Virus Detection System

网络病毒监控系统的体系与研究方法

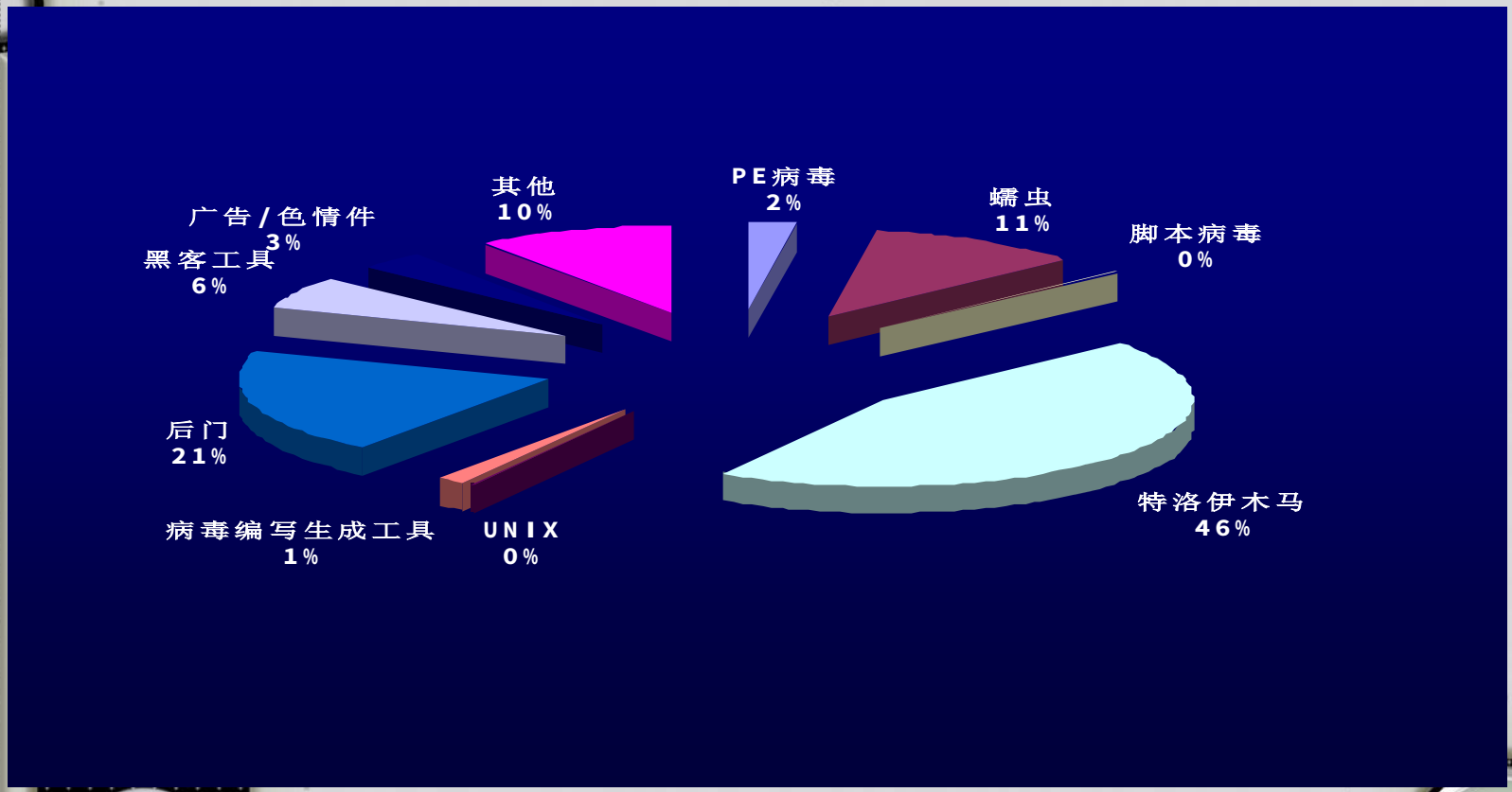
seak@antiy.net

X'con 2005

提纲

- 2004年度病毒趋势
- IDS体制的病毒考量
- VDS体制
- 数据处理方法

2004年新增病毒20047种



提纲

- 2004年度病毒趋势
- IDS体制的病毒考量
- VDS体制
- 数据处理方法

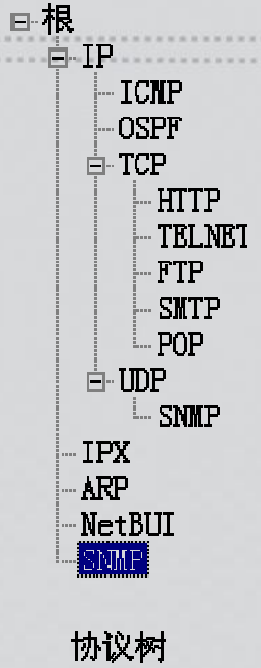
面对virus，IDS退缩了么？

- **2000**年，全年产生病毒**10350**种，后门**1029**种。
- **Snort x.x.x**
- **05/21/2001**
- **Backdoor.rules 127**条规则
- **Virus.rules 87**条规则，基本为邮件蠕虫规则，采用附件名、附件扩展名、邮件主题等检测方法。

