NIST Big Data Working Group (NBD-WD)

NBD-WD-2013/M0300

**Source: NBD-WG | All**

**Status: Information**

**Title: Meeting Notes and Chat Transcript**

**Author: Mark Underwood (Krypton Brothers)**

*Attendees: Arnab Roy, Mark Underwood, David Boyd, PW Carey, Orit Levin, Ilkay Altintas, Robert Whetsel, PavithraKenjige, Bob Marcus, William Miller, Eugene Luster, Tim Zimmerlin*

Caveat: Apologies for misinterpretations and misattributions (MU)

# Mark’s Notes

David Boyd – Re: Web traffic – Want to see what’s in it so we can look at what we would look at. Aspects: tracked individual, server (owner) demographics of the user demographics. Attack vector visits and honeypots.

Orit: How to prevent [] attack. Inferences can be made about what happens on the network. Also traffic engineering issues arise from weblog examinations.

David - [TPC](http://www.tpc.org/default.asp) – Generator tool – revisit this one.

Technologies – analyze the amount of data being leaked in response to queries. Studies have shown that if you think a DB is de-identified. What are the mechanisms that should be used to minimize it? Arnab responded: This is limited – you can only analyze some of this in real time. Emerging technology, but implementations are unlikely in the next few years. Some research is being done.

Risk management - Hadoop + relational data for analytical tools.

Robert: Technologies are blurred, but Data center clouds – e.g., Roy issue for federation – key management issues. It’s not just that data is in certain partitions but the ability of data to exist as well.

Review: cloud computing security work is complete and mature, but there is a subtopic on key management.

Robert: Has done some forensics events w/ Army Reserve – it’s for if we have to prosecute. Enter Big Data: “how are we going to crack that egg?” DISA day job. This is seen as a serious problem there.

PW: Anticipate the risk areas? – Fraud opportunity is limited, thinks fraud can be predicted. Proposal: this is a use case.

[DIACAP](https://en.wikipedia.org/wiki/Department_of_Defense_Information_Assurance_Certification_and_Accreditation_Process) and 503 (?) – the problem is that the audit log itself is the big data.

“A static defense is going to be broken.” The good guys have the advantage of looking back and seeing how things have changed.

**TODO** Review the existing S&P use cases to see which ones deal with defensive measures. Possibly not well covered.

Robert got some prototype with content – it’s a content system. It has variety and a lot of velocity.

Robert sent an abstract and PPT to Wo to distribute with his abstract.

* End of S&P Sub-Conversation

Ilke – reviewed what use cases and tools might be of interest in San Diego meeting, and how.

Robert or Eugene: We hope that we can gain a sense of the usefulness of the reference architecture by examining the tools chosen by the student team, the experiences they have.

Robert or Eugene: Suggests 1-3pp lead-in to the problems to be solved are? E.g., Motivation | Tips given the toy worlds – Geoffrey is fond of a third of these. We have the summaries that G did (1/3 pp) – to do more requires more about what’s needed. Geoffrey can provide the cloud platform, but additional guidance is needed. Geoffrey cites issues with student labor; they tend not to stick around. A “term paper” would be easier than a full implementation.

Diving the next level down is a challenge. The reference architecture is useful, but the next level down is unwieldy. Does it call for more detail in the use cases? Address this in San Diego; review which use cases would be most amenable to the project teams that we can feasibly stand up.

Request: Find out who’s going to the workshop. DISA would be interested in a stack (e.g., open stack on the grid). Geoffrey could offer a course that required one of these term projects. Ilke has some project-oriented potentials – but the work needs to be focused.

MOOC – Geoffrey has one but can develop. Non-credit or credit. People in the group and use case owners can participate. This will be on the agenda in San Diego. The MOOCs do attract a lot of professionals, Geoffrey observes. The professionals may be more likely to want to move toward Big Data projects.

Bob: suggests that that interfaces should be better described (drilled-down) and looking for potential standards. He believes that it’s premature to have students pursuing implementations in the absence of interface descriptions and standards.

Do the interfaces need to be described JSON or XML or . . . .?

Dave – it may not be so easy to get to the next level of detail without the problem exploding exponentially. E.g., Storm is different from initiating a map/reduce job.

Bob : Suggests that they feedback from experienced project practitioners. Thinks that a strategy is needed for the next 6 months, thinks we may a bit adrift.

Discussed [Future Grid](https://www.futuregrid.org/) – take VMs as defined . . .

Mark: Discussion is partly about how to move ahead with the right balance of practitioner and academic expertise and orientation.

Dave: A possible reset is to look at the RA and see what existing standards could be applicable. E.g., distributed file system extensions, SQL extensions, etc. What is needed on the roadmap in V2.

PW : NIST has a central role to play.

Geoffrey: Some practical and realistic plans to be considered

Following on the next page is the chat log.

# Chat Log

(9:59 AM) Mark Underwood (Krypton Bros) joined.

(10:00 AM) David Boyd (L-3 Data Tactics) joined.

(10:03 AM) Eugene Luster (R2AD/DISA CTO) joined.

