



AIS Data Quality and the Authoritative Vessel Identification Service (AVIS)

Dave Winkler

USCG Operations Systems Command

CGI Federal

david.winkler@cgifederal.com

david.m.winkler@uscg.mil



National GMDSS Implementation Task Force

10 Jan 2012

Arlington VA

Overview

- Data Quality in this context is focused on accuracy of AIS static data elements, not signal quality
- AIS Signal represents a radio, not necessarily a vessel
- Signal must be correlated to external knowledge bases (authoritative data) to confirm if a vessel is accurately, completely and properly identifying itself
 - Radio licenses (FCC/BoatUS/SeaTow/Shine/USPS)
 - MMSI/foreign radio registrations (ITU)
 - Certificate of Documentation
 - Lloyds Register (\$) / Equasis
 - Classification Society records (IACS)
 - Notice of Arrivals, etc.
- Techniques to improve identification must be leveraged to advise enforcement

AIS Signal Correlation

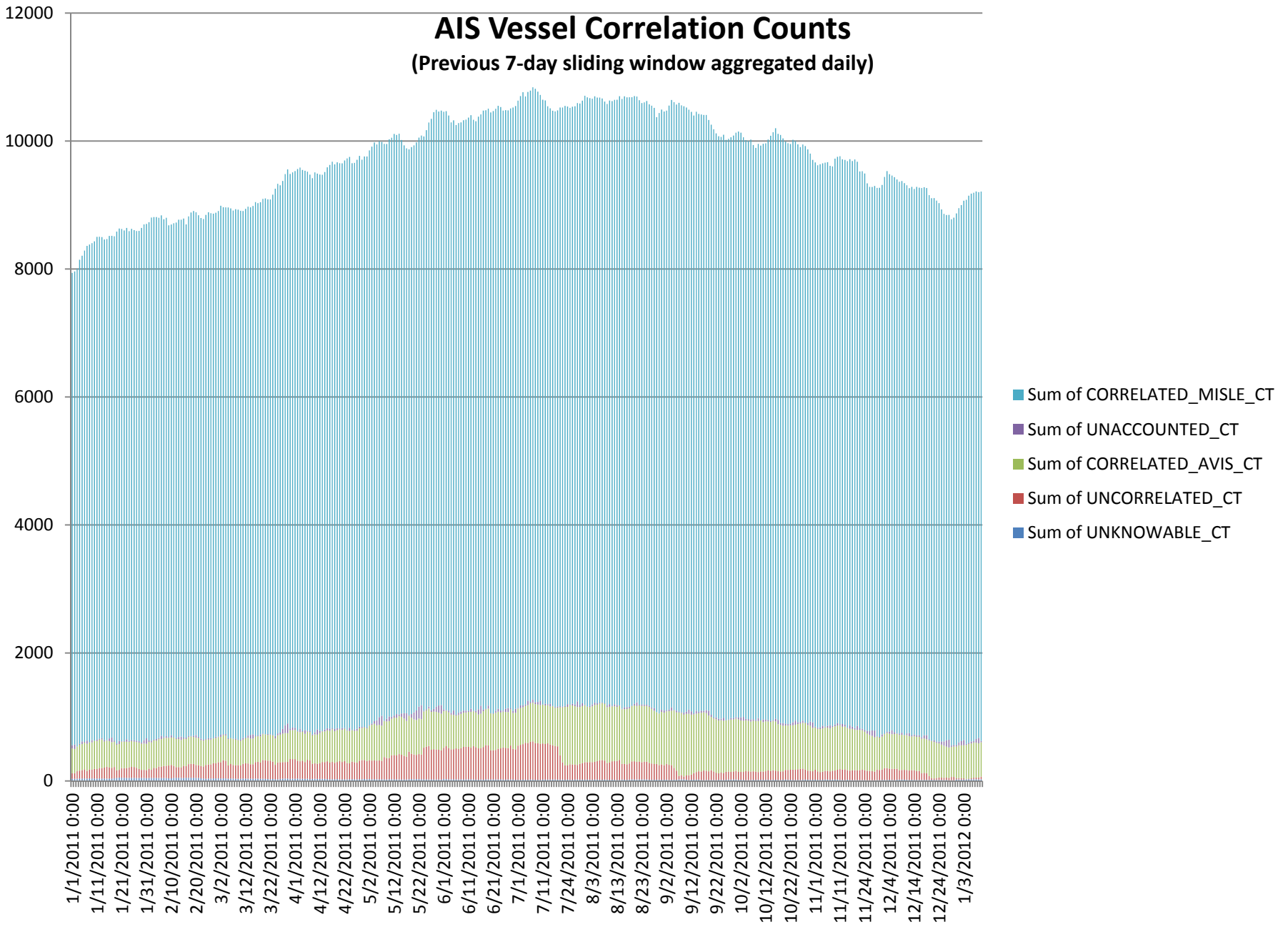
- Correlation to external knowledge base
 - For USCG, MISLE is system of record for vessel data
 - All documented vessels
 - Inspected
 - Boarded
 - Sighted
 - Foreign visits (Notice of Arrivals)
 - Correlation provides a fixed variable (permanent, unique vessel ID)
 - Allows continuous tracking as AIS identifiers
 - MMSI is not sufficient as a fixed variable
 - Often incorrectly reported
 - Vessels reissued new MMSIs upon license renewal and ownership change
 - Some administrations reissue same MMSI to different ships over time (Panama)

