(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 109740348 B
(45)授权公告日 2022.06.14

(21)申请号 201910084520.8

(22)申请日 2019.01.29

(65)同一申请的已公布的文献号
申请公布号 CN 109740348 A

(43)申请公布日 2019.05.10

(73)专利权人 福州大学
地址 350108 福建省福州市闽侯县上街镇福州大学城学院路2号福州大学新区

(72)发明人 董晨 张凡 郭文忠 陈景辉 贺国荣

(74)专利代理机构 福州元创专利商标代理有限公司 35100
专利代理师 蔡学俊

(51)Int.Cl.
G06F 21/56(2013.01)

(56)对比文件
CN 102469103 A,2012.05.23
CN 104330721 A,2015.02.04
CN 107886012 A,2018.04.06
CN 108052840 A,2018.05.18
CN 108154051 A,2018.06.12
CN 105893876 A,2016.08.24
CN 107703186 A,2018.02.16
CN 104850804 A,2015.08.19
US 2018089426 A1,2018.03.29
宋晨晨.基于侧信道分析的硬件木马检测技术.《万方数据学位论文库》.2017,

审查员 高航

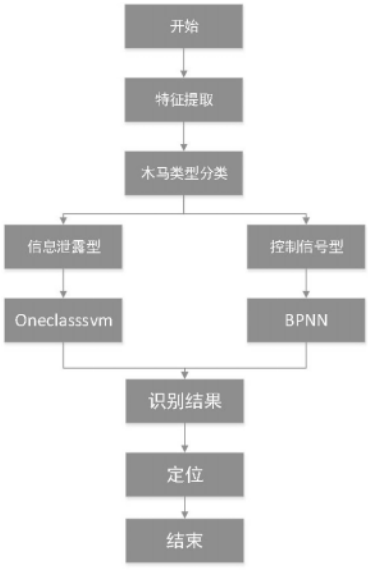
权利要求书2页 说明书6页 附图3页

(54)发明名称

一种基于机器学习的硬件木马定位方法

(57)摘要

本发明涉及一种基于机器学习的硬件木马定位方法,包括以下步骤:分析其网表结构,选择并提取电路结构特征;探究硬件木马类型,将硬件木马分为信息泄露型和控制信号型;从待测芯片中选择若干样本作为训练数据,剩下的芯片作为测试数据;对于信息泄露型木马使用oneclasssvm算法检测,对于控制信号型硬件木马使用BP神经网络进行检测;使用训练数据训练分类器,使用测试数据进行测试,统计结果;将最后的识别结果与理想结果进行对比,找到已经识别的木马线网。上述基于机器学习的硬件木马定位方法不需要复杂的实验环境、大量的时间和实验成本就可以定位出一个网表中的硬件木马。



CN 109740348 B

1.一种基于机器学习的硬件木马定位方法,其特征在于,包括以下步骤:

步骤S1:从若干待测芯片的门级网表中提取电路候选特征;

步骤S2:根据电路候选特征将待测芯片分为控制信号型芯片和信息泄露性芯片;

步骤S3:控制信号型芯片和信息泄露性芯片均随机选择一个芯片的电路候选特征作为训练数据,剩余的芯片电路候选特征作为测试数据;

步骤S4:构建一个BP神经网络,并使用控制信号型芯片的训练数据训练,得到训练后的BP神经网络;

步骤S6:构建一个Oneclasssvm分类器,并使用信息泄露性芯片的训练数据训练,得到训练后的Oneclasssvm分类器;

步骤S7:将控制信号型芯片的测试数据输入训练后的BP神经网络,将信息泄露性芯片测试数据输入训练后的Oneclasssvm分类器,得到测试结果;

步骤S8:将测试结果与理想结果对比,得到硬件木马的位置定位;所述电路候选特征包括木马线网特征和正常线网特征;

理想结果:就是用来测试的 电路中线网的实际情况,即 木马线网的理想结果是1,正常电路线网的理想结果是0。

2.根据权利要求1所述的基于机器学习的硬件木马定位方法,其特征在于:所述S4具体为:

步骤S41:对所有的层$2\leqslant l\leqslant L$,设权重$\Delta W^{(l)}=0$,设偏置$\Delta b^{(l)}=0$,这里$\Delta W^{(l)}=0$和$\Delta b^{(l)}=0$分别为全零矩阵和全零向量;

步骤S42:使用反向传播算法,计算各层神经元中节点i的权值$\nabla W_{(i)}^{(l)}$和偏置的梯度矩阵$\nabla b_{(i)}^{(l)}$:

1)计算$\triangle W^{(l)}=\nabla W_{(i)}^{(l)}=W_{ij}^{(l)}-\alpha\frac{\partial E}{\partial W_{ij}^{(l)}}$;

2)计算$\triangle b^{(l)}=\nabla b_{(i)}^{(l)}=b_i^{(l)}-\alpha\frac{\partial E}{\partial b_i^{(l)}}$.