(10:03 AM) Pw Carey, Compliance Partners, LLC joined.

(10:06 AM) Mark Underwood (Krypton Bros) disconnected.

(10:06 AM) Pw Carey, Compliance Partners, LLC: Sounds good from way over here.....

(10:06 AM) Orit Levin (Microsoft) joined.

(10:06 AM) Bob Marcus joined.

(10:07 AM) Pw Carey, Compliance Partners, LLC: Is there any updates regarding the 300 word March 5th deadline efforts.....?

(10:10 AM) Pw Carey, Compliance Partners, LLC: Over a year ago a security research company did an analysis of AWS Web logs regarding the accuracy of what they were reporting.....and it was amazing what the web logs had wrong.....including root kits and password captures, et cetera.....

(10:12 AM) David Boyd (L-3 Data Tactics): PW - I submitted one and asked Wo if there was a review process for acceptance and he said assume all were accepted.

(10:13 AM) Ilkay Altintas joined.

(10:14 AM) Pw Carey, Compliance Partners, LLC: Dear David: Thank you....&...Does that mean we won a PRIZE....?

(10:15 AM) Orit Levin (Microsoft): <https://www.w3.org/2014/strint/>

(10:15 AM) Pw Carey, Compliance Partners, LLC: We've heard that many times logging/tracking systems are set up re-configured, never looked at after awhile...until something goes South....then they become a hot item again.....

(10:16 AM) Pw Carey, Compliance Partners, LLC: No we didn't but we did see several publically available data sources......eg. National Data Archive, EU Data Archive et cetera......

(10:18 AM) Tim Zimmerlin joined.

(10:18 AM) Geoffrey Fox joined.

(10:19 AM) Robert Whetsel joined.

(10:19 AM) Orit Levin (Microsoft): The list of the papers are in the bottom of this page: <https://www.w3.org/2014/strint/report.html>

(10:19 AM) Robert Whetsel: sorry Im late my other meeting ran long

(10:20 AM) David Boyd (L-3 Data Tactics): ,what was the site they are discussing again?

(10:20 AM) Arnab Roy (Fujitsu): <https://www.opensciencedatacloud.org/publicdata/>

(10:20 AM) Pw Carey, Compliance Partners, LLC: Did you trip over any Fraud Data....?

(10:23 AM) PavithraKenjige joined.

(10:25 AM) Geoffrey Fox disconnected.

(10:25 AM) Geoffrey Fox joined.

(10:28 AM) Pw Carey, Compliance Partners, LLC: Some companies outsource the actual data analysis/analytics....loosing compliance control in the cheaper solutions over seas.....

(10:30 AM) Pw Carey, Compliance Partners, LLC: Currently, there is no regulation requiring the overseas analysts to identify themselves and/or the per centage of their responsibilities in the analysis....

(10:30 AM) Pw Carey, Compliance Partners, LLC: It's in the Standards process via the PCAOB....

(10:32 AM) Orit Levin (Microsoft) disconnected.

(10:32 AM) Orit Levin (Microsoft) joined.

(10:32 AM) Pw Carey, Compliance Partners, LLC: Be for we forget, is somebody taking notes, as we have to break away in 30 minutes.....?

(10:32 AM) Pw Carey, Compliance Partners, LLC: aka: before.....

(10:34 AM) Pw Carey, Compliance Partners, LLC: True...true....true....aka: GRC....

(10:35 AM) Pw Carey, Compliance Partners, LLC: We heard it was through a 3rd party vendor for tolet paper, et al....was how access was made....

(10:37 AM) Pw Carey, Compliance Partners, LLC: A speed bump non-the-less....but given the recent history with the NSA....this is just one solution the rest of the world is looking at.....

(10:42 AM) William Miller joined.

(10:42 AM) Pw Carey, Compliance Partners, LLC: Robert: Would like to chat with you regarding our efforts in Big Data Forensics.....?

(10:50 AM) William Miller: sounds like an access problem and that users, devices and applications do are not trusted to access the data

(10:51 AM) William Miller: we should discuss how to eliminate the exposure which is security relevant

(10:52 AM) Pw Carey, Compliance Partners, LLC: We're currently developing a tool in Java to do exactly that...forecast fraud all the way through the judicial system(s)....global....

(10:52 AM) Pw Carey, Compliance Partners, LLC: We agree this topic is open for further discussion

(10:53 AM) Pw Carey, Compliance Partners, LLC: Sorry our audio is breaking up....

(10:54 AM) Pw Carey, Compliance Partners, LLC: Can we chat off-line.....?

(10:55 AM) Robert Whetsel disconnected.

(10:56 AM) Pw Carey, Compliance Partners, LLC: We should reach out to Robert Whetsel for further discussion, Re: Fraud Analysis within the Cloud/Big Data Eco-Systems....

(10:58 AM) Pw Carey, Compliance Partners, LLC: Good meeting.....

(11:00 AM) PavithraKenjige disconnected.

