**Meeting Notes - UC & Requirements + Security & Privacy Subgroups telecom, 02/04/2014**

Prepared by ilkay Altintas

NIST and RDA Big Data infrastructure groups communicate. NIST cannot implement any of the technologies. Implementation of scenarios is a good task for RDA WG. NIST’s job is to pick (out of 51 general use cases plus 9 security & privacy use cases) unique Big Data characteristic scenario use cases. We will then pass these use cases to the RDA WG to scale it down, pick best technologies for implementation. Our reference architecture should be able to handle any big data scenarios. It is too big, but helps to identify what is unique about different use cases, e.g., a scenario with multi-source sensors in. Hopefully the RDA WG will scale the suggested use cases down so it could be implemented and test the interface between the architecture components. Maybe we can also involve some professors so that graduate students can implement/test using our Reference Architecture.

ATTENTION:

1. v1 seven documents are now under technical editing. They are available on our website under:

<http://bigdatawg.nist.gov/V1_output_docs.php>. If there are other comments, please have them in ASAP. The first two documents for technical editing are Use Cases & Requirements and Security & Privacy Requirements.

2. v2 process is in place and the general task is to define the common interfaces between RA components. By the time v2 is done around June, v1 should be done for public release.

3. Daniel Samarov (NIST), a colleague of Wo will actively participate in the NBD-PWG. In case Wo is not available (due to travel, etc.), Daniel can answer questions related to NIST’s position and policy.

Face to face meetings one day either Monday June 23rd or Tuesday June 24th. **No objections to both days so it is set for Tuesday June 24th**. June 24th is also better for out of states travel.

Before RDA starts working, our goal is to identify scenarios from our submitted use cases. Some use cases may be common between NBD and RDA (e.g., Earth science.) We are hoping the RDA will start 3-4 weeks from now.

Geoffrey Fox asked for volunteers to champion each scenario: helping to write-up scale down use case and identify patterns. Bob Marcus has ten scenarios listed from the Jan. 21 meeting minutes (<http://bigdatawg.nist.gov/_uploadfiles/M0286_v1_2320535937.docx>). All of us should look into and identify which use cases are appropriate for implementation. Priority will be given to those have datasets available. NBD should actively engage with the use case submitters to engage collaboration within the next 3-4 weeks before RDA WG kicks-in.

Pseudo random data can use random number generators with seeds. An existing benchmark example such as the TPC benchmark can be used. On the imagery/geospatial side there are good available sets (openstreet map and others). Also, I believe that Ernest was from NGA and they have some good unclassified imagery datasets. Ernest will look into their unclassified datasets and provide use cases around them.

Arnab presented M0283\_v1\_6114110303.xlsx (SRA Mappings). The spreadsheet captures the security & privacy (SnP) issues in column one and two and the rest of columns are the RA key-/sub-components. The notation for “f” is the “security fabric” defined from the RA diagram where “x” refers to components that SnP has special roles. Rainbow Series from Mitre may be worth to look into but interested party needs to contact Mitre directly to get hold of the specs.

There were questions from the members about “authentication’ and Akhil (??) answered it from the cell A11 “Identity and Access Management”. Provenance is also important as it gives all chain of evidence for the data.

ISO/IEC JTC 1 Study Group on Big Data (SGBD): Workshops will be for presenting the thoughts (via abstracts) that will then turn into longer papers (4 – 6 pages but due June 30). All seven V1.0 documents will be submitted for presentation across all three SGBD meetings namely from US, Europe, and Asia. Regional meetings can enable more feedback.

Next meeting is on Tuesday Feb 11th at the same time and the focus is on Reference Architecture + Technology Roadmap.

ACTION ITEMS:

1. Geoffrey will send a set of use cases that could guide the NBD for the rest of the unique Big Data scenario.
2. All: should look into and start identify which use cases are appropriate for implementation. Priority will be given to those have datasets available.
3. M0283\_v1\_6114110303.xlsx (SRA Mappings) will be added as a google document for everyone to edit/extend the table.
4. Interested parties should start preparing abstracts (~300 words) for the first workshop in March. Abstract deadline is March 11th.

