Data Analysis and Mining With DataSpy and IPA (Integrated Post Analysis)

ICS TechDay 2019 Presented by John Mitchell



April 30, 2019



Data Analysis and Mining with DataSpy and IPA

Topics Covered in Presentation:

- Review of PC Version of DataSpy
- Review of DataSpy running on Wireless NeoVI
- Definition of Integrated Post Analysis (IPA)
- Review architecture of IPA and operating modes
- Demo of FindInFiles and HistogramGenerator scripts on PC
- Review script configuration files for these scripts
- Review code in scripts
- Review main functions of IPA library
- Demo of FindInFiles and HistogramGenerator scripts on Wivi





Powerful plotting tool that provides accurate graphical and numerical display of large mdf data files.

- Easily handle files that are several Gb in size.
- Verified to provide accurate results using a range of data files logged using ICS products as well as those from ETAS.
- Create custom views of key signals and save configurations for use with other data files.







Users can zoom way out on large data files to view signal ranges or can zoom way in to view individual data points.

 User can select signal in legend to make it the active signal, then use arrow keys to step through each data point on signal.







Search for complex expressions in data file using Advanced Search feature.

- Provide a boolean expression of signals and then click on Find All
- Example below looks for WOT pedal launches from rest
- Expression parser can also be used to create calculated signals

	Find in files	
	Search In Chart Advanced Search	Enter expression
	Find What: ((693.)/ehicleSpeed@EILE1) < 1) && ((512.ActualRedalRos@EILE1) > 90) Find All Save	All to execute search
	Look in: Current File	
Search expressions can be built using	Current File Current Directory Selected Directories	
channel list or can be pasted into "Find	Look at file type: MDF Files	
What" text box	Expression Editor Existed Expression	
	Actual Sin asin log abs	Boolean
	86 ACTUAL SPK CYL2 cos acos log10 sqrt	operators
	312 RrAxleCouplingTrq_Actual	currently
	314 Ft_RrCouplingTrq_Actual + - * /	supported in
	424 ActualGear < > <= >=	expression
	512 ActualPedalPos	parser





Search results shown in output window below graphical view.

- User can click on each hit to jump to that point in data file.
- Tool places cursor 1 at the point the expression toggles true









DataSpy – Web Version Launches in Browser from Wivi

Some basic integration with Wireless NeoVI currently available

- Allows users to view data files directly from the server by clicking on link
- No need to install any software, works in any browser as long as user is logged into Wivi
- User authentication controls allow administrators to control which users can see which data

Dashboa	ard 📃 Devices 🗸	쇼 Organiza	ations 🚔 Server -	🚽 John Mitchell -					Coordinated	
ashboard / ICS TE	ST / Vehicle 400815 (20	016 neo VI) / I	Data							
earch Filters										
Date Range			off Collections	Search collection	ons	o# File	e Types			
Start Date		×	Chrysler200EntireBus			CSV		NR MAT	VSB YM	1
End Date		×	ChryslerOneShotData_			001	DAI		VOD AN	
lame			Chrysler200C_BusQuery							
Search names		×	Chrysler200_AllBusAndKV	VP2K_						
			ColdStartDarfMatrice		*					
data processing,	no data pending			F	ਤ 0 files, 0 B sele	ected 👻	CSV	DAT DB	MAT V	SB XML
rt	End	Name				View		Data Type		≡
-04-16 22:59:53	2017-04-16 23:38:45	DemoDataA	liBus				0.	DAT 118.65 MB	DB 278.42 MB	VSB *
7-04-16 21:24:27	2017-04-16 21:24:50	DemoDataA	llBus			Multi	iple Data	Views AT	DB 3.51 MB	VSB 610.92 KF
7-04-16 19:05:45	2017-04-16 19:42:10	DemoDataA	liBus				0.	DAT 111 21 MB	DB 261.04 MB	VSB
7-04-16 00:52:34	2017-04-16 00:52:48	DemoDataA	llBus			.	M V	iew DAT in D	ataSpy	VSB
7-04-15 23:52:03	2017-04-15 23:57:08	DemoDataA	liBus				E C	B Viewer		VSB
7-04-14 19:54:57	2017-04-14 19:55:01	DemoDataA	llBus				6.	14.95 MB	36.17 MB	7.95 MB
7-03-15 12:46:34	2017-03-15 12:49:49	Chrysler200		e03-14-17 v4			0	2.34 KB	44 KB	311 B
7-03-13 12:46:34	2017-03-15 12:40:40	Gnrysler200		503-14-17_VI			0.	345 B	28 KB	343 B
7-03-14 21:27:44	2017-03-14 21:58:43	Chrysler200	C_KWP2K_ColdStartParam	s03-14-17_v1		• •	0.	DAT 2.39 MB	DB 28.91 MB	VSB 7.9 MB
/-03-14 21:27:41	2017-03-14 21:27:41	Chrysler200	C_KWP2K_ColdStartParam	s03-14-17_v1		•	6.	DAT 345 B	DB 28 KB	VSB 325 B
7-03-14 20:06:57	2017-03-14 20:07:55	Chrysler200	C_KWP2K_ColdStartParam	is03-13-17_v2		.	0.	DAT 345 B	DB 28 KB	VSB 342 B
17-03-14 19:11:54	2017-03-14 19:20:01	Chrysler200	C_KWP2K_ColdStartParam	is03-13-17_v2			0.	DAT 345 B	DB 28 KB	VSB 336 B
	2017 02 14 16:52:27	Chrysler200	C KWP2K ColdStartParam	e03-13-17 v2			G .	DAT	DB	VSB

