翻译说明：

原著：《Multisensor Data Fusion for Next Generation Distributed Intrusion Detection Systems》

作者：**Tim Bass**

来源：https://www.researchgate.net/publication/2741235_Multisensor_Data_Fusion_for_Next_Generation_Distributed_Intrusion_Detection_Systems

翻译：derta

时间：2003-1-30

注释：红色为笔者认为重要的或较好的部分，蓝色为笔者认为翻译值得商榷的地方

---

# 应用于下一代分布式入侵检测系统的多传感器数据融合

## 摘要：

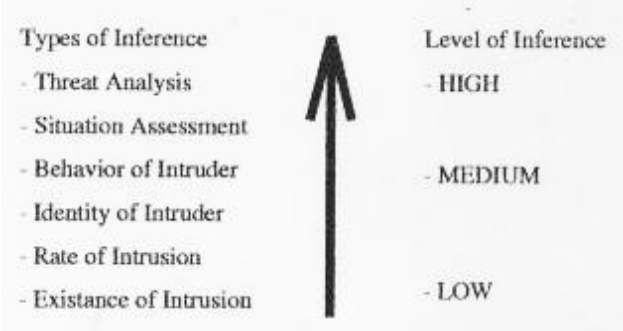
下一代计算机空间的入侵检测系统将从异质的分布式网络中的多传感器融合数据，以形成计算机空间的态势估计（*cyberspace situational awareness*）。本文初步提出了一些使用多传感器数据融合作为基层模型的工程要素；概括了当前基于Internet的入侵检测系统和基本的数据融合构架；使用TCP/IP模型开发传感器框架模型和数据库模型；推荐使用SNMP ASN.1 MIB结构表示依赖于内容的威胁和脆弱性数据库。

# 介绍

据估计安全评估工具的市场从1999年起以每年1.5亿美元的增长率增长，到2002年已经超过了6亿美元[1]。能源部（Department of Energy）近日召集网络安全专家就恶意代码、异常活动和入侵检测等领域向美国政府的R&D技术规划提供指导[2]。显然，快速增长的计算机空间入侵检测和态势评估市场面临严峻的技术挑战。

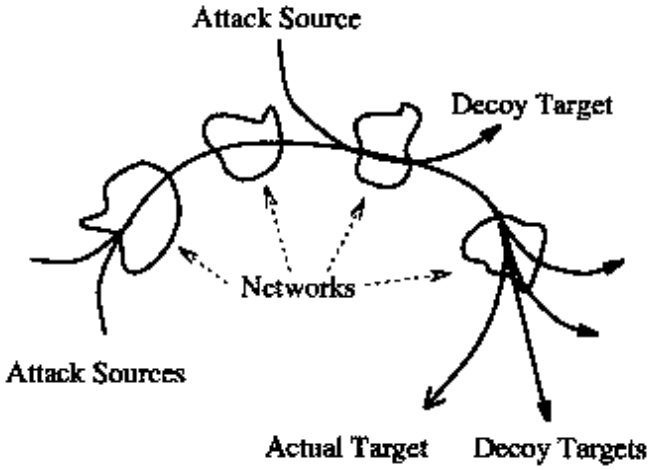
图1说明了态势评估推论的层次在计算机操作（cyberoperations）时既要支持作战管理又要支持网络管理[3]。商业和军事都需要对计算机空间的入侵检测和态势评估系统；在充满干扰的环境下精密的电子器件必须能够辨识对象，跟踪对象，计算速度，估计有计划的威胁。这都是需要一定技术的。

**Figure 1:** Hierarchy of IDS Data Fusion Inferences




网络安全技师认为当前的IDS在技术上还没有先进到足以抵御攻击的悄无声息（non-signature based cyberattacks）（下一章详述）。NATO的服务器遭受到来自于Serbian黑客的联合暴力攻击，他们使用了邮件炸弹和网管工具，通过消耗网络资源使服务器拒绝服务[4]。1997年，*The Langley Cyber Attack*的邮件炸弹事件在关键基础设施的保护方面向Marsh Commission证明了当前的IDS不能应对对重要计算机的软硬件严重威胁[5]。导致这方面不足的一个原因是IDS的误警率一直令人头疼。当技术资源不能部署到计算机系统或调查非入侵事件时安全资源被误导，误报警导致了严重的组织损失。实际上总是误警的系统对于用户的信心是毫无用处的。

