The Realities of Dealing with Data Retention Mandates

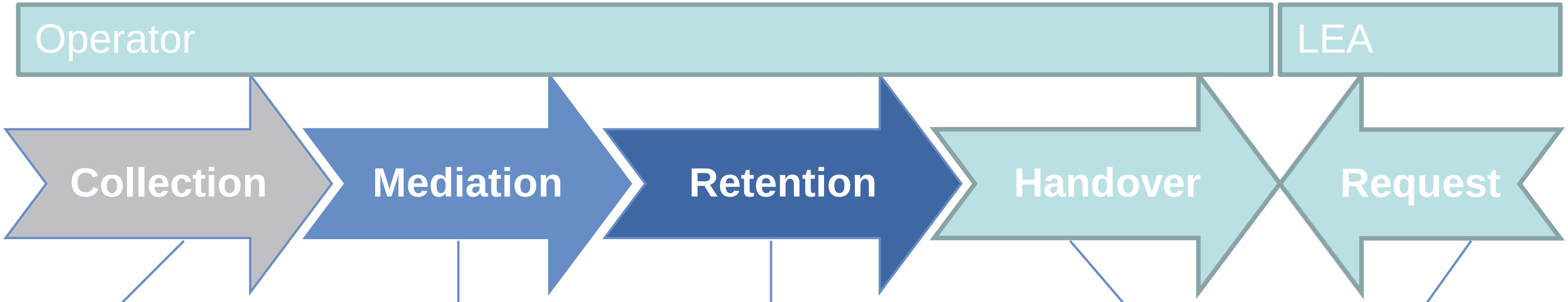
Shane O’Flynn
VP Client Services
Openet

OPENET
transactional intelligence

What is Data Retention?

- **Data retention (DR)** is the lawful storage of specific data sets associated with telephony and internet-based services
- Data that is stored is typically:
  - Subscriber info, Name, Address, Service Type, Subscription Dates, etc.
  - Telephone calls made and received (date/time from/to)
  - Emails sent and received (date/time from/to)
  - Location data
- Service Providers store records pertaining to specific service
- Service Providers retrieve formal requests for DR data for specified requests over specified interfaces from the Law Enforcement Agency (LEA)
- Objective to gather evidence on unlawful activities – to provide LEAs the information needed to investigate crimes

Process Flow



Convergent Mediation (collectors)

Data from Services
Data from Customers

Convergent Mediation

Aggregate
Correlate
Error Check

Match usage data with customer data

Storage (up to 2 years)

Data retention
Security
Backup
Data retrieval
Non-repudiation
Audit

Destruction

Data deletion
Backup deletion
Audit

Query and Retrieval (data lookup)

Query by LEA
Report Delivery to LEA

Request

Data Retention: New Regulation

- March 15, 2006 - the EU adopted Directive 2006/24/EC, for retention of data generated or processed in connection with “publicly available electronic communications services” and “public communications networks”.

- Member States must ensure that operators retain necessary data for between 6 months and 2 years, being able to:
  - Identify the source/destination of a communication
  - Identify the date, time, and duration of a communication
  - Identify the type of communication
  - Identify the communication device
  - Identify the location of mobile communication equipment

EU Directive: Objectives

- Ensure data available for investigation, detection, and prosecution of serious crime
- Harmonize Member State obligations
- Applies to traffic and location data needed to identify subscriber or registered users
- Permit LEAs to access and use such data without undue delay
- Retention period between 6 months and 2 years
- Makes no reference to technology

Challenges for DR – Commercial

- No business reason for "DR"
  - Circuit switched "retention" was simply CDRs
  - DR may in some countries not be used for business reasons
- Increasing volumes of traffic put pressure on operators seeking to retain data
  - Especially when many operators do not mediate, charge, or bill for IP data on a usage basis
  - Lack usage information and storage mechanisms
- New tools needed to ensure the same investigative abilities available in PSTN (e.g., telephone number identity and associated call records
- Unclear or missing legislation
- Feature creep

Billing Model Changing

- Move to flat rate billing for data services
  - Operators seeking to impose caps rather than try count view into the stream
  - No commercial driver to perform expensive DPI of data services

- Account structure are becoming more sophisticated
  - No longer just pre or post-paid subscribers
  - Expect hybrid, shared, family, corporate and transient accounts

