基于发布/订阅系统的安全管理平台设计

张继德　屈尔庆　贺志芳

（河北理工大学计算机与自动控制学院　唐山 063009）

摘　要　目前，安全管理平台建设向着大规模协同工作的方向发展。本文首先提出一个采用基于内容的发布/订阅系统设计的安全管理平台，使其具有良好的可扩展性和动态接入特性，为协同管理提供了良好的体系保证；然后提出一个发布/订阅传输协议，在有限隐私共享的假设前提下，运用 bloom filter 和信任等级解决发布/订阅路由协议与安全保护冲突，并对协议的保密性、匿名性、隐私性等安全特性做了分析；最后以模拟测试验证了此协议的可行性。

关键词　发布/订阅系统，安全，Bloom filter

Design of Security Management Platform Based on Publish-subscribe System

ZHANG Ji-De　QU Er-Qing　HE Zhi-Fang

(College of Computer and Automatic Control, Hebei Polytechnic University, Tangshan 063009)

Abstract　The development trend of Security Management Platform (SMP) is large-scale and cooperation. This paper first gives a design of scalable and dynamic access SMP based on content-based Pub/Sub system, then proposes a Pub/Sub transport protocol, using bloom filter and trust rank to solve the conflict between route and security requirement on the premises of limited privacy share, and provides a security analysis on confidentiality, anonymity and privacy. Finally, a simulate test verifies the feasibility of proposed protocol by performance optimization.

Keywords　Pub/Sub, Security, Bloom filter

1　引言

安全管理平台(Security Management Platform，SMP)是近年来对安全产品、安全事件、安全策略集中管理的新思路。目前安全管理平台的主要平台结构是基于代理的多层服务模式，然而日益增长的各类攻击呈现出地址来源分散，时间跨度长，攻击方法多样，更新速度快的特点，现有模式只适合封闭环境，因此，我们提出采用发布/订阅系统来满足开发的互联网环境下，大规模分布式的协同工作的需求。

本文采用基于内容的发布/订阅系统为消息传输体系，在此体系内，参与者要求提供数据保密性和身份匿名性，即参与者要求他们提供的数据在被恶意截获或发送到非订阅第三方后，数据中包含的敏感信息不被泄漏，同时订阅者和发布者无法获知对方身份信息。通常的解决手段是加密，但是加密的消息流很难实现基于内容的路由，因为系统要求每一个路由节点检查每一个传输包的内容。如果路由节点可以解密消息流包，则要求路由节点可信，文[1]在可信路由的基础上，加强了对发布者和订阅者的访问控制，提高系统的安全性，但是其代价是昂贵的。同时，采用由于隐私等级而引发的多重密钥共享问题也需要解决，现有的解决方案是基于 Shamir 的门限秘密共享方案[2]，文[3]提供了多个秘密共享，但其实现多等级的秘密共享也需要多次共享，而且需要随订阅者(共享参与者)的动态变化而重新秘密发布和计算。

由于事实上很难实现端对端的隐私策略，因此我们做出以下有限隐私共享假设，即发布者的隐私对于同一信任度的参与者共享的，但信任度没有高低区分，即 $\forall A, \forall B, (A \rightarrow \text{Privacy}(X)) \wedge (B \rightarrow \text{Privacy}(Y)) \wedge (\text{TrustLevel}(A) = \text{TrustLevel}(B)) \rightarrow (\text{privacy}(X) \leftrightarrow \text{privacy}(Y))$

在此前提下，本文首先提出一个基于内容的发布/订阅模型的安全管理平台结构，然后通过引入 bloom filter，实现具有隐私和安全保护的数据传送机制，并分析其安全性。最后通过模拟试验，通过性能指标证明其可行性。

以往的对于安全管理平台的研究重点主要集中在安全事件的关联分析，以及安全设备间的兼容互操作上，安全形式的可视化[4,5]等内容上。文[6]提出了一个基于 P2P 体系的 IDS，以一个无中心的分布协作式平台来解决扩展问题和瓶颈问题，这与本文的研究方向是比较类似的，发布/订阅系统具有异步、松散耦合、多对多通信的优点，适应动态多变的大规模分布式网络环境的需求，因此有着广阔的研究和应用前景[7]，文[8]提出了一种基于发布/订阅通信的动态数据收集模型。文[9]探讨了发布/订阅系统所涉及的安全问题，Bloom Filter 是一种优秀的数据结构，在数据库[10]、分布式文件系统[11]、文本检索[12]等很多领域广泛使用。