**Figure 2:** Cyberattack with Multiple Sources & Targets




另一个原因是态势估计技术在我们的关键电子基础设施方面还只是刚刚起步。网络中心的指挥员没有可信的工具以辨识、跟踪和估计“i-objects”的多重攻击，例如图2。在计算机空间非对称冲突中的对手占有优势是因为没人统治并且有权利者只有很初级的态势知识；这就导致了今天的计算机空间的空隙或权力真空。

下一代的IDS需要从各种异质的分布的网络传感器融合数据。第3章概述了这些高层的IDS融合需求，这也在我们最近在ACM上发表的论文[3]上有所论述。隐含的课题和挑战绝不仅仅是入侵检测系统；网络管理也是非常耗费的基础构造。通常，这些系统不能给网络工程师提供具体的态势信息，而只是大量的系统信息和底层数据。下一代的网络管理和入侵检测系统将在统一的模型下交互，把数据融合成信息和知识；这样网络操作员就能够对计算机空间中他们自己那一块的系统健康和实时安全做出有根据的决策[3]。

本文为如何使用多传感器的数据融合提高高级计算机空间管理系统的性能和可靠性提供一个功能性的概貌，涉及系统设计，并建议进一步的研究和发展领域。另外，我们认为传统的诸如“网络管理”的概念应该推广到“基于计算机空间态势评估的融合”。

# 入侵检测系统概述

以前Internet入侵检测系统检查操作系统审计日志和网络通信[6][7]以保护重要信息基础设施的有效性、保密性和完整性。入侵检测系统保护重要信息基础设施以防DoS攻击、未授权的信息公开、数据的篡改或破坏。对这些事件的自动化的监测和及时报告需要对针对网络和主机的信息攻击做出响应。现在的入侵检测手段可概括为已知模式模板、威胁行为模板、通信分析、统计异常检测和基于状态的检测。这些系统还没有成熟到可靠的检测、核对、评估新的以网络为中心的攻击。

八十年代为了完善计算机安全措施引入了计算机入侵检测系统。IDS设计者常引用Denning[6]在1987年构建的入侵检测模型，它是基于主机的主体profile、系统对象、审计日志、异常记录和活动规则。一般的入侵检测结构是指基于规则的模式匹配系统；审计跟踪应对于主体profile以检测基于登陆、程序执行和文件存取的计算机误用行为[3]。

主体异常模型应用于设计许多基于主机的IDS，例如*Intrusion Detection Expert System*

(IDES)[8]、*Network Intrusion Detection Expert System* (NDIX) [10] and *Wisdom & Sense* (W&S), *Haystack*, *Network Anomaly Detection and Intrusion Reporter* (NADIR)[11]。还有其他的基于挖掘模型的IDS，有关它们的优秀的调研可以在[7]中找到。在这些系统中使用的基本的监测算法包括[3]:

使用带有权重的函数检测背离正常模式的差异
基于对正常使用状况的剖面分析的协方差矩阵
基于规则的专家系统检测安全事件

第二个当今先导的入侵检测方法是基于网络的多主机的。Heberlein *et al.* 把Denning的模型拓展为带有网络安全监视(*Network Security Monitor*)框架的基于以太网的流量分析[12]。这又被进一步发展为结合了基于主机的IDS和网络流量监视[7][9]的分布式入侵检测系统（DIDS）。当今的商用IDS，例如*Real Secure*和*Computer Misuse Detection System* (CMDS)，都具有分布式的结构，他们使用了基于规则的检测或统计异常的检测，或兼而有之。

IDS的设计者面临严峻的挑战：把从大量异质的分布式代理、系统管理器和数据库融合来的数据和威胁信息绘制成一幅内聚的画面，这幅画面可以用来显示和评估计算机空间的安全。首先，让我们回顾一下多传感器数据融合的基本概念，在[3]中有相关介绍。我们需要一个科学的方法来研究高可靠性的计算机空间IDS，它可以辨识、跟踪、评估带有多重复杂威胁的计算机空间形势。

# Internet IDS 数据融合

在一个标准的军用指令控制系统(C2)中，数据融合传感器用于观察电磁辐射、声波和热敏的能量、核物质粒子、红外线辐射、噪声和其他信号。在计算机空间IDS中传感器有所不同是因为环境尺度不同。不像导弹发射和超声波在空气中传播，计算机空间传感器是用于观察网络中的数据流。但就像C2系统指令官对弹头的来源、速度、威胁和目标感兴趣一样，网络安全人员对恶意入侵者和罪犯的身份识别、攻击频率、威胁和目标感兴趣[3]。

