Secure Big Data Processing

@ dcypher, Utrecht, Oct 4, 2017

Cees de Laat

System & Network Engineering
University of Amsterdam

SARNET: Security Autonomous Response with programmable NETworks

Marc Lyonais, Leon Gommans, Rodney Wilson, Rob Meijer,
Frank Fransen Tom van Engers, Paola Grosso, Cees de Laat,
Ameneh Deljoo, Ralph Koning, Ben de Graaff, Gleb Polevoy, Stojan Travanovski.

UNIVERSITY OF AMSTERDAM

AIRFRANCE KLM

ciena

NWO
Netherlands Organisation
for Scientific Research

TNO

COMMIT/

Big Data: real time ICT for logistics
Data Logistics 4 Logistics Data (dl4ld)

Jan Burgmeijer, TNO, PI, Cees de Laat, UvA, Co-PI, Leon Gommans, KLM

simacan

TNO

AIR FRANCE KLM

ORACLE

ciena

THALES

evofenedex

NWO
Netherlands Organisation
for Scientific Research

Gemeente
Amsterdam

TRANSFIDES

SARNET: Security Autonomous Response with programmable NETworks

Marc Lyonais, Leon Gommans, Rodney Wilson, Rob Meijer,
Frank Fransen Tom van Engers, Paola Grosso, Cees de Laat,
Ameneh Deljoo, Ralph Koning, Ben de Graaff, Gleb Polevoy, Stojan Travanovski.

UNIVERSITY OF AMSTERDAM

AIRFRANCE KLM

ciena

NWO
Netherlands Organisation
for Scientific Research

TNO

COMMIT/

Big Data: real time ICT for logistics
Data Logistics 4 Logistics Data (dl4ld)

Jan Burgmeijer, TNO, PI, Cees de Laat, UvA, Co-PI, Leon Gommans, KLM

simacan   TNO   AIR FRANCE KLM

Z   ORACLE   ciena   THALES

evofenedex   NWO
Netherlands Organisation
for Scientific Research   Gemeente Amsterdam   TRANSFIDES

Cyber security program SARNET

Research goal is to obtain the knowledge to create ICT systems that:

- model their state (situation)

- discover by observations and reasoning if and how an attack is developing and calculate the associated risks

- have the knowledge to calculate the effect of counter measures on states and their risks

- choose and execute one.
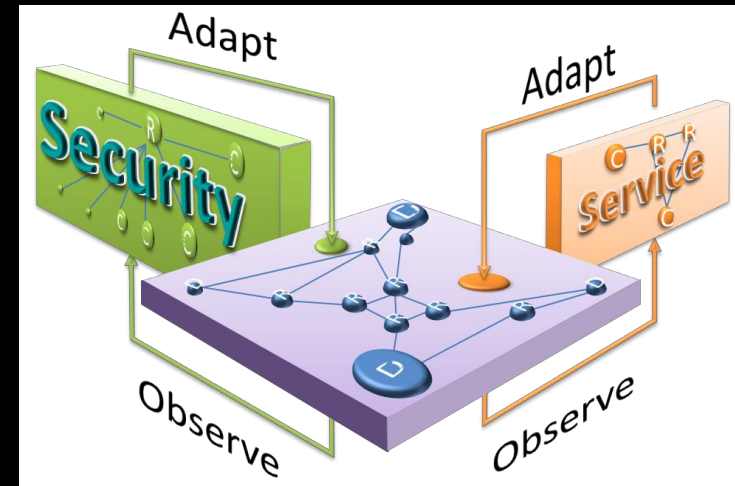
In short, we research the concept of networked computer infrastructures exhibiting SAR: Security Autonomous Response.