AIS Signal Correlation - Techniques

- Correlation – from what?
 - Single identifying elements of MMSI, IMO, Call Sign and Name cannot be relied upon
 - Individual data elements may be wrong or missing
 - Composite of 4 discrete identifiers are turned into an **“AIS Vessel Signature”**
 - AIS Signature is tokenized into a 45-character string
 - positions 1-9=MMSI
 - positions 10-18=IMO
 - positions 19-25=Call Sign
 - positions 26-45=Name
 - Use vertical pipe character to prefix (front pad) any identifiers which do not use all allocated characters
 - selected because it is not part of the NMEA 6-bit dictionary
 - Remove all non-alphanumeric characters
 - Usually variations in punctuation, spacing, or even inclusion of emoticons :-(
 - Can be regarded as non-significant differences
 - Only affects strings (Call Sign and Name)

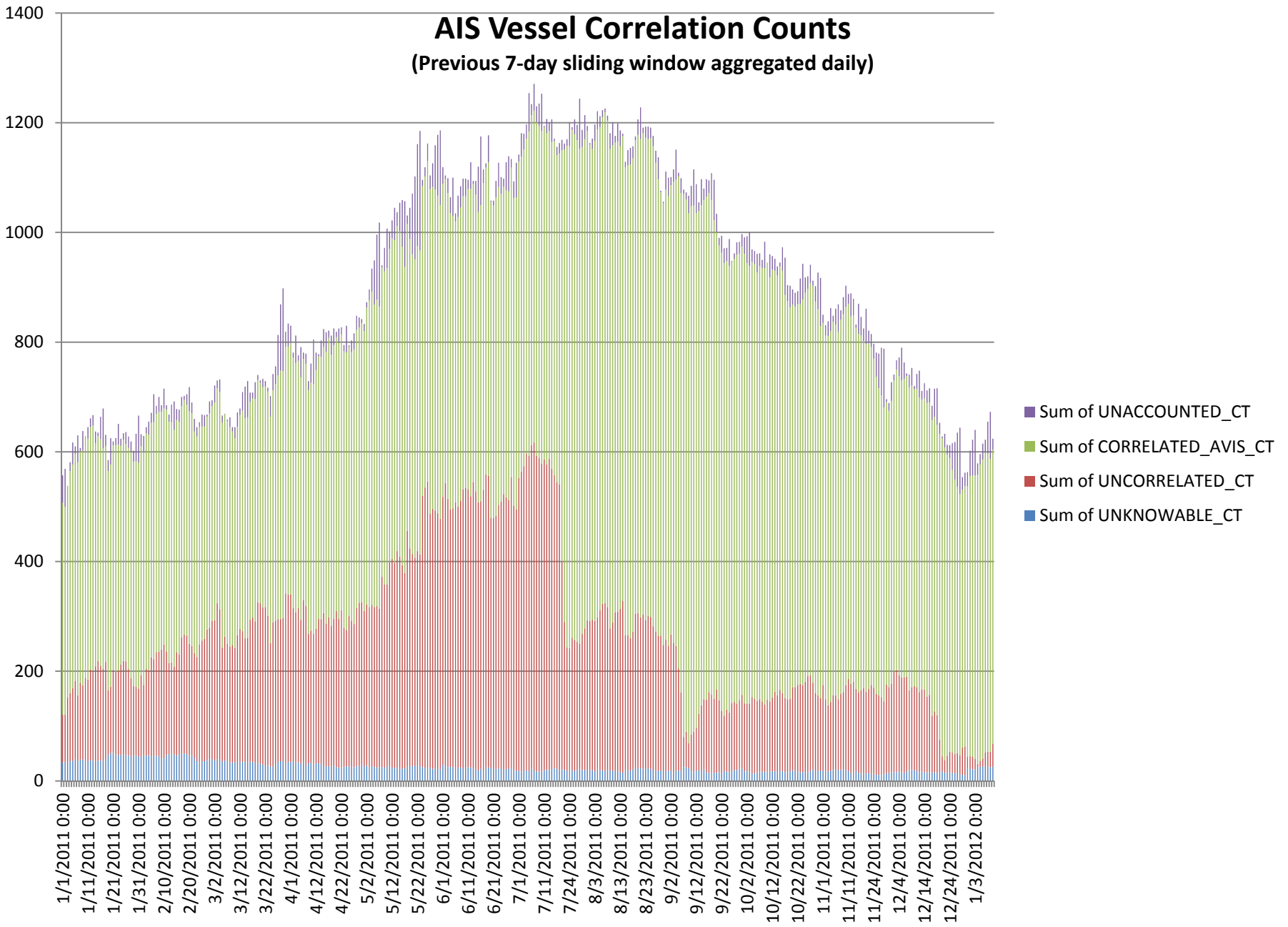
AIS Vessel Correlation Counts

(Previous 7-day sliding window aggregated daily)



