i-1$ 时隙的成功接入 MTC 终端的数量，它等于随机接入过程中第 4 步竞争释放消息的数量。另外，$\bar{B}_{[i]}$ 是 $B_{[i]}$ 的期望，由公式（3）我们得出：

$$\bar{B}_{[i]}=\begin{cases}0, & i\leqslant l\\ \frac{1}{r}\sum_{s=1}^{i-l}D_{[i]}\left(1-p_{[i]}\right), & l<i\leqslant l+r\\ \frac{1}{r}\sum_{s=1}^{r}D_{[i-l-s]}\left(1-p_{[i-l-s]}\right), & i>l+r\end{cases} \quad (13)$$

其中，$l=a_0T_0/T_m$，$r=b_0T_0/T_m$。

2.2 基于最大似然准则的实时检测算法

我们根据得到的当前时隙及过去时隙的到达节点数的估计值，基于最大似然准则估计网络内的总节点数并做出判决。在我们的实时入侵检测算法中，我们每 λ 个时隙做一次判决，比如当 $\lambda=20$，随机接入信道（RACH）的周期是 5 ms，我们每 0.1 s 做一次判决。$\hat{A}_k$ 表示在第 $k-1$ 次判决和第 k 次判决间的到达 MTC 终端的数量，其中，下标“k”表示第 k 次判决。对于第 k 次判决，我们已经得到了过去所有 $k\times\lambda$ 个时隙的到达终端的估计值。我们的流量模型是一个概率密度函数遵循 Beta 分布的随机到达过程。在过去的 $k\times\lambda$ 个时隙中，总的到达 MTC 终端数量服从二项分布，二项分布的参数为 N_0 和 $\tilde{g}_k$，$\tilde{g}_k$ 代表 $g(t)$ 的累计分布函数，如公式（14）：

$$\tilde{g}_k=\int_{t_0}^{t_{\lambda k}}g(t)\mathrm{d}t \quad (14)$$

至此，我们可以得到网络内的总节点数 N_0 的最大似然估计为：

$$\hat{N}_0=\frac{\sum_{z=1}^{k}\hat{A}_z}{\tilde{g}_k} \quad (15)$$

如果 $\hat{N}_0$ 大于合法终端数和允许的最大恶意终端数之和，eNodeB 会发送一个警告信息；否则，eNodeB 会发送一个空白信息，我们则认为目前的 MTC 网络处于正常情况。我们得到入侵检测问题的最大似然判决如下：

$$\hat{N}_0\underset{H_0}{\overset{H_1}{\gtrless}}N+N_1 \quad (16)$$

当累积的终端到达数量较小时，到达过程的随机性可能导致较高的错误检测概率。为了降低错误检测概率，我们设置 k_0 作为启动最大似然判决的门限。k_0 时刻需要满足错误检测概率小于 δ 的条件，δ 是一个很小的概率值。第 k 次判决时的错误检测概率 P_{F_k} 由式（17）给出：

$$P_{F_k}=\sum_{x=\tilde{g}_k(N+N_1)}^{N}\binom{N}{x}\tilde{g}_k^x\left(1-\tilde{g}_k\right)^{N-x} \quad (17)$$

可以利用 De Moivre-Laplace 定理化简公式（17）。De Moivre-Laplace 定理是中心极限定理的特殊形式，它指出在一定条件下正态分布可以作为二项分布的近似。因此可以将公式（17）写为：

$$P_{F_k}\simeq Q\left(\frac{N_1\tilde{g}_k}{\sqrt{N\tilde{g}_k(1-\tilde{g}_k)}}\right)-Q\left(\frac{N(1-\tilde{g}_k)}{\sqrt{N\tilde{g}_k(1-\tilde{g}_k)}}\right) \quad (18)$$

