Securing Carrier VoIP: Session Border Control

Hadriel Kaplan
hkaplan@acmepacket.com
NANOG 34
May 15-17th, 2005

Agenda

- Overview
- The Problem
- The Solutions
- Notes from the field

*Note this presentation is mostly from a VoIP Service Provider application viewpoint

The VoIP World Today

- Free p2p VoIP is nice, but…
  - most people need to reach wireline and wireless sets
    - most of the planet is still POTS or Cellular
  - Grandma doesn't keep her PC on all the time
- Carriers + Enterprise are deploying VoIP at an incredible rate
  - Class-5 replacement, peering, IP Centrex, Cable voice, PBX replacement, converged access, etc.
  - PCMM, IMS/TISPAN, MSF, etc.
- Almost every Tier 1-3 Carrier does VoIP today, in some form

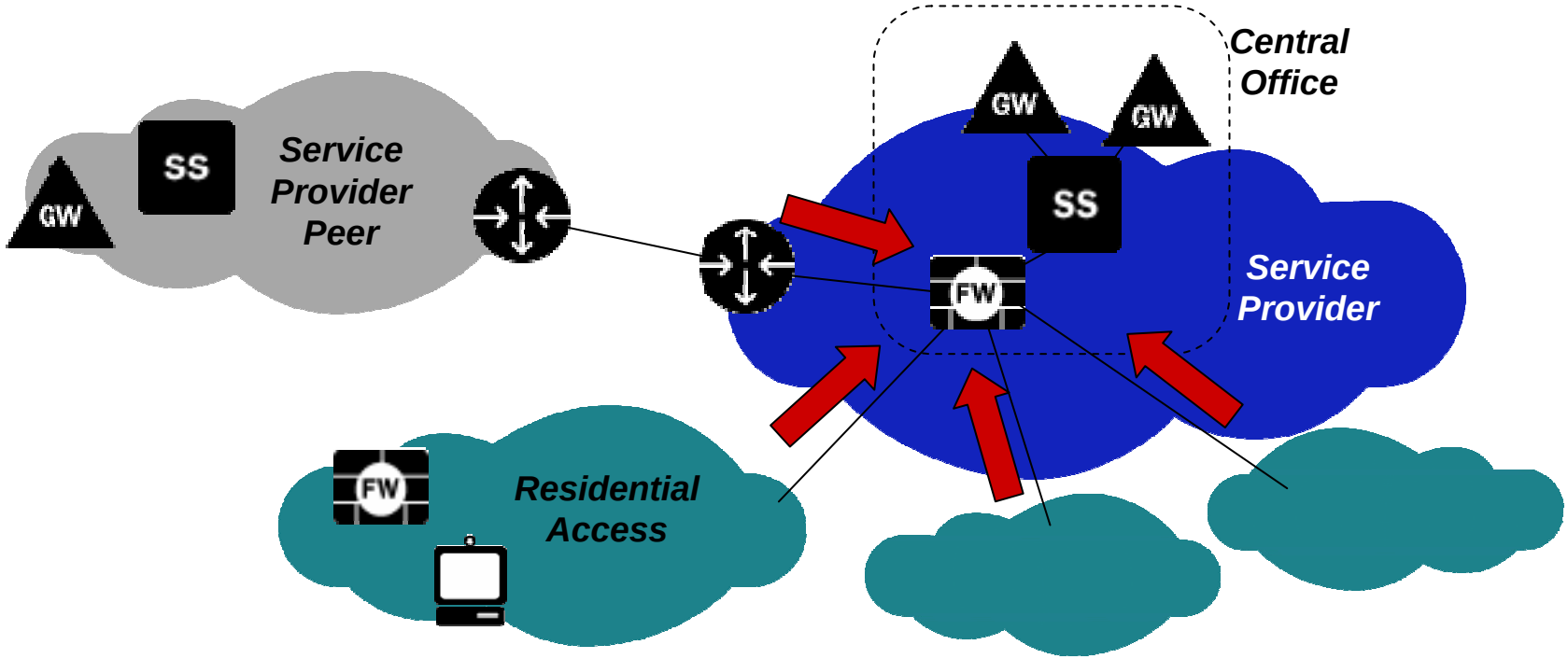
The Problem

- VoIP service becoming a more prominent target for attack
- Service provider multimedia infrastructure is susceptible to attack
  - Softswitches, proxies, gateways, app servers, etc.
  - software-based boxes have issues, but even PSTN gateways are susceptible
- And it’s not just “attacks” – it’s overloads too
- Loss of VoIP service is more than just loss of revenue
  - Customer defections, tarnished brand reputation, legal responsibility issues