Context & Goal

Security Autonomous Response NETwork Research
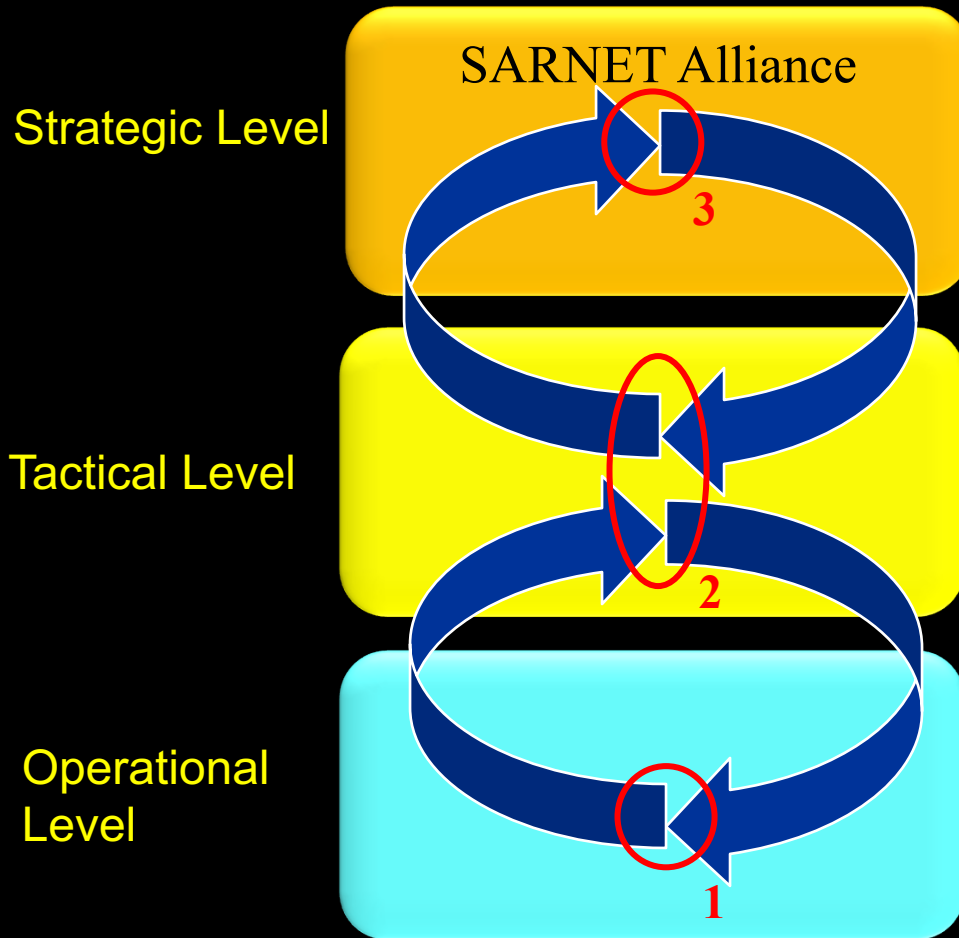
Ameneh Deljoo (PhD):
Why create SARNET Alliances?
Model autonomous SARNET
behaviors to identify risk and benefits
for SARNET stakeholders (**3**)

Gleb Polevoy (PD):
Determine best defense scenario
against cyberattacks deploying
SARNET functions (**1**) based on
security state, KPI information (**2**)
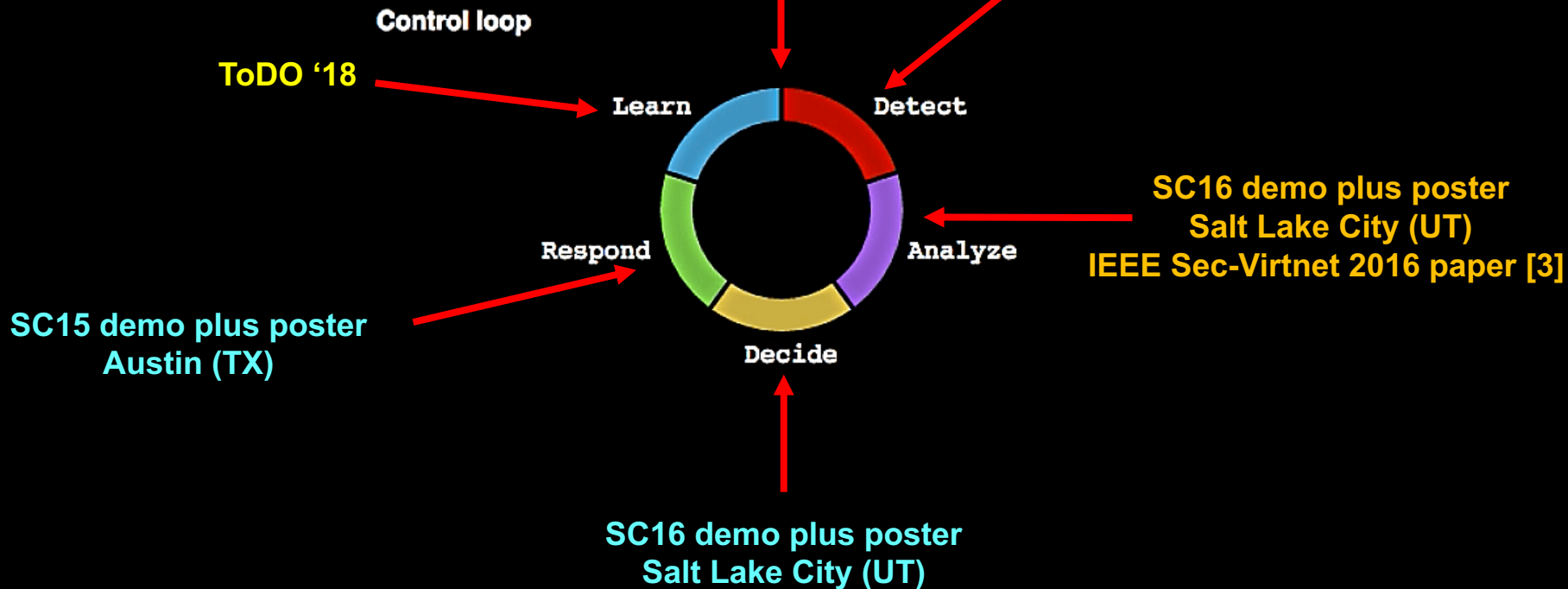keeping in mind strategic motifs (**3**).