Clicking on DataSpy link will show data file in DataSpy viewer



April 30, 2019



IPA – Data mining tool that can run on PC or Web

Product Definition:

- Library of mdf and db file i/o functions allowing users to develop data mining scripts to generate reports based on time based vehicle network data.
- Scripts can be used to mine through data on users PC and can be uploaded to our WirelessNeoVI server and have the server run the scripts on files directly on the server.
- Library of functions can be used with any language in PC mode. We have example scripts in Excel VBA, MATLAB, C#, and Python.
- Python scripts can be uploaded and executed by WirelessNeoVI.
- Other cross platform languages like R may be supported in the future.





Integrated Post Analysis – Product Architecture



Parallel scripting language set up so that users can develop scripts and script configurations on their PC and upload them to the Wivi server so that they can be run by the server. Example applications include:

- Generate reports based on data automatically on server without downloading any raw data.
- Generate complex histograms (ie: absorbed clutch power per shift type as a function of clutch volume) from user defined data sets
- Pull out chunks of data from multiple mdf files and append to a single file.



Integrated Post Analysis – Defining File Set

File Sets are named collections of files used to define what files to run script on.

- Use filters in WirelessNeoVI Data Archive then select files by clicking on file icon or you can select all file matching filter criteria
- After selecting desired list of file click on dropdown and header and select Add Data to File Set



www.intrepidcs.com

Integrated Post Analysis – Creating Report on WirelessNeoVI

Reports are created from the Reports page on WirelessNeoVI

- Click on + New Report button
- Define report inputs and click on Create (see next slide)

WIVI 3 ×				And a strength of the		0.00				
$\leftarrow \rightarrow \mathbf{C}$ \triangleq Secure https:/	//preview.wirelessneovi.com/ui/#!	/reports					☆	•	0	
Back Reports	0ata Archive IIII Reports →	📕 Devices 🗸	击 Organizations	💼 Server 🗸	1 John Mitchell ▼	+•	-0400 Am	erica/Ne	w_York	Click on + New Report button to
TechDaysDemoFindInFilesRer	port						+	New Re	eport	create a new report.
Created Script Config File Set	2018-04-29 09:22:13 FindInFiles_ExpInConfig ConfigForFindInFilesExpInConf TechDaysSampleFileSet		Process	sing completed a ndInFiles_04-29-1	t 2018-04-29 09:23:14 18_13-23-14.dsr	± IPA.log	I			Some output file types are recognized by Wivi and they
C Re-run								🛍 Dele	te	lauch views fron link.
TechDaysDemoHistogramRep	ort									
Created Script Config File Set	2018-04-29 09:20:26 HistogramGenDemoFiles_0322 IPAHistConfig_DemoData TechDaysSampleFileSet	18	Process	sing completed a	t 2018-04-29 09:20:47 29-18_13-20-47.xlsx	➡ IPA.log	•			All output files from a given report are listed on window of
C Re-run								🛍 Dele	te	report.







Integrated Post Analysis – Creating Report on WirelessNeoVI

Each report has 4 inputs

- Report Name
- Script file to use to generate outputs from report
- Config file lists signal names and other script configuration parameters referenced in the script
- File Set Named set of data files that the report is based on

New Report	
Report Name	
TechDaysDemoHistogramReport	✓
Post Analysis Script	
HistogramGenDemoFiles_032218	¥
Config	
IPAHistConfig_DemoData	T
File Set	
TechDaysSampleFileSet	•
Cancel	+ Create



April 30, 2019



Integrated Post Analysis – GenerateHistogram Script Output File

Script generates a histogram of each signal specified in the config using the bins specified in the config.