其中，$Q(\cdot)$ 是 Q 函数。给定 N，N_1 和 δ，利用式（18）可以由二分法求出 $\tilde{g}_{k_0}$ 和 k_0。假定 $N=30\,000$，$N_1/N=5\%$，当 $\tilde{g}_k=3\%,5\%,10\%$ 时，可以得到 $P_{F_k}=6.39\%,2.35\%,0.19\%$。当启动最大似然判决时，累积的到达节点数需要大于 $N\tilde{g}_{k_0}$。

3 入侵检测方案性能分析

3.1 估计算法的跟踪性能分析

2.1 节所述到达过程的估计算法第 1 步是依据最大似然准则估计 M。定义 M 的误差为：$\Delta M=M-\hat{M}$。在不引起歧义的情况下，我们省略下

标“$[i]$”，当涉及下标“$[i-1]$”时，则不会省略角标。对于 M 个通过ACB的终端随机选择导频，导频状态概率分布向量为：$\pi(M)$，若 $\pi(M)$ 中第 k 个元素不为零，则对应的第 k 个导频状态的概率为 $[\pi(M)]_k$。对于第 k 个导频状态，它的最大似然估计为：

$$\hat{M}_k = \arg\max_{0\le m\le N}\left\{[\pi(m)]_k\right\} \tag{19}$$

则 M 的估计误差的期望值 $E\{\Delta M\}$ 可以写为：

$$E\{\Delta M\} = \sum_{k=1}^{(K+1)(K+2)/2}[\pi(M)]_k * \left(\hat{M}_k - M\right) \tag{20}$$

其中，$E\{\cdot\}$ 表示期望。由式(20)，可以计算不同 M 下的 $E\{\Delta M\}$。

2.1节所述到达过程的估计算法第2步是根据式(11)计算 $\hat{A}$，定义 A 的估计误差为：$\Delta A = A - \hat{A}$。由公式(11)可知 ΔA 由 D，B，C 的估计误差构成，且有如下关系：

$$\Delta A = \Delta D - \Delta B - \Delta C \tag{21}$$

其中，$\Delta D = D - \hat{D}$，$\Delta B = B - \bar{B}$，$\Delta C = C - \hat{C}$。D 的估计误差由2部分构成：M 的估计误差和二项分布参数的最大似然估计值和实际值间的误差。经过化简，ΔD 的期望为：

$$E\{\Delta D\} = \frac{E\{\Delta M\}}{p} \tag{22}$$

B 的估计误差是 B 的期望值和实际值间的误差，所以有 ΔB 的期望为0。经过化简，C 的估计误差等于上一时隙 M 的估计误差，所以，我们得到 $\Delta C_{[i]}$ 的期望为：

$$E\left\{\Delta C_{[i]}\right\} = E\left\{\Delta M_{[i-1]}\right\} \tag{23}$$

综合式(21)、(22)和(23)，我们可以得到 ΔA 的期望：

$$E\{\Delta A\} = \frac{E\{\Delta M\}}{p} - E\left\{\Delta M_{[i-1]}\right\} \tag{24}$$

公式(24)说明：A 的估计误差的期望只和 M 及 $M_{[i-1]}$ 的估计误差的期望有关，而 $E\{\Delta M\}$ 和 $E\left\{\Delta M_{[i-1]}\right\}$ 可以由公式(20)得到。

3.2 实时检测算法的性能分析

我们引入对比方案1，利用它可以得出本文方案成功检测概率和错误检测概率的下界。对比方案1，它和本文方案的区别在于：对比方案1中eNodeB利用到达过程的估计值只在第10秒进行判决。所以，对比方案1称为保守的基于空口流量的入侵检测方案(简称为保守空口检测方案)。对于对比方案1，第10秒时 N_0 的估值为：