(11:01 AM) Bob Marcus: I don't think it will be possible to get students to do intensive work implementing the Use Cases

(11:01 AM) PavithraKenjige joined.

(11:02 AM) Bob Marcus: Puyppet and Github are orthogonal to Big Data issues. Cloud is relevant but most Big Data applications are currently not public Cloud-based

(11:02 AM) Orit Levin (Microsoft): Dropping from the call now... See you in San Diego!

(11:02 AM) Pw Carey, Compliance Partners, LLC: Is there any type of reward for them....Research, Grants, Grades....ok...take care....

(11:03 AM) Orit Levin (Microsoft) disconnected.

(11:03 AM) Bob Marcus: It is urgent to drill down in the Reference Architectures and define interfaces between the components.

(11:03 AM) Pw Carey, Compliance Partners, LLC: Can you offer up an example of same....?

(11:04 AM) Pw Carey, Compliance Partners, LLC: Just a thought....

(11:04 AM) Bob Marcus: What potential standards are going to be discussed at the ISO meeting? We haven't developed any cosensus possibilities yet in the NBD-PWG

(11:06 AM) Pw Carey, Compliance Partners, LLC: From within the PWG world....the globe is divided into two halves.....one side requires greater protection.....and the other does not.....

(11:07 AM) Pw Carey, Compliance Partners, LLC: Greater privacy protection for the individual....while the other half not so much.....

(11:07 AM) Bob Marcus: I think that we need to rethink and focus our strategy after the ISO meeting

(11:08 AM) Pw Carey, Compliance Partners, LLC: Could we offer up some basic assumptions (right or wrong)....?

(11:10 AM) Bob Marcus: I don't think that we are getting very useful feedback on the Reference Architecture and potential standards from graduate students

(11:11 AM) Bob Marcus: The RA and standards are not related to research or industry job opportunities

(11:11 AM) Pw Carey, Compliance Partners, LLC: Anything wrong with undergrads....?

(11:12 AM) Bob Marcus: Why not get feedback from people with extensive Big Data industrial or research project experience?

(11:13 AM) Pw Carey, Compliance Partners, LLC: Dear Bob: We agree...collaborate with Silicon Valley start-ups in this regard....

(11:15 AM) Eugene Luster (R2AD/DISA CTO): Registration ends today.

(11:15 AM) Eugene Luster (R2AD/DISA CTO): I will attend as much as I can.

(11:18 AM) Bob Marcus: What do you expect to learn from the Masters Students about the RA and standards?

(11:19 AM) Bob Marcus: Who is going to drill-down on the RA, define interfaces, map Use Cases to the RA, and identify possible standards?

(11:19 AM) Geoffrey Fox: I doubt you will learn too much unless you give them a lot of help and focus

(11:20 AM) Geoffrey Fox: I am sort of trying to do drill down

(11:22 AM) Geoffrey Fox: I can offer a MOOC either free or for IU credit

(11:22 AM) Pw Carey, Compliance Partners, LLC: Upon successful completion of their efforts would the students receive and 'official certification of achievement from NIST'....?

(11:22 AM) Ilkay Altintas: I will need to leave now. See you next week in person!

(11:22 AM) Pw Carey, Compliance Partners, LLC: Aka: Certificate of Achievement from NIST....

(11:22 AM) Ilkay Altintas disconnected.

(11:22 AM) Bob Marcus: I don't see how Big Data project implementors can use the RA until drill-down is available and interfaces are defined

(11:22 AM) Pw Carey, Compliance Partners, LLC: OK...have a safe trip...

(11:23 AM) Geoffrey Fox: Bob -- as you know I agree with you

(11:24 AM) Pw Carey, Compliance Partners, LLC: Then by golly, Pw agrees with both of you.....

(11:25 AM) Bob Marcus: In my brainstorming document, I tried to illustrate the mapping of Use Cases to the RA, describe the interfaces, and identify possible standards. I think this will be necessary for other Use Case if we are to make progress.

(11:29 AM) Pw Carey, Compliance Partners, LLC: Besides hardware HW and software SW interfaces...what others are there.....open question...?

(11:32 AM) Pw Carey, Compliance Partners, LLC: 0297, Ver. 2 we will too.....thanks...

(11:35 AM) Pw Carey, Compliance Partners, LLC: Good points.....

(11:36 AM) Pw Carey, Compliance Partners, LLC: Is this being captured, too...?

(11:40 AM) PavithraKenjige disconnected.

(11:46 AM) PavithraKenjige joined.

(11:49 AM) Bob Marcus: SQL to horizontally scale data stores (e.g. Hadoop) is a possibility for standardization

(11:54 AM) Geoffrey Fox: I think a standard that captures Hadoop plus Spark (or other iterative MapReduce) is good

(11:54 AM) Pw Carey, Compliance Partners, LLC: This has been a very good meeting....and will some attending San Diego represent us RADICALS....please...Respectfully yours, Pw

(11:54 AM) Geoffrey Fox: bye

(11:54 AM) Pw Carey, Compliance Partners, LLC: That would be 'someone'....take care