上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

基于元数据和安全事件的大数据分析

上海交通大学网络信息中心

姜开达

2013年12月28日

个人介绍

- 1997 ~ 2001 ~ 2013 (16 years)
  上海交通大学 网络信息中心

- 工作经历:
  学生宿舍网络建设与管理
  大中小网络运行与维护
  信息系统建设与管理
  近五年主要专注于
  网络信息安全领域

主要内容

一 MetaData的获取

二 大数据分析平台

三 安全事件关联分析

四 未来研究方向

安全的颗粒度

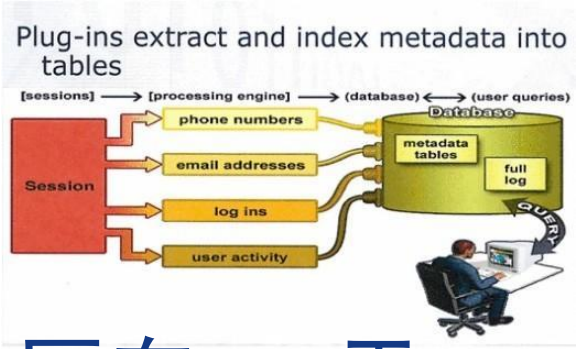
- NetFlow/sFlow 报文采样
  DDOS/Port Scan/异常流量发现

- 深度数据包检测（DPI）
  基于已知签名的识别，对加密流量作用有限

- 元数据（MetaData）介于两者之间
  http/smtp/pop3/ftp/dns/telnet……

美国情报收集系统 X-Keyscore

- 可针对邮件、网站内容等执行强大的查询
- 提供实时的目标活动信息



- 所有未过滤的数据可在缓冲区存 3 天
- 存储所监控网站的完整数据，为元数据建立索引

安全事件历史回溯

- Flow采样分析太粗，丢失信息较多
  基于五元组等信息
- 完整数据包长期存储，代价巨大
  1G/10G/100G　Mission: Impossible
- 依靠长期历史元数据是现实的选择
  在线分析漏过了，不能再错过离线处理

MetaData的生成

- 基于网络流量，生成需要的格式吐出

当前统计信息

当前序列号	3686508730
记录保存	4091739320/2366605198(成功/失败)
记录发送	4091739284/0(成功/失败)
数据包发送	3686508728/0(成功/失败)

当前配置信息

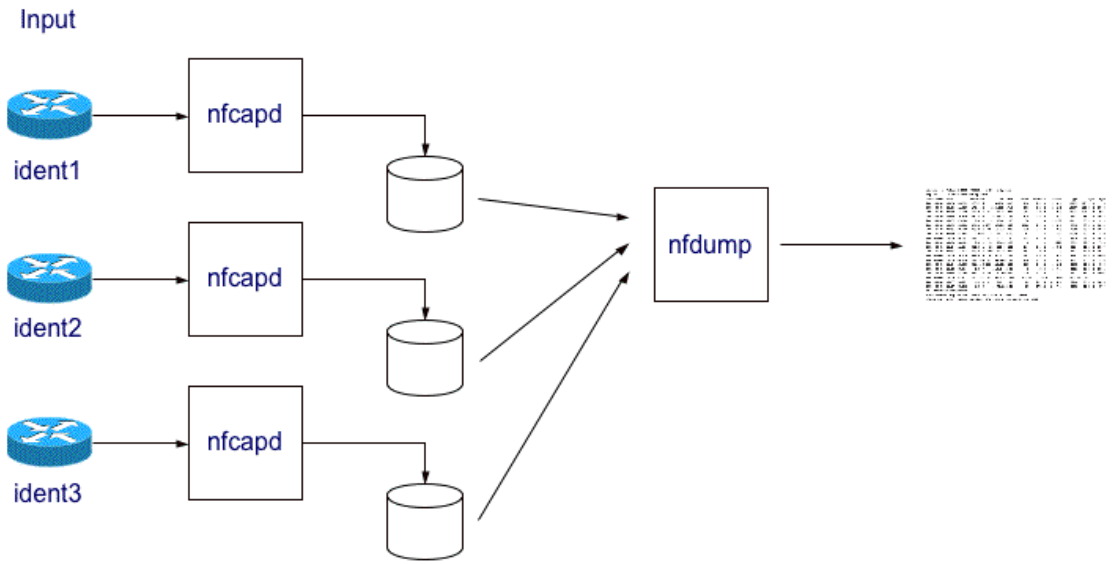
接收服务器IP	10. 20. 30. 40	(xxx.xxx.xxx.xxx)
接收服务器端口	514	(0~65535)
是否记录日志	记录	
是否记录HTTPHOST	记录	

nfdump 介绍

- http://nfdump.sourceforge.net/

- The nfdump tools collect and process netflow data on the command line.

- Web interface http://sourceforge.net/projects/nfsen/



CoralReef 介绍

- http://www.caida.org/tools/measurement/coralreef/

- CoralReef is a comprehensive software suite developed by CAIDA to collect and analyze data from passive Internet traffic monitors, in real time or from trace files.

- CoralReef supports monitoring of any standard network interface (via libpcap) on unix-like systems, as well as specialized high performance ATM, POS, and Ethernet devices at up to OC192 and 10 GigE bandwidths on Intel-based workstations running FreeBSD or Linux.