**Attendees:**

1. Wo Chang
2. Tim Zimmerlin
3. Pw Carey
4. Sanjay Mishra
5. Manoj Kumar
6. Geoffrey Fox
7. Ilkay Altintas
8. Orit Levin
9. K. Eric Harper
10. David Boyd
11. Mark Underwood
12. Akhil Manchanda
13. Thomas huang
14. Dan Samarov
15. William Miller
16. Nancy Grady
17. PavithraKenjige
18. Amy
19. Alicia Zuniga-Alvarado
20. Peggy
21. Ernest Smiley

**Web Log:**

(1:21 PM) Tim Zimmerlin (Automation Technologies): Hearing: that is not a valid passcode; please re-enter passcode...

(1:22 PM) Pw Carey\_Compliance Partners, LLC disconnected.

(1:23 PM) Sanjay Mishra (Verizon): Wo is trying to troubleshoot

(1:23 PM) Tim Zimmerlin (Automation Technologies): The problem with calling area code 206 is that is Seattle and so long distance.

(1:23 PM) Tim Zimmerlin (Automation Technologies): Can hear Wo now.

(1:25 PM) Manoj Kumar (CyberIQ): working now.

(1:27 PM) Geoffrey Fox: echo pretty bad!

(1:27 PM) Ilkay Altintas: There is quite some echo for me as well.

(1:28 PM) Manoj Kumar (CyberIQ): it seems some folks are on phone bridge as well.

(1:28 PM) Orit Levin (Microsoft) disconnected.

(1:28 PM) Manoj Kumar (CyberIQ): @Wo. problem could be you have both on - telephone and web.

(1:29 PM) Ilkay Altintas: I'm tryiong to take notes but I can't hear through the echo

(1:30 PM) K. Eric Harper (ABB) disconnected.

(1:31 PM) K. Eric Harper (ABB) joined.

(1:31 PM) Ilkay Altintas: Much better

(1:34 PM) Ilkay Altintas: The echo is back

(1:36 PM) Geoffrey Fox: Either day OK

(1:36 PM) David Boyd (Data Tactics): since we have our normal mtg on Tues that might make sense but either day works.

(1:36 PM) Mark Underwood (Krypton Bros): Tuesday better for NY commute

(1:37 PM) Ilkay Altintas: Is it Tuesday June 24th?

(1:37 PM) Geoffrey Fox: yes

(1:40 PM) Ilkay Altintas: I can hear very well still

(1:40 PM) Ilkay Altintas: not able to take very good notes

(1:41 PM) K. Eric Harper (ABB) disconnected.

(1:42 PM) Akhil Manchanda joined.

(1:44 PM) Ilkay Altintas: I will have to disconnect. I can understand a word through the echo.

(1:44 PM) Ilkay Altintas: Is someone else able to hear enough to take notes?

(1:46 PM) Thomas Huang/JPL disconnected.

(1:46 PM) Thomas Huang/JPL joined.

(1:47 PM) K. Eric Harper (ABB) joined.

(1:48 PM) Dan Samarov78 disconnected.

(1:48 PM) Dan Samarov joined.

(1:48 PM) william miller joined.

(1:49 PM) Nancy Grady (SAIC) joined.

(1:50 PM) Pw Carey, Compliance Partners, LLC joined.

(1:50 PM) Tim Zimmerlin (Automation Technologies): How do people think about using psuedo random data sets or streams???

(1:51 PM) Tim Zimmerlin (Automation Technologies): Psuedo random data can use random number generators with seeds.

(1:51 PM) william miller disconnected.

(1:53 PM) David Boyd (Data Tactics): Last mtg I suggested using an existing benchmark example such as the TPC benchmark.

(1:54 PM) Pw Carey, Compliance Partners, LLC disconnected.

(1:55 PM) Thomas Huang/JPL disconnected.

(1:55 PM) David Boyd (Data Tactics): On the imagery/geospatial side there are good available sets (openstreet map and others). Also, I believe that Ernest was from NGA and they have some good unclassified imagery datasets.

(1:55 PM) Tim Zimmerlin (Automation Technologies): David, yes, TPC benchmarks to push data into data sets.

(1:56 PM) Orit Levin (Microsoft) joined.

(1:57 PM) K. Eric Harper (ABB) disconnected.

(1:59 PM) David Boyd (Data Tactics): The following book may have some good ideas and data: Specifying Big Data Benchmarks: First Workshop, WBDB 2012, San Jose, CA, USA, May 8-9, 2012 and Second Workshop, WBDB 2012...

