Ingest more events feeds as new accesses come online

Scope and Aims

- Increase maturity and availability of QFDs
- Pull through more QFDs based on Ops priority
- Deliver QFDs capable of holding ‘Convergence’ data and wider event types
- Provide a data mining and collaborative QFD development facility (BLACK HOLE - part of ROUGH DIAMOND)
- Enable sharing of QFD data with 2nd and 3rd Parties
- Interface with visualisation services in FIRE STORM
**What is a QFD?**

**Question Focused Database**

- Designed to answer single analytic question (e.g. ‘where is my target?’)
- Simple table structure compared to traditional multi-function databases (e.g. HAUSTORIUM)
- Pioneered by ICTR, now developed by a community including Next Gen Events, ICTR, SD, GTE, ...
- Additional instances can easily be deployed at new locations or to increase capacity
- No specialised database technologies so simpler to develop and maintain
- Smaller size and lower complexity means easier and quicker to develop and change

SECRET STRAP1
What does each QFD answer?

- **Mutant Broth**: When was my target on line? Where was my target on line?
- **HRMap**: What websites has my target visited? Who's been visiting this dodgy websites?
- **Karma Police**: Who's been posting (vBulletin boards) to this forum?
- **GooBzs (QFD Query Federator)**: What alternative identifiers can I use to search for my target? What has my target been searching for on-line? What part of the world has my target been looking at?
- **Social Animal**: Who is my target interacting with on social networking sites?
- **AutoAssoc**: What is my target doing on-line right now?!
- **Marbled Gecko**: Who's been looking at this suspicious part of the world? What has my target been looking at?
- **Infinite Monkeys**: Who’s been posting (vBulletin boards) activity has my target been up to?
- **Memory Hole**: Who’s been searching for these suspicious things on-line?
- **Samuel Pepys (Coming soon!)**: What is my target doing on-line right now?!
Ingest roadmap

Internet Presence Events from 10G bearers
 eAD events

SALAMANCA Telephony Events
 Initial Converged Events from TERRAINS

MONOPOLY Special Source Events
 Broadband RADIUS events

CDMA2000 test events

Further 10G bearers at RPC1 (Bude)

Feb 10 Mar 10 Apr 10 May 10 Jun 10 Jul 10 Aug 10

Trial part 1 - MUTANT BROTH, INFINITE MONKEYS, HRMAP, MEMORY HOLE from mobile tunnels

Experiment

Explore

Deployed across CPC and RPC1

Trial part 2 - MMS, Blackberry, Google Maps, mobile Hotmail, mobile Gmail from mobile tunnels

Explore

Deployed across CPC and RPC1

Trial part 3 - Hotmail, Gmail, mail RU, Yahoo webmail from internet bearers

Experiment

Explore

Deployed across CPC and RPC1

Trial part 4 - Windows Live IM, Yahoo Mail, SIP from internet bearers

Explore

Deployed across CPC and RPC1

TPS are working with the NGE Project and SMO Mobile theme to produce internet presence and application usage events from within mobile phone ‘tunnels’ in internet bearers. These will be trialed before full operational rollout

‘QFD style’ events will also be produced for types of event traditionally fed into the older HAUSTORIUM and HARBOUR PILOT databases

SECRET STRAP1
Convergence QFDs

This major thread of work will:

- Store events where internet applications are accessed from a mobile device
- Allow analysts to relate mobile device identifiers to internet identifiers such as email addresses
- Enable QFDs to store other more diverse event types, such as telephony events (currently SALAMANCA), and email events (currently HAUSTORIUM / HARBOUR PILOT)
- Interface to LOOKING GLASS visualisation coming soon (in FIRE STORM work package)

Screenshots from evolved MUTANT BROTH web interface, and an export of it’s data to Google Earth
SAMUEL PEPYS QFD

Purpose: Provide a near-real-time diarisation of any IP address

Results

Prototyped by ICTR – Currently being pulled through by ROCK RIDGE, will be scaled to full 10G volumes by May 2010

SECRET STRAP1
What is BLACK HOLE?

- A flat file store housing all data from a wide range of feeds (events and content)
- Provides a set of tools for accessing that data.
- Intended to be the source of events (and limited content) for the development of new QFDs and analytics.
- Contains a rolling 6 months retention
- Part of ROUGH DIAMOND

What does it enable?

- New QFDs to be rapidly prototyped, then to be added to the operational QFD suite
- Trialling of new bulk analysis ideas
- New sources of data to be introduced quickly into existing QFDs.
- Users to look for particular patterns and behaviours (target discovery)
- TR, GTAC and GTE access to more data for research purposes, which may not be QFD related.
User Feedback

'Absolutely FABULOUS well done'
lain Lobban, ref SUPERDRAKE reporting

'it's amazing to see how the pace of delivery in TDB has increased and I have been impressed by your responsiveness to customer needs.'
Senior User

'Almost exactly a year ago I set you the challenge of delivering an upscaled massive events capability ... in order to support Internet Operations being conducted by GCHQ. Through your stripy team working on BLAZING SADDLES, BLUESHIFT and SUPPORTING INO you successfully met this challenge and delivered us a significant new capability in July.'
Deputy Director Cyber Operations

'it's working flawlessly'
Analyst, ref BLACK HOLE

'Bloody awesome'
Analyst, ref SUPERDRAKE QFD

SECRET STRAP1