Justniffer 介绍

- http://justniffer.sourceforge.net/

- Network TCP Packet Sniffer
- Reliable TCP Flow Rebuilding
- Optimized for "Request / Response" protocols.
- Can rebuild and save HTTP content on files

Example 1 Retrieving http network traffic in access_log format

```
$ justniffer -i eth0
```

output:

```
192.168.2.2 - - [15/Apr/2009:17:19:57 +0200] "GET /sflogo.php?group_id=205860&type=2 HTTP/1.1" 200 0 "" "Mozilla/5.0
(X11; U; Linux i686; en-US; rv:1.9.0.8) Gecko/2009032711 Ubuntu/8.10 (intrepid) Firefox/3.0.8)"

192.168.2.2 - - [15/Apr/2009:17:20:18 +0200] "GET /search?
q=subversion+tagging&ie=utf-8&oe=utf-8&aq=t&rls=com.ubuntu:en-US:unofficial&client=firefox-a HTTP/1.1" 200 0 ""
"Mozilla/5.0 (X11; U; Linux i686; en-US; rv:1.9.0.8) Gecko/2009032711 Ubuntu/8.10 (intrepid) Firefox/3.0.8)"

192.168.2.2 - - [15/Apr/2009:17:20:07 +0200] "GET /sflogo.php?group_id=205860&type=2 HTTP/1.1" 200 0
"http://justniffer.sourceforge.net/" "Mozilla/5.0 (X11; U; Linux i686; en-US; rv:1.9.0.8) Gecko/2009032711
Ubuntu/8.10 (intrepid)Firefox/3.0.8)"
```

HTTP MetaData 示例

1．GET 类型

Time|G|Host|URI|||Referer|SrcIP|SrcPort|DstIP|DstPort|User-Agent

2013-08-25 11:01:02|G|news.sjtu.edu.cn|/images/index1_1.jpg||
|http://news.sjtu.edu.cn/info/1021/127584.htm
|111.180.72.x|31869|202.120.2.102|80|Mozilla/4.0 (compatible; MSIE 7.0)

2．POST 等其他类型

Time|P|Host|URI|Type|Size|Referer|SrcIP|SrcPort|DstIP|DstPort|User-Agent

2013-08-25 11:01:29|P|
welcome.sjtu.edu.cn|/jdyx/memeber/login.php?action=checkandlogin|application/x-www-form-urlencoded|38|http://welcome.sjtu.edu.cn/jdyx/news/
|110.179.29.x|26605|202.120.63.4|80|Mozilla/5.0 (compatible; MSIE 7.0)

3．服务器返回类型

Time|HTTP Code|||Type|Size||SrcIP|SrcPort|DstIP|DstPort|

2013-08-25 11:03:06|404|||text/html|168||220.120.2.102|80|202.119.208.93|8189|

HTTP MetaData 存储

- 以上海交通大学校园网为例

2013.12.25
访问校内近千网站的原始HTTP元数据量为
18.1 G Bytes，78,962,897 条记录

存储校内网站一年的元数据量不超过 5T Bytes
但是记录条数在数百亿量级

目前可以实现单机每日20亿量级的元数据生成

Hadoop系统硬件概况

- 共 24 结点
  - 2个管理结点，一个作业提交结点，21个存储计算节点
- CPU
  - Intel Xeon(R) CPU E5-2670@2.60GHz，双CPU，开HT后32核
- Memory
  - 8 × 8 = 64GB
- Disk
  - 240GB(SSD)×2+2TB(SATA)×12
- Network
  - Intel 82599 万兆网
- HDFS总容量
  - 333TB/3(replication=3)

```
Configured Capacity: 366196897701888 (333.05 TB)
Present Capacity: 349462760841216 (317.83 TB)
DFS Remaining: 334980505710592 (304.66 TB)
DFS Used: 14482255130624 (13.17 TB)
DFS Used%: 4.14%
Under replicated blocks: 0
Blocks with corrupt replicas: 0
Missing blocks: 0
```

三个机柜 24 台服务器

Hadoop系统软件概况

- 使用Hadoop版本
  - Cloudera Standard 4.8.0
- 组件
  - CDH4.5.0+IMPALA 1.2.1+SOLR 1.1.0）
- 特点
  - 易部署（网页批量部署）
  - 性能良好（TestDFS IO：10*10GB文件写142MB/s，10*10GB文件读801MB/s）
  - 监控信息全面

    提供对每台主机及整个集群的监控信息
  - 维护方便

Cloudera Manager

正在显示 1 到 21，共 21 个条目 第一个 上一个 1 下一个 最后一个 显示 25 条目

名称	IP	机架	CDH 版本	群集	角色	状态	上一检测信号	维护模式
任何名称	任何 IP	任何机架	全部	全部	全部	全部	全部	全部
hadoopjob.hadoop.sjtu.edu.cn	192.168.0.12	/default	CDH4	Cluster 1 - CDH4	8 个角色	运行状况良好	13.67s ago	
master.hadoop.sjtu.edu.cn	192.168.0.10	/default	CDH4	Cluster 1 - CDH4	12 个角色	运行状况良好	13.23s ago	
namenode.hadoop.sjtu.edu.cn	192.168.0.11	/default	CDH4	Cluster 1 - CDH4	5 个角色	运行状况良好	4.82s ago	
slave01.hadoop.sjtu.edu.cn	192.168.0.13	/default						
slave03.hadoop.sjtu.edu.cn	192.168.0.15	/default						
slave04.hadoop.sjtu.edu.cn	192.168.0.16	/default						
slave06.hadoop.sjtu.edu.cn	192.168.0.18	/default						
slave07.hadoop.sjtu.edu.cn	192.168.0.19	/default						
slave08.hadoop.sjtu.edu.cn	192.168.0.20	/default						
slave09.hadoop.sjtu.edu.cn	192.168.0.21	/default						
slave10.hadoop.sjtu.edu.cn	192.168.0.22	/default						
slave11.hadoop.sjtu.edu.cn	192.168.0.23	/default						
slave12.hadoop.sjtu.edu.cn	192.168.0.24	/default						
slave13.hadoop.sjtu.edu.cn	192.168.0.25	/default						
slave14.hadoop.sjtu.edu.cn	192.168.0.26	/default						
slave15.hadoop.sjtu.edu.cn	192.168.0.27	/default						
slave16.hadoop.sjtu.edu.cn	192.168.0.28	/default						

slave01.hadoop.sjtu.edu.cn — 运行状况良好 (历史)