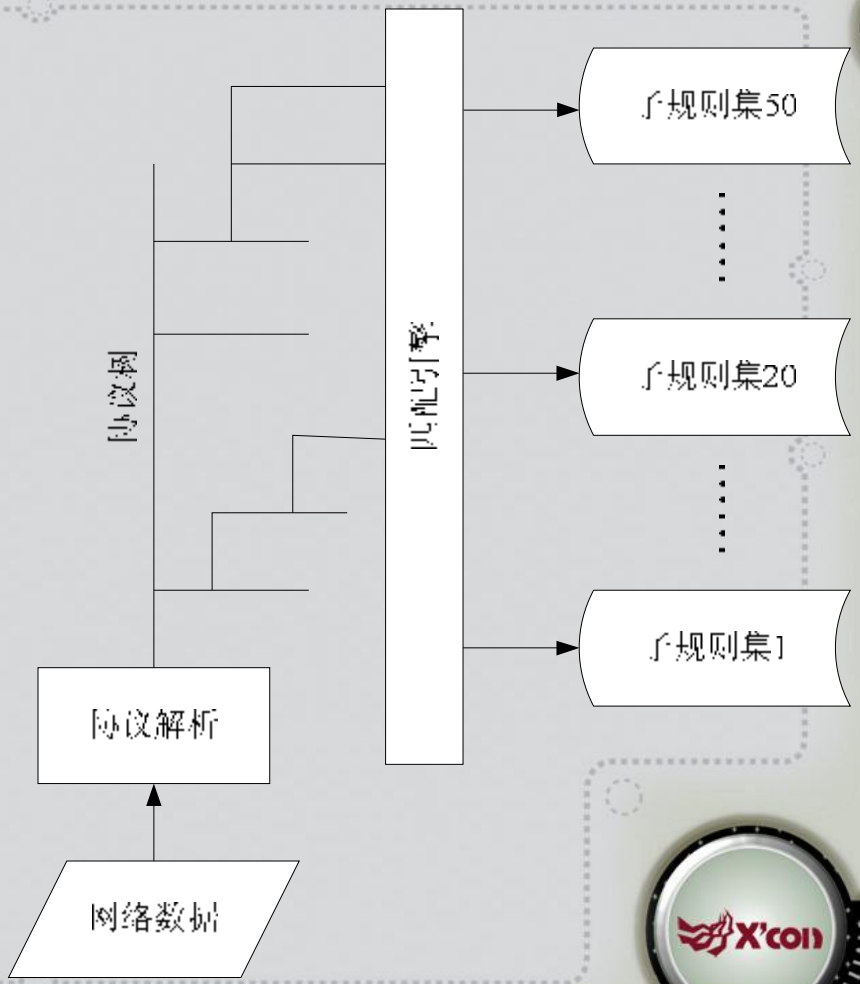
- **2004**年，全年产生病毒**20047**种，后门**4010**种。
- **Snort 2.3.3**
- **03/01/2005**
- **Backdoor.rules**，**82**条规则
- **Virus.rules 1**条规则，附件扩展名检测

“We don't care about virus rules anymore….”—— www.snort.org

传统IDS



协议树



- 精细协议解析
- 小规则集，短特征匹配
- 单规则集一般不超过**500**条记录

归一化软件方法

- 归一化方法，在面对大规模复杂事件处理的情况下，采取的对需要处理的事件进行归类，形成一个或一组的相应的处理模块和对应的可扩展数据结构与数据集合的规划方法。
- **AV:**检测对象格式分流（**BIN\MZ\PE\Macro\Script**）：降低误报率，减少匹配模式，激发预处理流程。
- **IDS:**协议分流（协议树），降低误报率，减少匹配数据，减少匹配模式

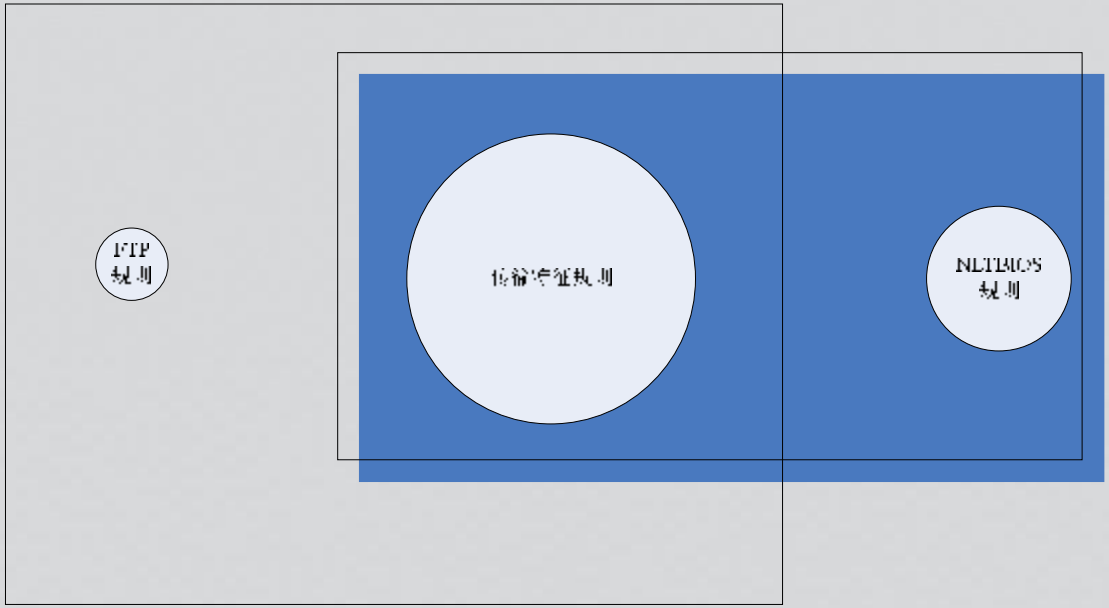
AVML与snort

- 在AV和IDS之间的归一化寻找界点
- Echo virus(id=″B00801″;type=″Backdoor″;os=″Win32″;format=″pe″;name=″bo″;version=″a″;size=″124928″;Port_listen=on[31337];content=|81EC0805000083BC240C05000000535657557D148B8424240500008BAC242005000050E9950500000F85800500008B|;delmark=1)

- alert tcp $EXTERNAL_NET any -> $HOME_NET 21 (msg:"Backdoor.bo.a Upload"; content: |81EC0805000083BC240C05000000535657557D148B8424240500008BAC242005000050E9950500000F85800500008B |;)

- alert tcp $EXTERNAL_NET any -> $HOME_NET 139 (msg:"Backdoor.bo.a Copy"; content: |81EC0805000083BC240C05000000535657557D148B8424240500008BAC242005000050E9950500000F85800500008B |;)。