α为学习速率,它的取值范围为(0,1);

E是m个训练样本的误差函数,

$$\mathrm{E}=\frac{1}{m}\sum_{i=1}^{m}E(i),$$

E(i)是单个样本的训练误差,

$$E(i)=\frac{1}{2m}\sum_{i=1}^{m}\sum_{k=1}^{n}(d_k(i)-y_k(i))^2.$$

$d_k(i)$为输出层第k个输出的期望值,$y_k(i)$为输出层第k个输出的实际值,m为训练样本数量;

步骤S43:更新权值和偏置:

1)计算$W^{(l)}=W^{(l)}+\frac{1}{m}\triangle W^{(l)}$

2)计算$b^{(l)}=b^{(l)}+\frac{1}{m}\triangle b^{(l)}$。

3.根据权利要求1所述的基于机器学习的硬件木马定位方法,其特征在于:所述Oneclasssvm分类器具体模型为:

$$\min_{\omega\in F,\xi\in R^{\xi},\rho\in R}\frac{1}{2}\|\omega^2\|+\frac{1}{vl}\sum_i\xi_i-\rho$$

约束于$(\omega\cdot\Phi(x_i))\geqslant\rho-\xi_i,\xi_i\geqslant0$;

Φ是x_i到F的映射,l是观察值的数量,i∈[l],ξ是非零松弛变量,ω和ρ是要求的值,v∈[0,1]为训练误差。

一种基于机器学习的硬件木马定位方法

技术领域

[0001] 本发明涉及硬件木马检测领域，具体涉及一种基于机器学习的硬件木马定位方法。

背景技术

[0002] 近年来，大部分关于信息安全的工作都集中在软件安全的开发上，而忽视了硬件的安全性。随着增加集成电路(IC)的复杂性以及设计和人工制造过程的全球化，事实上，集成电路的安全问题主要来自于被插入恶意电路的芯片。恶意电路通常是以硬件木马(HT)命名。硬件木马的标准定义是由IBM研究中心于2007年提出的：硬件木马指存在的原始电路的恶意电路或有害改动从芯片设计阶段的生命周期到封装测试阶段。一个硬件木马是一种设计好的将会在用户不知情的情况下发生电子设备中的电路。根据国际半导体技术路线图(ITRS)计划，到2020年，IC产量将增加十倍。但伴随的安全问题并不仅仅是十倍。

[0003] 随着超大规模集成电路(VLSI)电路的规模越来越大，一个芯片中可以容纳数百万个门，使得芯片变得越来越大极易受到HT攻击。依靠海上铸造厂进行IC制造是大规模生产微电路的一种经济有效的方法。但是，这样的外包方式可能会导致严重的安全威胁。这些威胁加剧了硬件木马用于关键应用时，例如车辆系统，通信系统，电力网络，运输系统或军事应用的危害。

[0004] 在芯片的制造过程中，很可能被插入芯片攻击者的恶意电路，这会导致一些问题如电路功能受损，关键信息被篡改甚至泄露。攻击者可以引入一个设计好的硬件木马在一个随机的时间来禁用或破坏系统，或者可能会泄露机密信息和密钥。

发明内容

[0005] 有鉴于此，本发明的目的在于提供一种基于机器学习的硬件木马定位方法，考虑的芯片中硬件木马的类型，采用不同的机器学习算法处理不同类型的硬件木马，实现的硬件木马的定位。