- Output file is *.xlsx
- Script uses Python xlsxwriter library to generate xlsx output
- Last tab lists all of the input files used for the report







Integrated Post Analysis – FindInFiles Script Output File (*.dsr)

FindInFiles script outputs a file with extension *.dsr which stands for DataSpy report

- DSR file is JSON file that lists all of the hits on events found in search
- Each hit has a Description, StartTime, and EndTime describing a period of interest in a data file

"HitList": [
<pre>{ "FilenameAndPath": "C:/Jmitchell/DataSpy/SourceCode/ICS_IPA_FORK/SampleDataFiles/DataSpySampleDataFi "Hits": [</pre>	 Filename is listed first, then all hits related to that file. Unlimited number of files and hits are supported.
<pre>{ "Description": "WOTLaunchFromRest", "EndTime": 39.58238410949707, "StartTime": 38.056864619255066 }, { }</pre>	 Each hit, has a Descrition, StartTime, and EndTime
<pre>"Description": "EngineStateVariableChangeFrom9To10", "EndTime": 39.94499200582504, "StartTime": 37.931904673576355 }, { []]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]</pre>	
"EndTime": 48.49747771024704,	

14





Integrated Post Analysis – FindInFiles Script Output File (*.dsr)

DSR files can be loaded into DataSpy allowing convenient organization and navigation of search results.

- Hits are displayed in report page on bottom of DataSpy window
- Users can click or use arrows to jump from hit to hit in DataSpy
- User can sort hits by clicking on header row of Report Page





Click on any hit in the report page to jump to that hit in the plot window.

15





Integrated Post Analysis – GenerateHistogram Script Config File

Script config file defines parameters used by the script.

- Most scripts will have input channels lists define in the config file
- Other parameters are based on design of script
- Allows users to control script performance without having to edit code.







Integrated Post Analysis – FindInFiles Config File

In addition to the signal list, the FindInFiles script also allows users to provide a list of EventDefinitions to search for. Each event has the following parameters:

- Description a name for the event which will show up in DataSpy
- StartExpression a Boolean expression using signal name and any valid Python operators describing the start of the event
- EndExpression a Boolean expression using signal name and any valid Python operators describing the end of the event. Once the StartExpression is found the script looks for the EndExpression. EndExpression can be not of StartExpression
- Putting a "Prev___" before a signal name in an expression refers to the value of the signal name in the previous loop
- TimeFromExpStart keyword can be used in the EndExpression to refer to the time since the current event started







Integrated Post Analysis – GenerateHistogram example script

Sample GenerateHistogram script loops through list of files from start to end and creates a *.xlsx file that has a time based histogram.

- Does not require Excel
- Histogram bins are defined in script config.
- Uses Python numpy data analysis library



xlHistWorkbook.AddFileInfoListSheet(dbFilePaths, IPAInterfaceLibrary.is_running_on_wivi_server())
xlHistWorkbook.CloseWorkbook()



April 30, 2019



Top section of script calculates 2D time_tallys[] list that has the numbers for the histogram for each signal.

Bottom section writes histogram data to an Excel worksheet.

Integrated Post Analysis – FindInFiles example script

Sample FindInFiles script loops through list of files from start to end looking for events:

- Generates a *.dsr (DataSpy report) file that lists the hits for events in config file
- A few dozen lines of script support the FindInFiles functionality
- Python eval function used to evaluate text from config as code at runtime



Script converts user defined events into executable code then uses the "eval" function in Python to run the code at runtime

GetNextRecord function advances the virtual cursor to next record; event conditions are checked for each record throughout file.19





Integrated Post Analysis Review – Main Script Functions

For the Python scripts, the IPA functions are wrapped in a Python class called ICSDataFile. A data filename and config filename are provided when the script creates and instance of this class.