The Real Problem

- SIP/UDP is open to all (we want it to be!)
  - TLS-based someday, but few today do that
    - Even TLS doesn't stop DoS Attacks
      - It's implemented in software or with CPU-level accelerator
  - IPsec is used in 3GPP, but rarely elsewhere in voip
  - Service provider rarely knows or limits source addresses
    - Except in peering, and even then it can be spoofed
  - Even with digest auth., it's not hard to attack the SIP port and overload a server on the front end
  - DoS or DDoS attack may not bring server down, but may bring *service* down (which is the same to users)

The First Solution

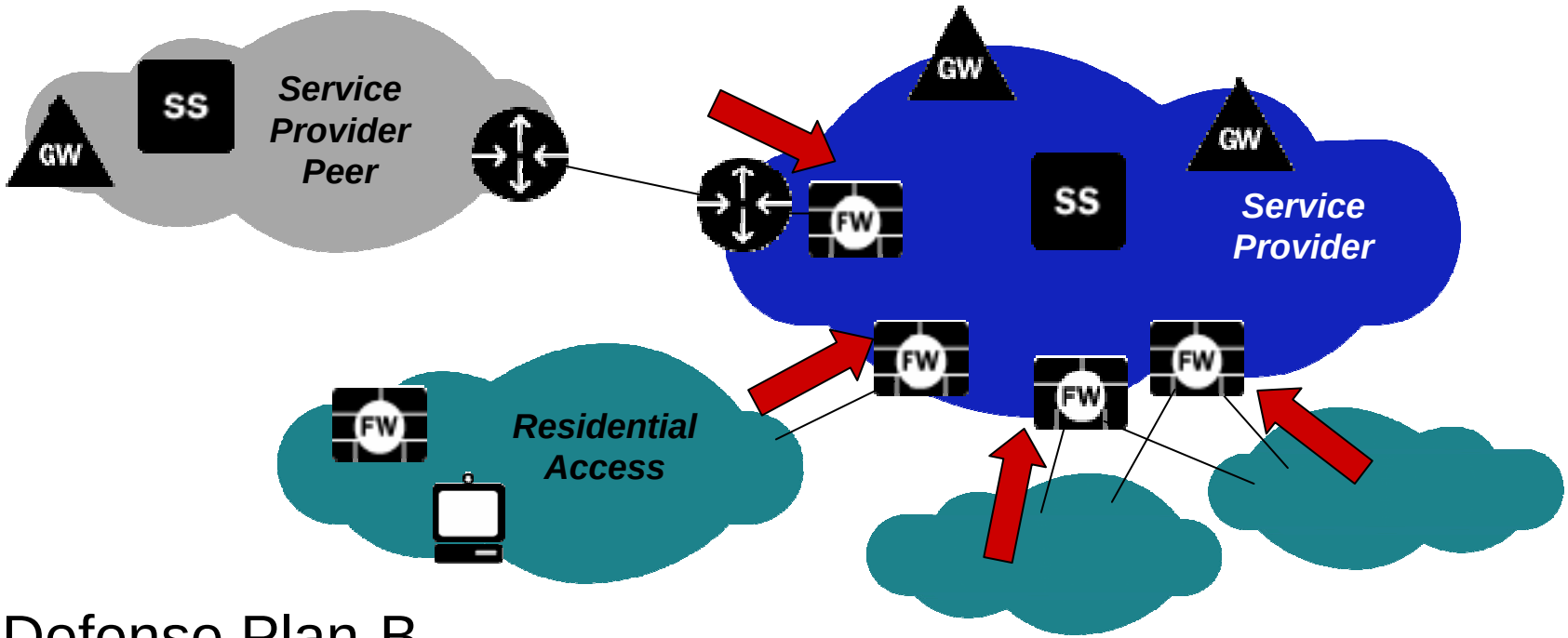


- VoIP Central Office model (3 years ago): Firewall + Router ACLs
  - But DDoS attacks, overloads, etc. very easy on firewall
  - Had problems if two or more paths out of CO (w/ 2 firewalls)
  - Very inflexible design constraints