[0006] 为实现上述目的，本发明采用如下技术方案：

[0007] 一种基于机器学习的硬件木马定位方法，包括以下步骤：

[0008] 步骤S1:从若干待测芯片的门级网表中提取电路候选特征；

[0009] 步骤S2:根据电路候选特征将待测芯片分为控制信号型芯片和信息泄露性芯片；

[0010] 步骤S3:控制信号型芯片和信息泄露性芯片均随机选择一个芯片的电路候选特征作为训练数据，剩余的芯片电路候选特征作为测试数据；

[0011] 步骤S4:构建一个BP神经网络，并使用控制信号型芯片的训练数据训练，得到训练后的BP神经网络；

[0012] 步骤S41:对所有的层$2\leqslant 1\leqslant L$,设权重$\triangle W^{(1)}=0$,设偏置$\triangle b^{(1)}=0$,这里$\triangle W^{(1)}=0$和$\triangle b^{(1)}=0$分别为全零矩阵和全零向量；

[0013] 步骤S42:使用反向传播算法,计算各层神经元中节点i的权值$\nabla W_{(i)}^{(l)}$和偏置的梯度矩阵$\nabla b_{(i)}^{(l)}$:

[0014] 1)计算$\Delta W^{(l)} = \nabla W_{(i)}^{(l)} = W_{ij}^{(l)} - \alpha \frac{\partial E}{\partial W_{ij}^{(l)}}$;

[0015] 2)计算$\Delta b^{(l)} = \nabla b_{(i)}^{(l)} = b_i^{(l)} - \alpha \frac{\partial E}{\partial b_i^{(l)}}$.

[0016] α为学习速率,它的取值范围为(0,1);

[0017] E是m个训练样本的误差函数,

[0018] $\mathrm{E} = \frac{1}{m}\sum_{i=1}^{m} E(i)$,

[0019] E(i)是单个样本的训练误差,

[0020] $E(i) = \frac{1}{2m}\sum_{i=1}^{m}\sum_{k=1}^{n}(d_k(i) - y_k(i))^2$.

[0021] d_k(i)为输出层第k个输出的期望值,y_k(i)为输出层第k个输出的实际值,m为训练样本数量

[0022] 步骤S43:更新权值和偏置:

[0023] 1)计算$W^{(l)} = W^{(l)} + \frac{1}{m}\Delta W^{(l)}$

[0024] 2)计算$b^{(l)} = b^{(l)} + \frac{1}{m}\Delta b^{(l)}$。

[0025] 步骤S6:构建一个Oneclasssvm分类器,并使用信息泄露性芯片的训练数据训练,得到训练后的Oneclasssvm分类器;

[0026] $\min_{\omega\in F,\xi\in R^t,\rho\in R} \frac{1}{2}\|\omega^2\| + \frac{1}{vl}\sum_i \xi_i - \rho$.

[0027] 约束于$(\omega \cdot \Phi(x_i)) \geq \rho - \xi_i, \xi_i \geq 0$.

[0028] Φ是x到F的映射,l是观察值的数量,i∈[1],ξ是非零松弛变量,ω和ρ是要求的值,v∈[0,1]为训练误差。

[0029] 步骤S7:将控制信号型芯片的测试数据输入训练后的BP神经网络,将信息泄露性芯片测试数据输入训练后的Oneclasssvm分类器,得到测试结果;

[0030] 步骤S8:将测试结果与理想结果对比,得到硬件木马的位置定位。

[0031] 理想结果:就是用来测试地电路中线网的实际情况,既木马线网的理想结果是1,正常电路线网的理想结果是0.

[0032] 进一步的,所述电路候选特征包括木马线网特征和正常线网特征本发明与现有技术相比具有以下有益效果:

[0033] 本发明考虑的芯片中硬件木马的类型,采用不同的机器学习算法处理不同类型的硬件木马,实现的硬件木马的定位,是一种全新的思维方式,具有高效,准确,低成本的特

点,对推进硬件木马检测有重大意义。

附图说明

[0034] 图1是本发明方法流程图;

[0035] 图2是是本发明一实施例中待测芯片的门级网表图;

[0036] 图3是本发明的一实施例中测试电路RS232-T1100;

[0037] 图4是本发明的一实施例中测试电路RS232-T1000;

[0038] 图5是本发明的一实施例中测试电路s38417-T100;

[0039] 图6是本发明的一实施例中测试电路s15850-T100。

具体实施方式

[0040] 下面结合附图及实施例对本发明做进一步说明。

[0041] 请参照图1,本发明提供一种基于机器学习的硬件木马定位方法,包括以下步骤:

