DBAPPSecurity 安恒信息

WEB应用安全和数据库安全的领航者

基于模糊粗糙集方法与威胁情报工作流的技术分享

刘志乐

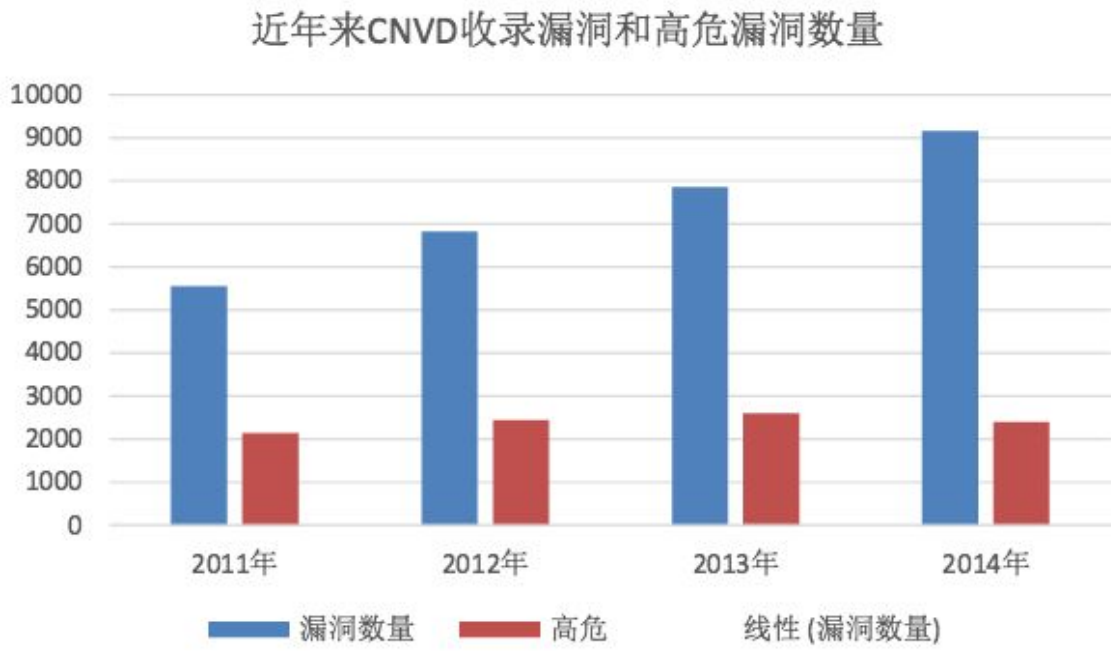
www.dbappsecurity.com.cn

内 容 提 要