分流体制造成的检测冗余

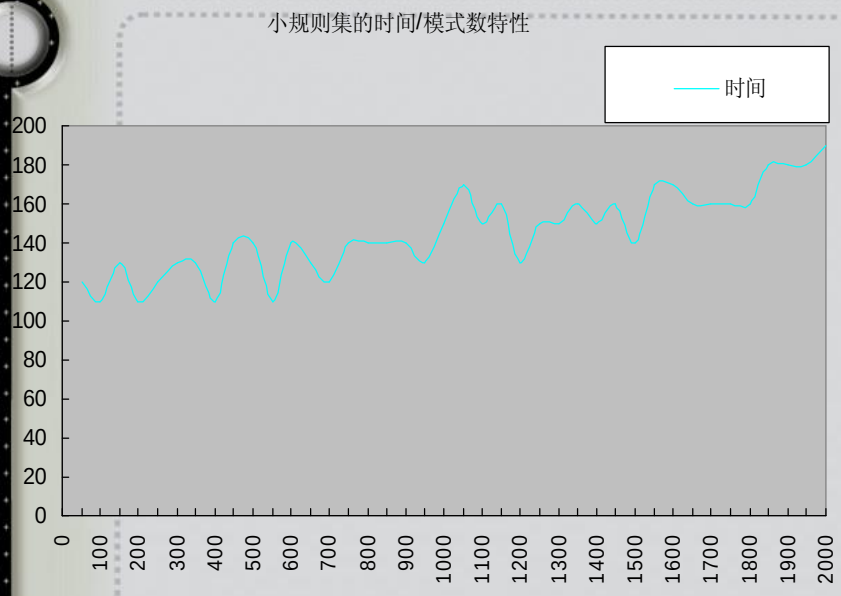


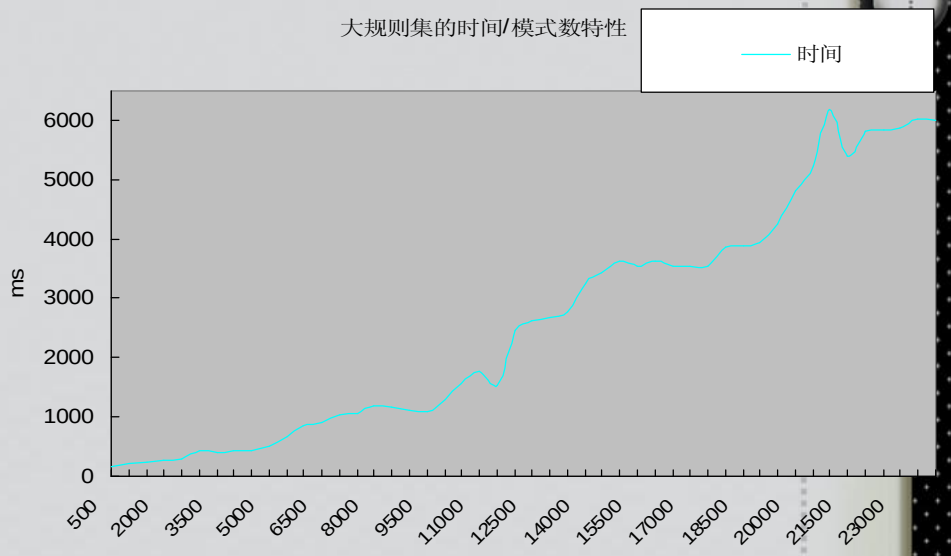
规则规模压力

分类	数量
邮件蠕虫	2807
即时消息蠕虫	172
P2P环境蠕虫	1007
IRC蠕虫	715
其他蠕虫	675
总计	5376

- 除蠕虫以外，与网络相关的木马、后门等超过2万种。
- 对应的规则数量可能超过3万种。

效率压力





用snort2.x所采用的BM算法测试，小规则集的良好时间特性，与大规则集的急剧性能衰减对比鲜明。

规则急剧增长所带来的效率压力，是IDS兼顾病毒体制的根本压力。

IDS网络性的视野、检测层次的考量、没有更细粒度的病毒定位期望，是IDS可以不承担效率压力的根本与原因。

提纲

- 2004年度病毒趋势
- IDS体制的病毒考量
- VDS体制
- 数据处理方法

分析主要矛盾

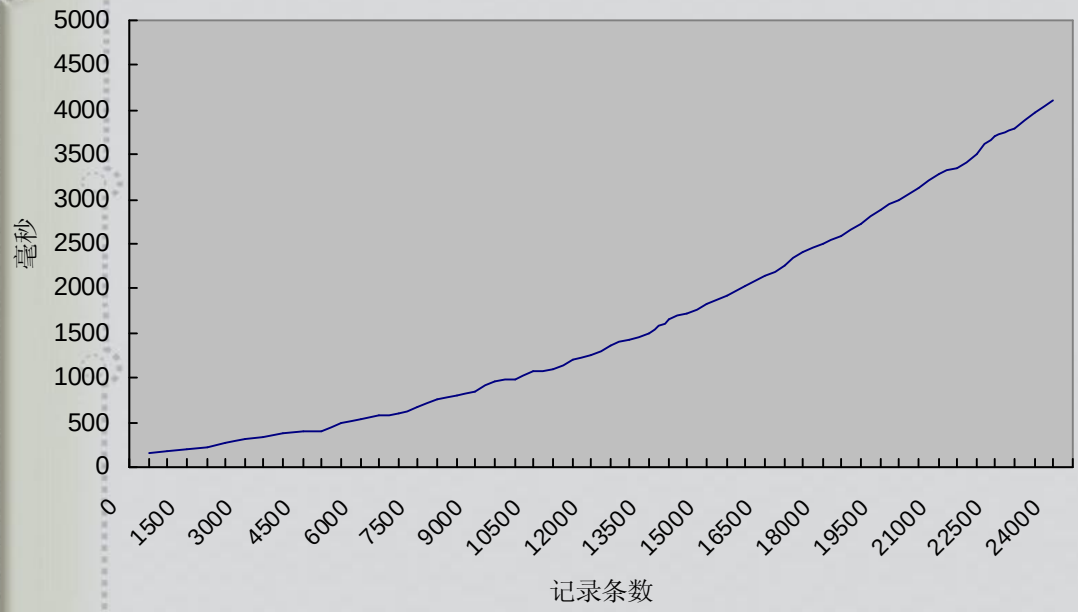
- 从上述分析看，在有足够规模维护规则的前提下，效率压力是网络病毒检测的主要压力。
- 传统归一化模型的根本在于事务分类，而事务分类的基础是检测对象的分类，由不同对象的分析，形成对应的方法。为了解决效率问题，能否从检测方法出发？形成方法类别范畴，再对对象进行归并。
- 所以：新的归一化模型以匹配速度和匹配粒度为核心，其构造是面向匹配算法的

