SSO Collection Optimization

Core SSO Team:
Address Books

- Email address books for most major webmail are collected as stand-alone sessions (no content present*)
- Address books are repetitive, large, and metadata-rich
- Data is stored multiple times (MARINA/MAINWAY, PINWALE, CLOUDs)
- Fewer and fewer address books attributable to users, targets
- Address books account for ~ 22% of SSO’s major accesses (up from ~ 12% in August)

<table>
<thead>
<tr>
<th>Access (10 Jan 12)</th>
<th>Total Sessions</th>
<th>Address Books</th>
</tr>
</thead>
<tbody>
<tr>
<td>US-3171</td>
<td>1488453</td>
<td>237067 (16% of traffic)</td>
</tr>
<tr>
<td>DS-200B</td>
<td>938378</td>
<td>3111113 (33% of traffic)</td>
</tr>
<tr>
<td>US-3261</td>
<td>94132</td>
<td>2477 (3% of traffic)</td>
</tr>
<tr>
<td>US-3145</td>
<td>177663</td>
<td>29336 (16% of traffic)</td>
</tr>
<tr>
<td>US-3180</td>
<td>269794</td>
<td>40409 (15% of traffic)</td>
</tr>
<tr>
<td>US-3180 (16 Dec 11)</td>
<td>289318</td>
<td>91964 (32% of traffic)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>3257738</td>
<td>712366 (22% of traffic)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Provider</th>
<th>Collected</th>
<th>Attributed</th>
<th>Attributed%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yahoo</td>
<td>444743</td>
<td>11009</td>
<td>2.48%</td>
</tr>
<tr>
<td>Hotmail</td>
<td>105068</td>
<td>1115</td>
<td>1.06%</td>
</tr>
<tr>
<td>Gmail</td>
<td>33697</td>
<td>2350</td>
<td>6.97%</td>
</tr>
<tr>
<td>Facebook</td>
<td>82857</td>
<td>79437</td>
<td>95.87%</td>
</tr>
<tr>
<td>Other</td>
<td>22881</td>
<td>1175</td>
<td>5.14%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>689246</td>
<td>95086</td>
<td>13.80%</td>
</tr>
</tbody>
</table>
Address Books

• Enabled in SCISSORS for various SSO sites:
  – JPMQ (metadata: QMPJ) - DS-200B (MUSCULAR) 29 Feb 2012
  – DGOD (metadata: DOGD) - US-3171 (DANCINGOASIS) 13 Mar 2012
  – SPNN (metadata: NNPS) - US-3180 (SPINNERET) 03 May 2012
  – EGLP (metadata: PLGE) - US-3145 (MOONLIGHTPATH) 08 May 2012
Selector Detasks

Emergency Detasks

TOP SECRET//SI//NOFORN
So What?

- Store less of the wrong data
  - 20% reduction (so far) in content to long-term repositories
  - Data still resides at site for SIGDEV
- Increase data variety
  - Hole left by “wrong data” filled with more “right data”
  - More signals and case notations can be tasked at site
- Shifting collection philosophy at NSA
  - “Memorialize what you need” versus “Order one of everything off the menu and eat what you want”

WIKI: https://wiki.nsa.ic.gov/wiki/Collection_Optimization
XKEYSCORE: fingerprint/defeats/atrouter and fingerprint/defeats/atxks
This is another presentation on problems with NSA overcollection.

**SCISSORS (p. 3)**

- JPMQ (metadata: CMPJ) - DS-200B (MUSCULAR) 29 Feb 2012
- DGOT (metadata: TOGD) - US-3171 (DANCINGOASIS) 13 Mar 2012
- DGOD (metadata: DOGD) - US-3171 (DANCINGOASIS) 13 Mar 2012
- SPNN (metadata: NNPS) - US-3180 (SPINNERET) 03 May 2012
- EGLP (metadata: PLGE) - US-3145 (MOONLIGHTPATH) 08 May 2012

**Ownerless address books blocked by SCISSORS (p. 4)**

**Ownerless address books blocked, by points of access (p. 5)**
**Emergency detasks** (p. 6)

![Emergency Detasks Chart]

**SIGDEV** (p. 7)

- Store less of the wrong data
  - 20% reduction (so far) in content to long-term repositories
  - Data still resides at site for SIGDEV

**“Shifting collection philosophy at NSA”** (p. 7)

- Shifting collection philosophy at NSA
  - “Memorialize what you need” versus “Order one of everything off the menu and eat what you want”