[0042] 步骤S1:从若干待测芯片的门级网表中提取电路候选特征;从如图二所有待测芯片的门级网表中提取硬件木马候选特征(如表一),在门级网表中都是以module开始,endmodule结尾,里面的内容是定义这个电路的输入输出线网有哪些,比如在图一中的门级网表,第一句逻辑结构描述是and g1(x,a,b);对应的是图一右边这个电路中叫做g1的那个与门,它的输入是a,b,输出是x。通过所有这样的语句可以把整个电路图描述出来。编码提取如表一中的51个特征中,待测芯片是通过这个芯片中的每一根线网表示的,这些线网在对应芯片的门级网表中都有表示,其中包含了木马线网和正常线网。采用编程的方式处理网表中的字符串来提取这些特征,举例说明这些特征,第一个特征fan_in_x:离线网n,x个等级的逻辑门输入的数量。这里的等级如图三图四所示,从电路的总输入或总输出开始,按顺序,第一个逻辑门/复用器/触发器为第一级,第二个为第二级,以此类推。离线网n,的x级上所有逻辑门的总输出的数量,其他特征类似.

[0043] 表一硬件木马候选特征

[0044]

硬件木马候选特征	x∈[1,7]
fan_in_x	离线网 n, x 个等级的逻辑门输入的数量
in_flipflop_x	从输入到输出方向，离线网 n，x 个等级的触发器的数量
out_flipflop_x	从输出到输入方向，离线网 n，x 个等级的触发器的数量
in_multiplexer_x	从输入到输出方向，离线网 n，x 个等级的复用器的数量
out_multiplexer_x	从输出到输入方向，离线网 n，x 个等级的复用器的数量
in_loop_x	从输入到输出方向，离线网 n，x 个等级的循环的数量
out_loop_x	从输出到输入方向，离线网 n，x 个等级的循环的数量
in_const_x	从输入到输出方向，离线网 n，x 个等级的常数的数量
out_const_x	从输出到输入方向，离线网 n，x 个等级的常数的数量
in_nearest_pin	从输入到输出方向，离线网 n 的最小等级
out_nearest_pout	从输出到输入方向，离线网 n 的最小等级
{in,out}_nearest_flipflop	从输入到输出方向或相反，最近的触发器所在的等级
{in,out}_nearest_multiplexer	从输入到输出方向或相反，最近的复用器所在的等级

[0045] 步骤S2:根据电路候选特征将待测芯片分为控制信号型芯片和信息泄露性芯片；如表二中RS232开头的芯片是中是控制信号型硬件木马,s开头的芯片是信息泄露型硬件木马。

[0046] 表2待测电路

[0047]

网表名称	木马网络数量	正常网络数量
RS232-T1000	44	211
RS232-T1100	44	212
RS232-T1200	45	211
RS232-T1300	31	222
RS232-T1400	50	205
RS232-T1500	48	209
RS232-T1600	39	216
s15850-T100	61	2371
s35932-T100	34	6368
s35932-T200	40	6359
s35932-T300	59	6365
s38417-T100	29	5772
s38417-T200	35	5769
s38417-T300	31	5802
s38584-T100	21	7271
s38584-T200	198	7274
s38584-T300	976	7275

[0048] 步骤S3:控制信号型芯片和信息泄露性芯片均随机选择一个芯片的电路候选特征作为训练数据,剩余的芯片电路候选特征作为测试数据;

[0049] 步骤S4:构建一个BP神经网络,并使用控制信号型芯片的训练数据训练,得到训练后的BP神经网络;

[0050] 步骤S41:对所有的层2≤l≤L,设权重$\triangle W^{(l)}=0$,设偏置$\triangle b^{(l)}=0$,这里$\triangle W^{(l)}=0$和$\triangle b^{(l)}=0$分别为全零矩阵和全零向量;

[0051] 步骤S42:使用反向传播算法,计算各层神经元中节点i的权值$\nabla W_{(i)}^{(l)}$和偏置的梯度矩阵$\nabla b_{(i)}^{(l)}$:

[0052] 1)计算$\triangle W^{(l)}=\nabla W_{(i)}^{(l)}=W_{ij}^{(l)}-\alpha\frac{\partial E}{\partial W_{ij}^{(l)}}$;

[0053] 2)计算$\triangle b^{(l)}=\nabla b_{(i)}^{(l)}=b_{i}^{(l)}-\alpha\frac{\partial E}{\partial b_{i}^{(l)}}$.

