TeleTrusT – Bundesverband IT-Sicherheit e.V.

Der IT-Sicherheitsverband.

24.06.2014

Positionspapier der TeleTrusT-Arbeitsgruppe "Cloud Security"

"Big Data"

1. Begriffsbestimmung

Big Data bezeichnet große Datenmengen (Volume) unterschiedlichen Typs aus vielfältigen unterschiedlichen Quellen, sowohl strukturiert wie auch unstrukturiert (Variety), die automatisch mit hoher Geschwindigkeit (Velocity) erzeugt und verarbeitet werden, um daraus neue Informationen zu gewinnen und ggf. Entscheidungen abzuleiten. Eine weitere Eigenschaft von Big Data ist die unterschiedliche Qualität der Daten hinsichtlich Vertrauenswürdigkeit und Korrektheit (Veracity).

2. Bedeutung

Zwei Faktoren tragen zu einer rasch wachsenden Bedeutung von Big Data bei:

1. Eine zunehmende Zahl von Systemen produziert immer mehr Daten. Beispiele sind automatische Abrechnungssysteme, bildgebende Diagnoseverfahren, digitale Motorsteuerungen, in Maschinen eingebaute Sensoren, Gebäudeautomation und Videoüberwachungen. Stichworte dazu sind u.a. "Internet of Things", "Industrie 4.0" und "Smart Meter".
2. Bislang wurden viele dieser Daten nur kurzzeitig gespeichert oder unmittelbar verarbeitet und danach verworfen. Die rasant wachsenden IT-Kapazitäten (Rechenleistung, Speicher, Netzwerke) ermöglichen jetzt aber die Speicherung und Verarbeitung einer schnell wachsenden Menge dieser Daten.

3. Chancen

In der Kombination einer schnell wachsenden Datenmenge einerseits und der schnell wachsenden Fähigkeit zur Speicherung und Auswertung dieser Daten andererseits können Informationen gewonnen werden, die mit herkömmlichen Verfahren bislang nicht oder kaum zu ermitteln sind, z.B. durch Herstellung von Querbezügen und Auswertung statistischer Zusammenhänge. Beispiele sind:

- Berechnung lokaler Wetterprognosen;
- Aufdeckung geplanter terroristischer Anschläge aus der Analyse von Kommunikationsdaten;
- Aufdeckung von Angriffen auf die IT-Sicherheit und automatische Reaktion darauf durch Analyse von Zugriffsversuchen, Verkehrsdaten, Logfiles usw.;
- Vorhersage von Verkehrsspitzen und dynamische Verkehrslenkung durch Auswertung von Daten aus Navigationssystemen;
- Rechtzeitige Maschinenwartung durch Analyse von Sensordaten;
- Bedarfsgerechte Steuerung und Speicherung von Energie basierend auf aktuellen Verbrauchsdaten und kurzfristigen Verbrauchsprognosen;
- Markt-, Meinungs- und Sozialforschung durch Auswertung von Social Media;
- Nutzerprofilierung im Web, Retargeting.

4. Risiken

Die durch Auswertung der großen Datenmengen gewinnbaren Informationen gehen oft über das hinaus, was der Eigentümer der Daten mit der Erzeugung dieser Daten eigentlich bezweckt. Beispiele dafür sind:

- Rückschlüsse auf das konsumierte Fernsehprogramm aus Schwankungen des Stromverbrauchs;
- Rückschlüsse auf eingesetzte Systeme aus statistischen Werten übermittelter Daten;
- Ermittlung und Verfolgung des Aufenthaltsortes von Personen aus Mobilfunkdaten;
- Erstellung eines sozialen Profils von Personen aus der Analyse öffentlich verfügbarer Daten (Wohnort, Arbeitgeber, soziales Netzwerk in Facebook, besuchte Webseiten usw.).

Problematisch ist bei Big Data vor allem, dass aufgrund der sehr großen Menge von Daten und den herstellbaren Querbezügen und statistischen Zusammenhängen eine nachträgliche Personalisierung zuvor anonymisierter Daten in vielen Fällen relativ einfach möglich wird. Die dadurch entstehenden Möglichkeiten gehen weit über die Besorgnisse von Datenschützern bezüglich Rasterfahndung und Volkszählung aus den siebziger und achtziger Jahren hinaus.

5. Forderungen

Aus Sicht von TeleTrusT muss angesichts der Gefahren von Big Data ein gesellschaftlicher Dialog darüber geführt werden, was bezüglich der Erzeugung und Auswertung von Daten künftig erwünscht und was unerwünscht ist. Basis der Überlegungen muss eine Chancen-/Risikenabwägung sein, wie sie auch in anderen Technologiebereichen stattfindet (z.B. Biotechnologie, Kernenergie).

Mit Blick auf die Risiken von Big Data sehen wir folgende Anforderungen:

1. Datensparsamkeit
   Es dürfen nur solche Daten gespeichert werden, die für die Erledigung einer Aufgabe tatsächlich benötigt werden. Daten dürfen nicht auf Vorrat gespeichert werden, im Hinblick auf einen später eventuell denkbaren Nutzen. Für jede Erzeugung und Speicherung von Daten in größeren Mengen müssen Nutzen und Notwendigkeit dokumentiert werden.

2. Verschlüsselung und Zugang
   Die Speicherung von Daten muss, soweit irgend möglich, verschlüsselt erfolgen. Der Zugang zu Daten durch Dritte muss eingeschränkt sein. Die Übertragung von Daten mit potentiellem Personenbezug in einen "unsicheren Drittstaat" muss verboten sein.

3. Verbot der nachträglichen Personalisierung
   Eine nachträgliche Personalisierung von zuvor anonymen Daten muss generell verboten sein und darf nur unter gesetzlich genau geregelten Voraussetzungen ausnahmsweise erlaubt werden.