Challenges for DR - Technical

- Sophisticated targets seek the “anonymity” of the internet or pre-paid
- Nomadic targets access many different types of networks with different identities (IP address, MAC address, SIP URL, email address, IMSI, TN, etc) , creating correlation challenges.
- Transform diverse network traffic into a useful record:
  - Support multiple networks – wireline, broadband IP, wireless, etc.
  - Collect from multiple data sources
  - Correlate data from multiple sources
  - Quickly store retained records
- Cost-effective DR management
  - Compliant to national regulations for Data Protection
- Store DR data securely and efficiently
  - Fully integrated to Service Provider O&M
- Prevent impacts on day-to-day network operations
- Manage and execute warrants in a timely manner
- Distributed networks (separation of access and service domains)

One of the Greatest Challenges: Volume and Complexity of Traffic and Data Formats

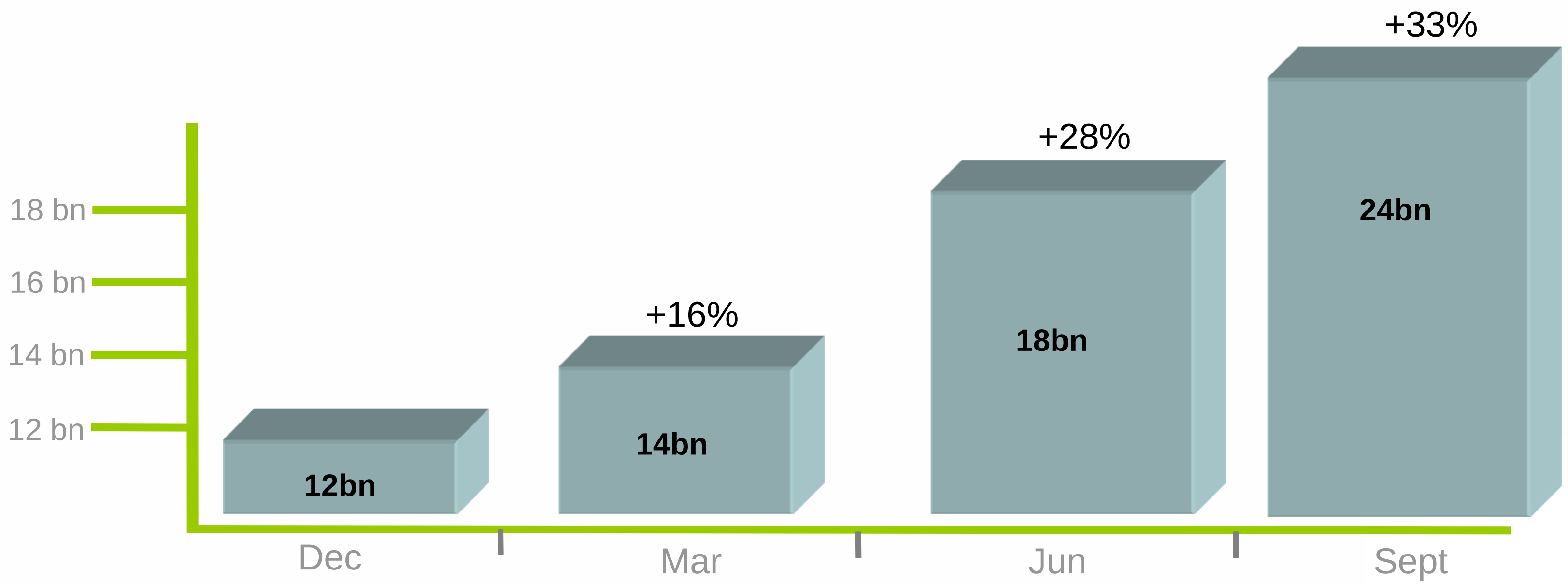
- Difficult to compile user transaction data for all activities and all services
- IP traffic generates at least an order of magnitude more records than circuit switched traffic
  - One phone call typically produces one call detail record, one IP-based session produces tens or hundreds of records
  - Records can arrive out of sequence and are regularly incomplete
  - The number of potential identifiers for each device may be different
  - Challenge to correlate the identifiers associated with an individual’s traffic across multiple wireline and wireless phone numbers, e-mail addresses, SIP addresses, MAC addresses, etc.