Ralph Koning (PhD)
Ben de Graaff (SP):
1. Design functionalities needed to
operate a SARNET using SDN/NFV
2: deliver security state and KPI
information (e.g cost)

Strategic Level

Tactical Level

Operational
Level

SARNET Alliance

Status SARNET Operational Level

Laboratory: ExoGeni & PRP
Fieldlab with KLM & CIENA
OSA-Optical Forum Conference paper [1]

CoreFlow
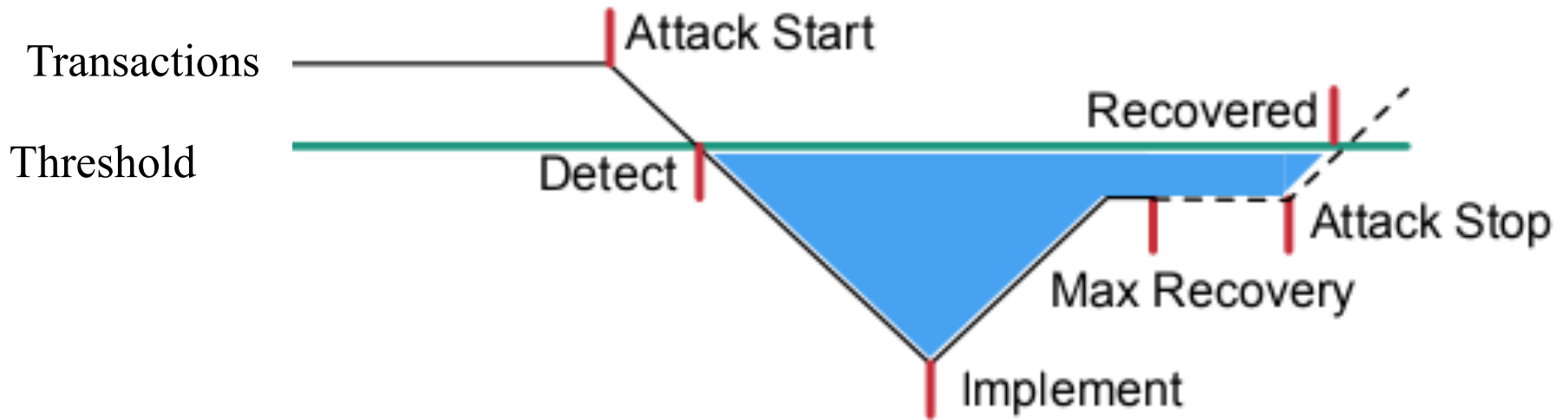**Berkeley Internship 2016**
SC16 INDIS workshop paper [2]

Control loop

ToDO '18

Learn Detect

Respond Analyze

Decide

SC16 demo plus poster
Salt Lake City (UT)
IEEE Sec-Virtnet 2016 paper [3]

SC15 demo plus poster
Austin (TX)

SC16 demo plus poster
Salt Lake City (UT)

1. Paper: R. Koning, A. Deljoo, S. Trajanovski, B. de Graaff, P. Grosso, L. Gommans, T. van Engers, F. Fransen, R. Meijer, R. Wilson, and C. de Laat, "Enabling E-Science Applications with Dynamic Optical Networks: Secure Autonomous Response Networks ", OSA Optical Fiber Communication Conference and Exposition, 19-23 March 2017, Los Angeles, California.
2. Paper: Ralph Koning, Nick Buraglio, Cees de Laat, Paola Grosso, "CoreFlow: Enriching Bro security events using network traffic monitoring data", SC16 Salt Lake City, INDIS workshop, Nov 13, 2016.
3. Paper: Ralph Koning, Ben de Graaff, Cees de Laat, Robert Meijer, Paola Grosso, "Analysis of Software Defined Networking defences against Distributed Denial of Service attacks", The IEEE International Workshop on Security in Virtualized Networks (Sec-VirtNet 2016) at the 2nd IEEE International Conference on Network Softwarization (NetSoft 2016), Seoul Korea, June 10, 2016.