这些系统的输入由传感器数据、指令、和以前建立起来的数据库的数据组成。例如，系统的输入可以来源于大量的分布式的sniffer包、系统日志、SNMP的trap和query、用户profile数据库、系统message、威胁数据库和操作指令。在图3中已经说明了这一点。

基于数据融合的IDS的输出包括对威胁源的识别（及其可能的定位）、恶意的活动、威胁的分类、攻击的频率、对映射目标潜在的严重性的评估[3]。在[3]中我们建立起Waltz[13]来描述计算机空间的IDS多传感器融合系统的传感器一般特性；对于这些阐述的进一步论述将在第四章给出。

**检测性能（Detection Performance）**

是在给定的以网络为中心的噪声背景下对入侵特征进行检测的特征，例如误警率、检测概率和范围。当检测到恶意活动时，非恶意活动就构成了噪音。

**空域或时域的分辨率（Spatial/Temporal Resolution）**

是在空间或时间上对两个或两个以上的以网络为中心的对象区分的能力。在第四章中将谈到一些基本的属性。

**空间覆盖度（Spatial Coverage）**

是传感器视野的覆盖度的跨度（例如系统日志的跨度是计算机系统的例行程序和监视的系统调用）。

**检测/跟踪模式（Detection/Tracking Modes）**

是传感器运行的模式，例如扫描、单个或多计算机对象跟踪；或是具有多模式运行的能力。

**目标再访率（Target Revisit Rate）**

是传感器对一个计算机对象或事件执行测量的再访频率。

**测量精度（Measurement Accuracy）**

是计算机传感器测量和观察的精确性和可信度的统计概率。

**测量粒度（Measurement Dimensionality）**

是各网络对象分类之间的变量参数。

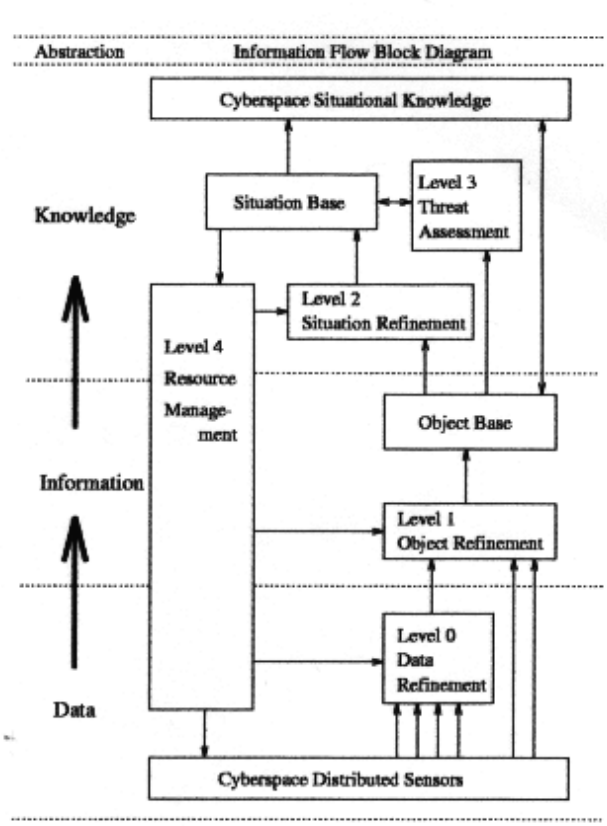
**软硬件数据报告（Hard vs Soft Data Reporting）**

是传感器报告的决策情况，例如，没有相关性可以下达一条指令决策？传感器需要验证？

**检测/跟踪报告（Detection/Tracking Reporting）**

是传感器的一种特性，用于报告个体的计算机事件或传感器对事件的时间序列的持续跟踪。

**Figure 3:** Intrusion Detection Data Fusion




在我们的模型中，态势数据通过网络传感器的初步观测的基元、标识符、次数和描述获得。原始数据需要校准过滤，参照图中的*Level 0 Refinement*。第1层的对象提取在时间（或空间）上相关联，其数据标以公制的权重。观测数据可以根据入侵检测基元(intrusion detection primitive)关联、配对、分类[3]。对象通过配位的行为、依赖、共同的源点、共同的协议、共同的目标、相关的攻击率或其他高层次的属性而被检测出，形成一个基于对象