归一化思路

- 将需要进行内容匹配网络数据划分为以下三种类型：

对象分类	例：
可直接性匹配的数据	一般性的扫描、攻击、传输
需要预处理的数据	URL（不区分大小写） 邮件（编码）
需要特定算法的数据	脚本

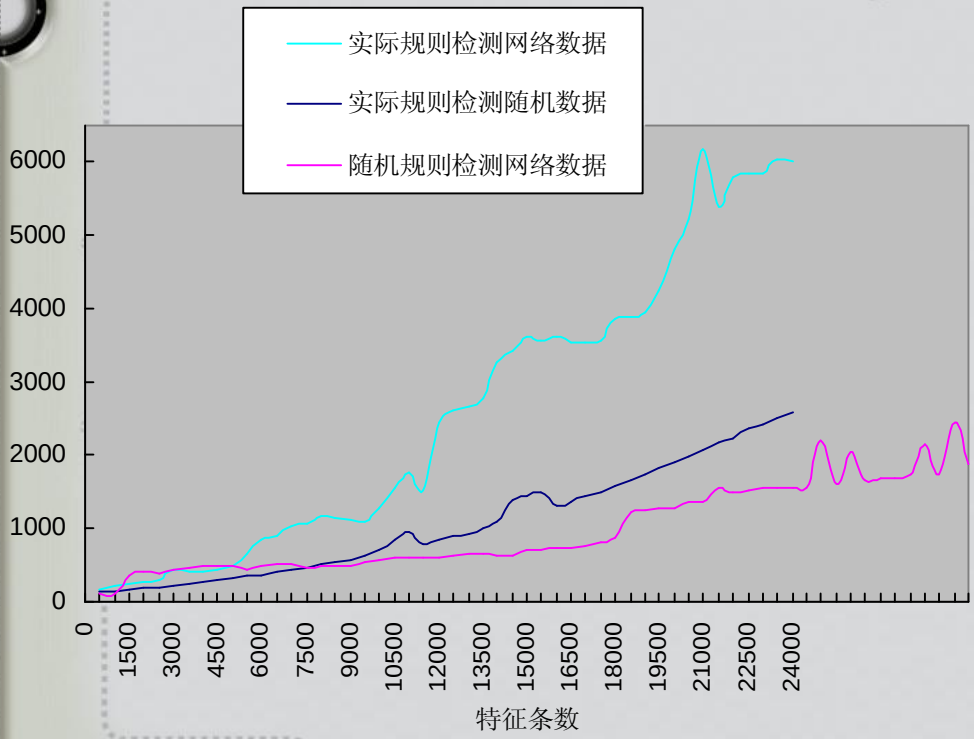
算法优化（一）



在特征规则小于6000条的情况下， 时间与记录条数的线性增长关系并不明显。但在10000条记录左右，开始出现反向的越升，导致性能的急剧下降，直至不可用。

记录条数的变化对匹配时间的影响

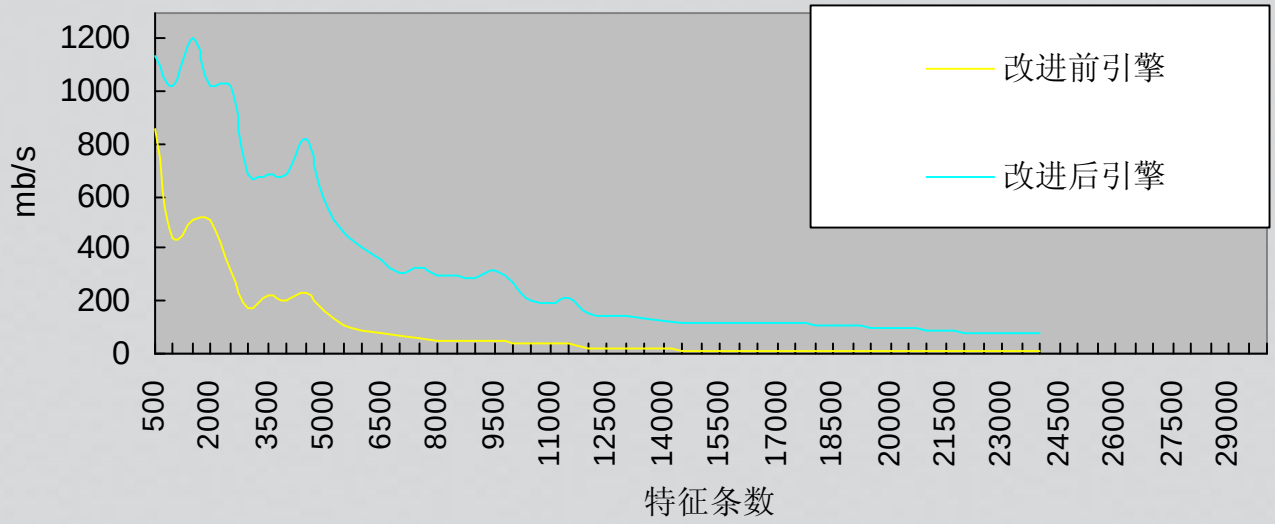
算法优化（二）



检测规则和检测对象对于数据的影响

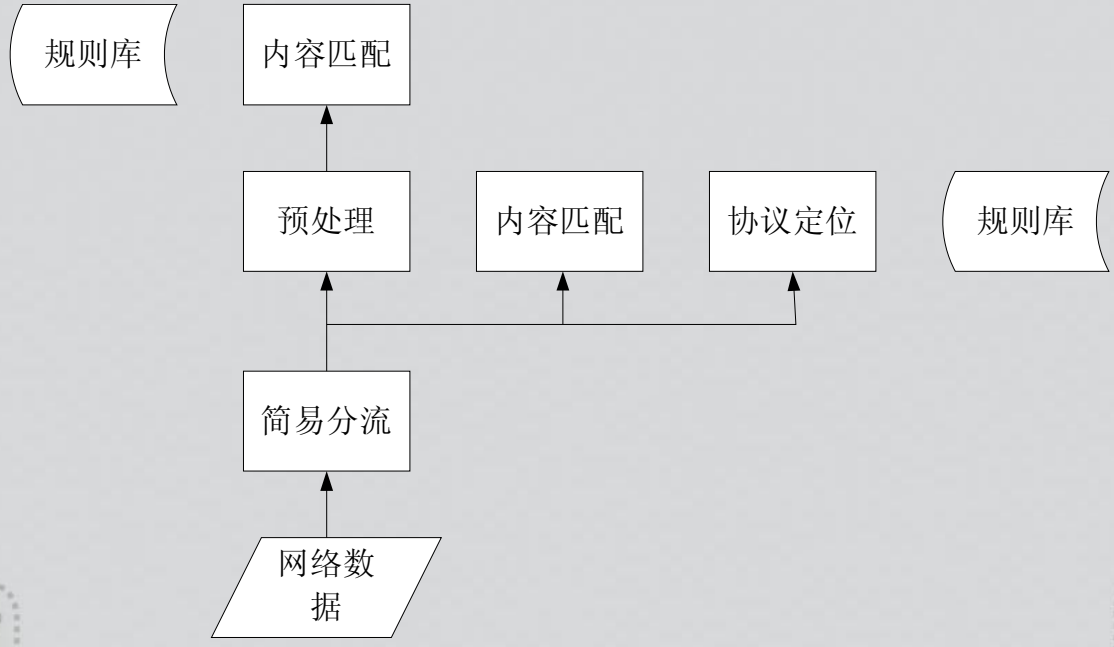
- 检测速度也与匹配对象和模式串质量存在密切关系，由于病毒特征之间存在着一定的近似性，并不能呈现随机分布的特点。同时网络数据也呈现出一定的分布特性，也并不呈现随机分布的特点。因此都对匹配情况构成影响