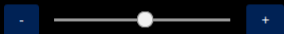# Effectiveness and Impact

Scenario

SC16 DEMO STARNET Operational Level

Example application: Spoofed Network Traffic
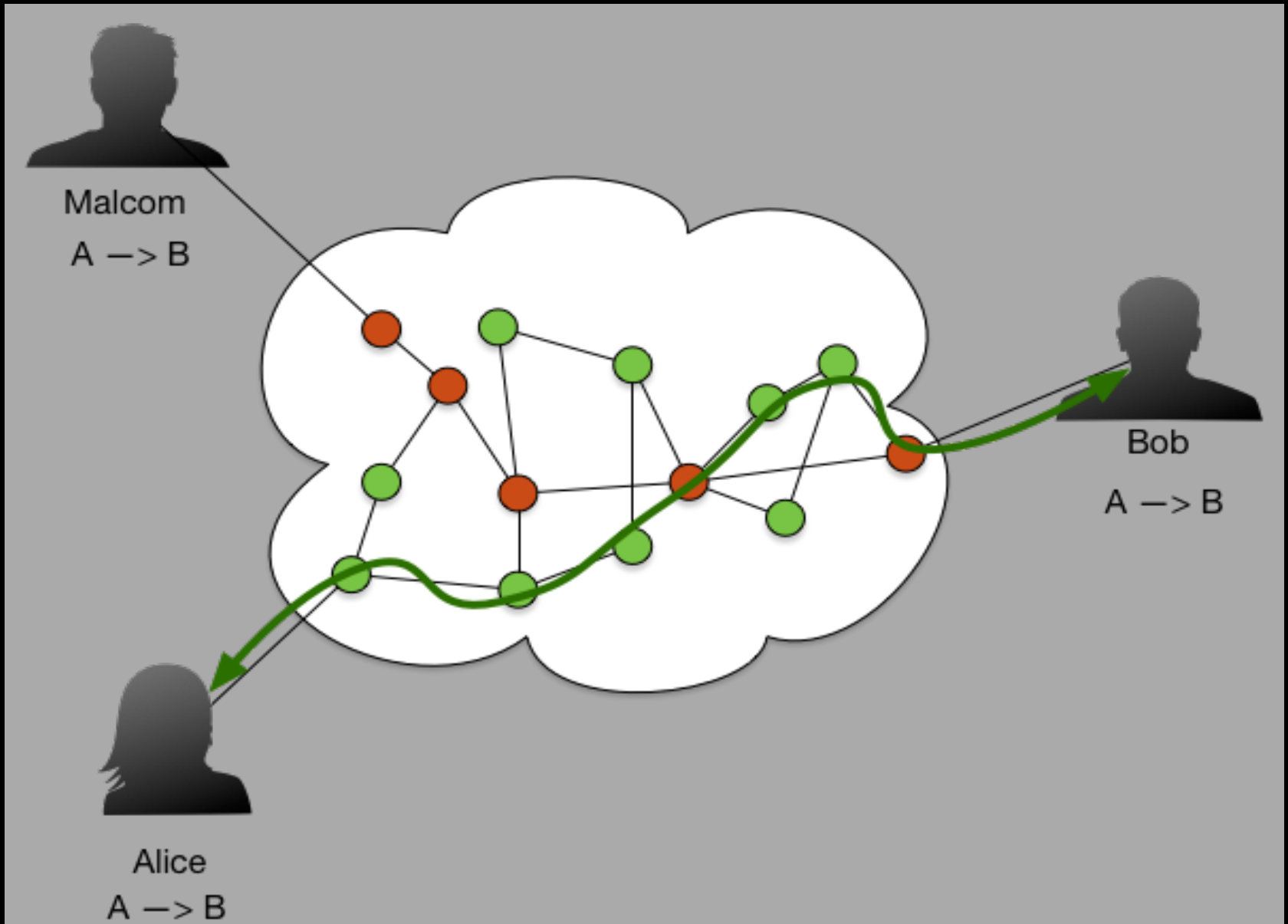
Analyse part: CoreFlow motivation

To effectively block attacks, the information from an IDS is not always sufficient

When an event triggers, the security team has to **manually collect** additional **data from different sources** to enrich the event to **create context and understanding** of the event.

