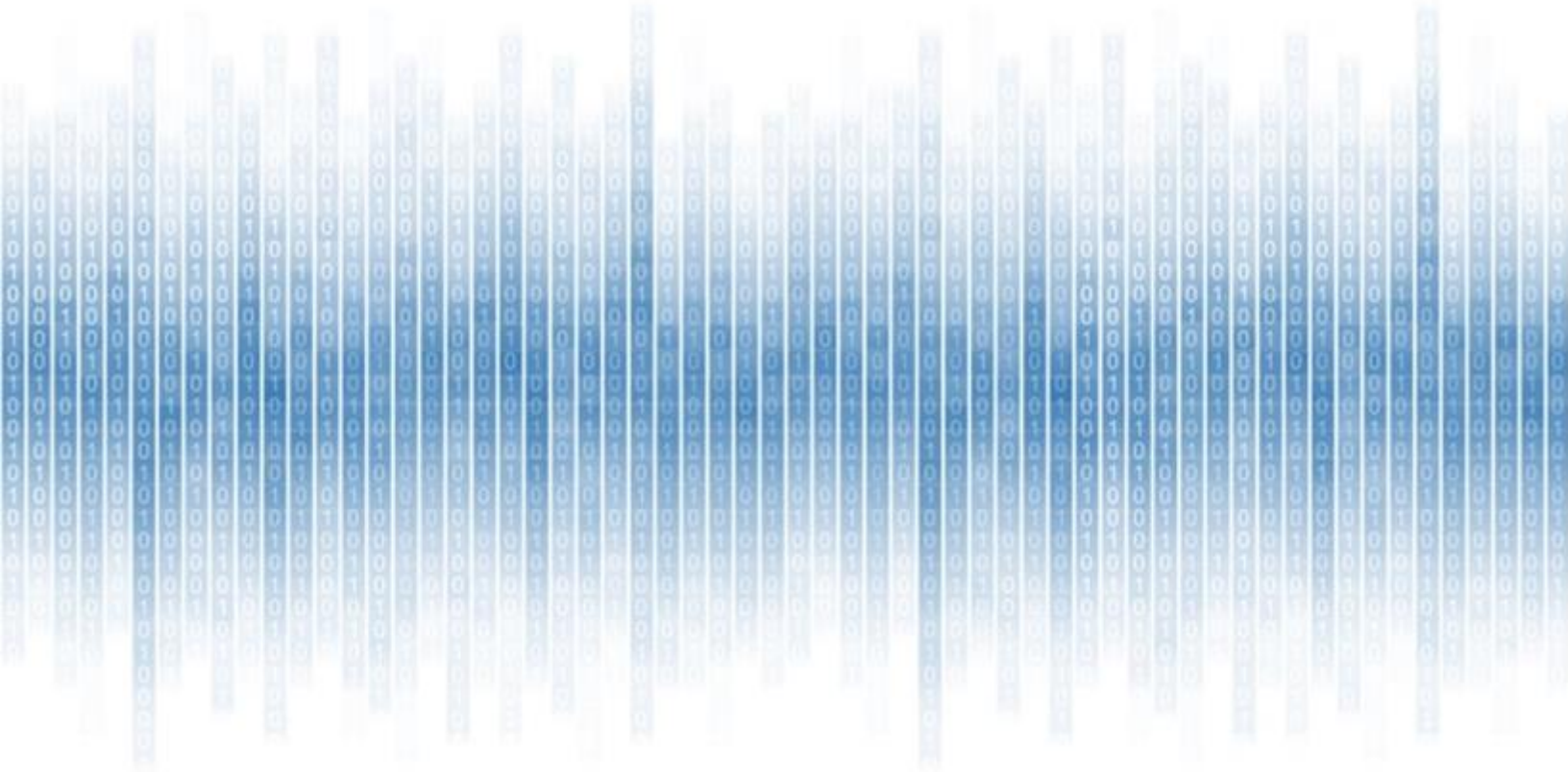
cloud security alliance℠



Big Data Working Group

Comment on Big Data and the Future of Privacy

March 2014

(1) What are the public policy implications of the collection, storage, analysis, and use of big data? For example, do the current U.S. policy framework and privacy proposals for protecting consumer privacy and government use of data adequately address issues raised by big data analytics?