AIS Vessel Correlation Counts

(Previous 7-day sliding window aggregated daily)



AIS signal-to-vessel correlation sample

AIS_UID	MMSI	IMO_NUMBER	CALL_SIGN	NAME	RECORDS	MIN_ONPLOT_DT	MAX_ONPLOT_DT	MISLE_VESSEL_ID
367051230 0WDC6095 LTSAMUELS COURSEN	367051230		0 WDC6095	LT SAMUEL S COURSEN				609844
367051230 1WDB3537 COURSEN	367051230		1 WDB3537	COURSEN	130,827	2007-11-08 22:40:16.000	2008-07-26 23:57:19.000	609844
367051230 1WDC6095 COURSEN	367051230		1 WDC6095	COURSEN	370,739	2007-11-08 22:43:00.000	2009-05-10 13:12:24.000	609844
1193046303174162WDC6095 COURSEN	1193046	303174162	WDC6095	COURSEN	15	2008-08-13 12:31:11.000	2008-08-13 12:52:30.000	609844
36705123 1WDC6095 COURSEN	36705123		1 WDC6095	COURSEN	7	2008-08-13 12:54:58.000	2008-08-13 12:55:06.000	609844
1193046303174162WDB3537 SAMUELS COURSEN	1193046	303174162	WDB3537	SAMUEL S COURSEN	57	2009-06-02 14:22:01.000	2009-12-28 18:52:58.000	609844
367051230 1WDB3537 SAMUELS COURSEN	367051230		1 WDB3537	SAMUEL S COURSEN	873,748	2009-06-02 14:31:00.000	2011-03-19 22:01:21.296	609844
1193046303174162WDB3537 SAMUELS CURSEN	1193046	303174162	WDB3537	SAMUEL S CURSEN	11	2009-06-02 14:47:26.000	2009-10-28 13:33:21.000	609844
367051230 1WDB3537 SAMUELS CURSEN	367051230		1 WDB3537	SAMUEL S CURSEN	38,391	2009-06-02 15:00:57.000	2009-12-28 15:06:00.000	609844
1 36701230WDB3537 SAMUELS COURSEN	1	36701230	WDB3537	SAMUEL S COURSEN	684	2009-09-04 11:20:10.000	2009-09-05 13:05:19.000	609844
367051230 D11233 NAUTICAST	367051230		1 D11233	NAUTICAST	114	2009-10-07 11:00:40.000	2011-03-19 22:21:36.430	609844
367051239 1WDB3537 SAMUELS COURSEN	367051239		1 WDB3537	SAMUEL S COURSEN	22	2009-10-09 11:29:20.000	2009-10-09 11:32:09.000	609844
367051230 1WSB3537 SAMUELS COURSEN	367051230		1 WSB3537	SAMUEL S COURSEN	60,665	2009-12-28 19:16:16.000	2010-03-27 22:28:46.176	609844
367051230 1WDC6095 SAMUELS COURSEN	367051230		1 WDC6095	SAMUEL S COURSEN	1,086	2011-03-19 21:11:59.186	2011-03-20 16:25:08.123	609844
367051230 0	367051230		0		4,498	2011-04-12 15:39:47.156	2011-08-09 14:32:15.810	609844
367051230 0WDC6095 SAMUELS COURSEN	367051230		0 WDC6095	SAMUEL S COURSEN	122,045	2011-04-12 15:42:16.730	2011-09-22 23:58:25.403	609844