$$\hat{N}_0^{b1} = \sum_{i=1}^{T/T_u}\hat{A}_{[i]} \tag{25}$$

引入对比方案1后，我们可以得到对比方案1和本文所提方案下成功检测概率 P_D 和错误检测概率 P_F 的关系为：

$$P_D > P_D^{b1}, P_F > P_F^{b1} \tag{26}$$

其中，上角标 $b1$ 表示对比方案1下的相应变量。如前所述，A 的估计误差的期望可以由公式(24)计算出。但是，A 的估计误差难以计算。为了便于分析，我们假设 N_0 的估计误差是一个均值为0、方差为 σ^2 的高斯随机变量。则 $\hat{N}_0^{b1}$ 可以写为：

$$\hat{N}_0^{b1} = N_0 + n, \tag{27}$$

其中，$n \sim Norm\left(0,\sigma^2\right)$，$Norm(\mu,\nu)$ 表示均值为 μ、方差为 ν 的高斯随机变量。当 $N_0 = N + N_1 + 0.5\%N$ 时，我们可以得到：

$$\frac{\hat{N}_0^{b1}}{N} \sim Norm\left(1.005 + \frac{N_1}{N}, \frac{\sigma^2}{N}\right) \tag{28}$$

对比方案1下的成功检测概率为：

$$P_D^{b1} = 1 - Q\left(\frac{0.005}{\sqrt{\sigma^2/N}}\right) \tag{29}$$

同理可得，当 $N_0 = N + 0.5\%N$ 时，对比方案1下的错误检测概率为：

$$P_F^{b1} = Q\left(\frac{N_1/N - 0.005}{\sqrt{\sigma^2/N}}\right) \tag{30}$$

当 σ^2 取不同值，可计算出对比方案1的成功检测概率和错误检测概率作为本方法成功检测概率和错误检测概率的下界。例如：当 $N = 30\,000$，$N_1/N = 5\%$ 时，σ^2 分别取2，4，6，则 P_D^{b1} 分别为72.99%，66.75%，63.82%，P_F^{b1} 分别为0%，0%，0.07%。

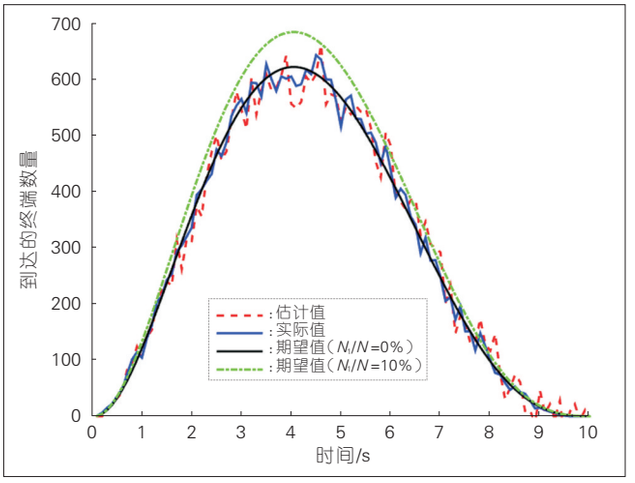
4 仿真评估

本节我们利用仿真来评估我们的入侵检测方案的性能。我们采用文献[10]中的仿真参数。仿真中假设在单个小区中有30 000个MTC终端需要进行数据传输，随机接入请求符合Beta分布，其中 $\alpha=3$，$\beta=4$，$T=10\,s$。ACB过程的参数为 $a_0=0.7$，$b_0=0.6$，$T_0=4\,s$。物理随机接入信道的配置索引为6，这意味着随机接入信道每隔5 ms出现一次，带宽则为180 kHz。我们假设一次随机接入中可用的导频总数 K 为54，最大似然判决的平滑因子 λ 设置为20。