[0054] α为学习速率,它的取值范围为(0,1);

[0055] E是m个训练样本的误差函数,

[0056] $$\mathrm{E}=\frac{1}{m}\sum_{i=1}^{m}E(i),$$

[0057] E(i)是单个样本的训练误差,

[0058] $$E(i)=\frac{1}{2m}\sum_{i=1}^{m}\sum_{k=1}^{n}(d_k(i)-y_k(i))^2.$$

[0059] d_k(i)为输出层第k个输出的期望值,y_k(i)为输出层第k个输出的实际值,m为训练样本数量

[0060] 步骤S43:更新权值和偏置:

[0061] 1)计算$W^{(l)}=W^{(l)}+\frac{1}{m}\triangle W^{(l)}$

[0062] 2)计算$b^{(l)}=b^{(l)}+\frac{1}{m}\triangle b^{(l)}$

[0063] 步骤S6:构建一个Oneclasssvm分类器,并使用信息泄露性芯片的训练数据训练,得到训练后的Oneclasssvm分类器;所述模型具体为:

[0064] $$\min_{\omega\in F,\xi\in R^{t},\rho\in R}\frac{1}{2}\|\omega^2\|+\frac{1}{vl}\sum_{i}\xi_i-\rho\,.$$

[0065] 约束于$(\omega\cdot\Phi(x_i))\geq\rho-\xi_i,\xi_i\geq0$.

[0066] Φ是x到F的映射,l是观察值的数量,i∈[l],ξ是非零松弛变量,ω和ρ是要求的值,v∈[0,1]为训练误差。

[0067] 步骤S7:将控制信号型芯片的测试数据输入训练后的BP神经网络,将信息泄露性芯片测试数据输入训练后的Oneclasssvm分类器,得到测试结果;

[0068] 步骤S8:将测试结果与理想结果对比,得到硬件木马的位置定位。

[0069] 理想结果:就是用来测试地电路中线网的实际情况,既木马线网的理想结果是1,正常电路线网的理想结果是0.