- 安全分析的现状和趋势
- 基于模糊粗糙集的态势感知
- 基于工作流的威胁情报分析

1 安全分析的现状和趋势

网络安全发展趋势



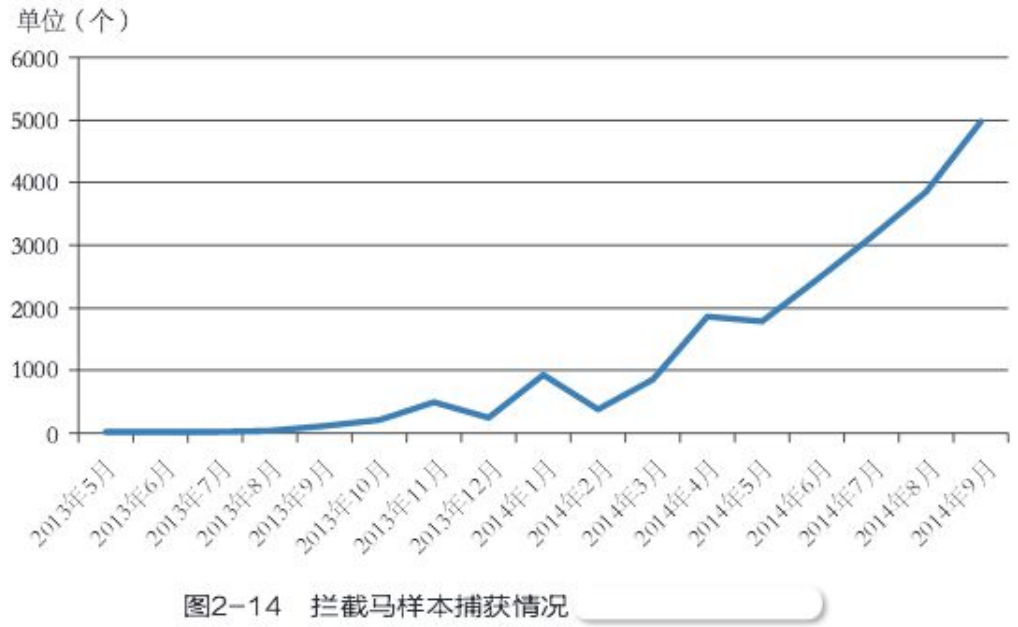
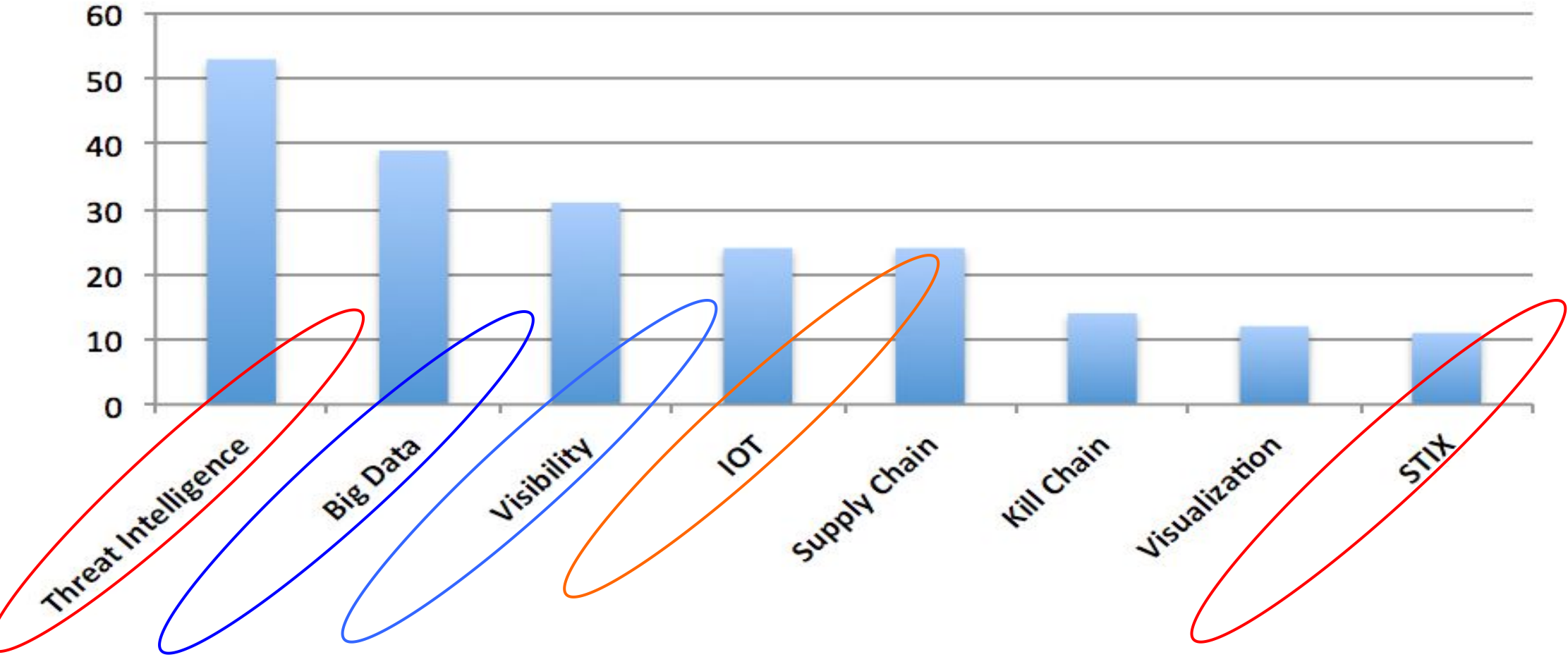