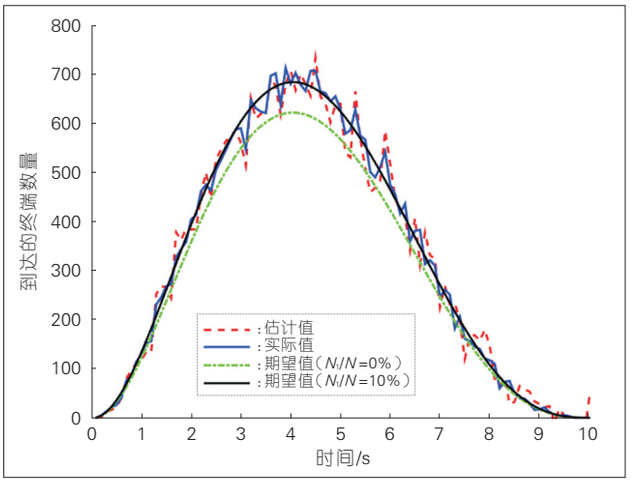
4.1 估计算法的跟踪性能

本文方案的第1步是基于机器类型终端随机接入过程中碰撞与成功的状态进行流量负载估计，所以估计算法的性能对最终的检测效果影响很大。为了评估2.1节到达过程的估计算法的性能，我们在图3和图4中分别给出了在不同情况下到达过程的实际值和估计值。如图3所示，正常情况下估计值和实际值之间的误差很小，可见我们的估计算法的跟踪性能是很理想的，这为我们的入侵检测方案实现较低的错误检测概率提供基础。同样地，观察图4，可以看到我们的估计算法的跟踪性能在入侵发生时也是很理想的，图中的估计值曲线和正常情况下的期望值曲线差异明显，这为我们的入侵检测算法实现较高的检测概率和较短的检测时间提供基础。

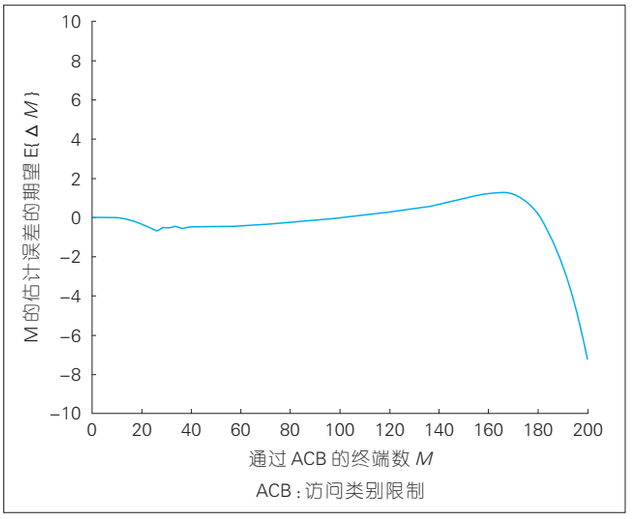
图5给出了不同的 M 取值下 M 的估计误差的期望值 $E\{\Delta M\}$ 的变化



◀图3
正常情况下到达过程的估计情况



◀图4
入侵时到达过程的估计情况



◀图5
M 的估计误差的期望 $E\{\Delta M\}$ 随 M 的变化情况

情况，可以看出在不同的 M 取值下 $E\{\Delta M\}$ 近似为0，说明我们的估计算法在不同的负载情况下都有准确的估计。注意到，当 M 接近200时估计误差的期望值较大。这是因为为了降低算法的复杂度，估计算法考察的每次竞争中的最大节点数为 M_{max}（参见2.1节）。所以，在逼近边界条件的时候，估计的误差较大。

4.2 入侵检测方案的性能

我们比较了本文方案和其他3种对比方案下的检测时间、成功检测概率、错误检测概率。3个对比方案中，eNodeB利用到达过程的估计值或者导频碰撞概率的观测值在第10秒进行判决，具体方案如下：