(1:59 PM) David Boyd (Data Tactics): From amazon: <http://www.amazon.com/gp/product/3642539734/ref=oh_details_o01_s00_i00?ie=UTF8&psc=1>

(2:00 PM) David Boyd (Data Tactics): I have not had time to read any of these papers yet but I am familiar with some of the work like the SWIM benchmark.

(2:01 PM) K. Eric Harper (ABB) joined.

(2:07 PM) K. Eric Harper (ABB) disconnected.

(2:09 PM) Orit Levin (Microsoft): Where does it show in the table?

(2:11 PM) Orit Levin (Microsoft): Where is the word "authentication" in the table?

(2:11 PM) K. Eric Harper (ABB) joined.

(2:11 PM) Orit Levin (Microsoft): Is this doc available?

(2:12 PM) PavithraKenjige disconnected.

(2:17 PM) K. Eric Harper (ABB) disconnected.

(2:18 PM) K. Eric Harper (ABB) joined.

(2:19 PM) David Boyd (Data Tactics): Some links to the rainbow series can be found here: <http://www.dynamoo.com/orange/links.htm>

(2:21 PM) Mark Underwood (Krypton Bros): Thanks David

(2:22 PM) David Boyd (Data Tactics): Here is actually a better link: <http://www.iwar.org.uk/comsec/resources/standards/rainbow/rainbow.html>

(2:23 PM) David Boyd (Data Tactics): Leave it to the brits to give us easier access to US documents.

(2:23 PM) Amy disconnected.

(2:24 PM) Tim Zimmerlin (Automation Technologies): Provenance includes the unbroken chain of evidence. The originator is only the start of the chain.

(2:26 PM) Orit Levin (Microsoft): Doc 283 has just this table. Text describing each of the taxonomy elements and the mapping between them and the high level taxonomy would be very helpful in order to move this VERY USEFUL taxonomy forward.

(2:27 PM) Mark Underwood (Krypton Bros): From looking at that site, it appears that parts of Rainbow are supeceded by later work I think FAS isn

(2:27 PM) K. Eric Harper (ABB) disconnected.

(2:27 PM) Mark Underwood (Krypton Bros): n't involved in reviewing these standards -- the last I checked

(2:28 PM) David Boyd (Data Tactics): Mark - correct

(2:28 PM) Nancy Grady (SAIC) disconnected.

(2:28 PM) Mark Underwood (Krypton Bros): Orit - agreed - we will try to do that

(2:28 PM) Mark Underwood (Krypton Bros): Our idea for successive versions of this document would be hyperlinked to the cell

(2:29 PM) Mark Underwood (Krypton Bros): so explanations would be easy to access

(2:35 PM) Tim Zimmerlin (Automation Technologies): Geoffrey, in your research including Apache, have you found any "identity mgt systems" for BD?

(2:35 PM) Mark Underwood (Krypton Bros): Since we are the S&P group, suggest we don't allow anonymous editing :)

(2:35 PM) K.Eric Harper (ABB) joined.

(2:37 PM) Geoffrey Fox: To Tim: Openstack has technologies in A&A area. We use on FutureGrid with an LDAP system to manage credentials

(2:39 PM) David Boyd (Data Tactics): We use a combination of LDAP, OpenSSO, and use the Accumulo element level security labels.

(2:40 PM) Mark Underwood (Krypton Bros): Tim - agree w/ your comments on provenance. It's one of my main specialization interests

(2:41 PM) Pw Carey, Compliance Partners, LLC joined.

(2:42 PM) PavithraKenjige joined.

(2:42 PM) Pw Carey, Compliance Partners, LLC disconnected.

(2:44 PM) K.Eric Harper (ABB) disconnected.

(2:47 PM) K.Eric Harper (ABB) joined.

(2:48 PM) Alicia Zuniga-Alvarado disconnected.

(2:50 PM) Tim Zimmerlin (Automation Technologies) disconnected.

(2:50 PM) Manoj Kumar (CyberIQ) disconnected.

(2:50 PM) Ilkay Altintas disconnected.

(2:51 PM) Orit Levin (Microsoft) disconnected.

(2:51 PM) David Boyd (Data Tactics) disconnected.

(2:51 PM) Mark Underwood (Krypton Bros) disconnected.

(2:51 PM) Dan Samarov disconnected.

(2:51 PM) K.Eric Harper (ABB) disconnected.

(2:51 PM) Peggy disconnected.

(2:55 PM) PavithraKenjige disconnected.