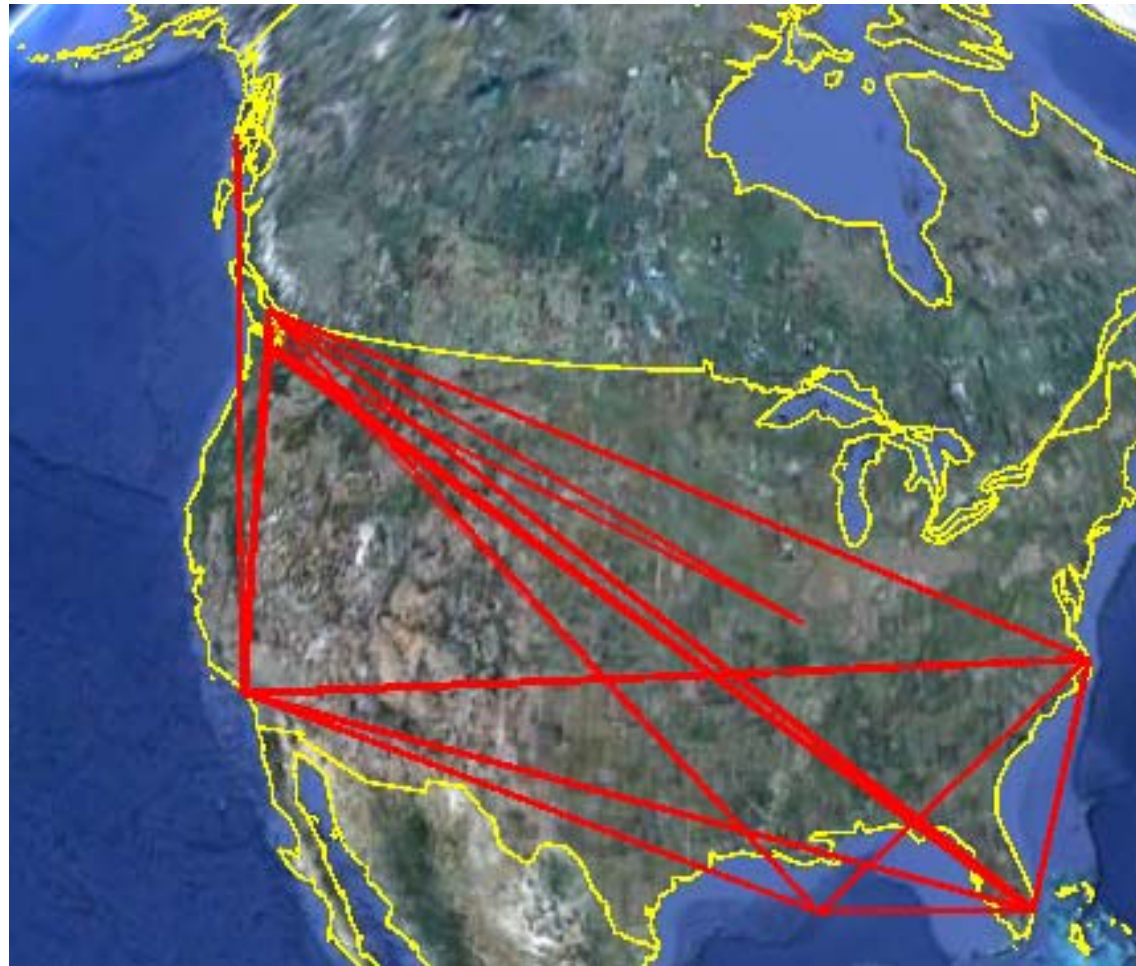
MMSI Duplication

- Largest problem for systems consuming unvalidated AIS data
 - Safety problem when multiple ships use same MMSI in same local region
 - For historical data analysis, often difficult to track history of a vessel which uses a duplicative MMSI
- Limited domain of duplicative MMSIs
 - Only approx. 150 MMSIs over the past 3 years
 - #1 problem: Nauticast X-Pack-US default MMSI 1193046
 - Why? Keeps coming back until operators repair or replace their transponder
 - Typical MMSIs: 111111111, 123456789, 987654321, 1, 5, etc.
 - If another data element is correctly configured its identity can be verified, but often spatial analysis (ports/facilities visited, nearest neighbor vessels) must be used to get “eyes on the target”

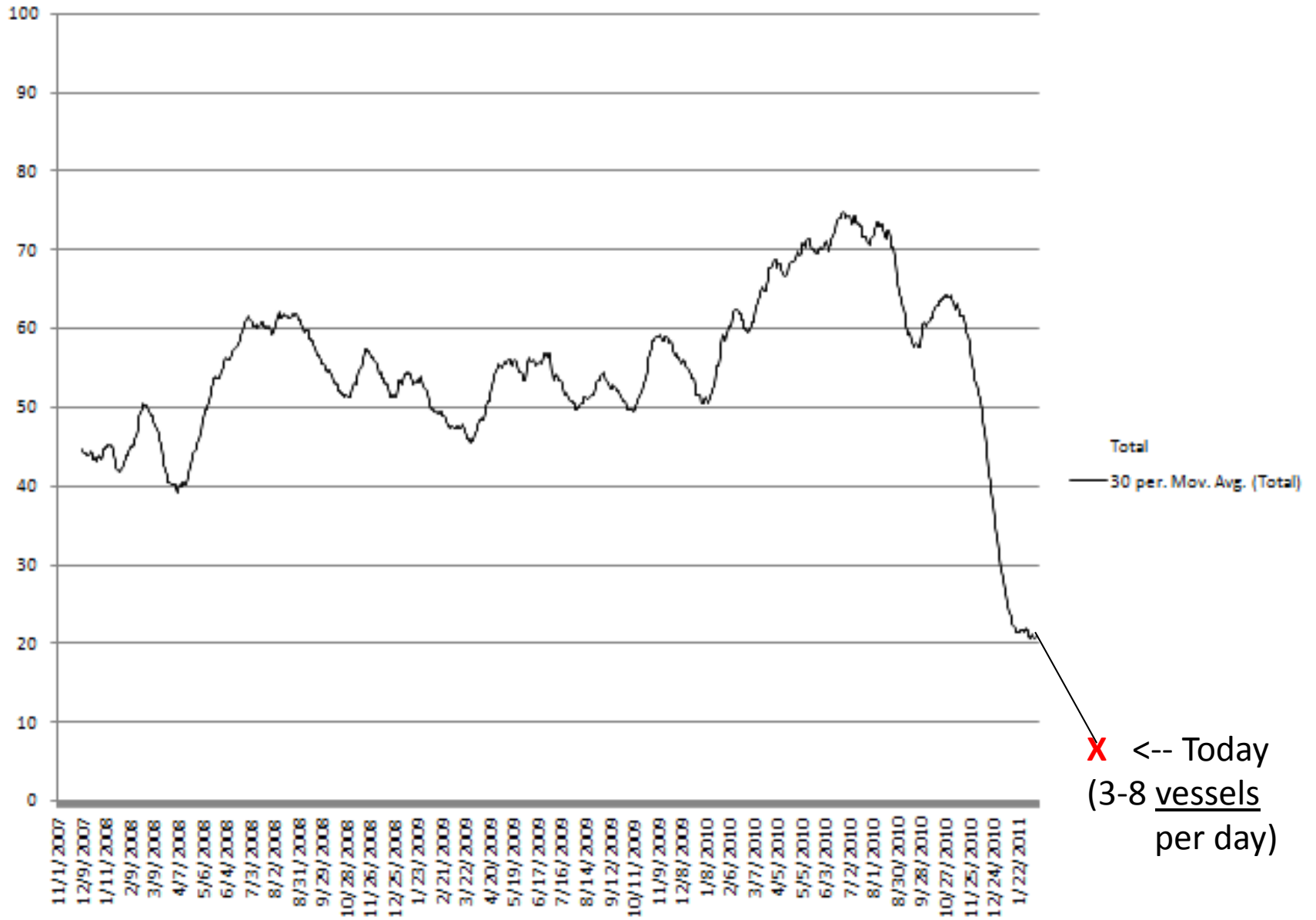
Effects of MMSI duplication on vessel tracking