（1）保守的基于空口流量的入侵检测方案：详见3.2节。

（2）基于导频平均碰撞概率的入侵检测方案（简称为平均碰撞概率检测）：

$$\bar{\eta} \underset{H_0}{\overset{H_1}{\gtrless}} \bar{\eta}_1 \tag{31}$$

其中，$\bar{\eta}$ 是10 s内观测到的导频平均碰撞概率，$\bar{\eta}_1$ 是当假设 H_1 为真时的导频平均碰撞概率。

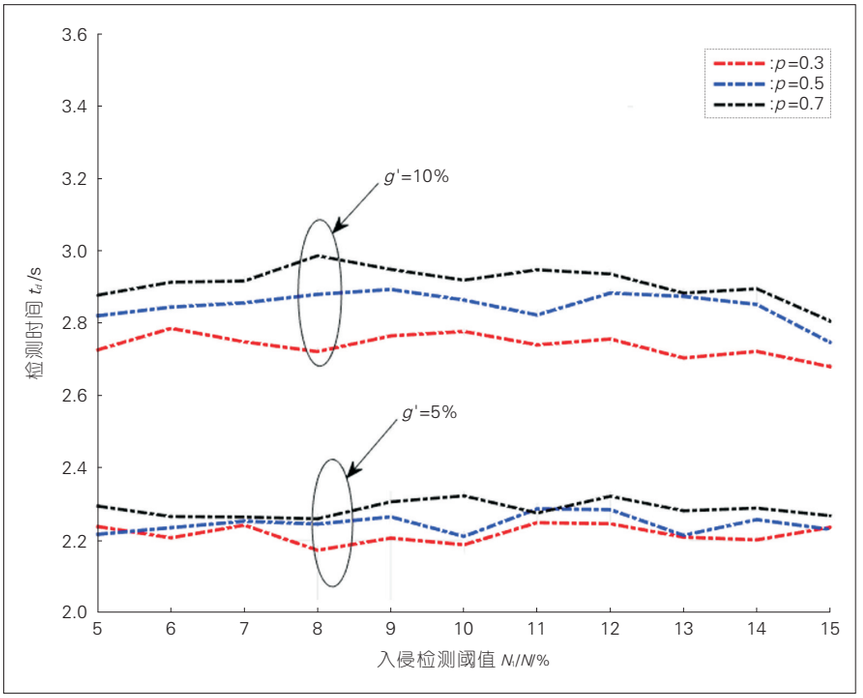
（3）基于导频平均碰撞概率偏差的入侵检测方案（简称为平均碰撞概率偏差检测）：

$$\bar{\eta} \underset{H_0}{\overset{H_1}{\gtrless}} (1+\varepsilon)\bar{\eta}_0 \tag{32}$$

其中，$\bar{\eta}_0$ 是当假设 H_0 为真时的导频平均碰撞概率，ε 是允许的最大导频碰撞概率偏差因子。

这3个对比方案都是在累积了大量的随机接入信息后进行判决，所以它们的检测时间均为10 s，而本文方案是低于10 s的。对比方案2和3利用了导频碰撞概率，所以它们只能用在ACB退避因子 p 是常数的情况。值得注意的是：在仿真中，我们假设入侵发生时的MTC终端总数 N_0 为 $N\times(1+N_1/N+0.5\%)$，作为所有 $N_0>N+N_1$ 的典型值。另外，我们假设正常情况下的MTC终端总数 N_0 等于 N，作为所有 $N_0<N+N_1$ 的典型值。

图6给出了检测时间 t_d 随系统被入侵的检测阈值（简称为入侵检测阈值）N_1/N 变化的情况，其中退避因子 p 取不同值。如图6所示，在同一 $\tilde{g}_{k_0}$（图中简写为 g'）下对于所有的