Provides a single unified view of each individual's data and

© Copyright 2008 Openet – Company Confidential –
For Use Under Non-Disclosure Only

Growth in Messaging Q-on-Q - SMS

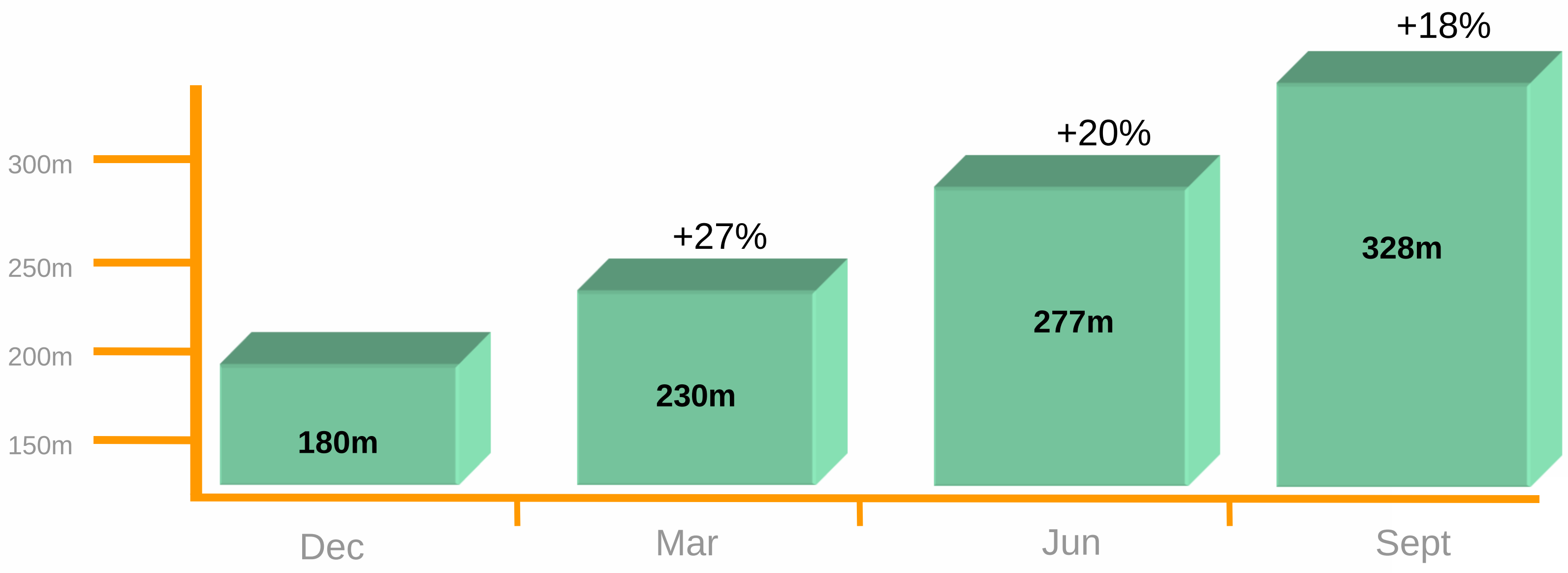
- 2007 SMS Q-on-Q Growth



Source AT&T

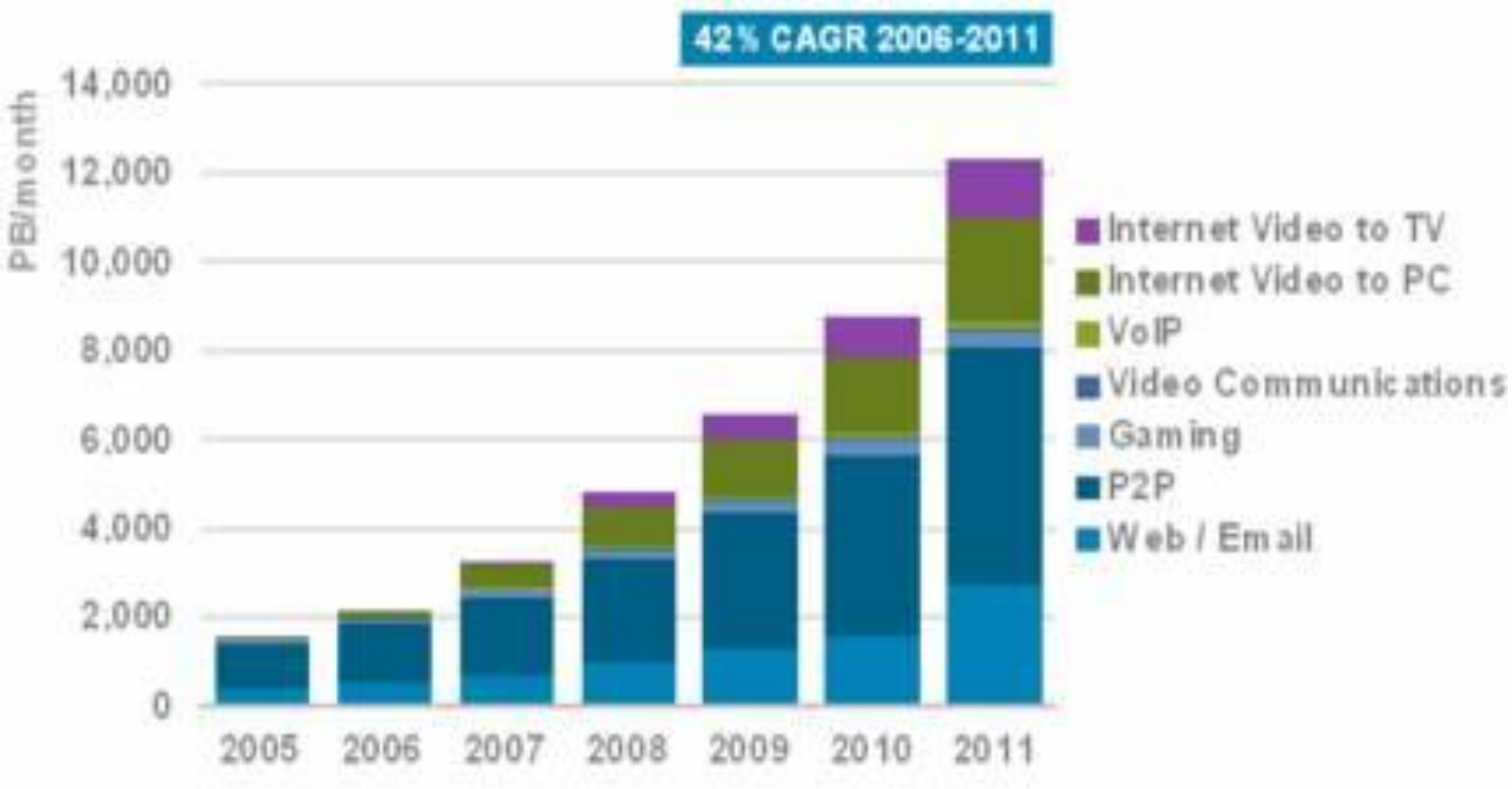
Growth in Messaging Q-on-Q - MMS

2007 MMS Q-on-Q Growth



Source AT&T

Data Volume Growth



Source Cisco

Growth In Number Of Applications

- Subscribers demanding pervasive services
  - Multiple user devices
- Success i-Phone and app Store
- Services becoming richer
  - Multimedia
  - Geospatial
- Applications driving change of network usage e.g., YouTube, BBC iPlayer