的聚集的集合。对象在这样的对象基上的上下文中排列、关联、置位后，态势提取(*Situation Refinement*)就可以提供态势知识和识别。读者关心的其他高层次的属性可以参考[3]以便于在此功能框架下的进一步讨论。下面我们从传感器、第0层数据和第1层对象提取过程看起。

# 基于融合的入侵检测系统

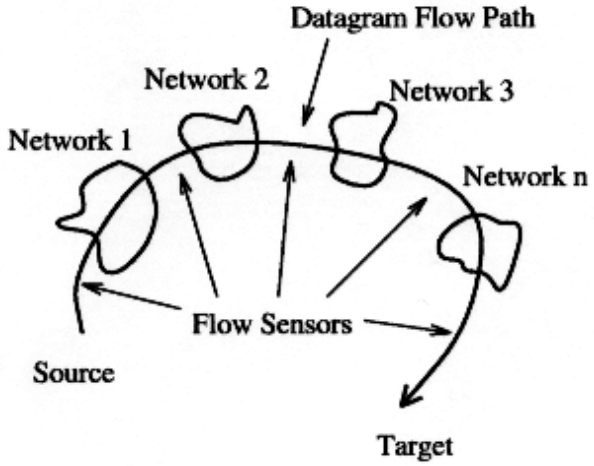
在谈及跟踪和辨识以网络为中心的对象的细节之前，我们提供一个穿越Internet对象的通用模型，如图4。基本的结构是IP数据报：它们在存储-提交(store-and-forward)的环境中从源到目的移动；路由基于不定的源地址确定的目的地址；经过每一个路由器数据报的生存期(TTL)衰减[19]。数据报途经大小网络、intranet、Internet交换式服务商。在本文中目的IP
地址的精确性的置信度与数据报的TTL一致。

在图4的Internet交互域传输模型 (interdomain transit) 中,交互域网关上的（或之间的）传感器辨识和跟踪数据流。在这里的讨论我们给出一个先验的临时的结构：计算机空间态势评估的瞬时清晰度 $\Gamma$(伽马)是数据报传输时间$D_t$与传感器融合推理时间$S_t$之比。伽马因子表达了方程1中基本的线性关系。它还提供了一种先验——数据报传输时间远远大于传感器融合时间。换句话说，跟踪辨识系统必须使感知、传输、处理、关联和对网络对象作出反应的时间快于对象到达目标的时间。

$$\Gamma = \frac{D_t}{S_t} \tag{1}$$

类似的，我们谈一下空间物体的跟踪，例如射弹。如果截获射弹的时间大于雷达跟踪系统和其他相关处理，那就不可能在导弹打击目标之前对其跟踪并做出反应。如果网络对象用30秒到达目标，那计算机空间的态势估计所需的决策融合处理必须小于30秒。

**Figure 4:** Internet Attack ID Sensor Placement

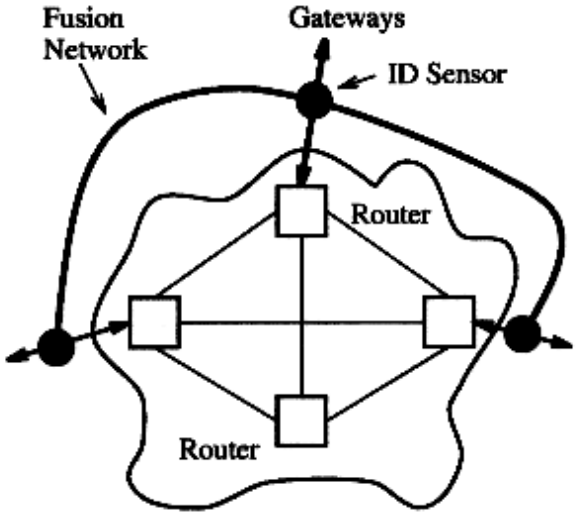



伽马因子对于未来的研究是值得注意的一个领域，因为其概念还需进一步探讨。在这里的讨论中，我们提供一些先验：

使用带宽的（或信道的）通讯不可能以数学的方式发展有效的计算空间态势估计系统（带有高效的空域和时域关联）。

这一结构表明图4中的传感器网络必须是带外(out-of-band)的，必须比被监督的网络快，在图5、6中概念性的作了说明。在这一结构中，一个带外的网络收集传感器信息并发布C2(command and control)指令给过滤器、防火墙和其他激活的网络设备。有了高比例的伽马因子就可以通过传感器（和可选的指令控制联结）获得高危态势估计。