[0070] 以上所述仅为本发明的较佳实施例,凡依本发明申请专利范围所做的均等变化与修饰,皆应属本发明的涵盖范围。

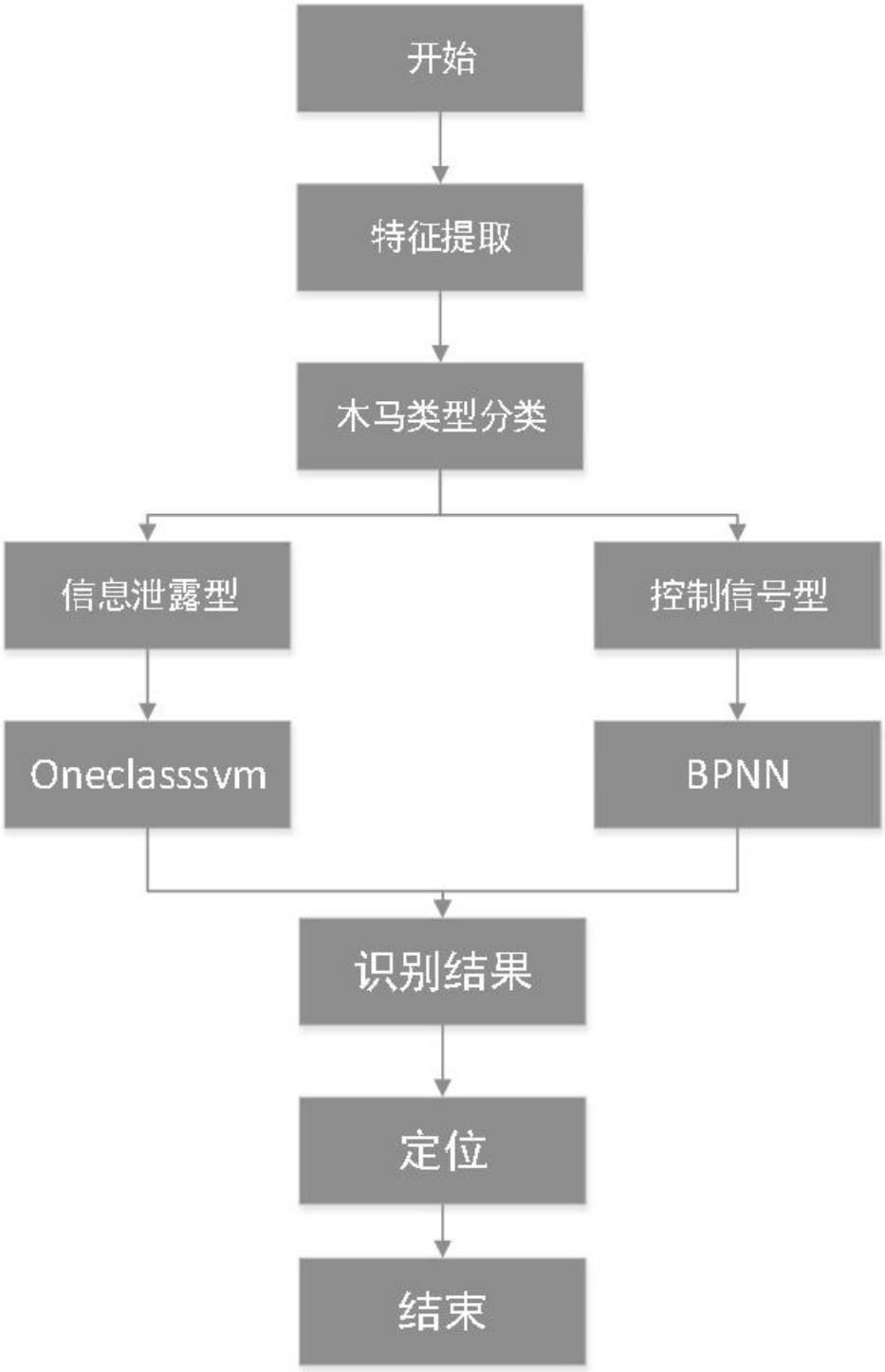


图1

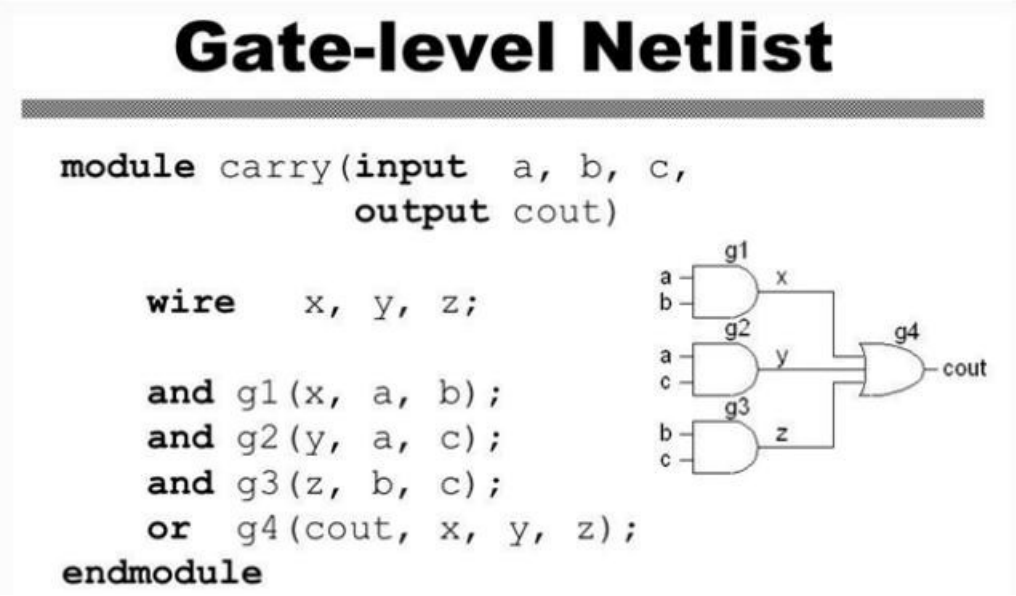

Gate-level Netlist
module carry(input a, b, c,
output cout)
wire x, y, z;
and g1(x, a, b);
and g2(y, a, c);
and g3(z, b, c);
or g4(cout, x, y, z);
endmodule
g1
g2
g3
g4
a
b
c
x
y
z
cout