Firewalls

- Provided rudimentary screening of SIP
  - Reversed the normal FW model
    - connections/sessions begin on outside (public) going into private, so filter rules had to be relaxed
  - Some had built in ALG to learn RTP pinholes to open (others had to leave ranges open)
    - But return routing path couldn't be guaranteed to flow through the same Firewall, which caused problems for pinholes
    - Media is not the major attack point – signaling is
  - If deployed at edge, created routing problems if not inline
    - But couldn't be inline due to performance
    - So carriers used static routes and policies

The Second Solution



- Defense Plan-B
  - Deployed Firewalls at borders
  - Better protection against simple attacks – divide the flood
  - Still couldn’t stop overload of infrastructure, or overload of ACLs
  - Created routing problems – couldn’t guarantee return
  - Softswitches and gateways still exposed (publicly reachable)

Firewalls (cont.)

- Some had built in parsers to verify SIP packets
  - But no throttle to slow down overload/attacks on servers
  - Couldn’t tell authorized/good users from bad
- Provided no screening of RTP
  - Didn’t know when to close pinholes, or handle 3PCC, or handle hair-pinned calls
  - Didn’t police the RTP+RTCP holes
- Couldn’t handle users’ home NATs
  - Internal SIP addressing was wrong

Traditional Attacks

- Unknown Protocol
- ARP Flood ("Poink Attack")
- IP Stream Option
- IP Spoofing
- IP Source Route Option
- IP Short Header
- IP Malformed Packet
- IP Bad Option
- IP Address Session Limit
- Fragments - Too Many
- Fragments, Large - Offset
- Fragments - Same Offset
- Fragments - Reassembly w/different offsets ("Tear Drop")
- Fragments - Reassembly w/ different offsets and padding ("Newtear Attack")
- Fragments - Reassembly w/ different offsets and oversize ("Bonk/Boink Attack")
- Fragments - Reassembly off by one IP header ("Nestea Attack")
- Fragments - flood initial fragment only ("Rose Attack")
- Fragments – Deny
- IGMP oversized packets ("Bomba Attack")
- IGMP oversized fragments ("Fawx Attack")
- IGMP TH_SYN and TH_ACK fragment flood ("Misfrag Attack")
- ICMP Source Quench
- ICMP Mask Request
- ICMP Large Packet (> 1472)
- ICMP oversized packet (> 65536) ("Ping of Death/SSPing Attack")
- ICMP Info Request
- ICMP incomplete Fragment ("Jolt Attack")
- ICMP Flood
- ICMP broadcast with spoofed source ("Smurf/Pong Attack")
- ICMP error packets flood ("Trash Attack")
- ICMP spoofed unreachable ("Click Attack")
- ICMP spoofed unreachable flood ("Smack/Bloop/Puke Attack")
- TCP Packets without Flag
- TCP Packet, Oversized
- TCP FIN bit with no ACK bit
- TCP Packet with URG/OOB flag ("Nuke Attack")
- TCP SYN Fragments - Reassembly with overlap ("Syndrop Attack")
- SYN Fragment
- SYN Attack w/IP Spoofing ("Land Attack")
- SYN Attack ("SYN Flood")
- SYN and FIN bits set
- Scan Attack – TCP Port
- UDP spoofed broadcast echo ("Fraggle Attack")
- UDP attack on diag ports ("Pepsi Attack")