Vessel Track of MMSI 111111111

(30 days, 22 Feb - 24 Mar 2011)



Duplicative MMSIs in use, Jan 2008 - Jan 2011



X <-- Today
(3-8 vessels
per day)

Multiple concurrent (active) MMSI Assignments to same vessel

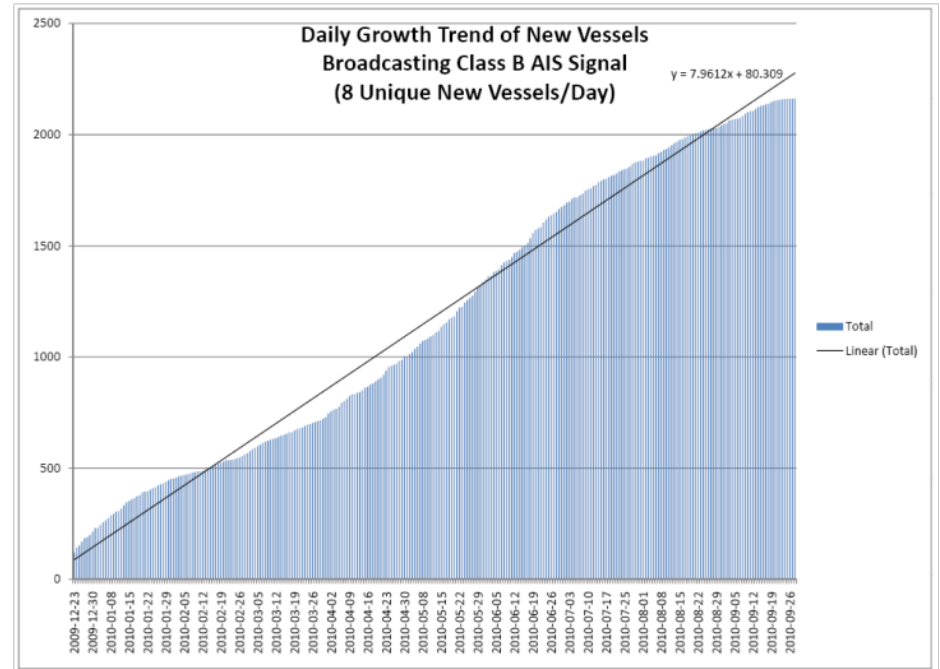
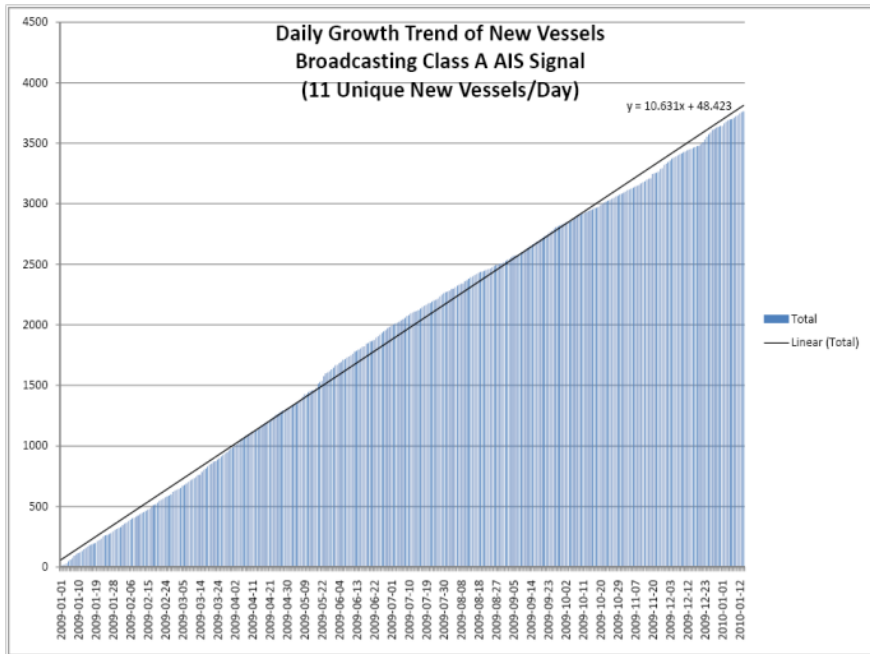
AIS_UID	MMSI	IMO	CALL SIGN	NAME
338066377	338066377			AGAPE
366813650	366813650		OWDA4836	AGAPE
338109284	338109284			AQUA THERAPY
366987190	366987190		OWDC2063	AQUA THERAPY
338053633	338053633	98000	19258	AUDREY K
367172810	367172810		OWDD6425	AUDREY K
338055716	338055716		0	BAREFOOT'N
366789550	366789550		OWDA2377	BAREFOOT'N
338103485	338103485			CAPT THANH II
367156270	367156270	8847662	WBC7796	CAPT THANH II
338105715	338105715			CAPTAIN MORGAN
366829590	366829590		OWDA6186	CAPTAIN MORGAN
338086233	338086233			CHEKARA
366800490	366800490		OWDA3632	CHEKARA
338098853	338098853			CHILANGUITA TOO
367448980	367448980		OWDF4187	CHILANGUITA TOO
338104023	338104023			CHULAMAR
368433000	368433000		OWDF7092	CHULAMAR
338104025	338104025			CROWD PLEEZER
366937780	366937780		OWDB7178	CROWD PLEEZER
338090025	338090025			DAY DREAMS
366877080	366877080		OWDB2543	DAY DREAMS
338105974	338105974			DIVINE INTERVENTION
367120140	367120140		OWDD2806	DIVINE INTERVENTION
338054427	338054427		0	DREAMIN ON
366722960	366722960		0	DREAMIN ON
338109066	338109066			FRAID KNOT
367478750	367478750		OWDF6877	FRAID KNOT