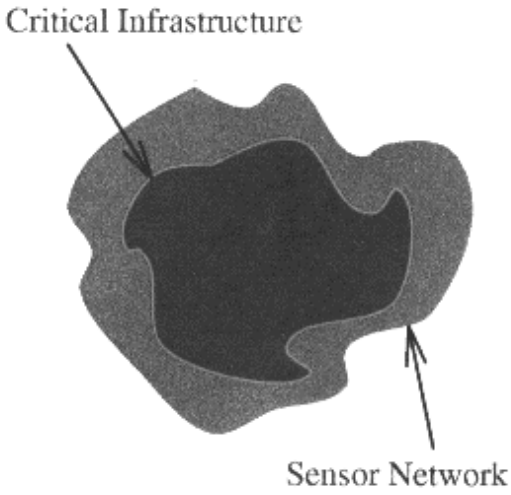
**Figure 5:** Gateway Sensors on CSA Fusion Network



# 传感器数据缩减和威胁对象

图4、5中繁忙的Internet网关处理的IP包数量可能非常大；从而，网关传感器获取并按比例向分析引擎和关联引擎提交大量数据。举个例子，在高速接口上一个路由器每秒处理100000个数据包，每个数据包记录14比特信息，这样没个传感器就产生了大约1.4 MBPS的数据。显然在以网络为中心的IP融合系统中的分布式传感器需要本地处理，如图7。因此，为了使中心融合处理和传输过载成本最小，传感器输出数据应该在传感器上就尽可能的缩减。本文不进一步讨论传感器数据的缩减、算法选择和$T_s$，这些留到其他的讨论或研究主题。

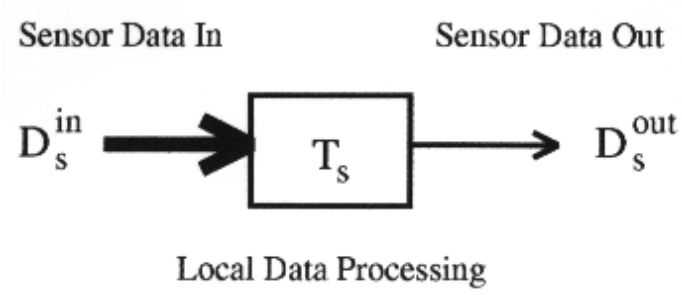
**Figure 6:** Critical Infrastructure - Sensor Network




传感器数据缩减率 (SDRR) $\Delta$ (delta) 是传感器输入$S_{in}$, 与输出 $S_{out}$之比。

$$\Delta = \frac{S_{in}}{S_{out}} \tag{2}$$

**Figure 7:** Sensor Data Reduction



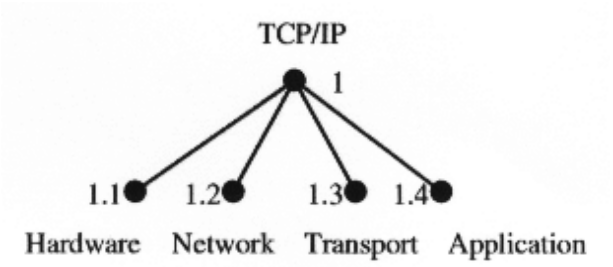

在本章剩下的讨论中我们关注传感器的输出$D^{out}_s$。我们以“shifting gears”来介绍并概述一个基于TCP/IP分类的威胁，这可以用作本地传感器处理和数据库需求框架设计。读者可以参考Antony在[17]中关于针对融合系统和态势知识结构的数据库需求的非常完整的讨论。知识即使陈述式的又是过程式的(*declarative* or *procedural*)。陈述式的知识是被动的知识或关系知识（例如文件）。过程式的知识陈述式的知识的一种特例，它以模式、算法和数学变换表示。一般认为陈述式的知识基的容量远大于过程式的知识基。

关系-实体图(*Entity-relationships*)是最基本的表达传感器数据库的陈述式模型。二进制数、家族树都是态势分析所需的基本的数据库关系的例子；大部分都可以以Sql语句表达：