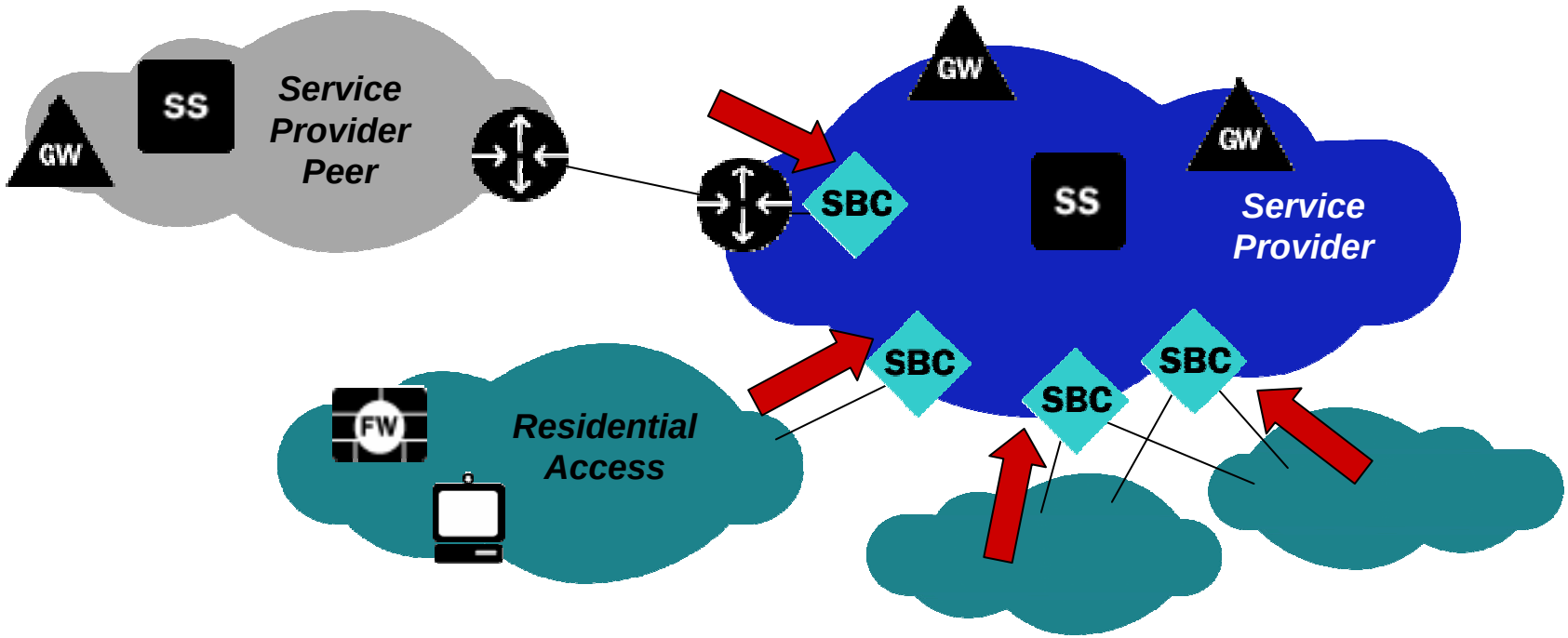
VoIP Attacks – the new threat

- UDP Short Header
- UDP Flood
- RTP rogue packets (after-call)
- RTP flooding during call
- RTP flooding attack
- RTP spoofing
- RTCP flooding
- RTCP spoofing
- MGCP RSIP malformed packet
- MGCP RSIP spoof
- MGCP RSIP flood
- MGCP CRCX malformed packet
- MGCP CRCX spoof
- MGCP CRCX flood
- MGCP malformed packet
- MGCP message spoof
- MGCP message flood
- SDP malformed contents ("Protos Test")

Not all SBCs can protect against all these, but many can at least mitigate these to only affecting them and not the infrastructure boxes

- SIP malformed packet ("Protos Test")
- SIP request message flood attack
- SIP response message flood attack
- SIP Invite spoof
- SIP Register spoof
- SIP Register flood attack
- SIP request spoof
- SIP response spoof
- SIP end-call attack
- H.323 H.225.0 malformed Setup packet ("Protos/NISCC Test")
- H.323 H.225.0 malformed packet
- H.323 H.225.0 Call Signaling spoofing
- H.323 H.225.0 Call Signaling flood
- H.323 H.245 malformed packet
- H.323 H.245 DTMF spoof
- H.323 H.245 DTMF flood
- H.323 H.225 RAS malformed packet
- H.323 H.225 RAS spoof
- H.323 H.225 RAS flood

Solution Phase 2 (Now)



- Session Border Controllers started appearing in Peering connections 2 years ago
  - They virtually dominate the VoIP Peering role now
  - Starting to dominate the Access role
  - Over a dozen vendors offer some form of SBC, for every market

What is an SBC?