CoreFlow can auto this process by automatically correlating the security events to available data sources and provide this context.

In this prototype we focus on the following sources:

- **Bro** - Generates the events

- **NetFlow** - To add network traffic information

- **Route Explorer** - To assist in determining paths

CoreFlow Route estimation algorithm

- It's able to fill in missing routers
- Flow traverse a router multiple times (loops)
- Finds potential 'shortest paths'

- Topology information from OSCARS
- Based on latest topology
- Does not account for policies or metrics
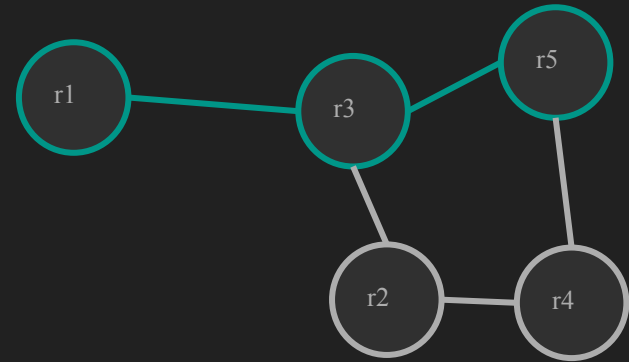
Unordered route:	Get possible routes from r3:	Reverse	Concat	Shortest
r3, r1, r5	r3, r1	r1, r3	r1, r3, r1	r1, r3, r5
	r3, r5	r5, r3	r1, r3, r5	r5, r3, r1
	r3, r2	r2, r3	r1, r3, r2	
	r3, r5, r4	r4, r2, r3	r1, r3, r2, r4	
	r3, r2, r4	r4, r5, r3	r1, r3, r5, r4	
			…	

Current workflow
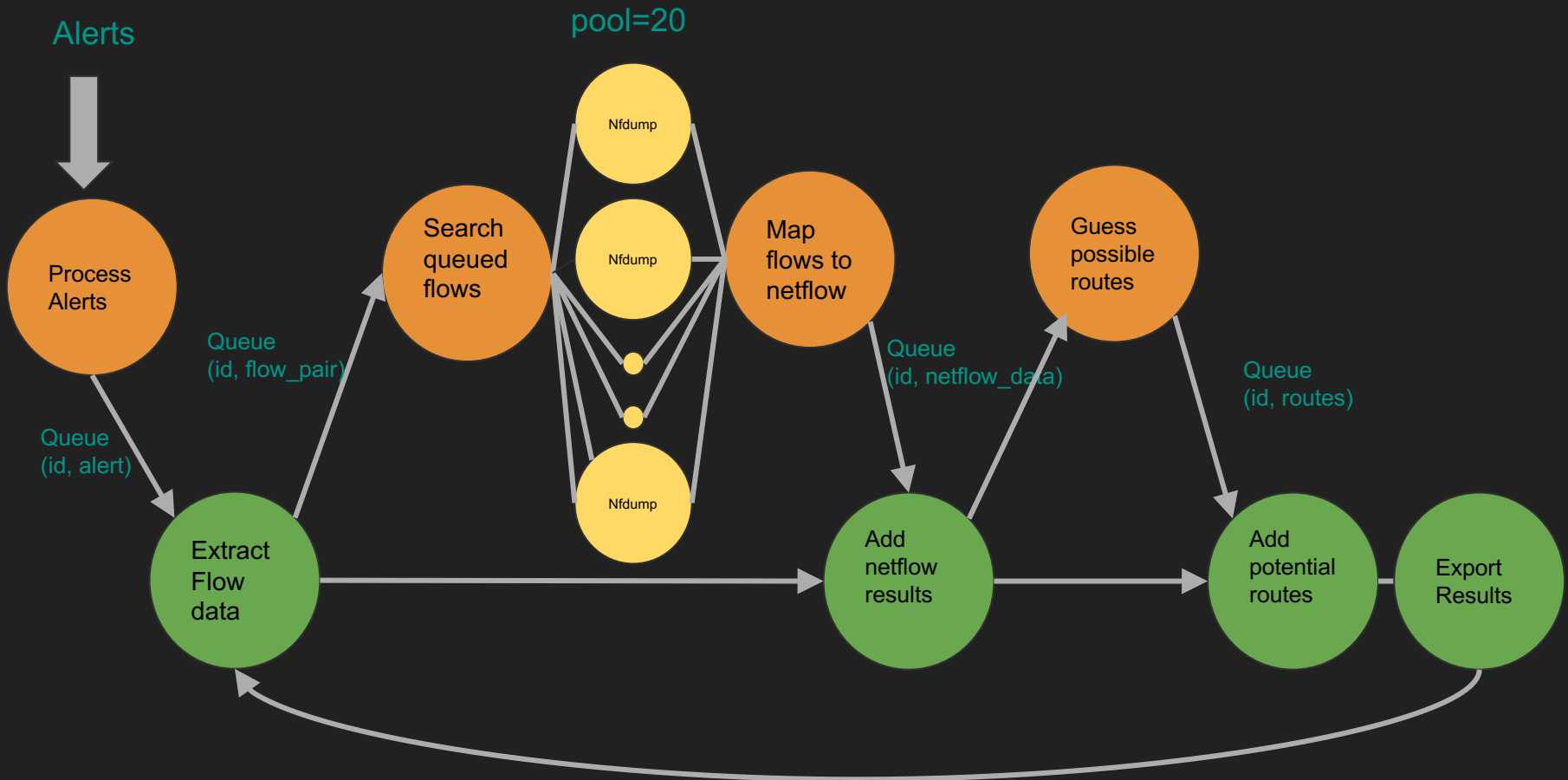
Agent Based Modelling Framework