Public policy implications have a bearing on Access, Ownership, Privacy, Liability, and Transparency.

Privacy protection has become an elusive goal in the big data era as researchers have shown that "linkability threats" can re-identity individuals. Due to the highly personal nature of data of individuals, the policy framework should lead to best practices to store and transmit the data. Existing practices focus on keeping data encrypted at rest and in transit with an infrastructure to ensure proper authorization and authentication of entities to get access to the data. With the advent of big data era, analytics tools require access to raw data for generating information of high value for both individuals as well as third party organizations. In practice, such data is shared after sufficient removal of unique identifiers by the processes of anonymization and aggregation. This process which has led to very many instances of re-identification based on big data linkability needs to be strengthened. The policy framework needs to address systematizing privacy preserving data disclosure mitigating linkability threats in the big data era. Specifically, the policy framework might have to lead to enforcement that all linkable data be encrypted. Furthermore, the policy framework needs to address the concerns on the geo location where the data is stored. As well as enforcement of transparency: individuals have the right to know which party has which access to their data, how the (raw) data is used and how it is protected.

(2) What types of uses of big data could measurably improve outcomes or productivity with further government action, funding, or research?

Following are examples of some of the uses of big data to improve outcomes:

1. Jobs data matching geo location data and education data will lead to better employment outcomes.
2. Sharing cyber threat intelligence among multiple businesses will lead to thwarting potential cyber threats to national infrastructure. There is a need for more funding in developing big data analytics techniques in cyber security.
3. Big data analytics on encrypted data to thwart linkability threats.
4. Improvements in health care, leading to more personalized medicine and treatment. This should also result in a more cost effective health care system.
5. Smarter city and transport infrastructure, leading to a most cost effective and greener environment, lesser and faster commute.

What types of uses of big data raise the most public policy concerns?

The following are the most public policy concerns:

1. Correlation of disparate data such as healthcare, financial, demographic and location data.

2. Tracking consumer behavior and sharing them with 3rd party without proper authorization for targeting and other purposes.

3. Big data storage in the cloud across multiple geo boundaries

4. Lack of transparency: who has access to which data, which data is collected and for what reason.

Are there specific sectors or types of uses that should receive more government and/or public attention?

Healthcare Education, Financial wellness, Employment, Mobility, and Information Access are some of the specific sectors that should receive more government/public attention.

(3) What technological trends or key technologies will affect the collection, storage, analysis and use of big data?

Predictive analytics, real time analytics, complex event processing, stream computing, high performance computing, cloud computing, deep machine learning algorithms, open source technologies, visualization and mobile apps usage are some of the technological trends that will affect the collection, storage, analysis and use of big data.

Are there particularly promising technologies or new practices for safeguarding privacy while enabling effective uses of big data?

(Somewhat) Homomorphic Encryption and Differential Privacy are some of the promising technologies for safeguarding privacy while enabling effective uses of big data. It should be noted that although technologies play an important role in the safeguarding of privacy, the approach should also include, amongst others, legal and administrative aspects.

The Cloud Security Alliance (CSA) Big Data Working Group (BDWG) has come up with 100 best practices to enhance the security and privacy of big data:
https://docs.google.com/document/d/1FqeHlA53sliNS3sd3ECy2hwyJu0UJDZT71zUs-02nX4/edit#

The top 10 best practices are listed below:

1. Authorize access to files by predefined security policy

2. Protect data by data encryption while at rest

3. Implement Policy Based Encryption System (PBES)

4. Use antivirus and malware protection systems at endpoints

5. Use big data analytics to detect anomalous connections to cluster

6. Implement privacy preserving analytics

7. Consider use of partial homomorphic encryption schemes

8. Implement fine grained access controls
9. Provide timely access to audit information
10. Provide infrastructure authentication mechanisms

(4) How should the policy frameworks or regulations for handling big data differ between the government and the private sector?

The policy frameworks and regulations for handling big data differ between the government and private entities in the context of the ability of the government to adjudicate -- for example, in policies governing demographics data in law enforcement and government investment of tax dollars.

(5) What issues are raised by the use of big data across jurisdictions, such as the adequacy of current international laws, regulations, or norms?

The question of where the data is stored, where the data is processed and where the data analytics results are distributed influence the cross boundary jurisdictions pertaining to privacy policies and regulations.