JumpBeforeTimestamp(TimeToJumpToInSeconds)

- Input is desired time in Datafile to place the virtual cursor
- Places virtual cursor at closest time prior to TimeToJumpToInSeconds where there is a complete record
- If TimeToJumpInSeconds is lower than time corresponding to the first complete record it places the virtual cursor at the first complete record
- Updates array of signal values corresponding to location of virtual cursor
- Returns the exact position of the virtual cursor

GetNextRecord()

- Advances the virtual cursor to the next chronological time where any of the active signals have updated values
- Returns the exact position of the virtual cursor
- Updates array of signal values corresponding to location of virtual cursor
- If virtual cursor is already at the last record in the file it returns MaxDouble (10^308)





Integrated Post Analysis – Main Script Functions (cont)

GetNextChangedRecord()

- Advances the virtual cursor to the next chronological time where any signal has a value that is different than the current records values
- Returns the exact position of the virtual cursor
- Updates array of signal values corresponding to location of virtual cursor
- If virtual cursor is already at the last record in the file it returns MaxDouble (10^308)

SetActiveMask(SignalMaskString)

- Used to control which signals the GetNextRecord() function considers when stepping to the next record.
- If only one signal is active, then GetNextRecord() will step through each timestamp that corresponds to this signal.
- If all signals are active GetNextRecord() will step through every new timestamp in every signal in the config file.
- Be default all signals are active.





Integrated Post Analysis – CurrentRecord Definition

One challenge in performing calculations on time based data is how to handle multiple asynchronous time vectors

- One common practice is to force the data to fit a fixed time rate. This method has detrimental effect on fidelity of data
- IPA method matches the way DataSpy displays values at cursor in legend. The legend reports the most recent values that precede the cursor.







Integrated Post Analysis – Used for cmProbe Validation

IPA Library used to validate cmProbe bench data on 40,000 simultaneous signals

- Bench ECU set up to generate 40,000 signals that follow known walking ones pattern
- Script written to search through large files and look for cases where recorded pattern did not match expected pattern like the one shown below





April 30, 2019



Integrated Post Analysis Review – Numpy vs Matlab

Indexing and accessing elements (Python: slicing)

MATLAB/Octave	Python	Description
a = [11 12 13 14	a = array([[11, 12, 13, 14],	Input is a 3,4 array
21 22 23 24 31 32 33 34]	[21, 22, 23, 24], [31, 32, 33, 34]])	
a(2,3)	a[1,2]	Element 2,3 (row,col)
a(1,:)	a[0,]	First row
a(:,1)	a[:,0]	First column
a([1 3],[1 4]);	a.take([0,2]).take([0,3], axis=1)	Array as indices
a(2:end,:)	a[1:,]	All, except first row
a(end-1:end,:)	a[-2:,]	Last two rows
a(1:2:end,:)	a[::2,:]	Strides: Every other row
	a[,2]	Third in last dimension (axis)
a(:,[1 3 4])	a.take([0,2,3],axis=1)	Remove one column
	a.diagonal(offset=0)	Diagonal

Assignment

MATLAB/Octave	Python	Description
a(:,1) = 99	a[:,0] = 99	
a(:,1) = [99 98 97]'	a[:,0] = array([99,98,97])	
a(a>90) = 90;	(a>90).choose(a,90) a.clip(min=None, max=90)	Clipping: Replace all elements over 90
	a.clip(min=2, max=5)	Clip upper and lower values

Transpose and inverse

MATLAB/Octave	Python	Description
a'	a.conj().transpose()	Transpose
a.' <i>Of</i> transpose(a)	a.transpose()	Non-conjugate transpose
det(a)	linalg.det(a) or	Determinant
inv(a)	linalg.inv(a) or	Inverse
pinv(a)	linalg.pinv(a)	Pseudo-inverse
norm(a)	norm(a)	Norms
eig(a)	linalg.eig(a)[0]	Eigenvalues
svd(a)	linalg.svd(a)	Singular values
chol(a)	linalg.cholesky(a)	Cholesky factorization
[v,1] = eig(a)	linalg.eig(a)[1]	Eigenvectors
rank(a)	rank(a)	Rank

NumPy is a Python library that supports large, multidimensional arrays and has many functions that mimic MATLAB.

NumPy includes ndarrays arrays which can be used as views into memory buffers allocated by C/C++ dlls so that data can be exchanged without the need to do a lot of copying which is slow.

"MATLAB is to Python what Encyclopedia Britannica is to Wikipedia" crowd based engineering





Integrated Post Analysis – Script Config File Generator

Excel tool used to generate signal list portion of config file

- Allows users to pick signals from files on Wireless NeoVI or on PC
- Supports prioritized optional list for each signal (multiple versions of Engine Speed)
- Excel tool uses functions from the IPA library