算法优化（三）

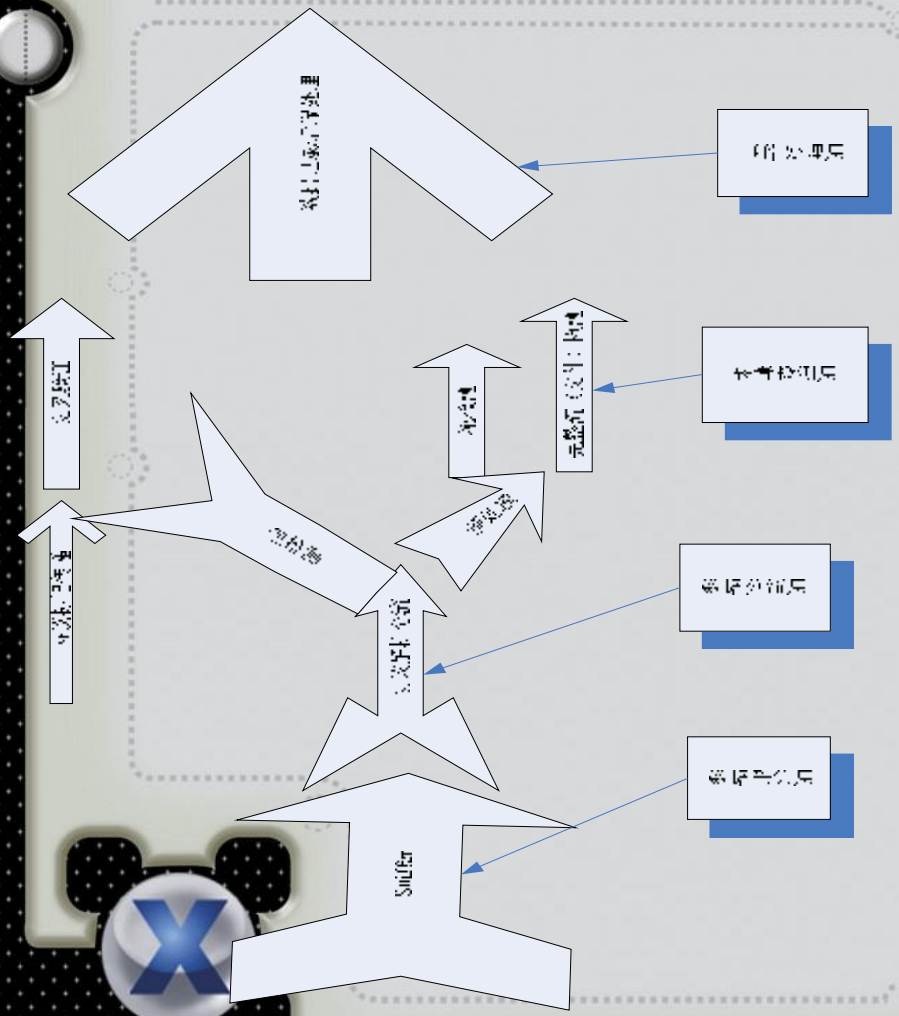


- 遏制病毒特征码近似性带来的效率影响

VDS核心架构思路

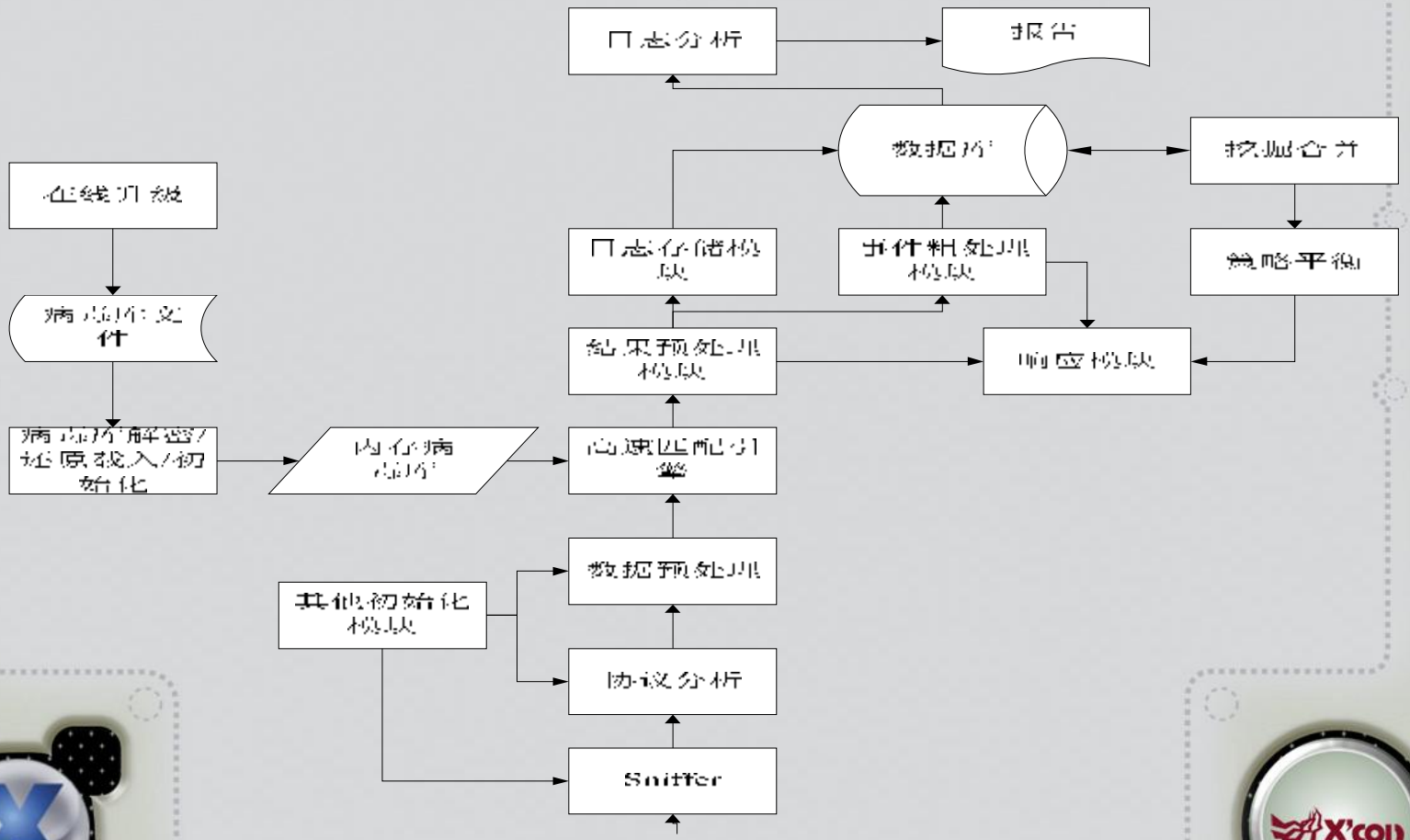


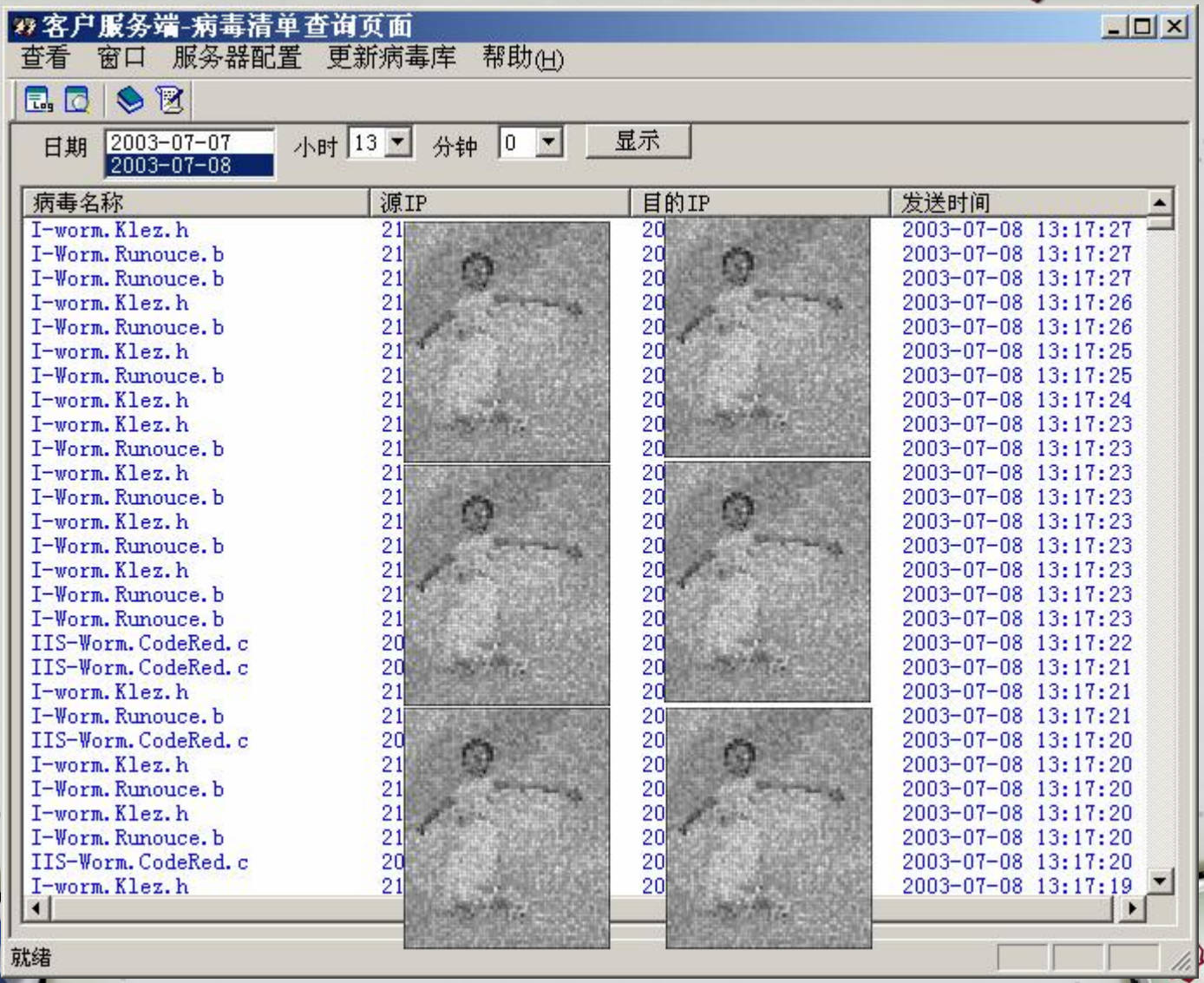
数据流转和病毒检测层次



- 划分为采集、分流、检测和处理四个层次
- 提供包、非完整流和完整流的三级别病毒检测

体系结构


日志分析
报告
数据库
报警合并
在线升级
日志存储模块
事件规则匹配模块
策略平衡
病毒库文件
结果预处理模块
响应模块
病毒库解密/还原载入/初始化
内存病毒库
高速匹配引擎
数据预处理
其他初始化模块
协议分析
Sniffer