图2-14 拦截马样本捕获情况

近年来的漏洞数量和木马样本走势
（来源：CNCERT/CC）

RSAC话题热度



STIX支持厂商（产品）达到40个，用户社区10个

创新沙盒的发展趋势

1. 大数据安全分析是初创公司扎堆的热点领域

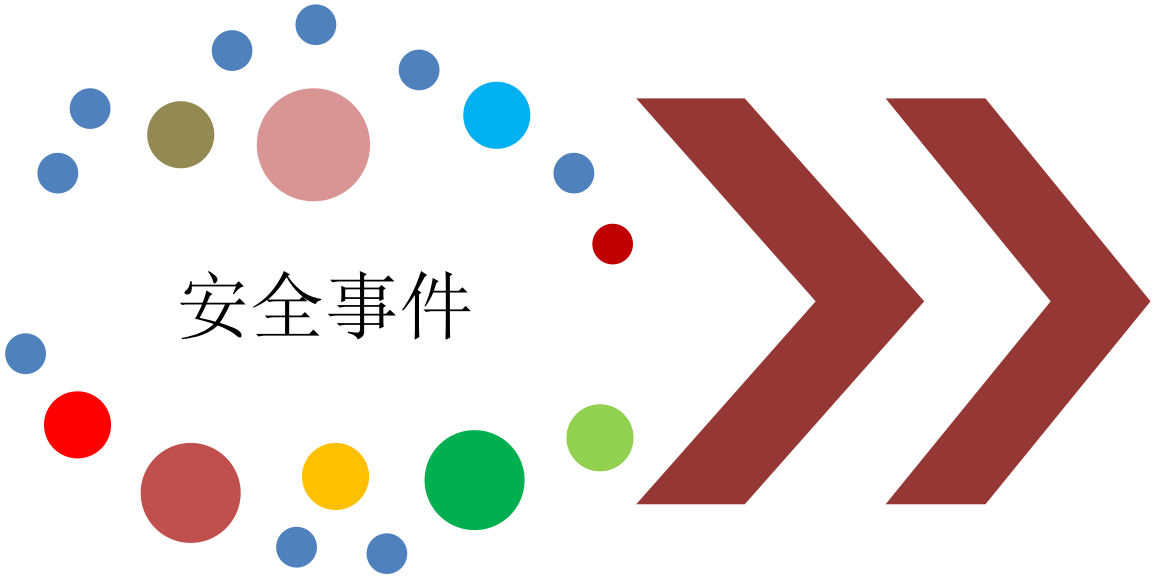
2. 基于机器学习的可疑行为检测技术成为研究焦点

3. 从客户环境中采集数据、汇总到云端提炼安全情报，再共享到客户环境中，成为安全情报产生和使用的重要情景

4. 以轻量级代理 + 弹性计算平台的方式提供安全能力，成 为初创公司青睐的业务模式，硬件盒子出现的越来越少

5. 众包模式作为一种新兴的生产组织形式，开始出现在网络安全服务中

网络安全发展趋势



- 速度越来越快
- 种类越来越多
- 数据量越来越大

海量数据

- 安全漏洞
- 攻击事件
- 状态日志
- 网络行为
- ……

2 基于模糊粗糙集的态势感知

粗糙集理论简述

粗糙集理论最初是由波兰科学家Pawlak提出的，是处理不确定、不完备和模糊信息的有力工具。

粗糙集理论建立在用等价关系（满足自反，对称，传递三个性质的关系，比如a和a本身是等价的，即自反，a和b等价，推出b和a等价，即对称，a和b等价，b和c等价，推出a和c等价，即传递）对全域（即所有元素的集合）进行划分（等价的分为一个集合，所有元素被分为多个集合）的基础上，可以定义一个集合的上近似和下近似，除数据集外不需要任何先验知识。

应用：
对决策信息系统进行属性约简

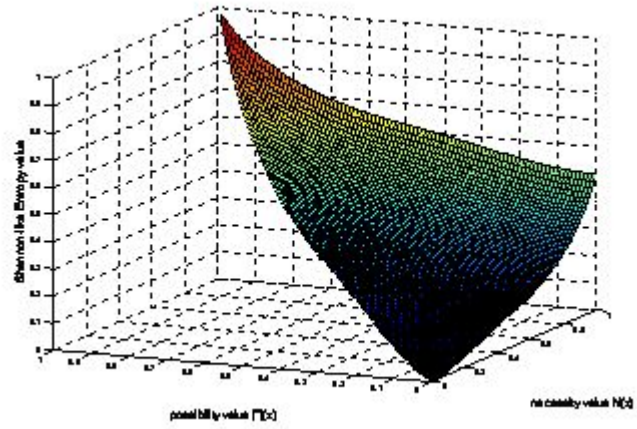
例：
态势因子和态势等级构成的表

模糊粗糙集简述

模糊粗糙集方法是粗糙集的扩展方法。
模糊集是用来表达模糊概念的集合。

对于一般的集合，一个元素只有属于这个集合和不属于这个集合两种情况。
然而在现实中，常常面对无法明确划分的情况，为描述这种情况，Zadeh提出了模糊集的概念，对传统集合的隶属关系进行了推广。

用0~1之间的一个数来表示元素与集合之间的隶属程度，当隶属度被限定为0和1时，模糊集就是一个传统集合。



态势感知的数学本质

- 态势理解
- ★ 态势评估
- 态势预测
- 态势可视化

网络安全态势感知的核心环节
数据融合领域的研究重点

即在融合各安全信息并进行简单处理的基础上，通过一些数学方法或者数学模型，经过分析，得到一个对当前网络安全状态的整体描述。

简言之，该过程即态势因子集合到态势集合的映射。

网络态势评估方法分类

基于数学模型的方法	层次分析法 集对分析法 ……
基于知识推理的方法	基于图模型的方法 基于证据理论的方法 ……
基于模式识别的方法	灰关联分析方法 粗糙集合方法 神经网络方法 ……

态势因子的定义

DBAPPSecurity 安恒信息

★ **态势评估**

粗糙集方法
(对应离散型的态势因子)