- Session: real-time, interactive communications using SIP, H.323, MGCP, H.248
- Border: IP-to-IP network “borders”
  - Service provider-customer/subscriber
  - Service provider-service provider
- Controller: authentication, authorization, admission, attack protection, overload protection, Lawful Intercept, interworking, protocol fixing, etc.
- SBCs control both signaling and media as a B2BUA, B2BGW, etc.
- SBCs offload some other proxy/GK work, and provide many other non-security benefits, but this presentation is on security (and only high-level part of that)

Why SBC?

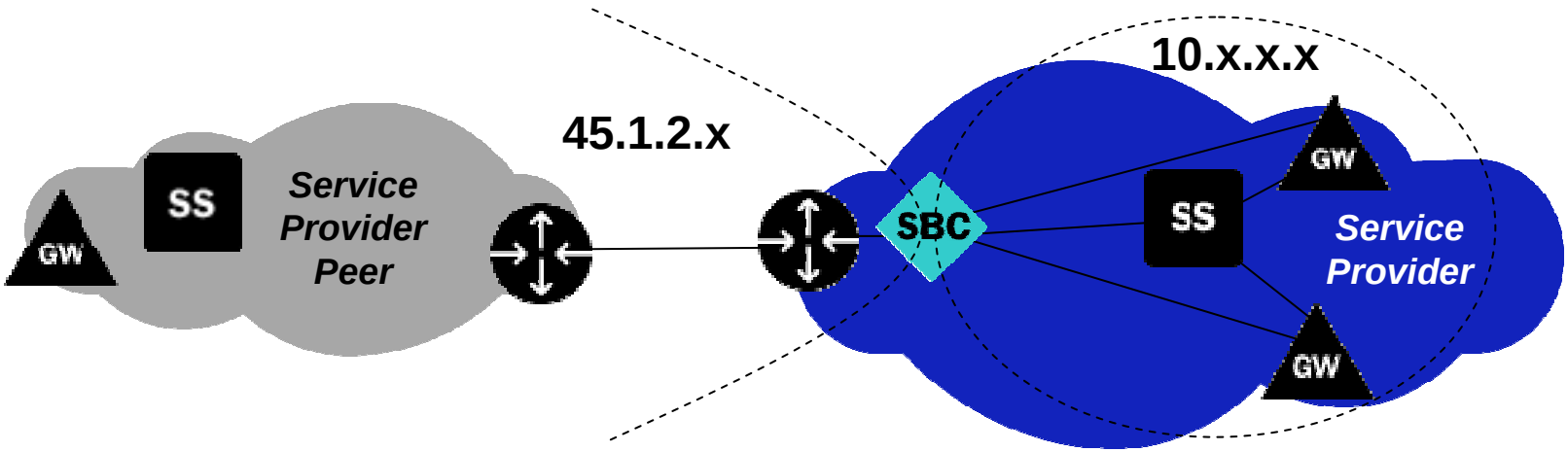
- The idea seems simple enough, why can't a Firewall do it?
  - Of course it could, if it had the right HW and SW
  - A Linksys gateway could also be a core router, if it only had the right HW + SW
- What's special about SBC Hardware?
  - Twice-NAT in HW at line-rate
  - Perform UDP-based DoS attack protection in HW
  - Police 10s or 100s of thousands of signaling and media flows, in HW at line-rate
  - Measure RTP audio/video quality, monitor RTCP reported values, perform rfc2833 translation, etc., in HW at line-rate
  - CALEA lawful intercept in hardware

What's So Special About SBC SW?

- Being a B2BUA for SIP provides far more security than an ALG
  - Keeping session state to enforce SIP behavior
  - Call-gapping to prevent overloading core
  - There is no deeper packet inspection ability than being the packet receiver/originator
  - Get chance to insert/remove/modify headers and fields
  - Complete topology hiding, even for BOTH sides
- Ability to handle overlapping address ranges, home NAT traversal, attack signature matching, fraud protection, audit trails, emergency override, CALEA wiretaps
  - All security functions for VoIP that need extra software

What's This Twice-NAT thing?