**SELECT(attribute) FROM (table) WHERE (condition)**

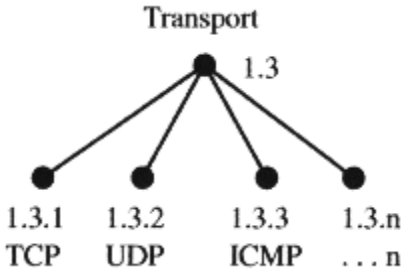
头脑里有了这样的基本的数据库模型和数据选择元语，本文提供了一个基于TCP/IP和SNMP MIB[18]管理的框架的TCP/IP威胁分类法，图8-10作了说明。SNMP MIB模型非常适于表达以网络为中心的威胁。

**Figure 8:** Example TCP/IP Threat OID



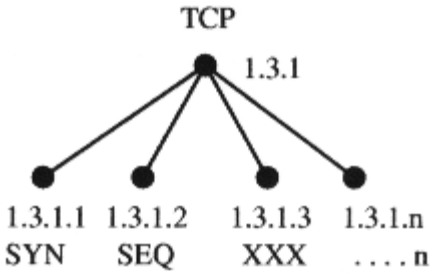

物理层的TCP/IP威胁是由一些自然灾害引起的服务中断，如大火、洪水、截断电缆、发射器故障和其它硬件失灵。在这一结构中我们关注IP网络层和TCP传输层。

**Figure 9:** Example IP Transport OID



在Inernet中主要有三种数据流（服务）：User Datagram Protocol (UDP), Transmission Control Protocol (TCP), and Internet Control Message Protocol (ICMP)[19]。域名服务器cache中毒和UDP端口洪泛拒绝服务攻击是使用UDP服务所暴露出的弱点的两个例子。Ping死和ICMP重定向炸弹是基于ICMP的Internet攻击的例子。TCP弱点攻击包括TCP序列号攻击和SYN洪水攻击，如图10所示。

**Figure 10:** Example TCP Transport OID




安全威胁可以用ASN.1 MIB表示法表示。例如，一个TCP SYN可以以下面的OID表示：

**tcpSYNFlood OBJECT IDENTIFIER ::= { iso 3.6.1.5.1.3.1.1 }**

另外*tcpSYNFlood* OID的一个子对象可以是源地址或目的地址带有恶意SYN的数据包：

**tcpSYNFlood.source OBJECT IDENTIFIER ::= { iso 3.6.1.5.1.3.1.1.1 }**

**tcpSYNFlood.dest OBJECT IDENTIFIER ::= { iso 3.6.1.5.1.3.1.1.1.2 }**

发展可扩展的TCP/IP安全MIB是发展Internet IDS融合系统重要的第一步。使用前面的模型框架作为一个起点，入侵检测本地处理的agent可以从管理系统控制。传感器信息应该在本地的传感器处理agent中以可行的、基于动态和静态的存储约束存储。

# 结束语

本文在[3]的基础上构建了多传感器的IDS框架，处理了特定的层次0和层次1的融合需求。建议SNMP形式的ASN.1 MIB模型作为交互式IDS的融合数据库模型。威胁数据库还需进一步讨论。无论怎样，这也只是发展计算机空间态势估计的分布式融合系统“漫漫长路”的起步。

身负使命的飞行员在敌对的环境下懂得如何以简练的规则应付目标。有了适当的定义好参数和位置指挥官就可以开火摧毁敌方的飞机。态势估计的信息越多指挥官越难决策[20]。

“当战斗空间是计算机时，攻击是活跃的、敌对的，在什么条件下信息员应该开火呢？”[20]

如今，当发生计算机攻击时网管人员手上没有适当的定义好的规则用于交战。没有高层次的计算机空间态势推论执行反击指令是鲁莽的，除非攻击源具有很大的可能性；但在

计算机空间敌对活动的源地址很容易伪装。因此，交战的反击信息规则的保真度是与计算机空间态势估计的程度和知识推理的质量直接成比例的。

本文只是在设置设计和发展计算机空间的态势估计系统的工程需求的过程中的几小步。在计算机空间中对动态的以网络为中心的对象进行识别和跟踪是多重计算机攻击的管理的核心技术能力。在这一复杂的基础构架中对以网络为中心的活动的使能和约束的辨识、跟踪、分类、评估可以使用多传感器数据融合作为一个模型。

本文中提出的模型和构架中的每一部分都还需要进一步发展。我们希望融合工程师和科学家已经发现本文中提及的研究发展领域既是饶有兴趣的又是出于一定动机的。我们还希望在一定程度上本文能对数据融合研究团体的最终目标有所贡献。

## 致谢

I am very appreciative to members of the USAF/SC for all of our stimulating network-centric discussions; especially Lt. Gen. Bill Donahue, Lt. Gen. Jack Woodward, Brig. Gen. Dale Meyerrose, Brig. Gen. (select) Bernie Skoch and Lt. Col. Dave Gruber. Also, I would like to thank Ed Waltz for his inspirational life work on multisensor data fusion and information warfare. .