图2

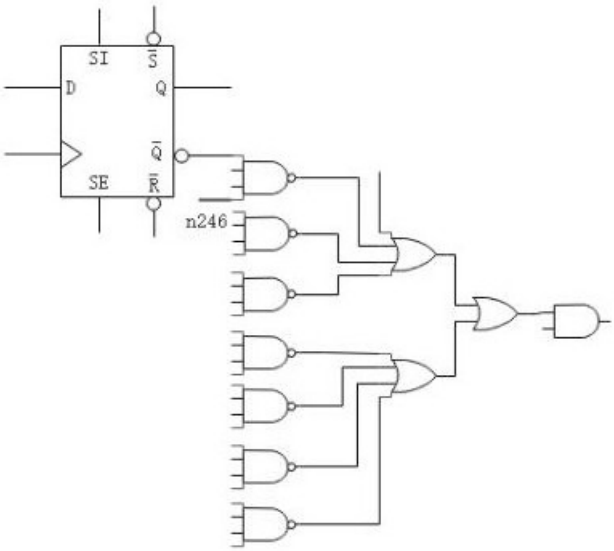

SI
D
SE
Q
n246

图3

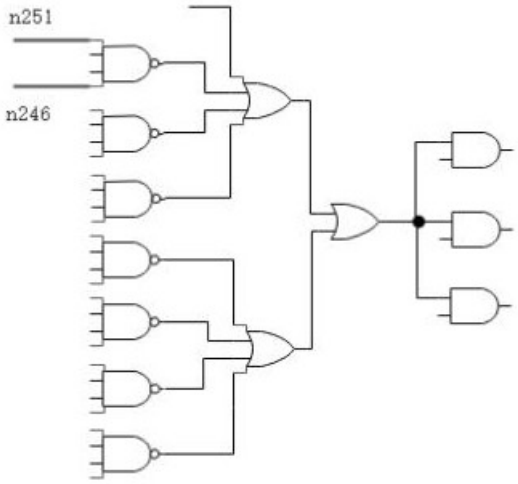

n251
n246

图4

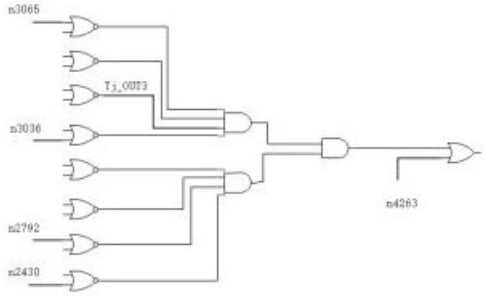

n3065
Tj_OUT3
n3036
n4263
n2792
n2430

图5

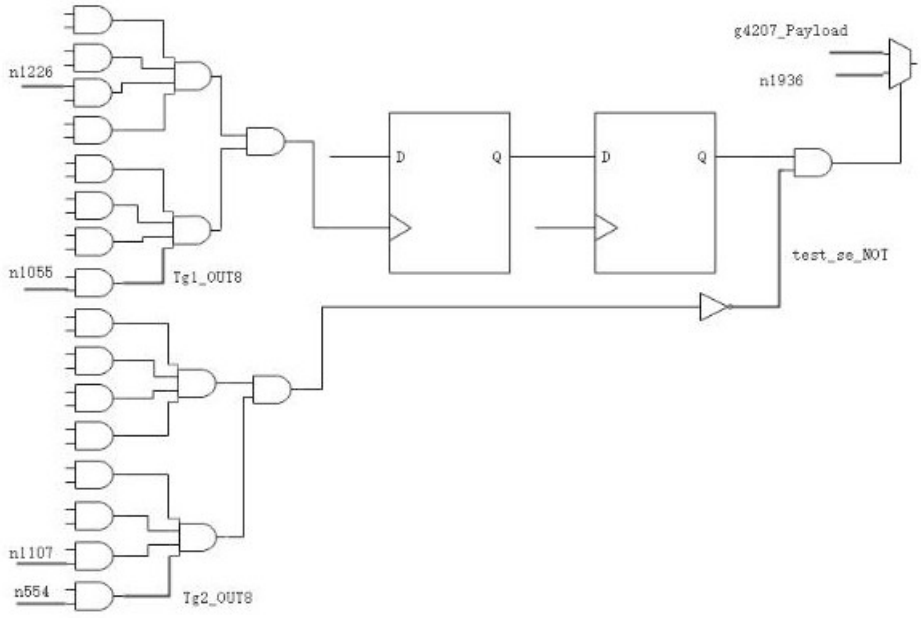

g4207_Payload
n1936
n1226
n1055
Tg1_OUT8
D
Q
D
Q
test_se_NOI
n1107
n554
Tg2_OUT8

图6