AVIS Accomplishments

Achieving Maritime Domain Awareness

- Virtually eliminate unidentified and uncorrelated AIS vessel signals broadcasting unregistered MMSI numbers -Specifically, 863 out of 866 vessels were corrected by September 2011, eliminating nearly 100% of incorrect broadcasts.
- On a daily basis demonstrated the capability to properly correlate > 90% of all AIS vessel positions to MISLE or AVIS in near real time.
- AVIS services identified another 9% of vessels operating within the U.S. AOR which are not recorded in the system of record.
- AVIS services provide capabilities to explicitly identify the last 1% of AIS broadcasts that were too ambiguous to be readily recognized by any existing information system.
- AVIS services provide capabilities to reduce the number of position reports broadcasting duplicate MMSI numbers from a historical 50 to 80 vessels per day to only 0 to 4 per day!

Near Complete MDA is achievable and maintainable at reasonable cost with existing resources!



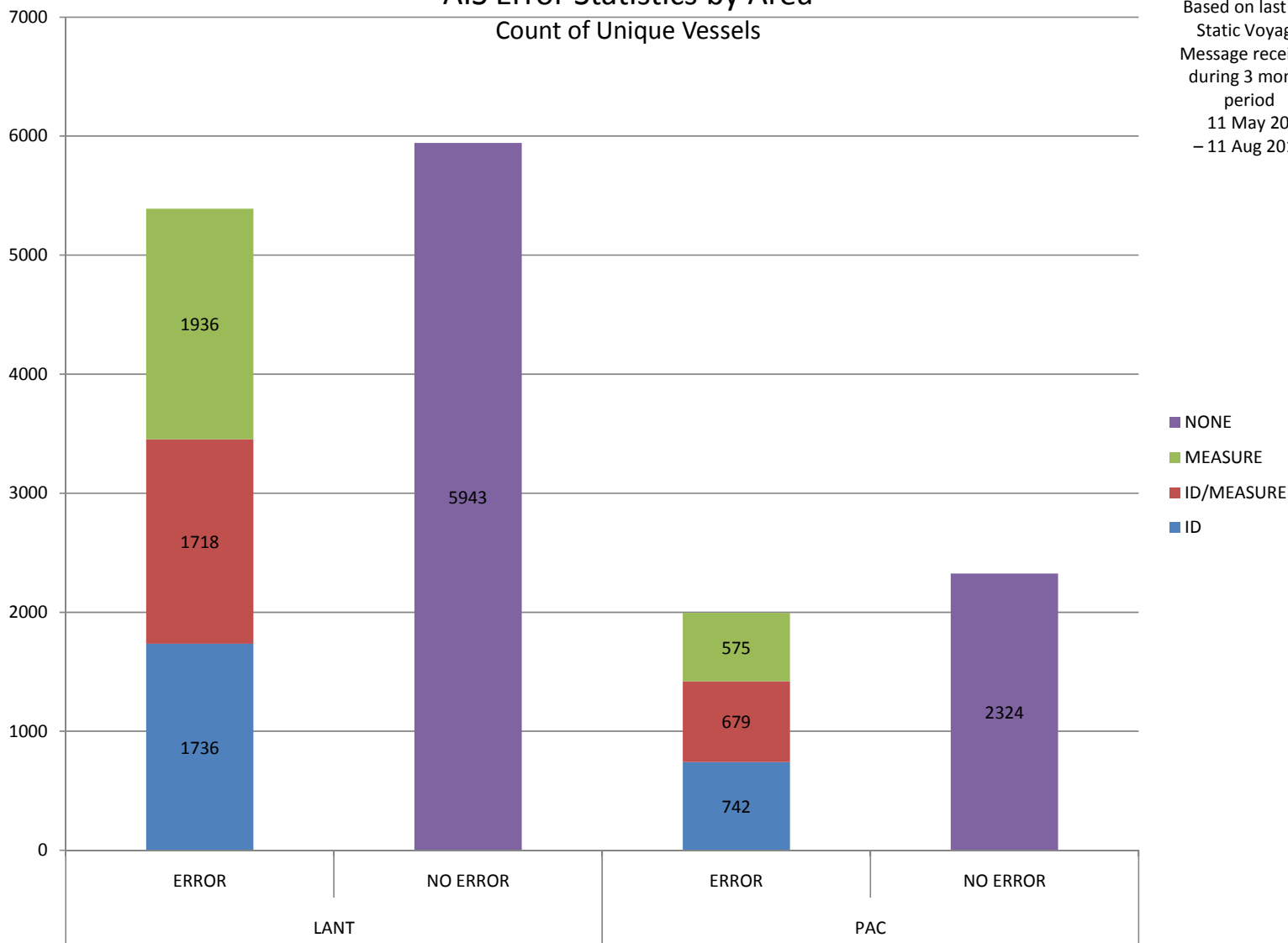
AVIS

Achieving Maritime Domain Awareness

AIS Error Statistics by Area

Count of Unique Vessels

Based on last AIS
Static Voyage
Message received
during 3 month
period
11 May 2011
– 11 Aug 2011