	Main component
Signal layer	Message / Act
Action layer	Action / Activity
Intentional layer	Intention
Motivational layer	Motive

In our model, we refer to four layers of components:

➢ the signal layer— describes acts, side-effects and failures showing outcomes of actions in a topology.

➢ the action layer—actions: performances that bring a certain result,

➢ the intentional layer—intentions: commitments to actions, or to build up intentions,

➢ the motivational layer—motives: events triggering the creation of intentions.

Simplified Eduroam case at signalling layer
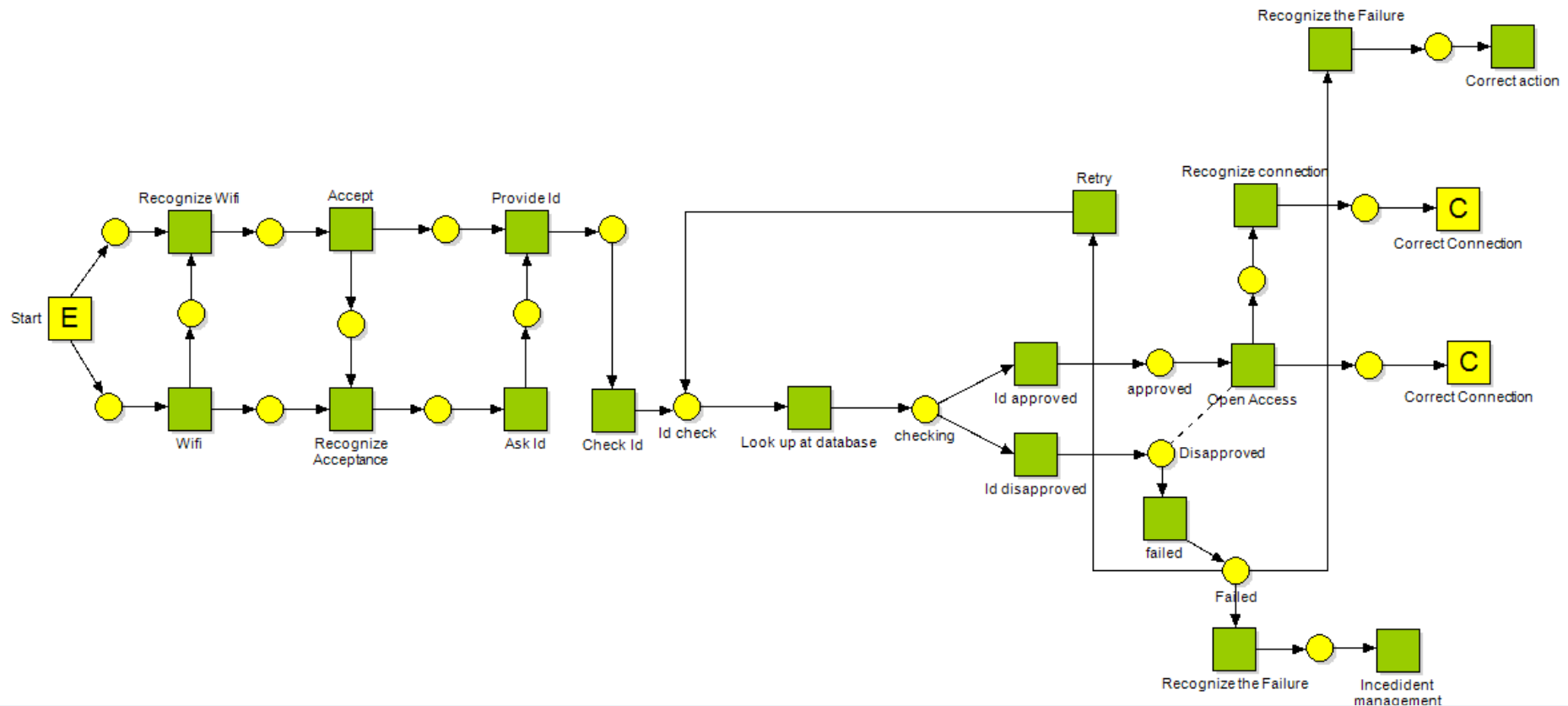
Petri net of EduRoam Case
(first step)