状态 | 进程 | 资源 | 命令 | 配置 | 组件 | 审核 | 图表库

详细信息 2013年12月27日, 8:08:07 早上 CST

IP	192.168.0.13	内核数	32
机架	/default	下载	0.02 0.01 0.42
上次更新	2.00秒 之前	物理内存	38 吉字节 / 62.9 吉字节
主机代理	详细信息	交换空间	0 字节 / 0 字节
CDH 版本	CDH4	事件搜索	警报, 严重, 全部
分配	centos 6.4		

运行状况测试 全部展开

- 10 良好。

运行状况历史记录

12月 26 下午12点40	良好	2 良好	显示
12月 26 下午12点39	未知	2 未知	显示
12月 25 8:14:46 晚上	良好	2 良好	显示
12月 25 8:13:46 晚上	未知	2 未知	显示
12月 25 8:08:47 晚上	良好	2 良好	显示
12月 25 8:08:46 晚上	未知	2 未知	显示

文件系统

磁盘	装入点	使用情况
/dev/sdh	/mnt/sdh	317.0 吉字节 / 1.8 天字节
/dev/sdi	/mnt/sdi	315.0 吉字节 / 1.8 天字节
/dev/sdj	/mnt/sdj	317.3 吉字节 / 1.8 天字节
/dev/sdk	/mnt/sdk	316.2 吉字节 / 1.8 天字节
/dev/sdl	/mnt/sdl	315.4 吉字节 / 1.8 天字节
/dev/sdm	/mnt/sdm	316.7 吉字节 / 1.8 天字节
/dev/sdn	/mnt/sdn	316.7 吉字节 / 1.8 天字节
/dev/sda	/mnt/sda	123 吉字节 / 234.7 吉字节
/dev/sdc1	/boot	41.7 兆字节 / 193.7 兆字节
/dev/sdd	/mnt/sdd	317.4 吉字节 / 1.8 天字节
/dev/sde	/mnt/sde	316.1 吉字节 / 1.8 天字节
/dev/sdf	/mnt/sdf	317.7 吉字节 / 1.8 天字节
/dev/sdg	/mnt/sdg	315.1 吉字节 / 1.8 天字节
/dev/sdc2	/	48.3 吉字节 / 78.4 吉字节
/dev/sdb	/mnt/sdb	123 吉字节 / 234.7 吉字节

图表 30 分钟 1 小时 2 小时 6 小时 12 小时 1 天

图形化界面 HUE

- **The open source Apache Hadoop UI (图形化用户界面)**

  提供 Hadoop 各种组件的Web UI操作

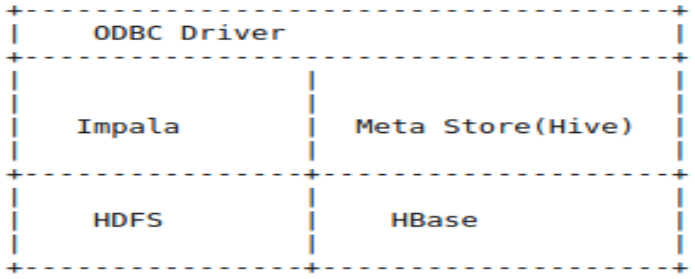
- Hive查询的UI, 查看作业进度、下载作业结果

类SQL工具——Hive

- Hive简介
  - 基于Hadoop的一个数据仓库工具
  - 把SQL语句转化为MapRedece任务运行
  - 元数据存储在数据库，数据存储在HDFS中
- 使用Hive的优势
  - 数据互通，简单的LOAD命令导入、查询结果可以用保存为cvs或者Excel格式
  - 学习成本低，可以通过类SQL语句快速实现简单的MapReduce统计
  - 基于HDFS的可扩展性
  - 可通过压缩、分区列建立索引优化

Cloudera其他组件介绍

- 全文索引——Cloudera Search
  - 核心：Hadoop+Solr
  - 可实现对HDFS中数据进行全文索引
  - 查询时间1~2s/TB的索引/1台服务器（目前还在测试）
- 实时查询开源软件——Cloudera Impala
  - Impala与其他组件关系



  - Impala(SQL on HDFS) VS Hive(SQL on MapReduce)
- 这些组件在后续的数据分析工作中很有用处

大数据分析应用

- 安全漏洞快速影响评估
- 安全事件综合分析，历史回溯
- 漏洞挖掘和0Day早期预警

Apache Struts 2 发现

select * from sjtu_in where url like '%.action?%' or url like '%.do?%'