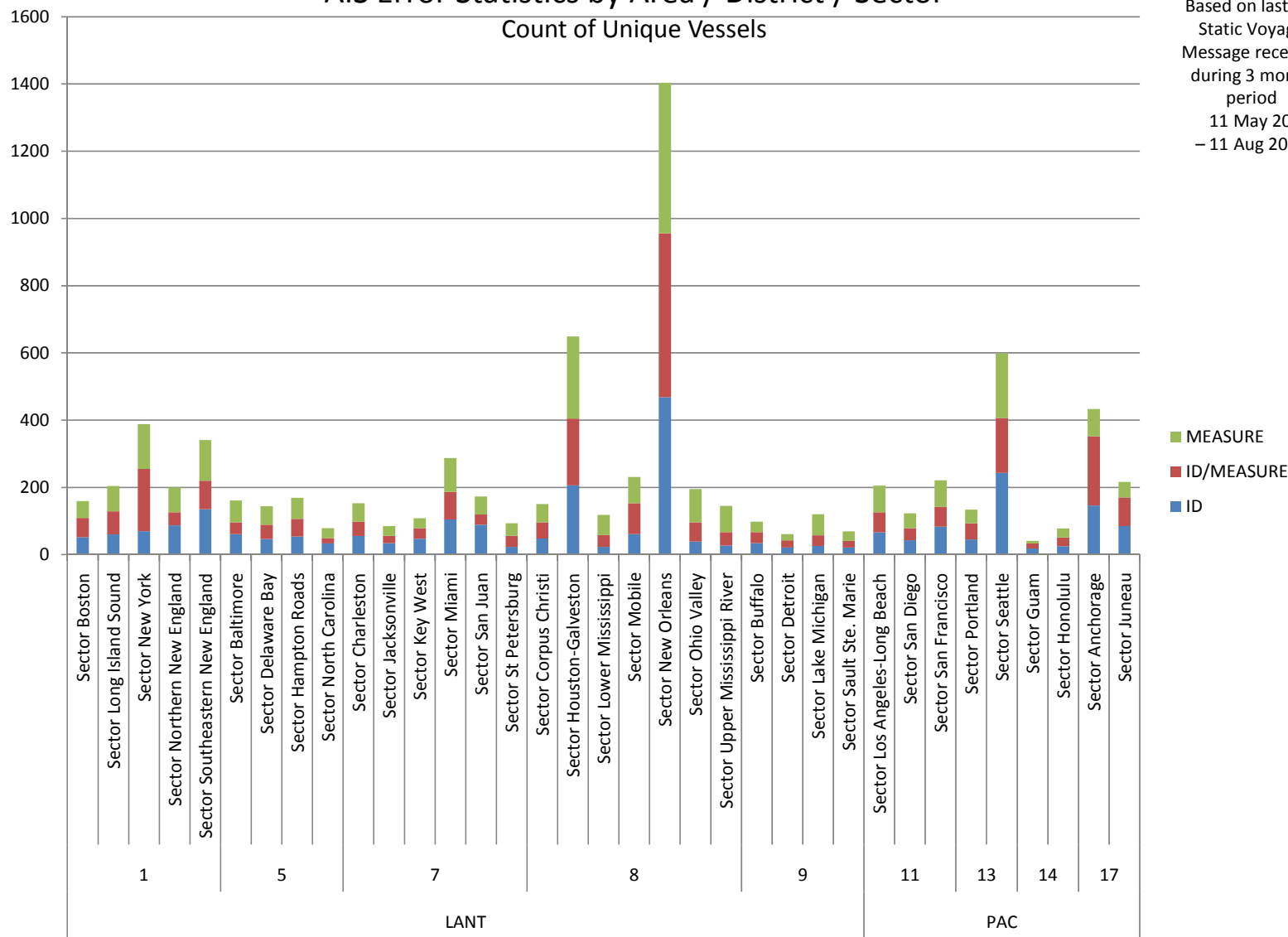


AVIS

Achieving Maritime Domain Awareness

AIS Error Statistics by Area / District / Sector
Count of Unique Vessels

Based on last AIS
Static Voyage
Message received
during 3 month
period
11 May 2011
– 11 Aug 2011



AIS Error – detailed error types

9,125 vessels had no known discrepancies with their AIS configuration (50.3%)

8,086 vessels had some known discrepancy with their AIS configuration (44.6%)

918 vessels which were not yet evaluated for their correctness (5.1%)

Focusing on vessels which had some error:

- 31.6% had both identity and measurement errors

- 34.1% had only an identity error

- 34.3% had only a measurement error

(Statistics as of November 2011)

AIS Error – detailed error types

5,330 total ID errors, consisting of:

=====

- 1,094 MMSI errors
 - 49 where wrong MMSI represents a different vessel (possible for authorities to identify as different vessel altogether)
 - 260 broadcasting previously issued MMSI
 - 785 broadcast a wrong MMSI number relating to no licensed vessel.

Many operating without currently licensed radio credentials.

- 1,771 Call Sign errors
- 2,474 IMO errors
 - 1,563 of which are IMO registered but:
 - 856 are broadcasting 0 for IMO number
 - 208 are broadcasting a 9-digit IMO number equal to their assigned IMO * 100
 - 200 are broadcasting their US documented official number instead of their IMO #
 - 299 are broadcasting some other number
 - 911 of which are not IMO registered but:
 - 297 are broadcasting their US documented official #
 - 614 are broadcasting some other number in the IMO field
- 1,813 Name errors

(note: these numbers are not additive as many vessels account for more than a single ID error type, but breakdown of MMSI and IMO type errors are additive)

AIS Error – detailed error types

5,312 non-ID Static Voyage errors (measurements, draft, ship type, etc.)

=====

755 incorrect lengths where AIS length is >25% different than documented

115 incorrect lengths where measurement was entered in feet instead of meters

115 incorrect lengths where beam > length

1,865 missing length and/or beam (A+B=0 and/or C+D=0)

2,022 missing draft

211 incorrect draft where it is deeper than overall length or beam

706 incorrect measurements where antenna placement is not specified (A or B or C or D=0)