±° ~	Wivi URL Username Password Start Date	https://ps00.w	rirelessneovi.co	Update File List From Wireless Neo Update Signal List Using Selected File	/I ID	After defi click on b	ning signals utton to		- - -
Intrenid	End Date	5/1/2018		Update Signal List from File on Your F	PC	generate	asl (aliased		
Intropid	Fleet Name	AllCinnala		Export Signal List * asl and * ny File		signal list			-
	Collection Name	Alisignais		Export Signal List last and ipy me		Signarinse			-
Vehicle	File Capture Date	File Size (KB)	FileID	Signal Name	Message Name	Network Name	Name In Script	Priority	-
John Mitchell	4/12/18 4:05 PM	9027	2 26804	ManufacturerSpecific8	ProprietaryA	HS CAN	ManufacturerSpecific8		-
John Mitchell	4/12/18 3:19 PM	8892	4 26792	ManufacturerSpecific7	ProprietaryA	HS CAN	ManufacturerSpecific7		-
John Mitchell	4/12/18 1:47 PM	494	4 26633	ManufacturerSpecific6	ProprietaryA	HS CAN	ManufacturerSpecific6		Pick signals hy
John Mitchell	4/10/18 11:09 PM	71355	6 25764	ManufacturerSpecific5	ProprietaryA	HS CAN	ManufacturerSpecific5		The signals by
John Mitchell	4/10/18 9:37 PM	5358	0 25613	ManufacturerSpecific4	ProprietaryA	HS CAN	ManufacturerSpecific4	1 -	putting priority
John Mitchell	3/24/18 6:05 PM	29132	8 25615	ManufacturerSpecific3	ProprietaryA	HS CAN	ManufacturerSpecific3		number next to
John Mitchell	3/24/18 12:52 AM	16579	6 25611	ManufacturerSpecific2	ProprietaryA	HS CAN	ManufacturerSpecific2		Humber next to
John Mitchell	3/23/18 10:54 PM	12975	6 25609	ManufacturerSpecific1	ProprietaryA	HS CAN	ManufacturerSpecific1		each signal
John Mitchell	3/23/18 10:07 PM	29874	8 25607	SourceAddress	ETC1	HS CAN	SourceAddress		roforoncod in th
John Mitchell	3/23/18 9:28 PM	19382	8 25605	InputShaftSpeed	ETC1	HS CAN	InputShaftSpeed		
John Mitchell	3/23/18 4:45 PM	12552	4 25603	ProgressiveShiftDisable	ETC1	HS CAN	ProgressiveShiftDisable	1	script. Use
John Mitchell	3/23/18 3:48 PM	8512	0 25601	MomentaryEngineOverspeedEnable	ETC1	HS CAN	MomentaryEngineOverspeed	dEnable	priority 1 if no
				PercentClutchSlip	ETC1	HS CAN	PercentClutchSlip		
				OutputShaftSpeed	ETC1	HS CAN	OutputShaftSpeed		optional list
				ShiftInProcess	ETC1	HS CAN	ShiftInProcess		, roquirod
				TorqueConverterLockupEngaged	ETC1	HS CAN	TorqueConverterLockupEnga	aged	l required.
				DrivelineEngaged	ETC1	HS CAN	DrivelineEngaged	1	L
				SourceAddress	EEC1	HS CAN	SourceAddress		[
				EngineSpeed	EEC1	HS CAN	EngineSpeed		1
				ActualEngine_PercTorque	EEC1	HS CAN	ActualEngine_PercTorque		f
				DriversDemandEngine_PercTorque	EEC1	HS CAN	DriversDemandEngine_Perc1	orque	f
				EngineRetarderTorqueMode	EEC1	HS CAN	EngineRetarderTorqueMode		
				RemoteAccelerator	EEC2	HS CAN	RemoteAccelerator		Z



April 30, 2019



Integrated Post Analysis – Road Map

Develop more example scripts

- Work with existing WirelessNeoVI customers to develop useful scripts to automate their current workflows
- Pull out key metrics from particular ECU feature and generate report
- Support more ways to visualize output data
 - View Excel output files directly on Wivi
 - Support other visualization utilities like ChartJS
- Multi-thread scripts to maximize performance
 - Currently scripts roughly process 1Gb/15sec
 - Multi-threading could speed that up to 1Gb/2sec
 - Custom C++ code cold be written with multi-thread for maximized performance.

Several features to implement on WirelessNeoVI

- Use Docker to create VMs to run each script
- Change permissions so that all users in an organization can view reports and run scripts. In order to view or download a report, a given user would have to have read access to all files in FileSet
- Support Public and Private reports



April 30, 2019



THANK YOU

For any queries or questions please feel free to contact:

John Mitchell Intrepid Control Systems, USA Madison Heights, MI, USA 48071 + 1 248 416 2848

Also, you can visit our website <u>www.intrepidcs.com</u> for more information



April 30, 2019