如攻击类型，只关心是哪一种攻击类型，
而攻击类型也没有谁大谁小之说
与其相对的：
连续型态势因子，如网络流量有大小，且是连续的

态势因子：
指能引起网络态势变化的因素，比如可通过处理监测数据、日志等原始数据生成态势因子。

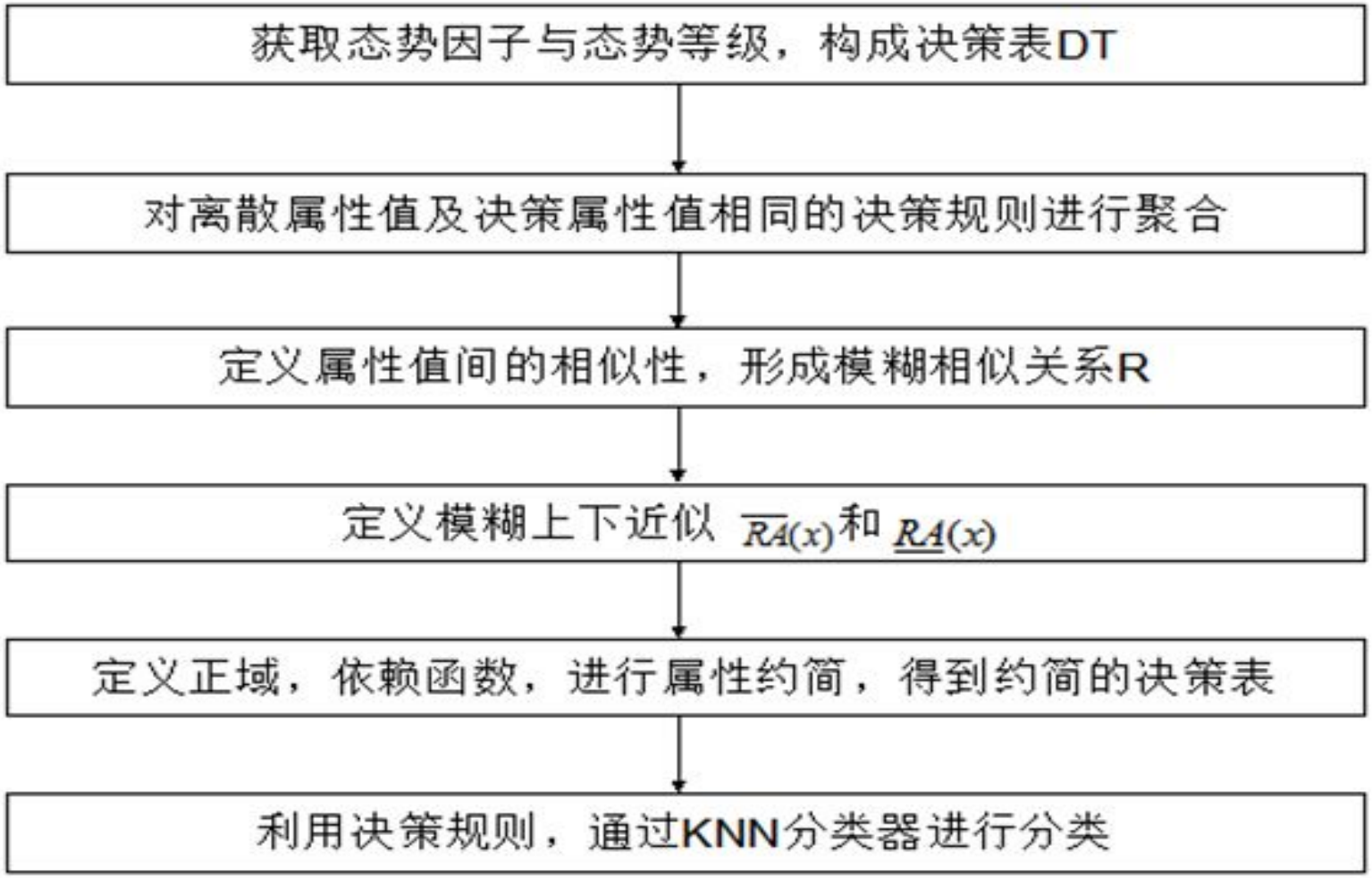
离散型态势因子：
即态势因子的取值是离散的，比如攻击者IP地址，攻击类型等，属于不可比较大小或者比较大小无意义的类型，且取值不连续

连续型态势因子的离散处理

例： 0-9 损失精度 0-3, 3-6, 6-9

原因：离散化处理后，例如7与8会被认为属于同一个部分

模型建立过程


获取态势因子与态势等级，构成决策表DT
对离散属性值及决策属性值相同的决策规则进行聚合
定义属性值间的相似性，形成模糊相似关系R
定义模糊上下近似 $\overline{RA}(x)$和 $\underline{RA}(x)$
定义正域，依赖函数，进行属性约简，得到约简的决策表
利用决策规则，通过KNN分类器进行分类

模糊粗糙集方法优越性证明

- 根据KDD Cup 99数据进行对比参照：

表 3 测试结果

类型 处理方法	Naive	Entropy	CACC	NCAIC	EF	Our approach
DoS 攻击	99.9652	99.9817	99.9670	99.9579	99.9780	**99.9943**
所有攻击	99.6918	99.7368	99.7489	99.7403	99.7160	**99.7806**