1,059 missing or invalid ship types

193 incorrect nav sensor types (not GPS or GLONASS)

(note: these numbers are not additive as many vessels account for more than a single static error type)

AIS Correction – ROI (per fleet)

Fleets with 3 or more MMSI errors (Tx within past 60 days)

Number of companies: 37

Number of vessels with MMSI errors: 151

Number or vessels with other errors: 603 Total number of unique vessels corrected by 37 phone calls: 754 (ROI = 20.4 vessels corrected per call)

Fleets with 2 MMSI errors (Tx within past 60 days)

Number of companies: 59

Number of vessels with MMSI errors: 118

Number or vessels with other errors: 396 Total number of unique vessels corrected by 59 phone calls: 514 (ROI = 8.7 vessels corrected per call)

Fleets with 1 MMSI error (Tx within past 60 days)

Number of companies: 256

Number of vessels with MMSI errors: 256

Number or vessels with other errors: 1,022 Total number of unique vessels corrected by 256 phone calls: 1,278 (ROI = 5.0 vessels corrected per call)

AIS Correction – ROI (per operator)

Non-fleet vessels (Tx within past 60 days)

Individual Fishing Vessels with MMSI errors: 68

Private Recreational Vessels with MMSI errors: 169

Total number of individual vessels with MMSI errors

corrected by 237 phone calls: 237 (ROI = 1 vessel corrected per call)

Remaining MMSI errors:

Federal Government fleets: 36 vessels

Vessels where owner is not yet determined: 114 (might be fleet, might be individuals)

Vessels Unevaluated for radio license accuracy: 372 (might have approx. 5% of these, ~20, with MMSI errors)

AIS Signal Validation

- After correlation to a fixed, immutable vessel identifier, the AIS elements must be validated for correctness
 - What was correct yesterday may not be so today
 - Authoritative data from flag administrations must be used
 - Problem with reference data – keeping it up-to-date
 - Vessel duplication also exist with reference data
 - Authoritative sources conflict each other
 - Ensure the data steward is used as the authoritative source, not an information aggregator (i.e., Lloyds)
- Generate a standard by which each AIS Signature may be measured against for correctness
 - Using n-sources is difficult for uniform comparison, better to synthesize a single authoritative dataset for AIS verification

Next – Interactive Demonstration