2003年7月8日哈尔滨工业大学出口实时病毒数据。

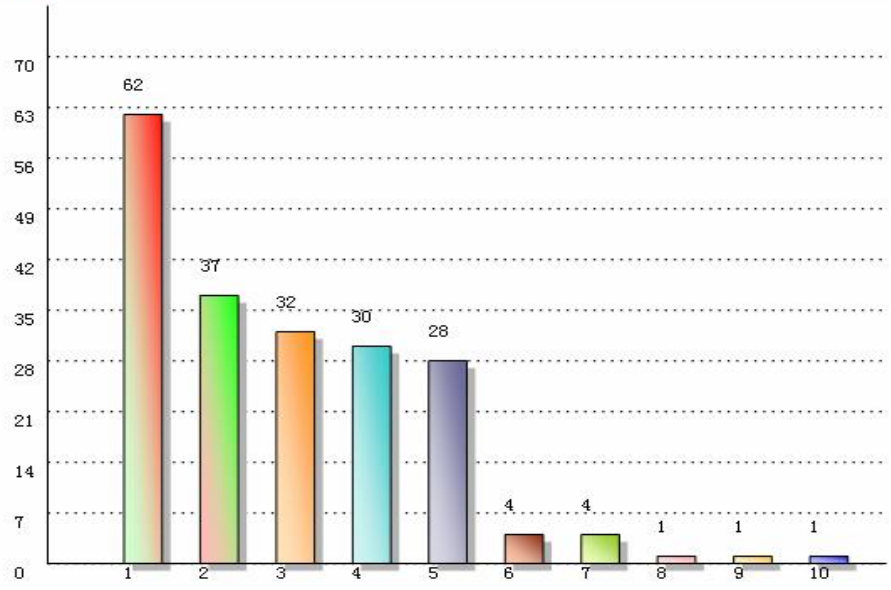
统计图表

2005年26周邮件蠕虫监测结果统计报告

检出次数排行榜

名次	病毒名称	进入内网比例	检出次数	病毒流量(byte)	感染列表
1	Email-Worm.Win32.Bagle.af	100%	62	0	受感染主机 受攻击主机
2	Email-Worm.Win32.LovGate.ad	100%	37	0	受感染主机 受攻击主机
3	Email-Worm.Win32.LovGate.ae	0%	32	0	受感染主机 受攻击主机
4	Email-Worm.Win32.LovGate.w	100%	30	0	受感染主机 受攻击主机
5	Email-Worm.Win32.LovGate.w	0%	28	0	受感染主机 受攻击主机
6	Email-Worm.Win32.NetSky.c	100%	4	0	受感染主机 受攻击主机
7	Email-Worm.Win32.LovGate.q	100%	4	0	受感染主机 受攻击主机
8	Email-Worm.Win32.Zafi.d	100%	1	0	受感染主机 受攻击主机
9	Email-Worm.Win32.Bagle.af	0%	1	0	受感染主机 受攻击主机
10	Email-Worm.Win32.NetSky.z	100%	1	0	受感染主机 受攻击主机
		总计	200	0	

检出次数统计图



未知预警

发现病毒体传输次数排行榜：

名次	病毒名	发现次数
1	[illegible]	[illegible]
2	I-Worm.UNKnow	2548
3	[illegible]	[illegible]
4	I-Worm.Nimda	2
5	Backdoor.Netbus.160.a	1
6	Trojan.Win32.HDBreaker	1

- **2003年8月18日发现I-Worm.Unknow**（未知邮件蠕虫）数量显著增长，**8月19日证明，该I-WORM.Unknow主要为为I-worm.sobig.f**。

体系本质和对象

- 粗架构，细粒度
- 全局病毒视图
- 源定位
- 测量与评估

提纲

- 2004年度病毒趋势
- IDS体制的病毒考量
- VDS体制
- 数据处理方法

检测对抗研究的现状与VDS的关联

- 基于**GrIDS**的网络蠕虫监测
- 基于**PLD**硬件的监测技术
- 基于**HoneyPot**的蠕虫检测
- 蠕虫对抗

- 如果没有对已知蠕虫的检测，一切蠕虫均为未知。
- **VDS**通过引入了工程化技术基础，使形成准确的（定性定量）的原始网络**virus**事件成为可能。
- 这种基础事件集合可以成为进一步的研究空间。

事件处理方法

- **DEDL，Detection Events Description Language**检测事件描述语言。
- 采用描述符的方式，将网络检测事件制定为一种规范的格式，并支持一般的条件推导。
- 定义了事件类型（**type**）、事件**ID**、源**IP**（**Source_IP**）、目标**IP(Target IP)**、事件时间等**20**多个事件要素。

- **处理方法**
- 内部技术合并
- 平行式合并
- 分析式平行合并
- 辐射式合并
- 聚合式合并
- 传导链式合并

事件处理方法（2）

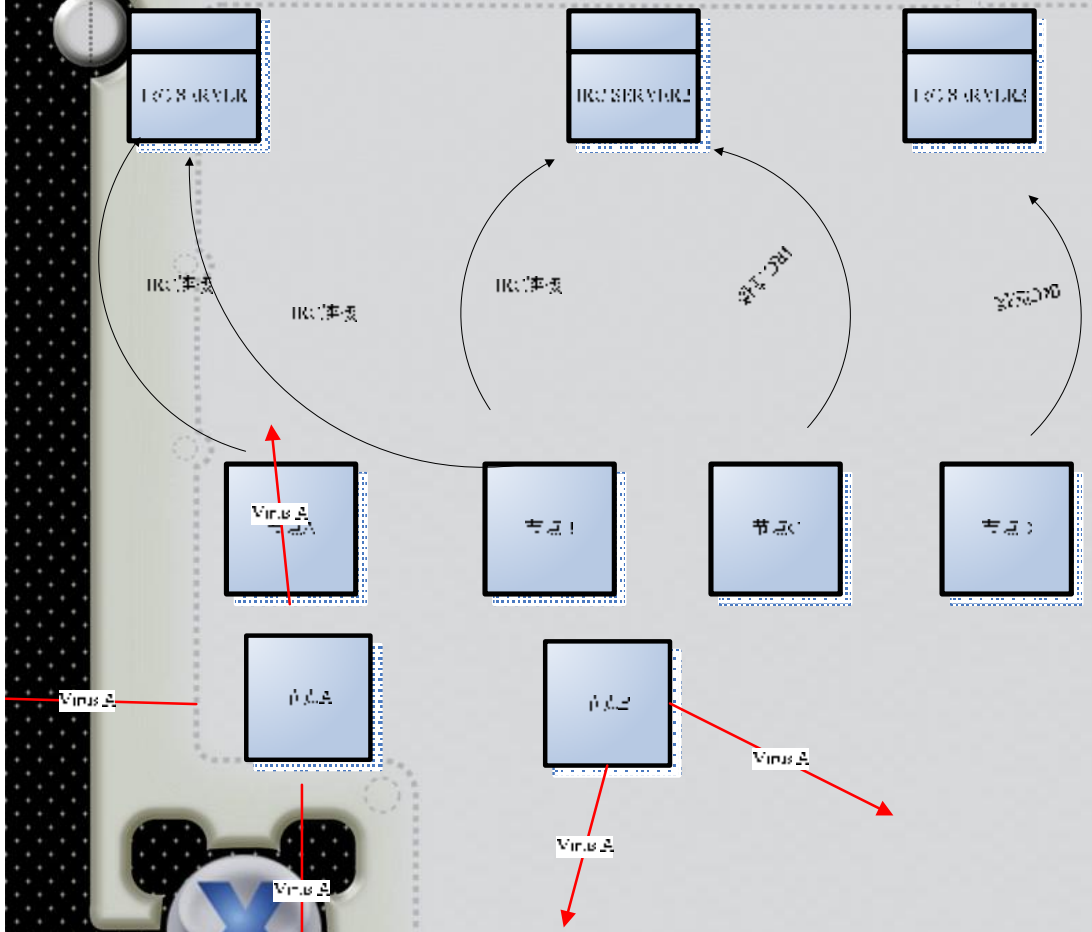
If exist
Net_Action(RPC_Exploit)[IP(1)->IP(2);time(1)]
Net_Action(RPC_Exploit) [IP(2)->IP(3) ;time(2)]
and
time(2)>time(1)
than
Net_Action(RPC_Exploit) [IP(1)-> IP(2) -> IP(3)]