使用四大类攻击类型进行测试，模糊粗糙集方法在Dos攻击上的结果以及总的结果上，相比于几种离散化方法，精度都要高。

网络流量异常评估

取得基础流量数据

41个条件属性类型：持续时间、协议、源、目的……
5个决策属性：正常+4个异常类型

41个条件规则聚合

聚合得到新的规则决策表

依据依赖关系，得到23个约简属性

持续时间、服务类型、状态、源、目的……

约简得到23个属性的决策表

依据该决策表评估流量异常情况状态

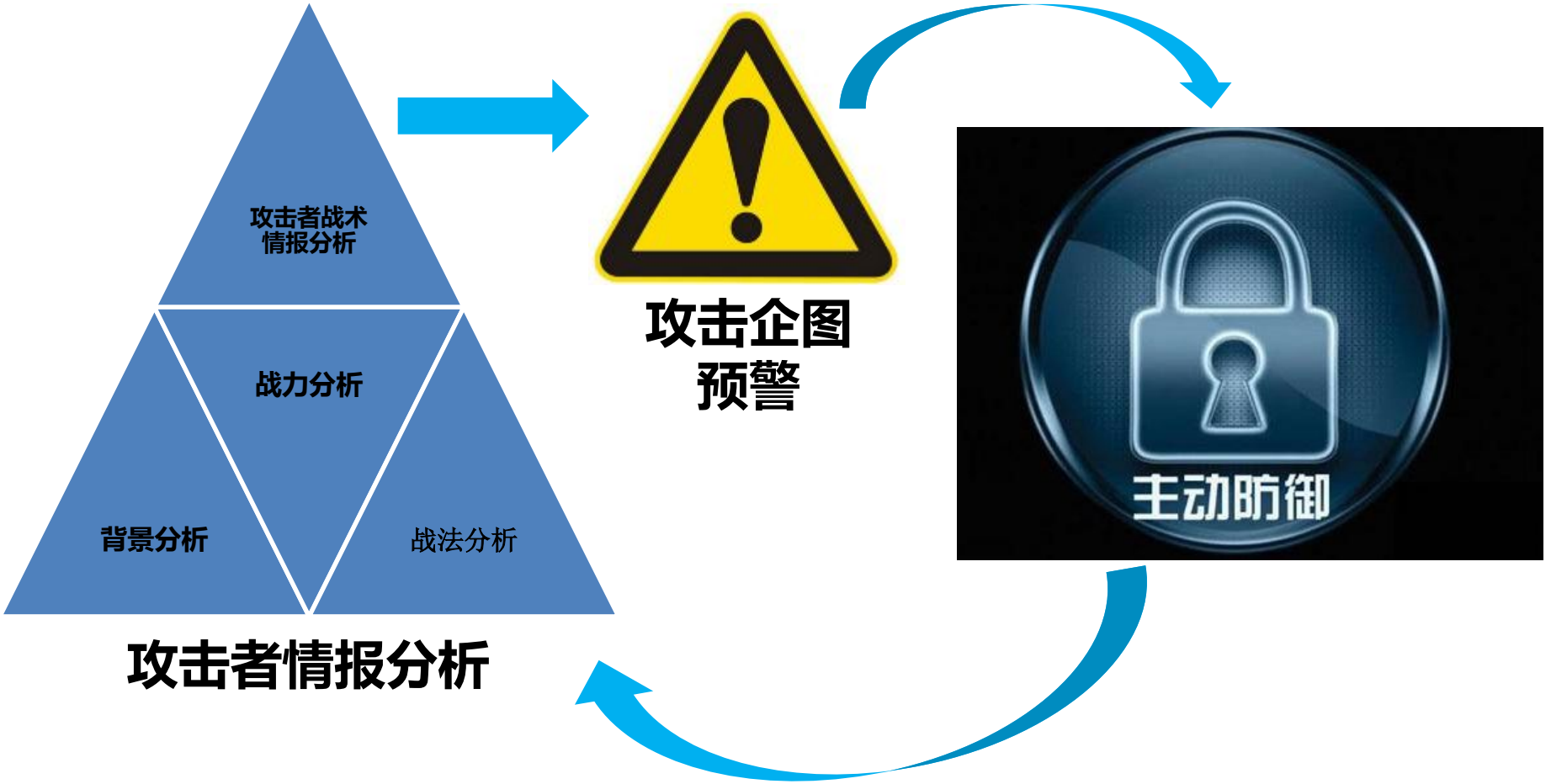
网络流量评估

对象	持续时间	服务类型	状态	来包字节	回包字节		决策状态
P1	2	Smtp,	SF	[900，1684]	[363,10 02]	……	0
P2	0	Private	REJ	0	0	……	1
P3	0	smtp	SF	[100,787]	[10,329]	……	1
Pn	1	smtp	SF	1600	213	……	?