Describing Intentions, Motivations and Actions
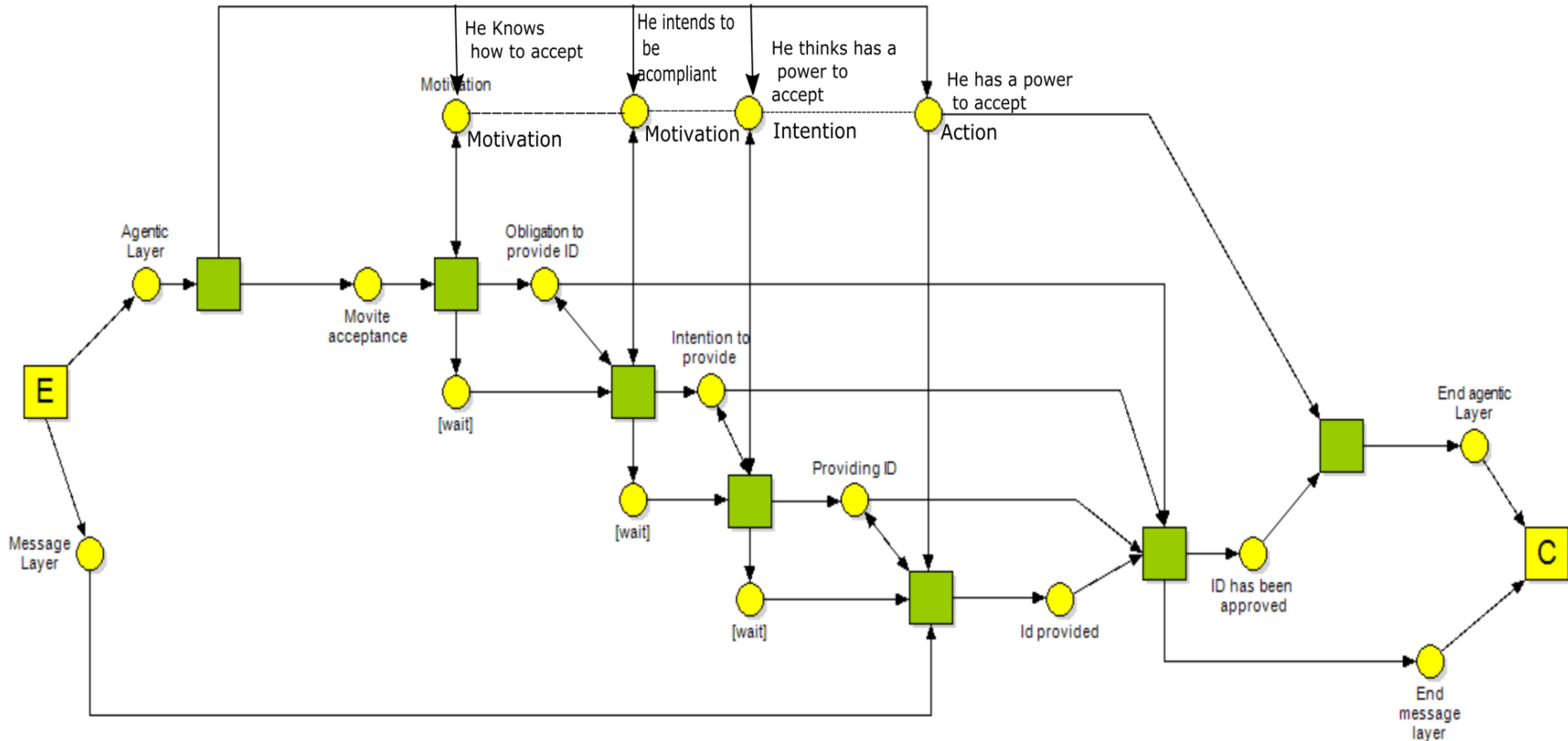
Petri net of EduRoam Case