- One of the first problems SBCs solved was how to fix/peg SIP+RTP to follow the same path, and at the same time provide basic security
  - Answer: completely replace both source and dest addresses in L3-7, and hide internal SIP/RTP topology
  - Internal VoIP equipment can be in private or un-advertised address space
  - External VoIP equipment gets represented by SBC's address, so internal equipment always sends packets back to it

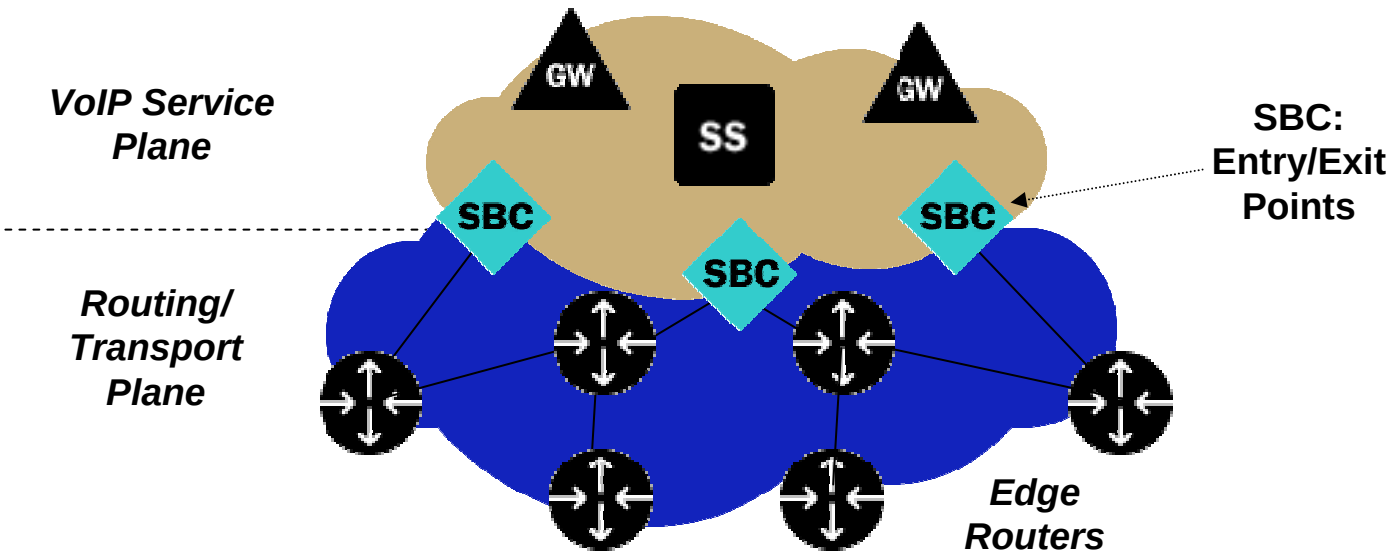


Twice-NAT: So What?

- With VoIP, a single session can move around to different internal/external nodes for both signaling and media
  - Firewalls and ALGs can't handle that – they need a DMZ
  - They also can't guarantee the return path flows through the same Firewall
- Twice-NAT means both sides use the SBC as their proxy/next-hop
  - Easy to monitor/manage/troubleshoot
  - IP or MPLS routing is only to/from the SBC addresses
- They never learn address of other side
  - Hard to attack what you don't know
- Internal VoIP addresses are not advertised
  - Harder to attack what's not reachable
- It also means your peering partners never learn the addresses of your customers, and vice-versa

Secure Entry Points

- The SBC provides the only path in/out of the VoIP Service plane
  - Creates a fixed security border, in a virtually open transport network



What about Router ACLs?