Pn 的决策状态因为可能有多条规则匹配，它们的决策可能不一致，
我们用最多的来代表最后的决策。

3 基于工作流的威胁情报分析

以威胁情报分析作为主动防御基础



分布式架构扩展感知与防御面

- ✓ SaaS，用户上传，分析共享数据和服务
- ✓ 安全社交网络，共享数据、服务和工作流
- ✓ 云计算资源共享
- ✓ 基于Hadoop和网格技术

分布式主动安全情报平台建设范例

全国32个计算分节点，日均处理数亿条安全数据

基于情报分析的工作流

安全可视化

安全情报分析

安全情报处置

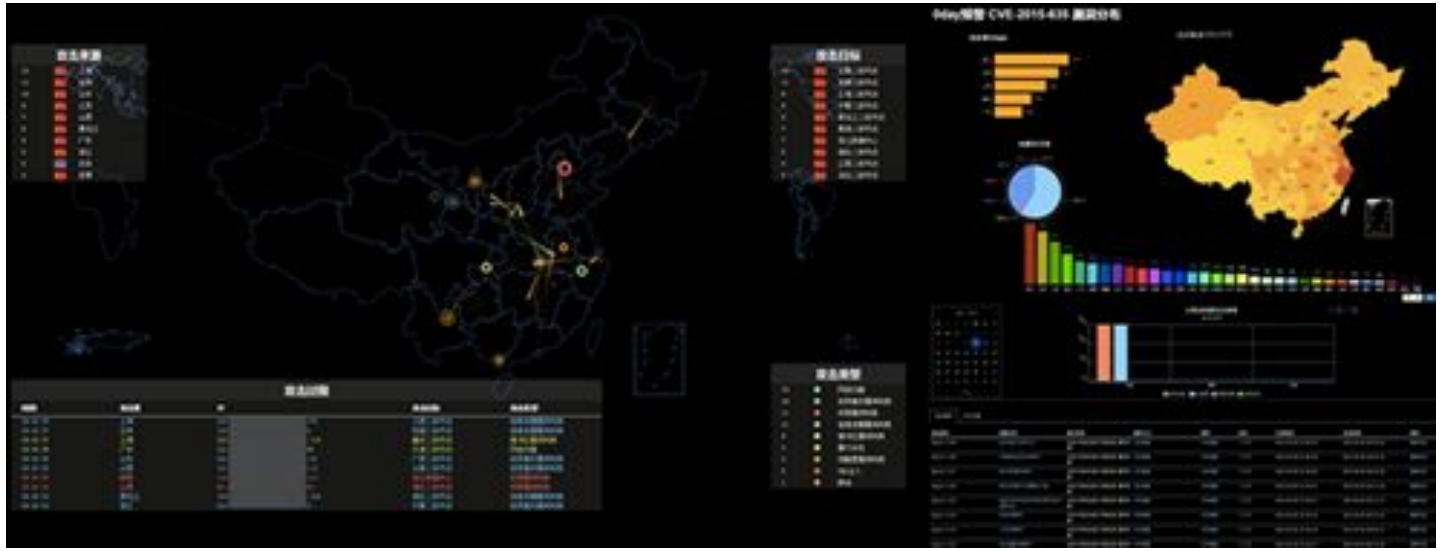
辅助分析

决策基础

安全可视化分析

数据多视角展示：

- ✓ 发现数据模式
- ✓ 模拟、预测、测试假设
- ✓ 信息抽取

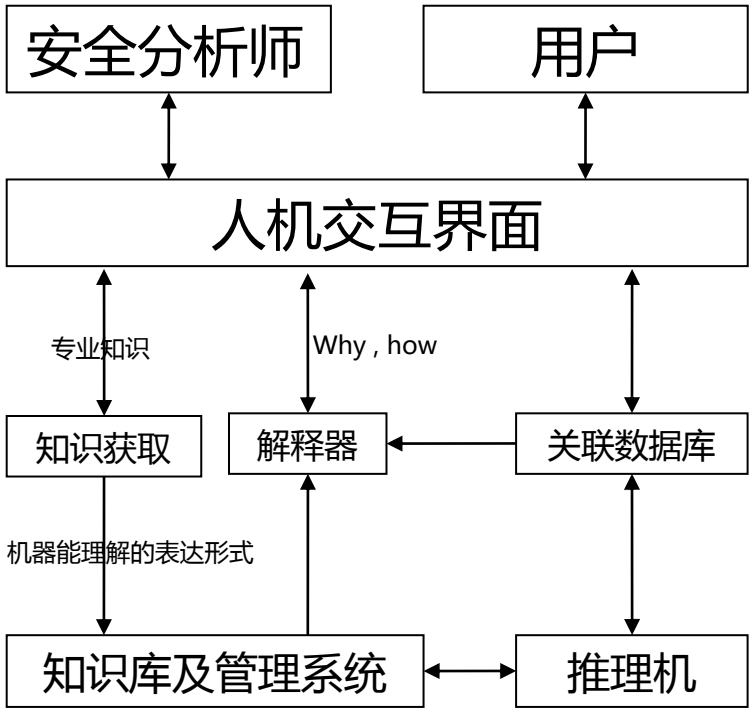


