**NIST Big Data Public Working Group (NBD-PWG)**

**NBD-PWD-2015/M0491**

**Source: NBD-PWG**

**Status: Draft**

**Title: Web chat from Meeting of Jan. 19, 2016**

**Chat Log D:\\_wo\1DMG\2015\\_BigDataWG\Docs\ChatLog 2016\_01\_19 15\_04.rtf**

**Cavan Capps (to Organizer(s) Only)**: 1:08 PM: Can you make me the presenter?

**Cavan Capps (to Organizer(s) Only)**: 1:13 PM: didn't mean to ask to be presenter

**Russell Reinsch (to Everyone)**: 1:16 PM: agenda 2.a: overlap day sounds like a good possibility. Could be a problem of facility availability.

**Bob Marcus (to Everyone)**: 1:17 PM: I agree that an overlap day would be very valuable. I think that we have enough output to make the discussion valuable.

**Russell Reinsch (to Everyone)**: 1:17 PM: Wo: Take 2 months [from now] to Asses group readiness.

**Bob Marcus (to Everyone)**: 1:20 PM: I think that external stakeholders will be disappointed if the NIST Cloud PWG and Big Data PWG outputs are not harmonized by the end of the year.

**Tim Zimmerlin (to Everyone)**: 1:21 PM: I agree with Bob. Today, the major elements of cloud and big data have been demonstrated repeatedly. Harmonizing is easy.

**Tim Zimmerlin (to Everyone)**: 1:22 PM: 1) HDFS; 2) Geoffrey's Apache BD stack; 3) Spark; 4) Containers. ...

**Tim Zimmerlin (to Everyone)**: 1:23 PM: 5) orchestrators like Puppet, Mesos, Kurbernettes, etc.

**Tim Zimmerlin (to Everyone)**: 1:24 PM: 6) Java Virtual Machine glues HL SW code to HW.

**Tim Zimmerlin (to Everyone)**: 1:24 PM: 7) SW shims like SPIDAL and VOLK.

**Bob Marcus (to Everyone)**: 1:25 PM: Big Data and Cloud perspective from industry (Intel)

**Bob Marcus (to Everyone)**: 1:25 PM: http://www.intel.com/content/dam/www/public/us/en/documents/product-briefs/big-data-cloud-technologies-brief.pdf

**Mark Underwood (to Everyone)**: 1:29 PM: Agree- overlap concept may be best, though we get painted with the cloud brush as a side effect

**Mark Underwood (to Everyone)**: 1:32 PM: From the perspective of the S&P fabric, there is a tendency to think of this bottom-up, which will slight app level and organizational issues we want to highlight

**Russell Reinsch (to Everyone)**: 1:33 PM: [Tim]: does Apache stack map to RA. [Dave]: primary interfaces are so diverse / tough to describe generalized way.

**Mark Underwood (to Everyone)**: 1:37 PM: Anyone gone to an Apache Flink intro meeting yet? (Got invited to one in NYC but cannot attend)

**Tim Zimmerlin (to Everyone)**: 1:42 PM: Co-evolution is a requirement. Cannot waste time & money throwing whole verndor systems away at regular time intervals.

**Tim Zimmerlin (to Everyone)**: 1:43 PM: HDFS is a common foundation. XML also is common. SQL views can thrive on top of both. RDF also can.

**David Boyd (to Everyone)**: 1:44 PM: Feel bad - I think I derailed the conversations.

**Tim Zimmerlin (to Everyone)**: 1:44 PM: Two weeks ago I listened to a Yahoo top technical lead explain how "it is impossible to keep up with real time". One week later, Yahoo announced major layoffs. Yahoo suffers from ETL blinders.

**Tim Zimmerlin (to Everyone)**: 1:45 PM: David, blame me...

**Tim Zimmerlin (to Everyone)**: 1:46 PM: Russell, yes, Apache stack does map well enough to use.

**Tim Zimmerlin (to Everyone)**: 1:50 PM: Wo, NIST could use github and yum for the six test BD datasets. People everywhere would use NIST's datasets and write reports. One key value is the variety of data in those six datasets.

**Tim Zimmerlin (to Everyone)**: 1:51 PM: Support for regression testing and challenge teams is another key value.

**Tim Zimmerlin (to Everyone)**: 1:58 PM: Cavan, IPsec is how certs & creds are managed and shared…for authentication, authorization, audit.

**Tim Zimmerlin (to Everyone)**: 1:59 PM: Hadoop was not designed for multi using SnP. Do not generalize to newSQL, though.

**Tim Zimmerlin (to Everyone)**: 2:00 PM: Java and JVM are commonly used with IPsec, in most all use cases…including IBM.

**Cavan Capps (to Everyone)**: 2:07 PM: Tim, I am used to thinking about IPsec in network based security, I am not used to thinking about it used between Python, R, Hive (or not), hbase, cassandra etc. (I guess if each of the software layers acted by message passing through the network - real or virtual - it could be done) Is that your understanding of how it could work now? Can it be fully enforced that way? - Thanks

**Tim Zimmerlin (to Everyone)**: 2:08 PM: Cavan, you raise an open question with noSQL open source apps.

**Tim Zimmerlin (to Everyone)**: 2:09 PM: My typical approach is to insert monitors and filter between untrusted part and the rest. One thing to note is reliance on reputation as a security motivator.

**Cavan Capps (to Everyone)**: 2:09 PM: thanks

**Tim Zimmerlin (to Everyone)**: 2:15 PM: Cavan, I also filter data streams for specific users, as a sub stream that advanced users can play with.