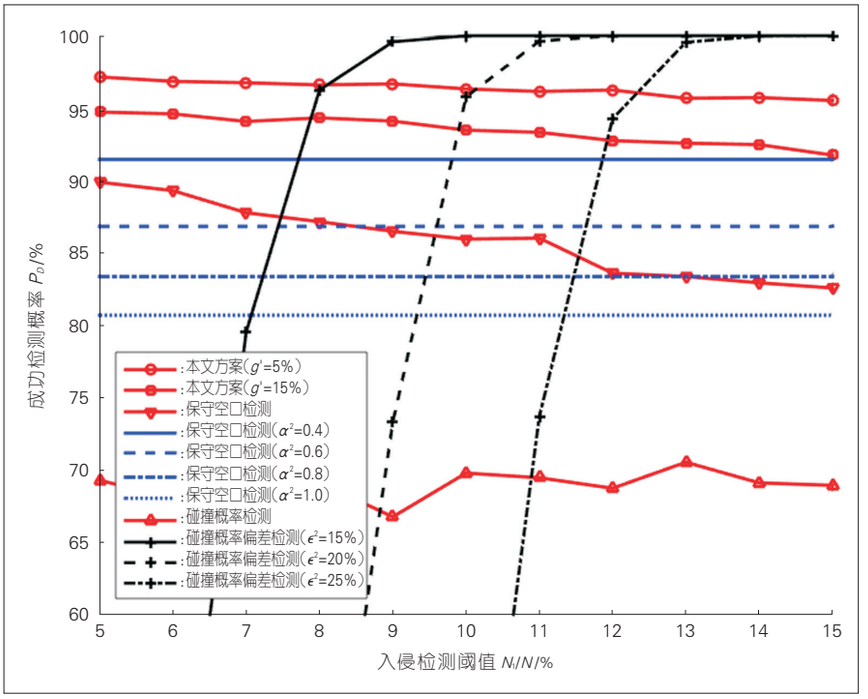
▲图6 检测时间 t_d 随入侵检测阈值 N_1/N 的变化情况

N_1/N，本方案下的检测时间是基本相同的。这是因为对于不同的 N_1/N，我们设置了相应的恶意终端数量 N_1，所以不同 N_1/N 下检测的难度相当。另外，也可以看到对于不同的 p，本方法具有稳定的性能。对于不同 $\tilde{g}_{k_0}$，由于启动最大似然判决的时刻不同，所以平均检测时间不同。如前文所述，3种对比方案的检测时间为10 s，所以在图6中省略了。图7给出了成功检测概率 P_D 随系统被入侵的阈值 N_1/N 变化的情况。本文可以实现95%以上的成功检测概率。对比方案1的检测概率低于本文方案是因为本方案是实时检测，而对比方案1只在第10秒检测。对比方案2的检测概率较低，是因为碰撞概率不能很好地描述少量入侵者存在时的空口流量变化。对比方案3中，随着横坐标增大检测概率上升，这是因为随横坐标增加，入侵者数量增多，碰撞概率增加。