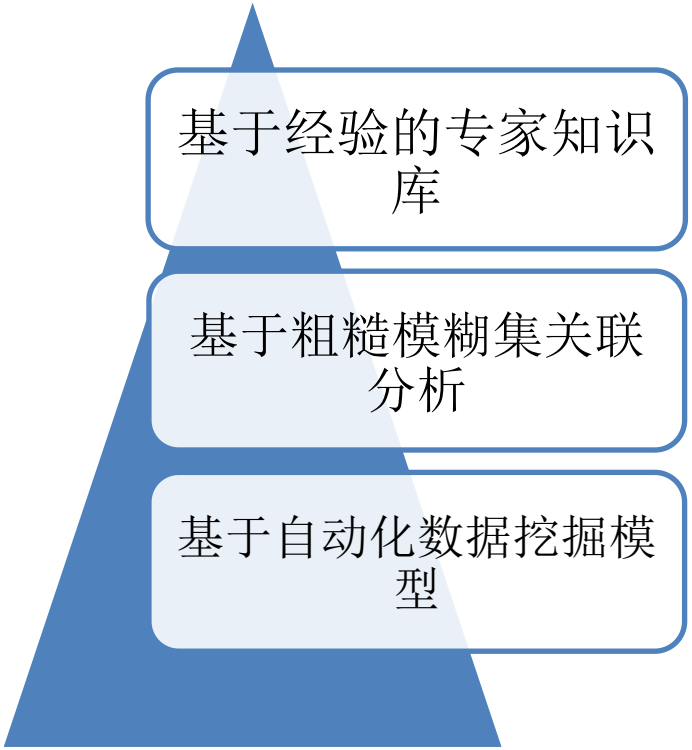
可视化与挖掘一体化：

- ✓ 黑箱整合：通过参数将用户与模型连接
- ✓ 白箱整合：用户参与算法中间步骤，决定如何构建模型
- ✓ 灰箱整合：专家协助或提供模型构建建议，用户与专家共同完成分析

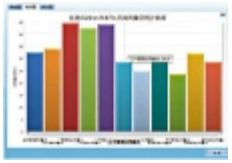
安全可视化的目的并非是为了单纯展示，而应以可视化辅助数据挖掘与情报分析

基于安全分析师的智能决策





协作情报分析



业务应用

定制应用

威胁感知设备

基于云的安全情报协同分析系统

安全分析师

自动化系统

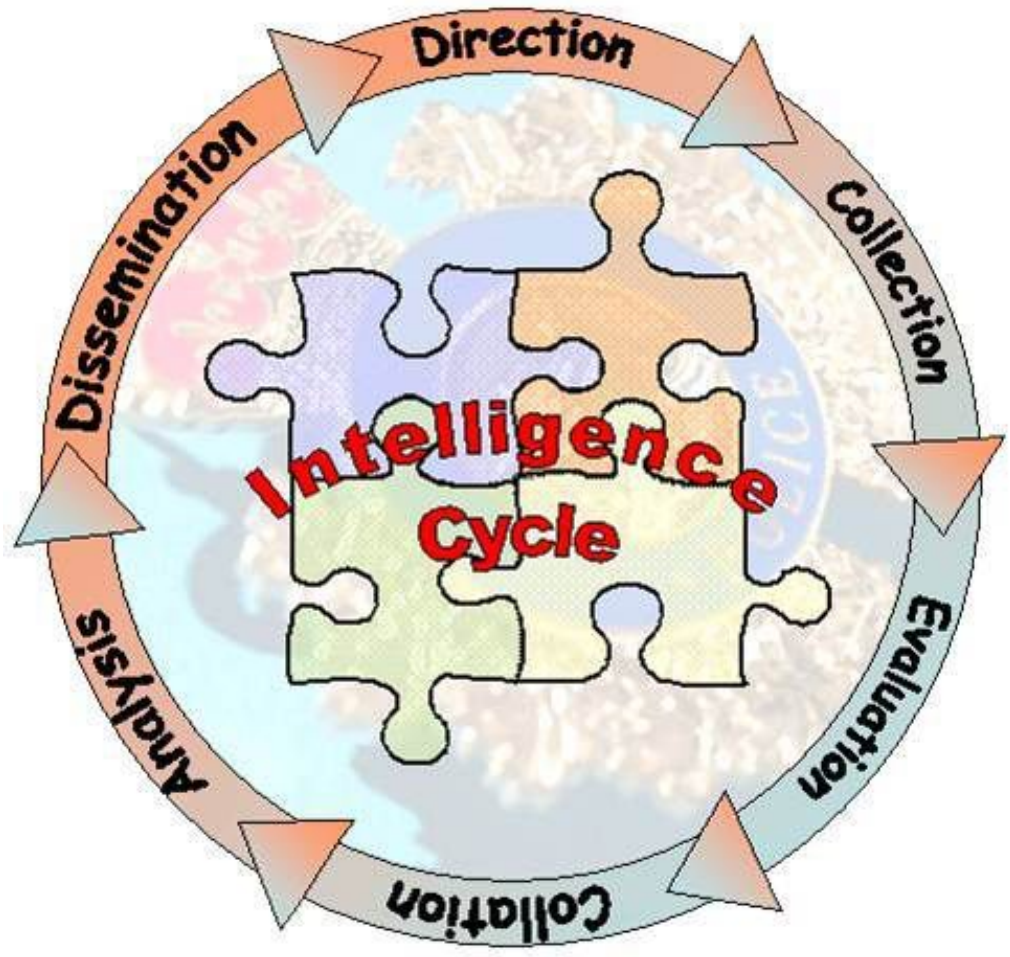
- ✓ 中央分析师与远程分析师协作加强
- ✓ 基于资源共享的协作：分析资源共享与分析产出共享
- ✓ 基于内容层面的协作
- ✓ 基于功能层面的协作：工作流协作