结果 查询 日志 列

	time	type	domain	url
1	2013-08-25 19:32:31	G	www[redacted].sjtu.edu.cn	/mis/degree/showLunWenInfo.do?xh=0071309005
2	2013-08-25 19:32:38	G	www[redacted].sjtu.edu.cn	http://www.lib.sjtu.edu.cn/list.do?articleType_id=53
3	2013-08-25 19:32:41	G	nvc.[redacted].edu.cn	/JournalX_nvc/authorLogOn.action?mag_Id=1
4	2013-08-25 19:32:51	G	www[redacted].sjtu.edu.cn	/mis/degree/showLunWenInfo.do?xh=0042912007
5	2013-08-25 19:32:51	G	www[redacted].sjtu.edu.cn	/mis/courseView.do?KCDM=X200544
6	2013-08-25 19:32:53	G	www[redacted].sjtu.edu.cn	/mis/courseView.do?KCDM=X090503
7	2013-08-25 19:32:55	G	www[redacted].sjtu.edu.cn	/mis/courseView.do?KCDM=C415001
8	2013-08-25 19:32:59	G	www[redacted].sjtu.edu.cn	/mis/degree/showDbInfo.do?xh=0071309005
9	2013-08-25 19:33:02	G	nvc.[redacted].edu.cn	/JournalX_nvc/author/AuthorImanuDetail.action?id=3165632607
10	2013-08-25 19:33:08	G	www[redacted].sjtu.edu.cn	/mis/showStudents.do?KCDM=B140701&JXBH=B1407011109M14
11	2013-08-25 19:33:10	G	www[redacted].sjtu.edu.cn	/mis/degree/showLunWenInfo.do?xh=0071309007
12	2013-08-25 19:33:13	G	nvc.[redacted].edu.cn	/CN/article/downloadArticleFile.do?attachType=PDF&id=302
13	2013-08-25 19:33:37	G	www[redacted].sjtu.edu.cn	/mis/degree/showLunWenInfo.do?xh=0071309012
14	2013-08-25 19:33:54	G	xueb[redacted].sjtu.edu.cn	/CN/article/showSupportInfo.do?id=9506
15	2013-08-25 19:33:54	G	se.s[redacted]edu.cn	/download.action?descriptionid=461
16	2013-08-25 19:33:56	G	www[redacted].sjtu.edu.cn	/list.do?articleType_id=53

查询性能分析

19个计算节点，未做分区列等优化
2 分42秒 查询完 56 亿条原始记录

作业：201312061201_0045 - Job Browser

作业 ID
201312061201_0045

用户
kaida

状态
RUNNING

日志
日志

MAPS:
38%

REDUCES:
无

DURATION:
1m:7s

操作
停止此作业

任务 | 元数据 | 计数器

近期任务

日志	任务	类型
	m_000269	MAP
	m_000303	MAP
	m_000307	MAP
	m_000309	MAP
	m_000310	MAP
	m_000311	MAP
	m_000313	MAP
	m_000326	MAP
	m_000328	MAP
	m_000332	MAP
	m_000335	MAP

PHP-DOS 发现

select * from sjtu_in where url like '%host=%' and url like '%port=%'

结果 | 查询 | 日志 | 列

	time	type	domain	url
158	2013-08-29 04:11:55	G	202.120.[redacted].142	/webdav/udp.php?act=phptools&host=94.23.204.149&time=500&port=80
159	2013-08-29 04:11:55	G	202.120.[redacted].238	/webdav/udp.php?act=phptools&host=94.23.204.149&time=500&port=80
160	2013-08-29 04:11:55	G	202.120.[redacted].142	/webdav/udp.php?act=phptools&host=94.23.204.149&time=500&port=80
161	2013-08-29 04:11:55	G	202.120.[redacted].238	/webdav/udp.php?act=phptools&host=94.23.204.149&time=500&port=80
162	2013-08-29 04:11:55	G	202.120.[redacted].238	/webdav/udp.php?act=phptools&host=94.23.204.149&time=500&port=80
163	2013-08-29 04:11:55	G	202.120.[redacted]20	/webdav/udp.php?act=phptools&host=94.23.204.149&time=500&port=80
164	2013-08-29 04:11:56	G	202.120.[redacted]33	/webdav/udp.php?act=phptools&host=94.23.204.149&time=500&port=80
165	2013-08-29 04:11:56	G	202.120.[redacted].142	/webdav/udp.php?act=phptools&host=94.23.204.149&time=500&port=80
166	2013-08-29 04:11:56	G	202.120.[redacted].238	/webdav/udp.php?act=phptools&host=94.23.204.149&time=500&port=80
167	2013-08-29 04:11:56	G	202.120.[redacted]33	/webdav/udp.php?act=phptools&host=94.23.204.149&time=500&port=80
168	2013-08-29 20:00:25	G	202.120.[redacted].142	/webdav/udp.php?act=phptools&host=85.236.109.28&time=800&port=12397
169	2013-08-29 20:00:25	G	202.120.[redacted].238	/webdav/udp.php?act=phptools&host=85.236.109.28&time=800&port=12397
170	2013-08-29 20:00:25	G	202.120.[redacted].238	/webdav/udp.php?act=phptools&host=85.236.109.28&time=800&port=12397
171	2013-08-29 20:00:25	G	202.120.[redacted].142	/webdav/udp.php?act=phptools&host=85.236.109.28&time=800&port=12397
172	2013-08-29 20:00:25	G	202.120.[redacted].238	/webdav/udp.php?act=phptools&host=85.236.109.28&time=800&port=12397
173	2013-08-29 20:00:25	G	202.120.[redacted].238	/webdav/udp.php?act=phptools&host=85.236.109.28&time=800&port=12397

部分WebShell 发现

黑名单，特殊Path，特殊URL，特殊文件名，POST频率

结果 查询 日志 列

	time	type	domain	url
47	2013-06-20 10:20:59	G	e.sjtu.edu.cn	/UserFiles/Image/1.asp;1(1).jpg
48	2013-06-20 10:21:03	G	e.sjtu.edu.cn	/UserFiles/1.asp;1(1).jpg
49	2013-06-20 10:21:08	G	e.sjtu.edu.cn	/UserFiles/File/1.asp;1(1).jpg
50	2013-06-20 10:21:14	G	e.sjtu.edu.cn	/UserFiles/Image/1.asp;1(1).jpg
51	2013-06-20 10:21:27	G	e.sjtu.edu.cn	/UserFiles/1.asp;1(1).jpg
52	2013-06-20 10:21:32	G	e.sjtu.edu.cn	/UserFiles/1.asp;1(1).jpg
53	2013-06-20 10:21:37	G	e.sjtu.edu.cn	/UserFiles/File/1.asp;1(1).jpg
54	2013-06-20 10:21:42	G	e.sjtu.edu.cn	/UserFiles/Image/1.asp;1(1).jpg
55	2013-10-16 15:09:47	G	scwr.sjtu.edu.cn	/admin/upload.asp?fuptype=db&fupname=akt.asp;.asp&frmname=akt.asp
56	2013-09-30 10:19:36	G	scwr.sjtu.edu.cn	/admin/upload.asp?fuptype=db&fupname=akt.asp;.asp&frmname=akt.asp
57	2013-06-07 16:35:19	G	sjtu.edu.cn	/a.asp;a.jpg
58	2013-06-07 16:35:19	G	sjtu.edu.cn	/1.asp;1.jpg
59	2013-06-07 16:35:19	G	sjtu.edu.cn	/1.asp;.jpg
60	2013-06-07 16:35:19	G	sjtu.edu.cn	/1.asp;jpg

