Define shaping, please?

- **Working definition:** Active implant copies traffic and directs a copy past a passive collector
  - Issues arise when collector is also processing passive traffic simultaneously

- **Current:** Implants on network infrastructure devices, not user endpoints

- **Two types:**
  - **Physical/link layer:**
    - an implant copies and shapes an entire link (E1, STM1) without selection; passive midpoint does selection
  - **Network layer:**
    - an implant performs *targeted* copying based on IP or application parameters and exfils only the targeted traffic; passive collector may or may not do further selection.
Examples

• Link layer: BRAVENICKEL project (optical Muxes)
  – Copied link is not disguised, just routed on an unused layer 2 path that a passive collector can monitor
  – Selection happens in the passive collector

• Network layer: APEX for HAMMERMILL (routers)
  – Router is tasked to select and exfil targeted traffic (perhaps all of a particular protocol)
  – Exfil is disguised (“munged”, encrypted) to avoid detection
  – Passive collector looks for IP source/destination address in order to detect the traffic
  – If further selection/processing is to be done in collector, the exfil must be “unwrapped” (unmunged, decrypted)
  – Exfil can be directed to passive or to TAO by changing the destination address
So Why does Jane the Analyst care?

- TAO implants have collection parameters that are put on exfil received thru TAO backend
  - case notation, SIGAD, PDDG, classification/legal authority
- The passive collector has another set of these:
  - Site has a SIGAD, collector has a PDDG, the link it sees the traffic on has a case notation, and the access has a classification floor/legal authority
- Current backend repositories and presenters weren’t designed to expect TWO of these!!!
- Which gets put on the data?? And where?
- And (drum roll) … how do we solve this problem CONSISTENTLY across the enterprise?
Example: APEX IPSEC VPN collection

• IPSEC VPN:
  – First packets between the devices establish the parameters and encryption keys (IKE)
  – Following this setup, “content” packets are encrypted and transmitted packet by packet (ESP)
  – CES wants the IKE exchange and maybe the ESP (content)

• TURMOIL passive capability:
  – Passive capability to detect IKE and ESP
  – Metadata record produced for *every* IKE exchange
  – IKE for *targeted* VPN forwarded directly to CES database
  – For *targeted* VPN, real-time decryption is performed IF CES can provide a key in time
  – Decrypted IP traffic is processed by TURMOIL apps for normal selection (VoIP, webmail, etc, etc)
APEX capture of active exfil:

- HAMMERMILL implant collects VPN traffic and sends “past” TURMOIL
  - Spin 15: All IKE for SIGDEV (TOYGRIPPE)
  - Beyond: All IKE plus targeted content
- TURMOIL detects traffic, unwraps TAO wrapper
- Unwrapped IP packets are fed back into the front of the VPN processing
• Now app streams (VoIP, webmail, etc) extracted from the tunnel carry two case notations
• Which gets put into metadata records?
• Both can be carried to PWV – but what happens after that?
• Not to mention...
  – Metadata records about VPN being stored in TOYGRIPPE
  – CES database storing IKE exchange
Example: TOYGRIPPE metadata record

• Current fields:
  – caseNotation – searchable field
  – sourceID – “The SIGAD of the site that provided the data”

• APEX proposed extension: add
  – Agent CaseNotation
  – Agent ID (UUID)
  – Passive CaseNotation

• Which caseNotation goes into searchable field?
  – Passive records won’t have the APEX block
  – TAO-collected records (returned via TAO, not passive) won’t have the APEX block
Shaping is happening now

- Operational (or coming soon) shaping:
  - HAMMERSTONE - TCP traffic to FORNSAT, soon SSO
    - No TURMOIL involvement
  - BRAVENICKEL – one operational flow – past SSO site
  - APEX – VPN metadata by end of June

- *Independent* decisions being made about how to stuff the double metadata into legacy databases
So what is your job here?

• How do you want to identify the source of your data?
  – Does CaseNotation still make sense in this new world?

• You need to drive processes, systems, & databases toward a CONSISTENT answer

• Transformed systems and tools (METAwave, Marina, etc.) need to be designed to do more than accommodate
  – do “the right thing” (whatever you the analysts think that is)
  – Let me guess – you want everything, don’t you?
Questions?