数据 → **情报** → **行动**

智能情报分析系统工作流

1. 分析人员输入需求
2. 系统根据需求关键词给出相关概念
3. 系统自动关联概念并允许分析人员连接、分类与组织
4. 系统根据概念扩展、自动收集信息资源
5. 系统根据分析人员反馈，推荐相关新资源
6. 系统将推送与检索资源整合存储
7. 分析人员处理数据
8. 分析人员直觉分析，处理输出报告
9. 自动化结构化分析
10. 横向智能化分析产出结果

完整的安全情报工作闭环



1. 初始分析
2. 散播式情报搜集
3. 定向
4. 定向搜集
5. 定向分析
6. 定向评估
7. 情报验证

全网暗链源分析

干扰因子多：

正常页面链接隐藏

可靠域名被劫持成为暗链源

正常网站被黑成为隐蔽暗链源

暗链？

无明显特征：

隐藏样式多样化

暗链内容多样化

暗链域名无特征

全网暗链源分析

全网暗链接监测　*全网监测有隐藏链接的站点*

初始分析

分析获取暗链植入的内容　*提取分析暗链主要内容*

散播式全网情报搜集

匹配全网有暗链的站点

采用提取的暗链内容进行全网匹配，发现海量植入暗链的站点

对各站点里暗链源进行汇聚分析

汇聚提取情报

获取高频度暗链源进行深入分析

定向分析、评估

全网暗链源分析（二）

获取高频度暗链源进行深入分析

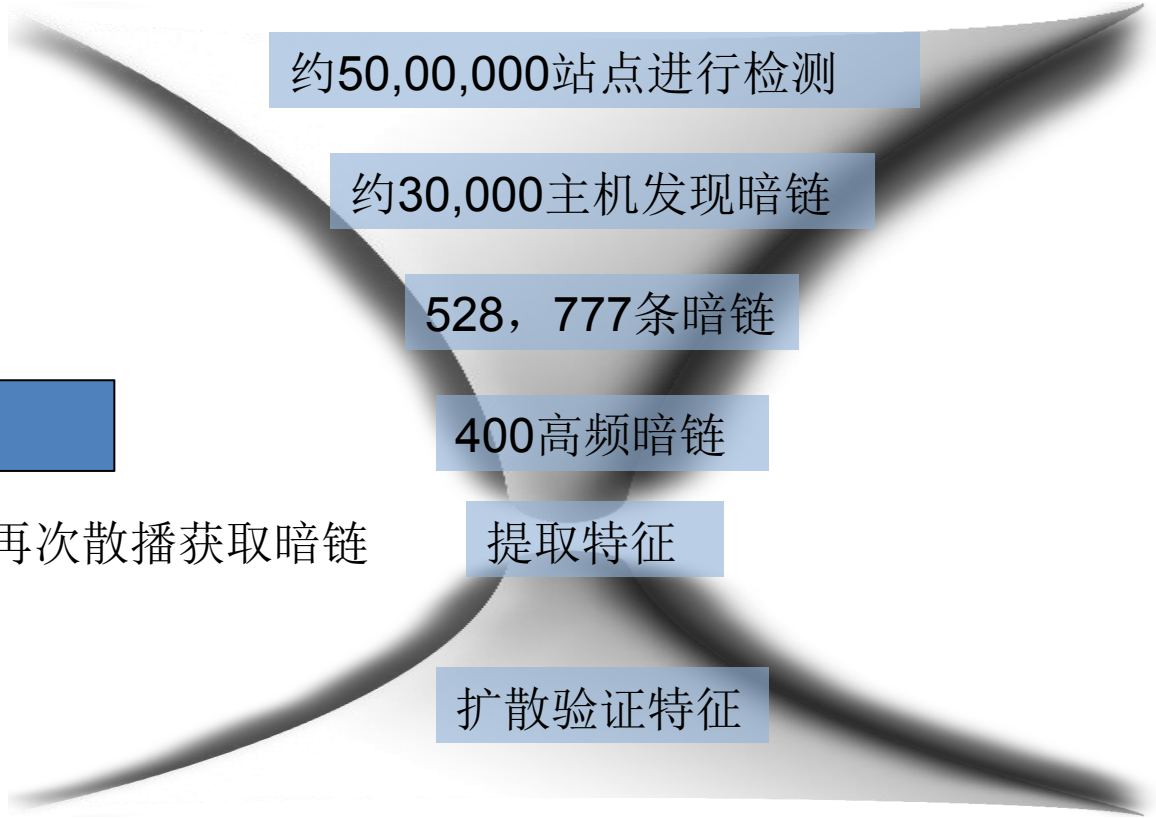
定向深入分析

专向分析暗链源形成特征原因

情报验证

全网特征验证并再次散播获取暗链

exp：
地址解析特征：海外gov站点
域名特征：伪造gov edu站点
内容特征:色情、博彩、游戏等



DBAPPSecurity 安恒信息

WEB应用安全和数据库安全的领航者

THANK YOU

www.dbappsecurity.com.cn