Zimbra Mail - 0day exploit

This script exploits a Local File Inclusion in

/res/I18nMsg,AjxMsg,ZMsg,ZmMsg,AjxKeys,ZmKeys,ZdMsg,Ajx%20TemplateMsg.js.zgz

which allows us to see localconfig.xml

that contains LDAP root credentials wich allow us to make requests in

/service/admin/soap API with the stolen LDAP credentials to create user

with administration privlegies

and gain acces to the Administration Console.

LFI is located at :

/res/I18nMsg,AjxMsg,ZMsg,ZmMsg,AjxKeys,ZmKeys,ZdMsg,Ajx%20TemplateMsg.js.zgz?v=091214175450&skin=../../../../../../../../../opt/zimbra/conf/localconfig.xml%00

Example :

https://mail.example.com/res/I18nMsg,AjxMsg,ZMsg,ZmMsg,AjxKeys,ZmKeys,ZdMsg,Ajx%20TemplateMsg.js.zgz?v=091214175450&skin=../../../../../../../../../opt/zimbra/conf/localconfig.xml%00

Zimbra 漏洞影响范围

12月6日，exploit-db.com网站披露了攻击代码。根据国家互联网应急中心（CNCERT）和上海交通大学网络信息中心的监测情况，从12月6日下午开始，出现了大量利用该漏洞对境内政府和高校用户发起的攻击。

CNVD对互联网上使用Zimbra的邮件服务器进行了检测。截至12月13日9时，共检测发现互联网上有1814个IP主机为Zimbra邮件服务器，详细统计见下表所示。此外，根据抽样测试结果，约有50.0%的Zimbra邮件服务器存在所述高危漏洞。

国家和地区	数量	占比
巴西	252	13.89%
美国	233	12.84%
中国	181	9.98%
法国	148	8.16%
中国香港	105	5.79%
印尼	68	3.75%
西班牙	58	3.20%
意大利	57	3.14%
德国	51	2.81%
俄罗斯	45	2.48%
印度	42	2.32%
南非	31	1.71%
英国	28	1.54%
智利	24	1.32%
土耳其	23	1.27%
其他	468	25.80%
总计	1814	100.00%

HTTP 关联分析


收藏夹
校园网Web分析系统 v2.0
WEB分析系统
上海交通大学网络信息中心
活跃服务器 | POST统计 | 热搜词统计
202.120.63.180
202.120.63.180
afo.sjtu.edu.cn
amat.sjtu.edu.cn
biomasschem.sjtu.edu.cn
cciidi.sjtu.edu.cn
clrc.sjtu.edu.cn
colp.sjtu.edu.cn
e2-chemicals.sjtu.edu.cn
engst.sjtu.edu.cn
gemba.sjtu.edu.cn
hbchu.sjtu.edu.cn
hcgu.sjtu.edu.cn
iah.sjtu.edu.cn
iasec.sjtu.edu.cn
icectc.sjtu.edu.cn
iconip2011.sjtu.edu.cn