2　背景知识

2.1　发布/订阅系统

发布/订阅系统(Pub/Sub)是一个分布式的消息路由系统，消息发布者(Publisher)发布信息后，消息通过系统路由给对消息感兴趣的消息订阅者(Subscriber)。与组播群通信的概念不同，Pub/Sub 不需要将组成员地址与组成员关系绑定，而是通过动态评估的方法决定组或者群的成员。通过图 1，我们定义发布订阅模型的通信协议如下：

1. 消息订阅者 S_1、S_2、S_3 根据各自规则 f_1、f_2、f_3、订阅系统中消息发布者产生的各类消息。

张继德　副教授，从事人工智能、高级数据库技术及计算机控制技术等教学研究工作；屈尔庆　工程师；贺志芳　助教。

2. 消息发布者 P_1 产生某消息 d 后，通过某一系统接入点将消息发布至系统，消息 d 满足 $f_1(d) \wedge \neg f_2(d) \wedge f_3(d)$，即 d 满足规则 f_1、f_3，但不满足规则 f_2。

3. 系统根据规则，将消息 d 与订阅者进行匹配，并将 d 传送至 S_1、S_3。

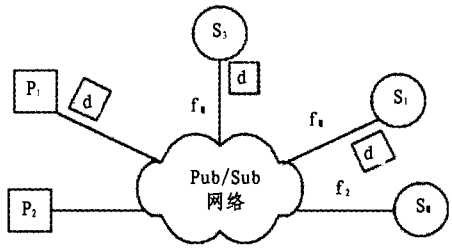

图 1　Pub/Sub 模型

2.2　Bloom filter

Bloom filter[13]使用长为 m 的比特向量 B_m 对集合 C_n 中 n 个元素的编码($m>n$)。我们称 B_m 为 C_n 的一个 Bloom 编码。编码方法首先选择包含 k 个独立的散列函数的散列集合 H_k，满足 $\forall C_i \in C_n, H_i(C_i) \in \{1,2,\cdots,m\}$，开始将 B_m 置零，然后对 C_n 中每个元素，用 k 个函数散列，得 k 个位置值，将 B_m 中这些位置上的比特值置 1，要检验一个元素是否属于集合，将其经过 k 个函数散列，得到 k 个值，若 B_m 中相应位置都是 1，则认为其属于集合。由于均匀的散列函数只能减少冲突，不能避免，因此存在某元素不属于集合而被错判为集合中元素的可能，其概率为 $\left(1-\left(1-\frac{1}{m}\right)^{kn}\right)^k \approx (1-e^{-\frac{kn}{m}})^k$。给定 m 与 n，最优 k 值为 $\ln 2 \times m / n$。应用时需权衡 m、n、k，可参看文[14]。