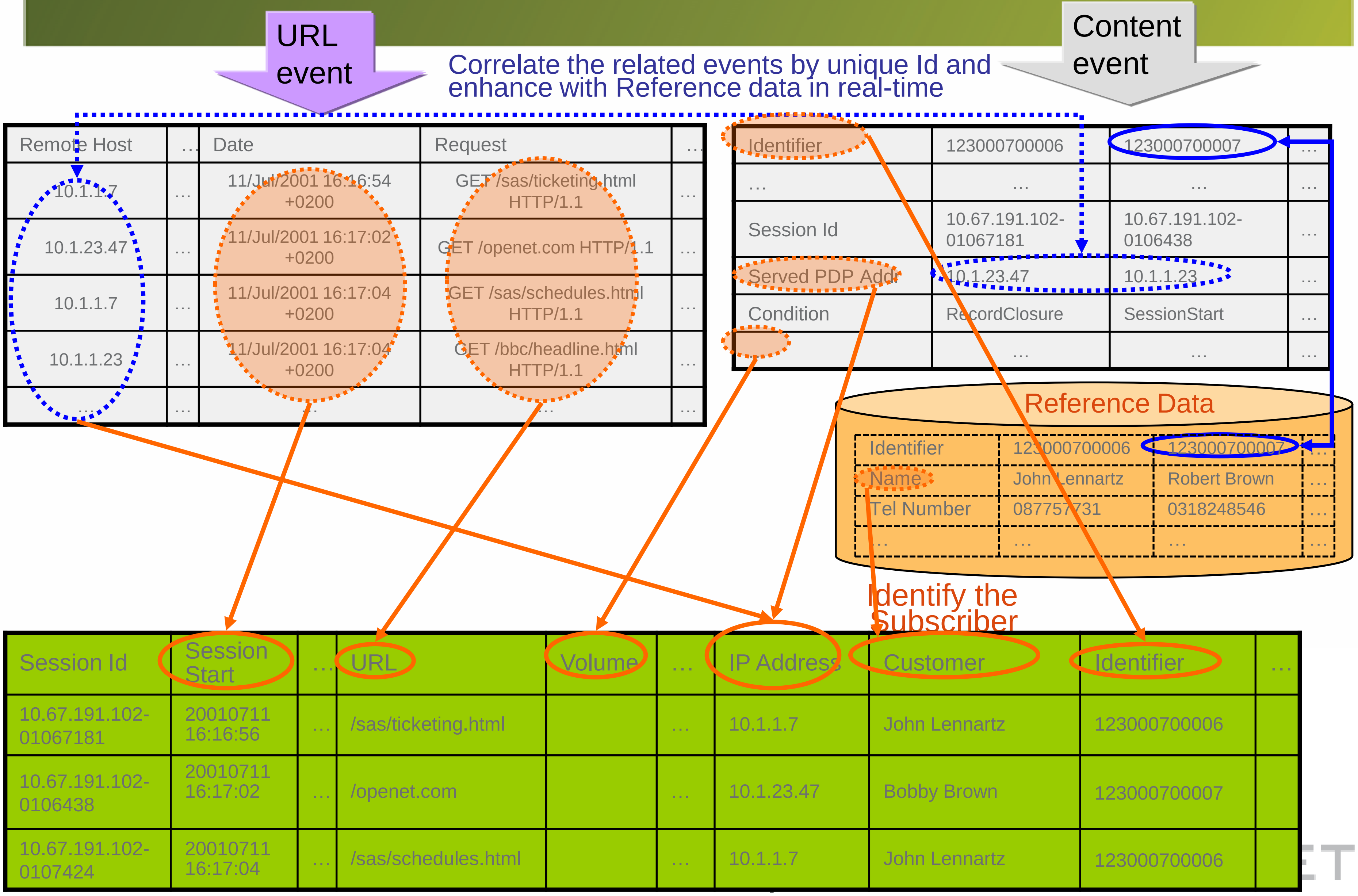
Storage Volume Challenges

- With increased volumes comes increased requirements for storage
- Billions of events per day at large mobile operators
  - How do I store this?
  - How do I search it?
  - How do I retrieve it quickly?
- Data storage is an important aspect of this solution because of:
  - The volumes of data to be stored (multi terabyte, possibly petabyte range)
  - The length of time data is to be retained (up to two years) - or longer
  - Speed of data retrieval
  - Non-volatility of data
  - Non-repudiation of data
  - Security of data (access rights)
  - Auditability
  - Cost issues

Identity Management Challenges

- Heterogeneous networks with multiple identifiers
  - IP address
  - SIP URL
  - IMSI
  - MSISDN
  - E-mail address
  - Application handle e.g. Skype
  - MAC address

Content Identification and Correlation Challenges



URL event

Content event

Correlate the related events by unique Id and enhance with Reference data in real-time

Remote Host	…	Date	Request	…
10.1.1.7	…	11/Jul/2001 16:16:54 +0200	GET /sas/ticketing.html HTTP/1.1	…
10.1.23.47	…	11/Jul/2001 16:17:02 +0200	GET /openet.com HTTP/1.1	…
10.1.1.7	…	11/Jul/2001 16:17:04 +0200	GET /sas/schedules.html HTTP/1.1	…
10.1.1.23	…	11/Jul/2001 16:17:04 +0200	GET /bbc/headline.html HTTP/1.1	…
…	…	…	…	…

Identifier	123000700006	123000700007	…
…	…	…	…
Session Id	10.67.191.102-01067181	10.67.191.102-0106438	…
Served PDP Addr	10.1.23.47	10.1.1.23	…
Condition	RecordClosure	SessionStart	…
	…	…	…

Reference Data

Identifier	123000700006	123000700007	…
Name	John Lennartz	Robert Brown	…
Tel Number	087757731	0318248546	…
…	…	…	…

Identify the Subscriber

Session Id	Session Start	…	URL	Volume	…	IP Address	Customer	Identifier	…
10.67.191.102-01067181	20010711 16:16:56	…	/sas/ticketing.html		…	10.1.1.7	John Lennartz	123000700006	
10.67.191.102-0106438	20010711 16:17:02	…	/openet.com		…	10.1.23.47	Bobby Brown	123000700007	
10.67.191.102-0107424	20010711 16:17:04	…	/sas/schedules.html		…	10.1.1.7	John Lennartz	123000700006	

Security

- Retained data must be of the same quality and subject to the same security and protection as data that are on the network
- Technical and organisational measures must protect data against destruction, loss or alteration, and unauthorised or unlawful storage, processing, access or disclosure
- Ensure that only specially authorised personnel have access to the data
- The data must be destroyed at the end of the period for retention

Thank You

Shane O’Flynn
VP Client Services
shane.oflynn@openet.com

OPENET
transactional intelligence