**Ann Racuya-Robins (to Everyone)**: 2:19 PM: Agree Privacy and Confidentially are different.

**Ann Racuya-Robins (to Everyone)**: 2:20 PM: Both have to be constructed

**Ann Racuya-Robins (to Everyone)**: 2:21 PM: I would say it is not entirely policy based but are technically constructed as well.

**Tim Zimmerlin (to Everyone)**: 2:21 PM: Traditionally, confidentiality is a security concern ala Confidentiality, Integrity, and Availability.

**Tim Zimmerlin (to Everyone)**: 2:22 PM: Traditionally, privacy is a legal and contractual concern.

**Ann Racuya-Robins (to Everyone)**: 2:23 PM: What do you mean by traditionally? Privacy has barely arrived on the scene?

**William Miller (to Everyone)**: 2:23 PM: I have reviewed multiple Big Data systems for the Internet of Things and all appear to have confientiality and privacy problems since the owner of devices have no contol over who has access to the data. IT is also posisble a 3rd party to take control of devices without their consent.

**William Miller (to Everyone)**: 2:24 PM: IIoT Cyber Physical Systems are vulnerabile to attack if using a Big data system as it is today.

**Ann Racuya-Robins (to Everyone)**: 2:25 PM: That would be helpful Cavan. I wasn't aware of the statistical perspective?

**Tim Zimmerlin (to Everyone)**: 2:25 PM: Ann, privacy has been around since the beginning of law and legal systems and legal standards and courts. Hundreds, if not thousands, of years.

**Tim Zimmerlin (to Everyone)**: 2:27 PM: William, experts claim the industrial IoT is secure today, with occasional cases of malfeasance.

**Tim Zimmerlin (to Everyone)**: 2:27 PM: The solution to privacy is rooted in consequences. The law has to burn people who violate privacy.

**William Miller (to Everyone)**: 2:30 PM: These are not firewall policies they myst define by the owner of a device or the data to be submitted. The subscriber must trust the data from a relaible data source. The provider or aggregator of the data is just faciliatating the transport but they need to provide the mechanisms to protect the confidentiality, privacy, and trust.

**Cavan Capps (to Everyone)**: 2:31 PM: I appreciate William's concern

**Tim Zimmerlin (to Everyone)**: 2:32 PM: William, firewalls are obsolete. Restricting routes, network probes, air gaps, etc.

**William Miller (to Everyone)**: 2:32 PM: yep

**Cavan Capps (to Everyone)**: 2:32 PM: I think this should be policy agnostic, not every nation will have the same policy; and I think privacy policies are still evolving

**Tim Zimmerlin (to Everyone)**: 2:33 PM: Security is a system of mechanisms. Privacy is an outcome over a time window like years.

**William Miller (to Everyone)**: 2:33 PM: There need to be government oversight to define the acceptable policies and enforcement of this policies.

**William Miller (to Everyone)**: 2:34 PM: This willb e discussed at George Washington University the ffirst week of Febrary 2016 with IEEE, NSF, and Internet 2.

**Tim Zimmerlin (to Everyone)**: 2:34 PM: William, who trusts the government? Who watches the watchers? And, I will be there.

**Cavan Capps (to Everyone)**: 2:34 PM: The question might be are there technology implimentations that limit the adoption of particular policies

**William Miller (to Everyone)**: 2:35 PM: This is guidance but may require penalties

**Frank Farance (to Everyone)**: 2:38 PM: Sorry need to drop suddenly. Really good discussion.

**William Miller (to Everyone)**: 2:41 PM: I belive a bottoms up approach versus top down is better since the privacy and confidnetiality discuss begins with the way data is collected. Web services are vulnerable and current IoT protocls are centralized exposing the users data. Control of the devices is another problem sicne it is easy to exploited the data which is open not encrypted in the Big Data systems.

**Ann Racuya-Robins (to Everyone)**: 2:42 PM: I agree with you on this William

**Tim Zimmerlin (to Everyone)**: 2:46 PM: The 2020 Census will use the Internet. What could go wrong?

**William Miller (to Everyone)**: 2:46 PM: How about dgitial signing of sensor data and use of tokenization for sessions, devices, and users access that can restrict privileges of access to sensor or acturoator or particular Big Data information.

**William Miller (to Everyone)**: 2:47 PM: It also states with IoT Discovery and PRovisoning of devices and data sources.

**Mark Underwood (to Everyone)**: 2:49 PM: I have to go to another meeting in 5 min

**Cavan Capps (to Everyone)**: 2:50 PM: William, I am also interested in the delimas of discovery and then generalized provisioning (with security integrated)

**Ann Racuya-Robins (to Everyone)**: 2:53 PM: If you can send to the list that would be helpful Mark

**Mark Underwood (to Everyone)**: 2:54 PM: Doc title was "NBDWG Big Data Privacy Vocabulary"

**Mark Underwood (to Everyone)**: 2:55 PM: Thanks everyone - Ciao

**William Miller (to Everyone)**: 2:56 PM: Check out XEP-0347 IoT Discovery and XEP-0324 Provisioning on http://www.xmpp.org

**William Miller (to Everyone)**: 2:56 PM: Also learn more at http://www.sensei0iot.org

**Arnab Roy (to Everyone)**: 2:57 PM: I need to go too. Good feedback - thanks everyone!

**William Miller (to Everyone)**: 2:57 PM: IoT Discovey and RPovisonign are beign discussed with W3C to address these issues.

**William Miller (to Everyone)**: 3:02 PM: Correction Provisioning sorry for spelling

**Cavan Capps (to Everyone)**: 3:02 PM: William, thanks for the info on the Iot meetings

**Ann Racuya-Robins (to Everyone)**: 3:03 PM: Got to go. Thank you all.