在本文中，定义运算符|表示按位或运算，∩表示连续|运算，& 表示按位与运算，则 Bloom filter 的基本运算表示如下

生成编码：

$$B_m = \text{Bloom}(C_n, H_k) = \bigcap_{i=0}^{n-1} \text{Bloom}(C_i, H_k)$$

检验元素 x 是否属于 C_n：

$$\text{Check}(B_m, x) = (B_m \& \text{Bloom}(x, H_k)) == \text{Bloom}(x, H_k)$$

3　平台设计

3.1　平台结构

本文提出的基于发布/订阅模型建立的平台结构见图 2。

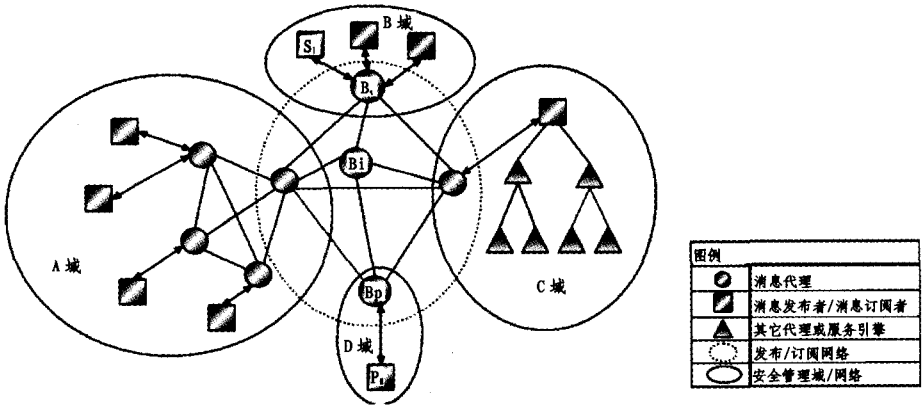

图 2　基于发布/订阅模型的安全管理平台结构

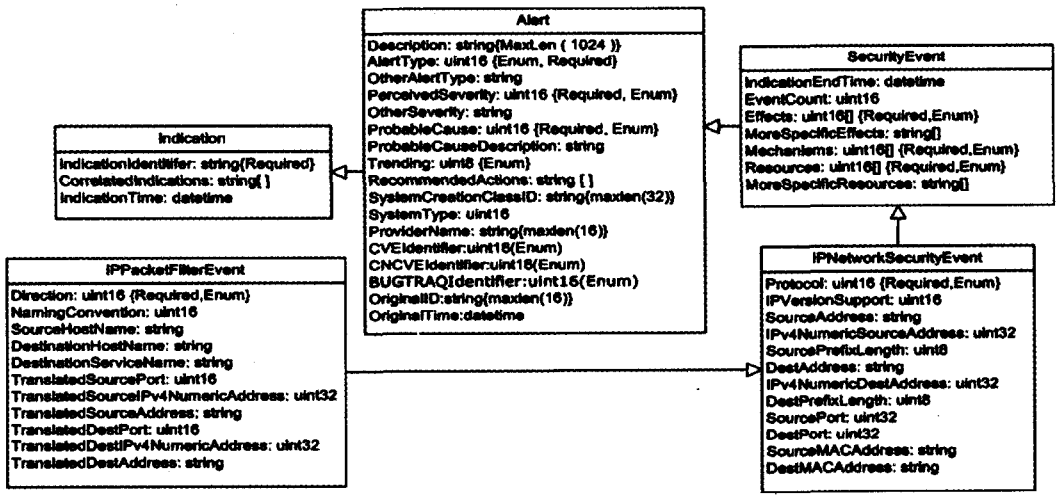

图 3　安全管理平台传输数据类图

平台根据划分为局部域和发布/订阅域两大类，局部域是基于安全管理域/网络的不同形成的，各局部域内部的网络拓扑结构可以完全不同，原有的安全管理平台结构可以全部保留，如 C 域，局部域也可以采用基于发布/订阅模式的分布式结构构建安全管理子平台，如 A 域。各局部域通过某个消息代理接入全局的发布/订阅域网络。各子域或子模块，首先作

为消息订阅者通过消息代理发布订阅规则，然后按模块间接口规则生成消息后，以消息发布者的身份将消息通过消息代理接入发布/订阅网络系统。消息通过订阅匹配，路由到相应模块的消息代理接入点，进而传递给订阅者上层服务模块。

平台结构通过匹配订阅规则而实现发布者与订阅者之间的松散耦合性解决了中心控制一分散采集型结构固有的协作能力弱的问题，避免了僵硬的层次体系结构对各安全设备间的所需的快速、灵活的信息交换的制约，同时减少了集中化处理的单点失效以及瓶颈约束，降低系统对计算处理设备的性能要求。

3.2 数据结构

为了便于设备间协同，使其兼容不同安全设备所产生的不同安全事件记录，需要定义平台间协作交换的安全事件的数据结构，其次，对于要记录字段的取值进行范式化，为计算和统计监控创造有利条件，同时减少存储空间。本文参考分布式管理任务组(DMTF)提出的公共信息模型(CIM)[15]设计数据类，类图见图3，通过多次继承实现敏感数据与通用数据的多层分隔，枚举类型属性的取值范围即是对所有可能取值的范化映射。数据格式表现方法采用 Map 方法[8]，即每个数据由多个“属性=值”组成的集合构成。

4 传输协议

4.1 数据包格式与路由匹配

协议数据包由路由数据与内容数据两部分组成。路由数据由 Bloom Filter 选择的 m 位的比特串 Bm 和校验信息组成。内容数据由数据包类型，数据内容和校验信息所组成。路由代理在收到数据包时，根据存储的路由表数据 m 位比特向量 Vm 与路由数据 Bm 进行 & 计算，当返回结果为 Vm 时，则此数据包按照 Vm 对应路径传输。