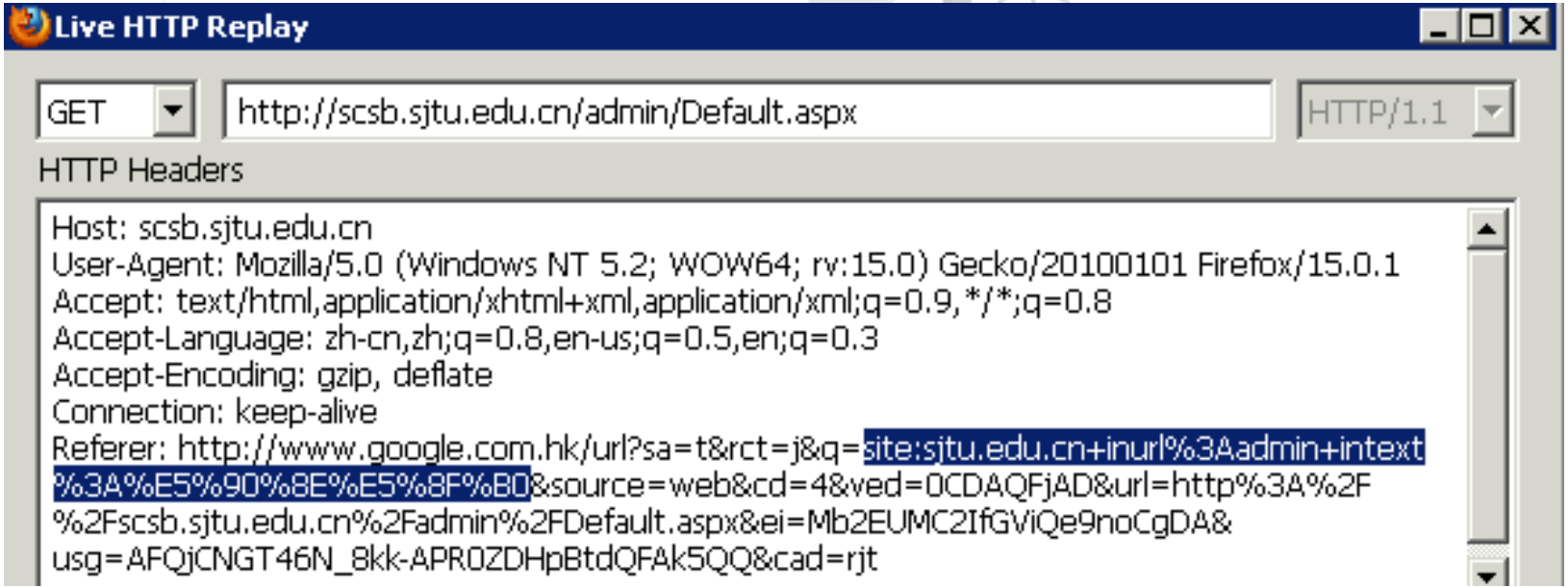

收藏夹
校园网Web分析系统 v2.0
WEB分析系统
上海交通大学网络信息中心
活跃服务器 | POST统计 | 热搜词统计
1 ourex.lib.sjtu.edu.cn 16374
2 news.sjtu.edu.cn 11799
3 gsa.sjtu.edu.cn 10049
4 m.sjtu.edu.cn 8390
5 pqdt.lib.sjtu.edu.cn 6097
6 electsys.sjtu.edu.cn 4733
7 electsys0.sjtu.edu.cn 2891
8 www.jdcw.sjtu.edu.cn 2878
9 xuebao.sjtu.edu.cn 2486
10 www.gs.sjtu.edu.cn 1938
11 mail.lib.sjtu.edu.cn 1408
12 ecard.sjtu.edu.cn 1360
13 cc.sjtu.edu.cn 1304
14 www.sjtu.edu.cn 1299

Referer 分析


文件(F) 编辑(E) 查看(V) 历史(S) 书签(B) 工具(T) 帮助(H)
scsb.sjtu.edu.cn/admin/Default.aspx
系统生物医学研究院网站后台管理系统


== 系统生物医学研究院网站后台管理系
用 户：
密 码：
登陆
取消


Google - Mozilla Firefox
Google
Hong Kong
site:sjtu.edu.cn inurl:admin intext:后台
Google Search
I'm Feeling Lucky


Live HTTP Replay
GET
http://scsb.sjtu.edu.cn/admin/Default.aspx
HTTP/1.1
HTTP Headers
Host: scsb.sjtu.edu.cn
User-Agent: Mozilla/5.0 (Windows NT 5.2; WOW64; rv:15.0) Gecko/20100101 Firefox/15.0.1
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Language: zh-cn,zh;q=0.8,en-us;q=0.5,en;q=0.3
Accept-Encoding: gzip, deflate
Connection: keep-alive
Referer: http://www.google.com.hk/url?sa=t&rct=j&q=site:sjtu.edu.cn+inurl%3Aadmin+intext
%3A%E5%90%8E%E5%8F%B0&source=web&cd=4&ved=0CDAQFjAD&url=http%3A%2F
%2Fscsb.sjtu.edu.cn%2Fadmin%2FDefault.aspx&ei=Mb2EUMC2IfGViQe9noCgDA&
usg=AFQjCNGT46N_8kk-APR0ZDHpBtdQFAk5QQ&cad=rjt

Referer 暴露搜索引擎关键词

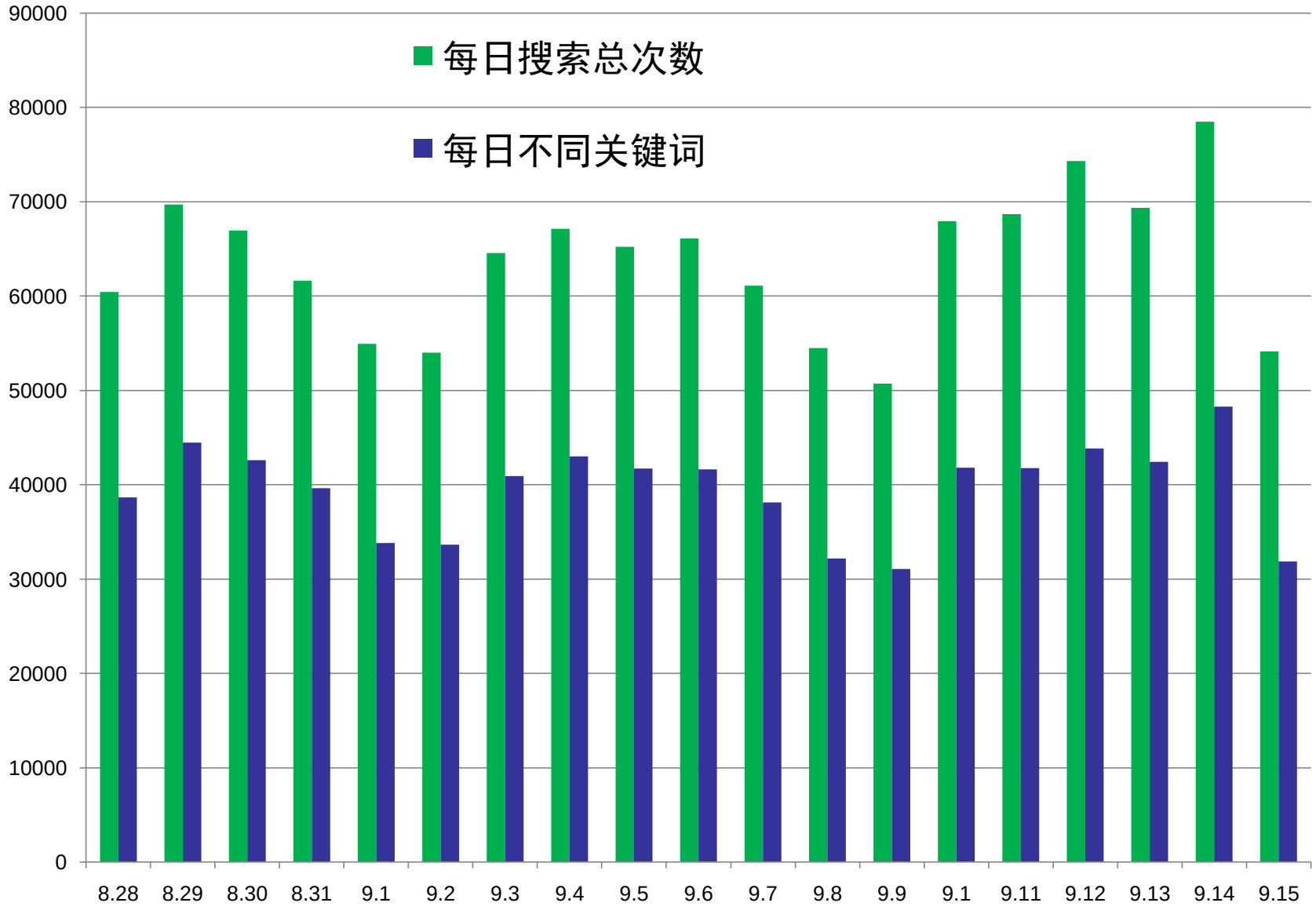
http://www.google.com/url?sa=t&rct=j&q=**world+university+rankings**&source=web&cd=6&ved=0CEcQFjAF&url=http%3A%2F%2Fwww.arwu.org%2F&ei=zluEUP7TNomnhAf_mYGQCQ&usg=AFQjCNHUhmS82v9GFC7ruy03eVrRYQFEkQ

