

TOP SECRET//SI//NOFORN



Yahoo Webmessenger

- Update data sent to individuals logged into Yahoo's Instant Messenger service online
 - Online contact status, unread emails in Yahoo inbox
 - Usually small sessions (2-4kB)
- Sporadic collection (30,000 – 60,000 sessions per day)
- Intermittent bursts of collection against contacts of targets
 - Large numbers of sessions (20,000+) against a single targeted selector
 - Not collected against the target (online presence/unread email from target)
 - No owner attribution (metadata value limited to fact-of comms for emails, online presence events for buddies)
- Over a dozen selectors detasked in two weeks
 - Because a target's contact was using/idling on Yahoo Webmessenger
 - Several very timely selectors (Libyan transition, Greek financial related)

TOP SECRET//SI//NOFORN

TOP SECRET//SI//NOFORN



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)

Access (10 Jan 12)	Total Sessions	Address Books	Provider	Collected	Attributed	Attributed%
US-3171	1488453	237067 (16% of traffic)	Yahoo	444743	11009	2.48%
DS-2008	938378	311113 (33% of traffic)	Hotmail	105068	1115	1.06%
US-3261	94132	2477 (3% of traffic)	Gmail	33697	2350	6.97%
US-3145	177663	29336 (16% of traffic)	Facebook	82857	79437	95.87%
US-3180	269794	40409 (15% of traffic)	Other	22881	1175	5.14%
US-3180 (16 Dec 11)	289318	91964 (32% of traffic)				
TOTAL	3257738	712366 (22% of traffic)	TOTAL	689246	95086	13.80%

TOP SECRET//SI//NOFORN

TOP SECRET//SI//NOFORN



Buddy Lists, Inboxes

- Unlike address books, frequently contain content data
 - Offline messages, buddy icon updates, other data included
 - Webmail inboxes increasingly include email content
 - Most collection is due to the presence of a target on a buddy list where the communication is **not** to, from, or about that target
- NSA collects, on a representative day, ~ 500,000 buddylists and inboxes
 - More than 90% collected because tasked selectors identified only as contacts (not communicant, content, or owner)
- Identifying buddylists and inboxes without content (or without useful content) an ongoing challenge

TOP SECRET//SI//NOFORN

TOP SECRET//SI//NOFORN



Address Books

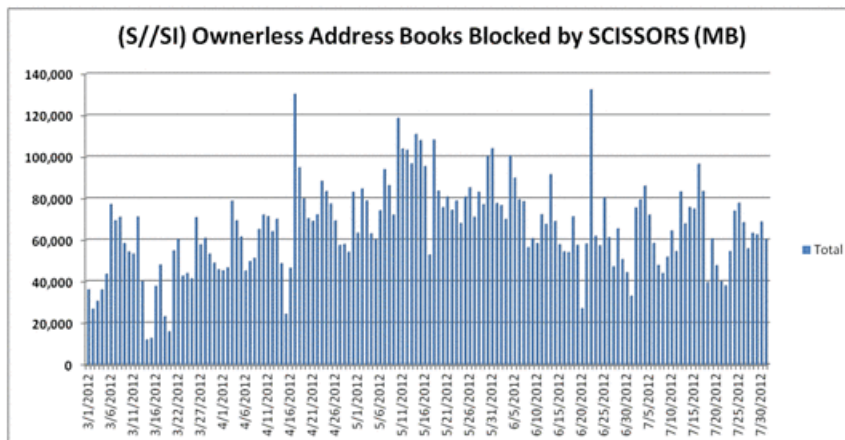
- Enabled in SCISSORS for various SSO sites:
 - JPMQ (metadata: QMPJ) - DS-200B (MUSCULAR) 29 Feb 2012
 - DGOT (metadata: TOGD) - US-3171 (DANCINGOASIS) 13 Mar 2012
 - DGOD (metadata: DOGD) - US-3171 (DANCINGOASIS) 13 Mar 2012
 - SPNN (metadata: NNPS) - US-3180 (SPINNERET) 03 May 2012
 - EGLP (metadata: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

TOP SECRET//SI//NOFORN

TOP SECRET//SI//NOFORN



Address Books

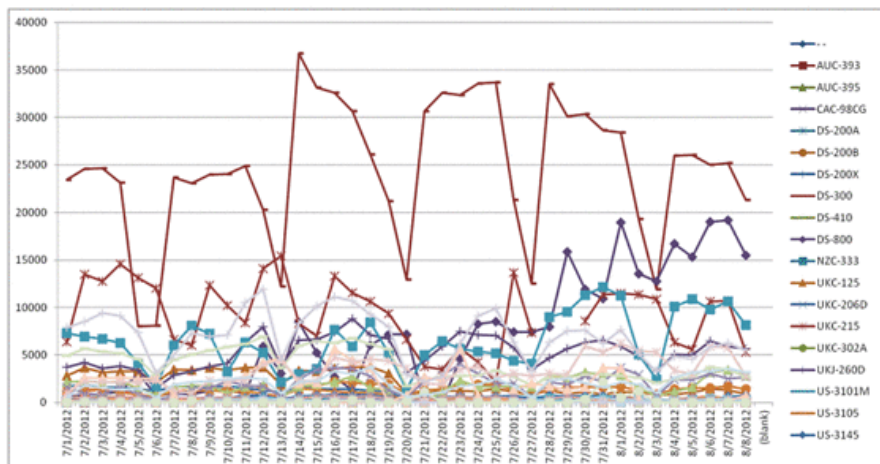


TOP SECRET//SI//NOFORN

TOP SECRET//SI//NOFORN



Address Books

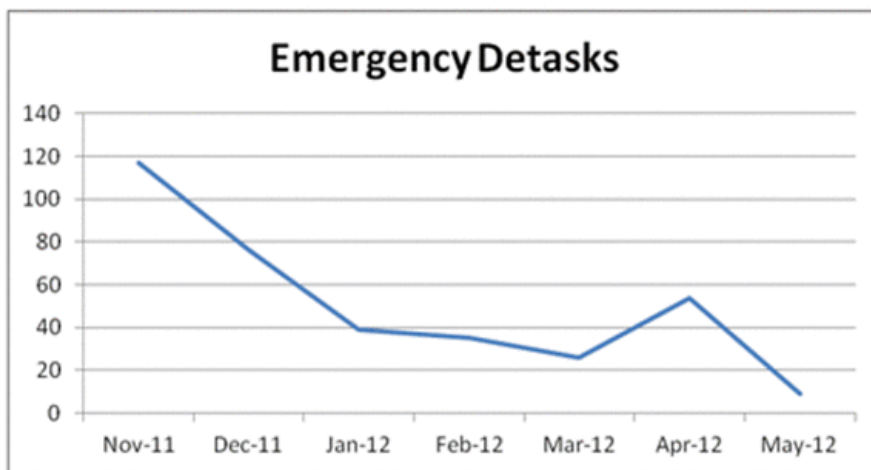


TOP SECRET//SI//NOFORN

TOP SECRET//SI//NOFORN



Selector Detasks



TOP SECRET//SI//NOFORN

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

TOP SECRET//SI//NOFORN