4.2 协议内容

以图2为例，消息发布者 $P1$ 通过代理 Bp 接入网络，消息订阅者 $S1$ 通过代理 Bs 接入网络，代理 Bi 连接 Bp，Bs。传输过程中的加减密运算采用下述表示方法：将明文数据 m 通过对称加密算法(如 AES，DES)，以 key 为密钥得到密文 c 的加密过程用 $c=E_{key}(m)$ 表示，相应的解密过程为 $m=D_{key}(c)$。将明文数据 m 通过非对称加密算法(如 RSA，ECC)，以公钥 PK 加密成密文 c 的过程定义为 $c=Enc_{PK}(m)$，用私钥 SK 解密定义为 $m=Dec_{SK}(c)$。下面结合图5说明整个发布订阅流程的传输协议。

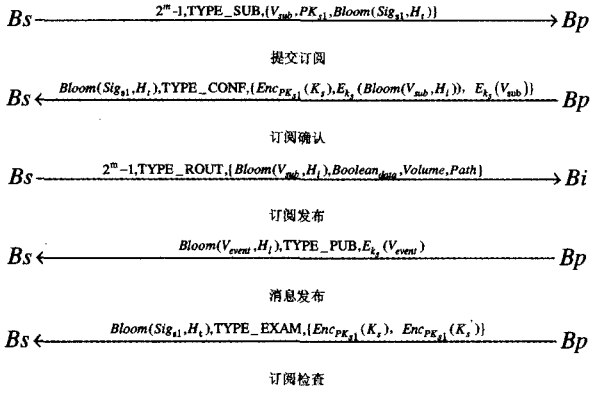

图5 协议内容

提交订阅：订阅者 $S1$ 通过 Bs 向网络广播订阅请求，广播路由地址 $V_{broadcast}=2^m-1$，即向量 V 所有位设为1，数据包类型为 TYPE_SUB，内容信息中包含订阅条件的集合 V_{sub}，$S1$ 的公钥 PK_{s1}，以及对 $S1$ 签名 Sig_{s1} 用 H_t 散列函数组生成的 m 位 Bloom 向量 B_{s1}，以此作为 $S1$ 的唯一地址。

订阅确认：消息发布者 $P1$ 收到订阅请求后，对不同信任度名单列表 TrustList 进行检验运算，如果 Check(Bloom(TrustList$_i$，H_t)，Sig_{s1})= true，则该 Sig_{s1} 的对应的 $S1$ 的信任度为 TrustList$_i$ 对应的 TrustLevel$_i$，将 TrustLevel$_i$ 的密钥 K_s 用 $S1$ 的公钥 PK_{s1} 以非对称加密算法加密得到 Ks 的密文 $C_{ks}=Enc_{pks1}(K_s)$，同时，将 V_{sub} 用 H_l 散列函数组生成 m 位 Bloom 编码 B_m，以此作为以后针对订阅条件 V_{sub} 匹配项。同时用对称加密算法将此订阅匹配项和订阅条件 V_{sub} 加密，得到的结果 $C_{Bm}=E_{ks}(B_m)$，$C_{Vsub}=E_{ks}(V_{sub})$ 与对应 K_s 的密文一起作为内容数据向地址 B_{s1} 发送类型为 TYPE_CONF 的确认数据包。在此步骤时，$P1$ 通过 $S1$ 的签名信息，也可以进行统计和审计信息的处理。

订阅发布：Bs 收到订阅确认后，首先用私钥 SK_{s1} 解密 $Des_{SKs1}(Cks)$，获得 Ks，然后用 Ks 解密 $D_{ks}(C_{Bm})=Bm$，$D_{ks}(V_{sub})=V_{sub}$。得到订阅条件和相应的匹配项。Bs 然后将 Bm 以及其它代理建立路由表时所需的路径信息 path，和一个原文订阅开关 $Boolean_{data}$，一个流量阈值 Volume，发送给每个代理，数据类型为 TYPE_ROUT。$Boolean_{data}$ 通常设为开启状态。

消息发布：$P1$ 在信息发布前，将订阅条件可选择属性对应的值形成集合 V_{event}，然后用 H_l 散列函数组将 V_{event} 编码生成 m 位 Bloom 编码 B_{event}，然后将发布消息中的属性值按照对外信任度的不同，用不同的信任度密钥 K_s 进行对称加密，最后以 B_{event} 为路由地址，$TYPE\_PUB$ 为数据包类型，通过 Bp 向系统发送。

订阅检查：$P1$ 通过 Bp 将新旧密钥以 $S1$ 的公钥 PK_{s1} 加密后，向 $S1$ 的唯一地址 B_{s1} 发送数据包类型为 TYPE_EXAM 的数据包。$S1$ 接受到新密钥后，还需再次发布订阅，使代理重置路由规则。