http://www.baidu.com/s?wd=**%C9%CF%BA%A3%B9%FA%BC%CA%B8%BE%D3%A4%B1%A3%BD%A1%D4%BA**&rsv_bp=0&rsv_spt=3&rsv_sug3=9&rsv_sug1=9&rsv_sug4=575&oq=%C9%CF%BA%A3%B9%FA%BC%CA%B8%BE%D3%A4&rsp=0&f=3&rsv_sug2=1&inputT=8751

http://so.360.cn/s?q=**%E6%9E%97%E5%AE%97%E5%88%A9**&pn=5&j=0&_re=0

http://www.bing.com/search?q=**dominus%20winery%20plan**&ie=utf-8&from=360&FORM=WENY01&mkt=zh-CN

搜索引擎关键词来源分析



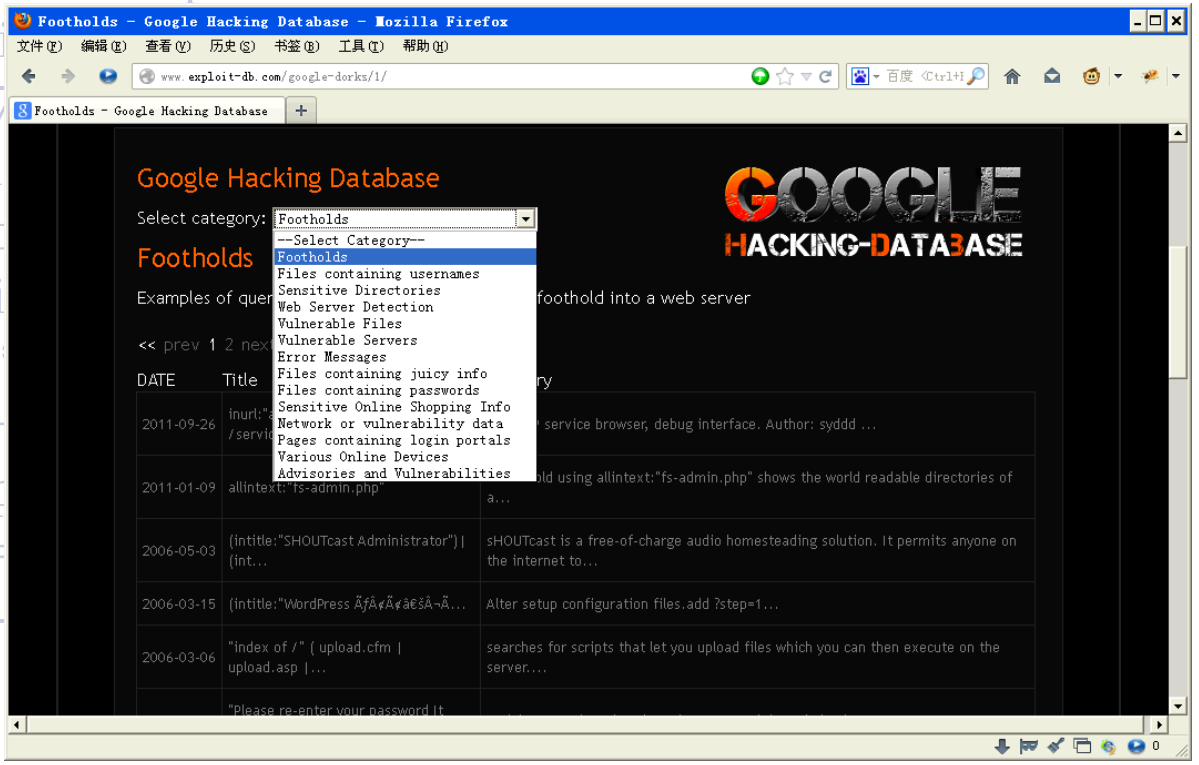
Apache Struts2 漏洞搜索关键词

- inurl:index.action
- inurl:(.action) site:.edu.cn
- site:.edu.cn index.action
- inurl:edu.cn filetype:action
- inurl:index.action
- allinurl:index.action
- intitle:登录 inurl:action site:edu.cn
- intitle:成绩 inurl:action site:edu.cn
- intitle:教学管理 inurl:action site:edu.cn
- inrul:main.action site:edu.cn

