Distributed Big Data Assets Sharing & Processing

Trusted Data Processing in Untrusted Environments.

C. de Laat (moderator), L. Gommans, R. Wilson

System & Network Engineering, University of Amsterdam
AirFrance KLM
CIENA
Fading Trust in Internet Dependency Trust

Research Gap!

1980 2017
Main problem statement

- Organizations that normally compete have to bring data together to achieve a common goal!
- The shared data may be used for that goal but not for any other!
- Data may have to be processed in untrusted data centers.
  - How to enforce that using modern Cyber Infrastructure?
  - How to organize such alliances?
  - How to translate from strategic via tactical to operational level?
  - What are the different fundamental data infrastructure models to consider?
Networks of ScienceDMZ’s & SDX’s

ISP
NFV
Networks of ScienceDMZ’s & SDX’s

cclient 1
cclient 2
cclient 3
cclient 4

c…
cclient n

SDN
Internet
Peer ISP’s

Supercomputing centers (NCSA, ANL, LBNL)

ISP
NFV

ISP
SDX
DTN
SDX

ISP
NFV

ISP
SDX

ISP
SDX

ISP

downloads

Ownership/trust relation

Petabyte email service 😊

contains DTN

Supercomputing centers (NCSA, ANL, LBNL)
Secure Policy Enforced Data Processing

- Bringing data and processing software from competing organisations together for common goal
- Docker with encryption, policy engine, certs/keys, blockchain and secure networking
- Data Docker (virtual encrypted hard drive)
- Compute Docker (protected application, signed algorithms)
- Visualization Docker (to visualize output)
Ambition to put capabilities into fieldlab

Re-enforcing ICT preconditions:
Each envisaged site has similar elements
Program:
15h00 Cees de Laat, University of Amsterdam
   Trusted Data Processing in Untrusted Environments.
15h05 Leon Gommans, Air France KLM
   Trusted Big Data Sharing.
15h25 Rodney Wilson
   Programmable Supernetworks, Science DMZ based Networking.
15h30 Panel of stakeholders  Flash talks (~3 min each):
   Inder Monga - ESnet - Data Science Driving Discovery.
   Matt Zekauskas - Internet2 - Thoughts on Internet2 and Trusted Large Data Transfer.
   Jerry Sobieski - NORDUnet - Issues of Big Data Sharing in a Global Science Collaboration.
   Adam Slagell – NCSA - What are we trusting?
15h45 Panel discussion moderated by Cees de Laat
16h00 End of session.