5 协议分析验证

5.1 安全性分析

本文在密钥发布以及用户身份认证上引入基于非对称加密的公钥体系(PKI)，PKI 的安全性已经通过理论和实践的证明，本文不再详细阐述，下面是从隐私保密角度对协议的安全性进行分析：

本协议体系中，密钥的安全强度由加密算法决定，同时通过订阅检查，密钥更新机制抵御密钥的丢失或者泄漏。在有限隐私共享的假设下，即使从路由代理处对数据包恶意解析，则其最多只能解密其信任等级内的信息，而这可以通过正常的订阅渠道获取，失去了破解的意义。

本协议路由匹配规则只是基于两个比特向量的比较，不存在内容解析而造成隐私破坏，订阅条件和订阅匹配向量之间的匹配只有发布者和订阅者知道，由于散列函数的单向特性，订阅者无法由获取具体的散列函数集合。如果订阅者恶意订阅大量规则，试图通过统计分析的方法获取匹配规则与订阅条件之间的联系，散列函数集合可以由消息发布者随机选择，以此破坏不同订阅规则之间的联系。由此可见系统的隐私保护是较为严密的。

对于目前流行的 DDOS 攻击，体系提供的方案是当代理接受数据包速度高于一定阈值 Volume 时，代理可以不转发

超过路由规则中阈值小于 Volume 的数据包，提前提前预警与处理。同时通过订阅检查，清除超时没有重置的订阅规则，防止恶意增加路由路径导致网络带宽消耗的拒绝服务攻击。

协议对于伪造，篡改和重发消息内容类型的攻击防范尚不严密，数字签名或时间戳的方法会因服务器资源耗费大造成拒绝服务攻击。

5.2 测试验证

为验证协议的可行性进行模拟测试，模拟测试的环境包含 5 个安全数据采集模块作为消息发布者，2 个计算模块作为数据采集模块信息的订阅者和显示模块的消息发布者，分别部署在 7 台 P4 2.2G，内存 256MB 的 linux 平台服务器上，Bloom filter 算法中的参数选取为 $m=140, n=8, k=12$，采用选定散列函数组，采用预先映射属性值与对应的 Bloom 编码，实际 Bloom 编码作|运算。在每秒 100 条安全事件的流量下的情况下，CPU 占用率为 13%～18%，当流量增大时，CPU 占用率增长不明显，只有内存占用增加，这是可以接受的。如果全部原文订阅开关设置为关闭状态，则当接收流量超过 51000 条/秒时，带宽占用才达到 1M。由于 Bloom filter 本身的出错概率限制，实际发送记录与接受记录相比大概在 1:004:1，在订阅开关关闭的情况下，这影响到了部分计算的准确率，在实际应用时需要在准确率和带宽占用之间做权衡。

总体来说，模拟试验从性能上验证了本文提出协议的可行性。

结束语 本文提出了一种适合大规模分布式的安全管理平台体系，其消息传输路由协议在其他应用场合也有着良好的应用前景，在一定程度上是一个泛应用平台。同时，作为一个安全管理平台，其所具备的安全、隐私保护特性、良好的可扩展性和动态适应性，是对于现有平台体系的一个革新。