……

用正则表达式来发现新型攻击

inurl:printpage.asp?ArticleID=|34------inurl:.*\.asp.*id=
MORPHEUS+inurl:/browse.php?id=|10------inurl:.*\.php?.*id=
inurl:/privacy.html|10------inurl:.*html
intitle:欢迎使用ExtMail|9------intitle:.*mail
inurl:/listguest.asp?no=|8------inurl:.*list
inurl:/content.aspx?id=|7------inurl:.*\.asp.*id=
chanel+2.55+jumbo+flap+bag+
inurl:/members/glenquiny.html|
MORPHEUS+inurl:/post.php?a
inurl:/bloglocal|7------inurl:.*blo
inurl:/forums/thread-view.asp?t
inurl:/sitemap.html|7------inurl:
site:.edu.cn+intext:转到父目录
inurl:edu.cn++毕业设计+filety
inurl:edu.cn+inurl:news.php?id
site:edu.cn inurl:index.action|2-



网络流日志查询

日志格式

App|StartTime|EndTime|SrcIP|SrcPort|DstIP|DstPort|Size

http.tcp 1320155721-1320155731 202.120.2.102:54285-8.8.4.4:80 374 24021

每天超过 4亿条流记录，350 GB

平均 5000 条/秒， 峰值 12000 条/秒

SSH/Telnet/HTTP/HTTPS/FTP/LDAP/远程桌面……

数百种应用类型

Hbase schema设计

- 一个索引表多个冗余表
  - App look-up index table
  - Time table
  - srcIP&app table
  - dstIP&app table
  - srcIP&dstIP&app table
  - App&time table
- 各个表用途
  - App look-up table：索引，用于查找指定srcIP，指定dstIP或指定srcIP+dstIP组合所使用过的所有app。
  - Others：根据指定条件检索log内容。

GUI 设计和交互操作

- GUI设计 Syslog网络日志分析系统——查询模块

- 交互操作
  - 用户填写开始时间（必填）、结束时间（必填）、客户端IP（选填）、服务器IP（选填）在相应的文本框；
  - 浏览器实时在APP的下拉框中返回满足文本框中所有条件的所有APP；
  - 用户选择一种APP类型（也可以不选）；
  - 用户按submit键提交；
  - 浏览器返回给用户该查询条件下的所有日志记录。
  - 如果输入某指定条件后，APP菜单中没有下拉选项，则说明系统中没有满足这一条件的记录。

借鉴 IPAudit 的思路

- http://ipaudit.sourceforge.net/

下面为互联网上的站点演示，可以查看其功能。

IPAudit - Log Search

Home

)ata Available from 2013-06-29-00:00 to 2013-12-27-10:30.

Search Form

Submit	Submit Form	
Start Date:	2013-12-26-00:00	*Eg: 2002-03-13-12:30*
End Date:	2013-12-27-00:00	
IP Address:		
Local Port:		*Eg: 21,23*
Remote Port:		*Eg: 21,23*
Max Lines Displayed:	100	*Eg: 200*
Print Incr:	2	*Eg: 2*
Min Session Size:		*Eg: 200, 2k, 1G*
Max Session Size:		*Eg: 200, 2k, 1G*
Protocol:	any	
First Talker:	any	
Last Talker:	any	

Local IP	Remote IP	Proto-col	Local Port	Remote Port	Incoming Bytes	Outgoing Bytes	Incoming Packets	Outgoing Packets	First Packet Time	Last Packet Time	First Talker	Last Talker
192.168.203.104	210.71.222.236	tcp	63593	80	781.6k	10.4k	538	198	00:00:01.2016	00:00:35.0680	L	R
192.168.102.69	118.163.106.191		0	0	17.36M	9.88M	75996	86729	00:00:01.2016	00:29:02.5450	R	R
192.168.108.96	199.9.249.168	tcp	51342	80	32.55M	783.3k	22547	14854	00:00:01.2017	00:01:18.1252	R	R
192.168.209.135	172.168.10.1	udp	19784	9900	0	189.0k	0	3225	00:00:01.2018	00:30:01.3748	L	-
192.168.103.96	182.203.86.215	udp	27139	51803	336.7k	43.1k	309	418	00:00:01.2018	00:00:44.2767	R	R
192.168.103.96	218.84.90.194	tcp	63931	13837	25.49M	941.0k	19476	9387	00:00:01.2019	00:07:52.9522	R	R
192.168.202.166	119.160.254.197	tcp	60420	443	11.8k	1.6k	13	10	00:00:01.2019	00:00:15.7537	L	R
192.168.202.166	119.160.254.197	tcp	60419	443	9.9k	1.1k	12	7	00:00:01.2019	00:00:15.7545	R	R

目前正在改进的工作

- 白名单过滤，降低存储分析压力

  基于现有数据的深度统计分析

- 重大安全漏洞的早期大规模预报

  CNVD 国家信息安全漏洞共享平台 CHINA NATIONAL VULNERABILITY DATABASE

  CN ERT/CC

  WooYun.org

  Sebug安全漏洞库

- Web 应用漏洞挖掘和安全威胁发现

展望未来

- 大规模网络安全态势感知和监测
- 多方来源安全事件的关联分析
- 满足百亿级的安全数据处理实用化
- 和安全研究机构的多方交流，情报交换
- 和云计算团队的继续深度合作

谢 谢！

kaida@sjtu.edu.cn