## 参考目录

**1**

Silverman, R.,
*Intrusion Detection Systems Sniff Out Digital Attack*,
The Wall Street Journal, pg. B6, February 4, 1999.

**2**

Schultz, G., Chairman,
*Detection of Malicious Code, Intrusions, and Anomalous Activity Workshop*,
Department of Energy, National Security Council & Office of Science and Technology Policy,
February 22-23, 1999.

**3**

Bass, T.,
*Intrusion Detection Systems and Multisensor Data Fusion: Creating Cyberspace Situational Awareness*,
Communications of the ACM (to appear), 1999.

**4**

de Bony, E.,
*NATO reinforces against Net attack from Serbs*,
InfoWorld Electric,
Posted at 9:40 AM PT, Apr 2, 1999.

**5**

Bass, T., Freyre, A., Gruber, D. and Watt., G.,
*E-Mail Bombs and Countermeasures: Cyber Attacks on Availability and Brand Integrity*,
IEEE Network, pp. 10-17, Vol. 12, No. 2., March/April 1998.

**6**
Denning, D.,
*An Intrusion-Detection Model*,
IEEE Transactions on Software Engineering,
Vol. SE-13, No. 2, pp. 222-232, February 1987.

**7**
Mukherjee, ., Heberlein, L., and Levitt, K.,
*Network Intrusion Detection*,
IEEE Network Magazine, Vol. 8. No. 3,
pp. 26-41, May/June 1994.

**8**
Denning, D. et al.,
*A Prototype IDES: A Real Time Intrusion Detection Expert System*,
Computer Science Laboratory, SRI International,
August 1987.

**9**
Snapp. S. et al.,
*A System for Distributed Intrusion Detection*,
Proceedings of IEEE COMPCON,
pp. 170-176, March 1991.

**10**
Bauer, D. and Koblentz, M.,
*NDIX - An Expert System for Real-Time Network Intrusion Detection*,
Proceedings of the IEEE Computer Networking Symposium,
pp. 98-106, April 1988.

**11**
Hochberg, et al.,
*NADIR: An Automated System for Detecting Network Intrusion and Misuse*,
Computers & Security, Elsevier Science Publishers,
pp. 235-248, 1993.

**12**
Heberlein, L. et al.,
*A Network Security Monitor*,
Proceedings of the IEEE Computer Society Symposium,
Research in Security and Privacy,
pp. 296-303, May 1990.

**13**
Waltz, E. and Llinas, J.,
*Multisensor Data Fusion*,
Artech House, Boston, MA, 1990.

**14**
Waltz, E.,
*Information Warfare Principles and Operations*,
Artech House, Boston, MA, 1998.

**15**
Hall, D.,
*Mathematical Techniques in Multisensor Data Fusion*,
Artech House, Boston, MA, 1992.

**16**
Varshney, P.,
*Distributed Detection and Data Fusion*,
Springer-Verlag, New York, NY, 1996.

**17**
Antony, R.,
*Principles of Data Fusion Automation*,
Artech House, Boston, MA, 1995.

**18**
Rose, M.,
*The Simple Book*,
Prentice-Hall, Englewood Cliffs, NJ, 1994.

**19**
Stevens, R.,
*TCP/IP Illustrated, Volume 1: The Protocols*,

Addison-Wesley, Reading, MA, 1994.

**20**

Bass, T.,
*Cyberspace Situational Awareness and Cyber Rules f Engagement*,
Silk Road, December 8, 1998.

**21**

Graham, B.,
*Cyberwar: A New Weapon Awaits a Set of Rules*,
The Washington Post, pp. A1, A10, July 8, 1998.

# 关于本文...

**Multisensor Data Fusion for Next Generation Distributed Intrusion Detection Systems**

Invited Paper:

*1999 IRIS National Symposium on Sensor and Data Fusion*
The Johns Hopkins University Applied Physics Laboratory

24-27 May 1999