参 考 文 献

1 Khurana H, Koleva R. Scalable security and accounting services for content-based publish/subscribe systems. In: proceedings of the 2005 ACM symposium on Applied computing, 2005. 801～807
2 ShamirA . How to share a secret[J]. Communications of the ACM, 1979, 22(11): 612～613
3 李慧贤，程春田，庞辽军. 一般访问结构上的多秘密共享方案. 华南理工大学学报(自然科学版)，2006，34(6)：95～98
4 穆成坡，黄厚宽，田盛丰. 入侵检测系统报警信息聚合与关联技术研究综述. 计算机研究与发展，2006(1)：1～8
5 张慧敏，钱亦萍，郑庆华，董世杰，管晓宏. 集成化网络安全监控平台的研究与实现. 通信学报，2003(7)：155～163
6 Locasto M E, Parekh J J, Keromytis A D, Stolfo S J. Towards Collaborative Security and P2P Intrusion Detection. In: proceedings of the 2005 IEEE Workshop on Information Assurance and Security, 2005. 30～35
7 张志伟，郭长国，曹贺锋，王伟球，王睿. 基于发布/订阅通信的动态数据集成模型. 计算机科学，2005，25(31)：124～126
8 马建刚，黄涛，汪锦岭，徐肛，叶丹. 面向大规模分布式计算发布订阅系统核心技术. 软件学报，2006，17(1)：134～147
9 Fiege L, Zeidler A, Buchmann A, Kehr R K, Mhl G. Security Aspects in Publish/Subscribe Systems. In: proceedings of the Third International Workshop on Distributed Event-Based Systems (DEBS), 2004
10 Koloniari K, Pitoura E. Bloom filters for hierarchical data. In: Proc. of the 5th Int'l Workshop on Distributed Data and Structures (WDAS), 2003
11 Little MC, Speirs NA, Shrivastava SK. Using bloom filters to speed-up name lookup in distributed systems. The Computer Journal, 2002, 45(6): 645～652
12 Chin-Chen C, Tian-Fu L, Jyh-Jong L. Partition search filter and its performance analysis. Journal of Systems and Software, 1999, 47(1): 35～43
13 Bloom B. Space/time trade offs in hash coding with allowable errors. Communications of the ACM, 1970, 13(7): 1970. 422～426
14 肖明忠，代亚非. Bloom Filter 及其应用综述. 计算机科学，2004，31(4)：180～183
15 DMTF. CIM Specification. http://dmtf.org/spec/cims.html

(上接第 287 页)

```
{
   for (j=0;j<N/2-1;j++)
      for(k=0;k<16;k++)
      { 取浮点数的第 j 字节的第 15-k 位 c
        把 c 存到字符串的 B[j*16+k]中
      }
}
```

算法 2 二进制变成十进制浮点数

```
REAL BinToFloat(unsigned char B[])
{
   求符号 S
   if(B 是规格化数) /*规格化*/
   {
     求 Bias
     求阶码 E
     E=E-Bias
     求 EV=2^E
     计算尾数 Fv
   if((是扩展双精数且前导符为 1)||(不是扩展双精度数)) Fv=1
   +Fv;
   计算浮点数 result=s*Ev*Fv;
   }
   else /*计算非规格化数*/
   { 求出尾码表示的整数 F
     计算浮点数 result=S*F/2^k/*k 分别为 -149、-1074、
     -16445*/
     }
     return result;
}
```

4.3 结果分析

程序分别在 Turbo C 2.0(TC)和 Visual C++ 6.0(VC)环境中运行，在 Turbo C 中的运行结果与在 Visual C++中运行基本一致。由于 VC 的 int 型为 4 字节，因此，程序要稍作调整后才能正常运行。实验结果验证了 C 语言的 float 和 double 型符合 IEEE754 的格式，每种格式的精度和数据的取值范围、0、NaN、无穷数等都与表 1 总结的结果相同。

在 VC 中虽然可以使用 long double 型数据，但是经测试不是扩展双精度数，实质还是 double 数据类型。由于在文[5,6]等很多参考文献中没有对于 long double 存储格式的描述，所以根据文[5]的描述，经过反复试验，得到 3.3 节的 long double 格式。在解码时，除要计算出尾码对应的定点小数外，还要显式地加上前导位，即第 64 位的值，其它计算方法与 float 或 double 数据类型相同。

结论 本文深入剖析了 IEEE754 浮点数格式，以及浮点数在表示十进数时的精度、取值范围，以及对浮点数在运算时的舍入误差进行了分析。最后通过实验验证了 C 语言浮点数的格式、取值范围，对 long double 扩展双精度的长度、存储格式以及二进制转换成十进制的方法等进行了研究，对高级语言在浮点数科学计算中可能碰到的问题进行了深度探讨，对于编写高安全性和高可靠性的程序有一定的参考意义。

参 考 文 献

1 王书增. 计算机原理[M](第 2 版). 北京：电子工业出版社，2006，9
2 李敬章. 计算机原理与体系结构[M]. 广州：中山大学出版社，1998，2
3 谭浩强. C 语言程序设计. 北京：清华大学出版社，2000
4 郑莉，等. C++语言程序设计(第 3 版). 北京：清华大学出版社，2003，12
5 IEEE Standard for Binary Floating-Point Arithmetic. ANSI/IEEE Standard 754-1985. Institute of Electrical and Electronics Engineers, August 1985
6 Goldberg D. What Every Computer Scientist Should Know About Floating-Point Arithmetic [M]. Association for Computing Machinery. Issue of Computing Surveys, March 1991