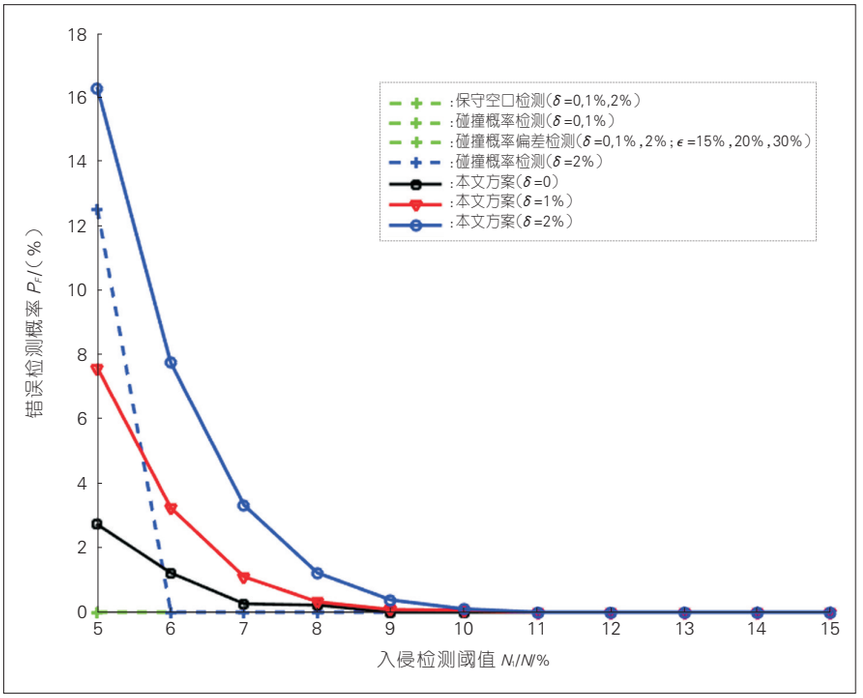
图8给出了入侵检测阈值 N_1/N 变化时错误检测概率 P_F 的变化情况。当 N_1/N 增加时，P_F 迅速下降。这是因为对于固定的 N_0，更大的入侵检测阈值意味着更松弛的安全要求。为了衡量本方案的稳健性，我们引入参数 δ 作为系统中MTC终端总数 N_0 的误差系数，系统中MTC终端总数 N_0 等于 $(1+\delta)N$。当 $\delta=1\%,2\%$ 时，P_F 在大多数 N_1/N 下低于5%，可见本方案在 N_0 有一定误差时的错误检测概率也是较低的。对于少数 P_F 大于5%的情况，此时的 N_1/N 较小，例如 $N_1/N=5\%$，这意味着网络的安全性要求是较高的。当有一定数量的恶意MTC终端出现时，即使未达到阈值，此时发出警报也是有益的。从图8中还可以看出：3个对比方案在大多数参数设置下，P_F 总是0，这是因为这3个对比方案都很保守，只在第10秒判决，相应的错误概率会很低。但是，对比方案2在 $\delta=2\%$ 时错误检测概率较高，这是因为它直接利用导频碰撞概率进行判决，所以对 N_0 的误差较敏感。

5 结束语

本文提出了基于空口负载特征学习的入侵检测体系与方法。基站通过分析mMTC节点随机接入过程中的空口信号，可以智能化学习接入负载特征。在此基础上，结合常态流



▲图7 成功检测概率 P_D 随入侵检测阈值 N_1/N 的变化情况



▲图 8 错误检测概率 P_F 随入侵检测阈值 N_1/N 的变化情况

量负载统计信息，我们设计了的入侵攻击检测的框架与实时检测方案。分析与仿真结果表明本文所提方法可以较准确地跟踪接入负载变化。与基准方案相比，可获得较高的检测概率和较短的检测时间。本文方案可以作为现有安全协议的补充，同时不会造成新的信号开销，可以用于低成本 mMTC 终端的智能管理和未来 IoT 安全防护的参考方案。下一步，我们将关注如何整合更多、更深入的信息来服务于 mMTC 和物联网中的入侵检测。