If exist
Net_Action(Trans,Worm.Win32.Dvldr)[IP(1)->IP(2);time(1)]
Net_Action(Trans,Worm.Win32.Dvldr)[IP(3)->IP(4);time(2)]
and
NET(a)∈{IP(2), IP(3)}/IP2,IP属于内网NET(a)
and
time(2)>time(1)
than
Net_Action(Trans,Worm.Win32.Dvldr) [IP(1)-> IP(2) -> IP(3) ->IP(4)]

行为规并

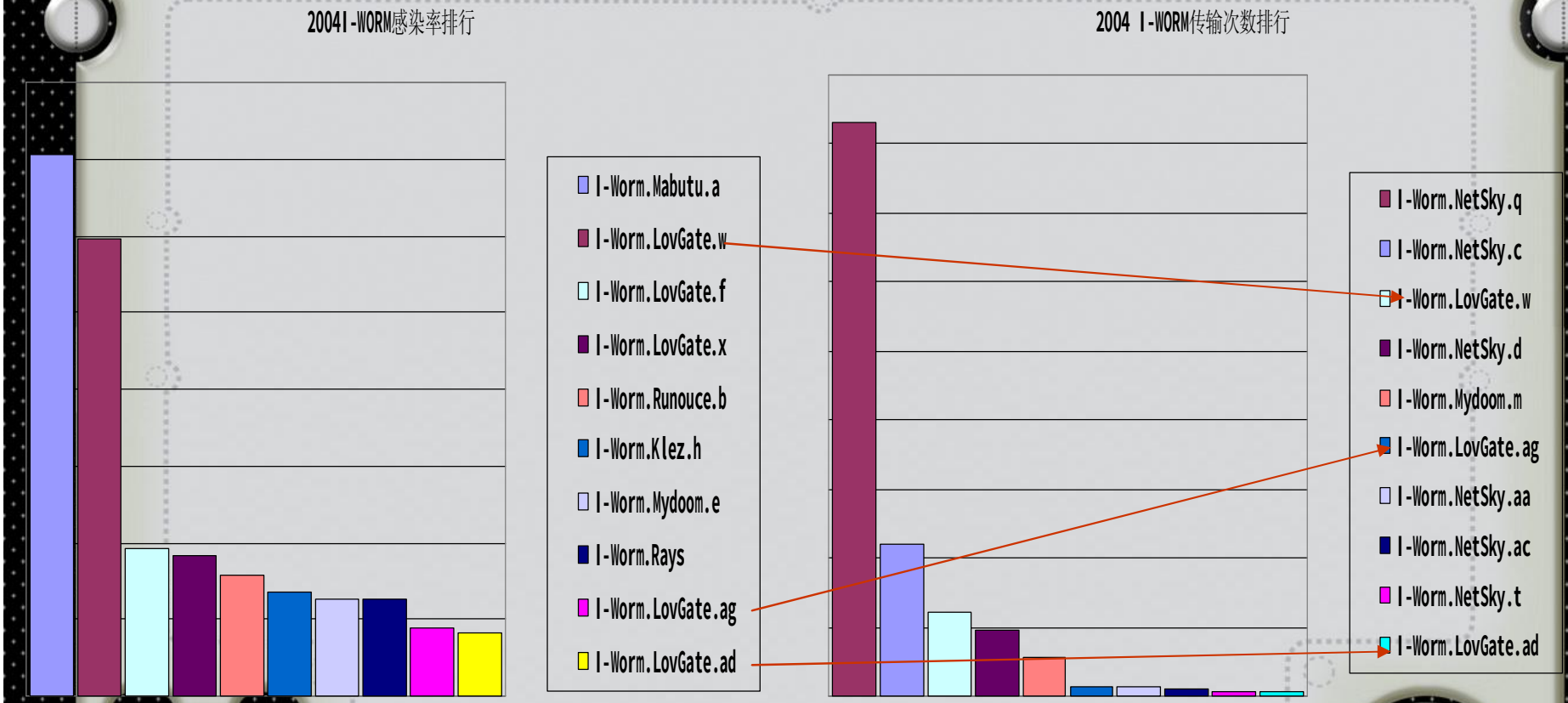
DEDL事件	AVML行为特性规则
Net_Action(act)[IP(1),IP(2):445; ;time(1)] Net_Action(act)[IP(1),IP(3):445; ;time(1)] .... Net_Action(act)[IP(1),IP(12):445; ;time(1)] Net_Action(Trans,Worm.Win32.Dvldr)[IP(1)->IP(12);time(1)]	Virus_act_lib Virus seek(id=″W02872″;dport=139,445;trans=netbios)

数据处理式发现



在传统的连接视图的基础上，进行行为冲销，则有效提高类似GrIDS方法的发现效率。

数据分析意义



传播次数与感染节点数关系不明显

信任链传播最有效

用数据反击以毒攻毒

谢谢大家

- 网络病毒监控多年的学术化探索和工程化尝试延伸成为一个新的技术体制。
- 病毒攻防之路是必然王国走向自由王国之路——我们在路上。