- They’re nice, but fairly useless
  - Protecting signaling boxes:
    - You can and should block all traffic except for port 5060 (or 1719/1720, etc.)
    - But that’s the exact port that will be attacked
      - In fact, if you can, use some port other than 5060 for SIP
    - So you can throttle it, but that will just make it easier to fill up with a DoS attack, or in call overload cases (very Busy Hour)
  - Protecting media:
    - As Microsoft has said on their Knowledge Base since 1999: “To establish outbound NetMeeting connections through a firewall, the firewall must be configured to do the following: Pass through primary TCP connections on ports 522, 389, 1503, 1720 and 1731. Pass through secondary UDP connections on dynamically assigned ports (1024-65535).” Same would be true of a router ACL.

A New Security Model

- Most Firewalls are public device agnostic:
  - They allow most anyone to communicate with the DMZ or default private servers
  - Remember this is the untrusted side creating the “connection” in to the trusted
  - SIP/UDP doesn’t really have a “connection” other than a session/dialog, which is SIP layer
- SBCs can actually create a trust relationship with public devices
  - E.g., based on their successful Registrations with a Registrar
  - The “trusted” public devices can then be given access to make calls, or more calls, or whatever
  - This is just one example of how being a B2BUA and having SIP intelligence adds security value

Notes from the field

- Some generalized notes slides, rather than specific deployment lessons because:
  - Security functions + designs are not discussed publicly by most carriers
  - No one wants to advertise attack risks, occurrences or problems with VoIP
  - From a marketing perspective, security is the last topic you want customers to worry about with voice service

Peering Notes

- Traffic volume growing, but still not huge per PoP
  - About 1-4k simultaneous calls avg. per PoP
  - Both SIP and H.323 still used
- Most dialing/routing plans still configured statically
  - TRIP still not used by anyone, nor ENUM much
  - So many peers use SBCs, that it's often only one or two VoIP next-hops per peer
- Attacks on Peering points not frequent
  - Peering points allow for simpler ACLs+policies
  - SIP often done over TCP at peering points or over a VPN/IPSEC tunnel if not directly adjacent
  - Most people don't know where they are and can't reach them
    - Use of SBCs obscure the "hops" info so it's not easy to find out
    - Rare for users to see their addresses in any messages

Access Notes

- SIP growing, but H.323 and MGCP still used
- Traffic+user volume booming
  - Vonage-style service demands are great
    - 20k subscribers + 2k simult. calls today per PoP average
    - Some PoPs at 10 times that already today
  - Home user NAT issues:
    - NATs have varying cache timeouts, from 30s to 30min.
    - Symmetric/restricted NATs do exist at home
    - STUN/TURN/ICE still not common (and have many issues)
    - Receive-only phones don’t open the NAT hole
    - Ironically, Home-NATs actually help protect against spoof attacks, because the NAT can change the source port from 5060 to something ephemeral – which is harder to spoof

Access Notes (cont.)

- The good news: most reported “attacks” so far have been unexpected overloads
  - Apartments or whole neighborhoods coming back online from power outage
  - American-Idol busy hour calls (someone still watches that show?)
  - Badly implemented device behaviors
- The only widely-reported (CERT) vulnerabilities have been with malformed packet handling
- The bad news: many carriers don’t want to publicly report VoIP service security issues

Where to go from here?

- Talk to the vendors
  - Ask about what's in SW vs. HW, attacks protected against + how, etc. (software "hardening" is not enough)
- VOIPSA (VoIP Security Alliance) just getting started – www.voipsa.org
  - Will create test plans, security requirements, and best common practices
- We need to start collecting attack statistics
  - Anonymize the targeted carrier/customer
  - Need better attack/test tools (don't wait for the hackers to write them)

Test it: tools available to test VoIP can also be used to attack it

- SipP: open-source for unix+windows
  - Meant to be a protocol test tool, not threat
  - Can generate ~2500 Invites/sec.
- Nessus/NeWT: open-source for unix+win
  - Vulnerability analysis tool (very well known)
  - Can generate ~500 Invites/sec.
- Others: Protos, Sivus, Sipbomber, SipSak
- Anyone can write a program to generate 25,000 Invites/sec. (I did on my Windows P3 notebook, and I'm not a programmer)
  - I don't know of a softswitch that can handle 25k Invites/sec
  - They may not crash, but they will be 100% busy – same effect
- Plus all the publicly available IP/UDP attack scripts can fill port 5060

Q & A