参考文献

[1] Ericsson. More than 50 Billion Connected Devices[R]. 2011
[2] BARKI A, BOUABDALLAH A, GHAROUT S, et al. M2M Security: Challenges and Solutions [J]. IEEE Communications Surveys & Tutorials, 2016, 18(2):1241–1254.DOI: 10.1109/COMST.2016.2515516
[3] CHENG Y, NASLUND M, SELANDER G, et al. Privacy in Machine–to–Machine Communications A Sate–of–the–Art Survey [C]//2012 IEEE International Conference on Communication Systems (ICCS). USA:IEEE, 2012:75–79.DOI: 10.1109/ICCS.2012.6406112
[4] LAI C, LU R, ZHENG D, et al. Toward Secure Large–sScale Machine–to–Machine Communications in 3GPP Networks: Challenges and Solutions[J].IEEE Communications Magazine, 2015, 53(12): 12–19.DOI: 10.1109/MCOM.2015.7355579
[5] 3rd Generation Partnership Project; Technical Specification Group Service and System Aspects; 3GPP System Architecture Evolution (SAE); Security architecture (Rel. 12): 3GPP TS 33.401 V12.13.0[S]. 2014
[6] LI J, WEN M, ZHANG T. Group–Based Authentication and Key Agreement With Dynamic Policy Updating for MTC in LTE–A Networks [J]. IEEE Internet of Things Journal, 2016, 3(3): 408–417.DOI: 10.1109/JIOT.2015.2495321
[7] CAO J, LI H, MA M. GAHAP: A Group–Based Anonymity Handover Authentication Protocol for MTC in LTE–A Networks[C]// 2015 IEEE International Conference on Communications (ICC). USA:IEEE, 2015: 3020–3025.DOI: 10.1109/ICC.2015.7248787
[8] LANER M, SVOBODA P, RUPP M. Detecting M2M Traffic in Mobile Cellular Networks[C]// IWSSIP 2014 Proceedings. Croatia, 2014: 159–162
[9] HE H, DU Q, SONG H, et al. Traffic–Aware ACB Scheme for Massive Access in Machine–to–Machine Networks[C]//2015 IEEE International Conference on Communications (ICC). USA:IEEE, 2015:617–622
[10] 3rd Generation Partnership Project, Study on RAN Improvements for Machine–type Communications; (Rel. 11)[S]: 3GPP TR 37.868, Vl 1.0.0. 2011
[11] GUPTA A K, Nadarajah S. Handbook of Beta Distribution and Its Applications[J]. Biometrics, 2004 , 62 (1): 309–310
[12] PHUYAL U, KOC A T, FONG M H et al. Controlling Access Overload and Signaling Congestion in M2M Networks[C]//2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), Pacific Grove. USA:IEEE, 2012:591–595
[13] 3rd Generation Partnership Project; Evolved Universal Terrestrial Radio Access (E–UTRA); Medium Access Control (MAC) Protocol Specification,Third–Generation Partnership Project, Sophia AntipolisCedex, France[S]:TS 36.321 V12.7.0. 2015
[14] HE H, DU Q, SONG H, et al. Traffic–Aware ACB Scheme for Massive Access in Machine–to–Machine Networks[C]//2015 IEEE International Conference on Communications (ICC).USA:IEEE, 2015: 617–622.DOI: 10.1109/ICC.2015.7248390
[15] DAVID T, VISWANATH P. Fundamentals of Wireless Communication[M]. Beijing: Posts & Telecom Press, 2009

作者简介

卢楠，西安交通大学研究生在读；主要研究领域为无线网络中的物理层安全技术；在 SCI 期刊及 IEEE 通信领域会议上发表多篇论文；获得授权发明专利 1 项。

杜清河，西安交通大学副教授；主要研究领域为 5G 系统关键技术、无线网络物理层安全技术、无线网大数据与机器学习技术、无线多媒体传输技术、无线通信信道建模与仿真技术等；先后主持和参加国家自然科学基金项目、“863”项目、国家科技重大专项项目 10 余项；获得 2 项科技成果奖，2007 年获 IEEE GLOBECOM 最佳论文奖，2017 年获《China Communications》最佳论文奖；已发表论文 100 余篇，其中被 SCI 检索期刊论文 50 余篇，获授权专利 10 项。

任品毅，西安交通大学教授；主要研究领域为无线网络物理层安全技术、5G 系统关键技术、无线网大数据与机器学习技术、无线认知网络、无线自组织网络；先后主持和参加国家自然科学基金项目、“863”项目、国家科技重大专项项目 30 余项；2010 年获 IEICE 通信协会最佳论文奖，2017 年获《China Communications》最佳论文奖，并获得 2 项科技成果奖；已发表论文 150 余篇，其中被 SCI 检索期刊论文 60 余篇，获授权专利 30 余项。