Agent Model evaluating Trust

Trustworthiness

Observation

Expectation

Risk Assessment

Outcome

Impact

Trust Assessment & Risk taking

Power

Legal Ability → Controlability

Impact   Uncertainty   Vulnerability

Trustee → Ability
Responsiblity
Credibility
Predictibility

Risk Assesment → Trust → Outcome

Risk Taking   Expectation

Trustor Propensity
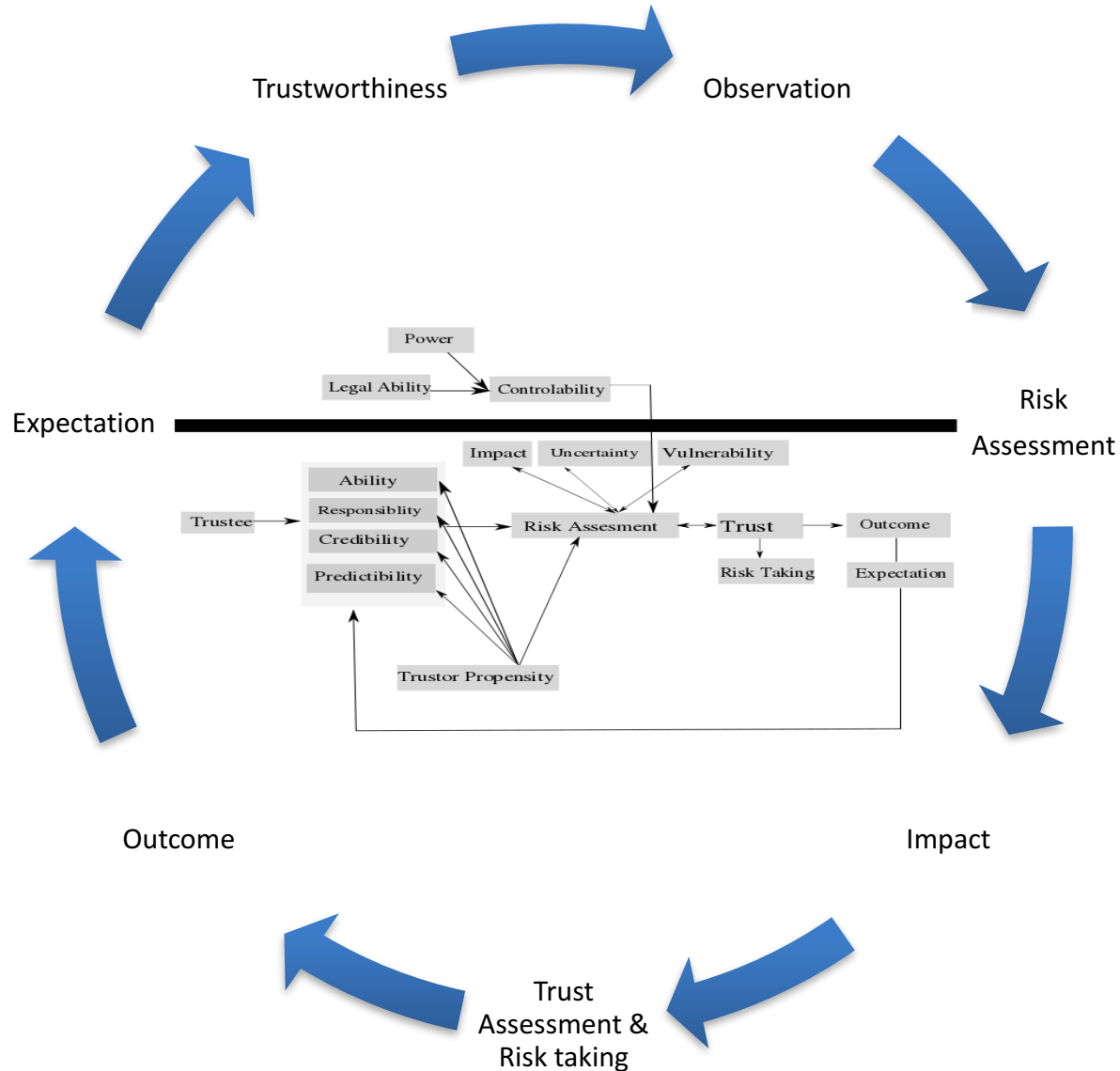
First step: Evolutionary Prisoners Dilemma using ABM Simulation
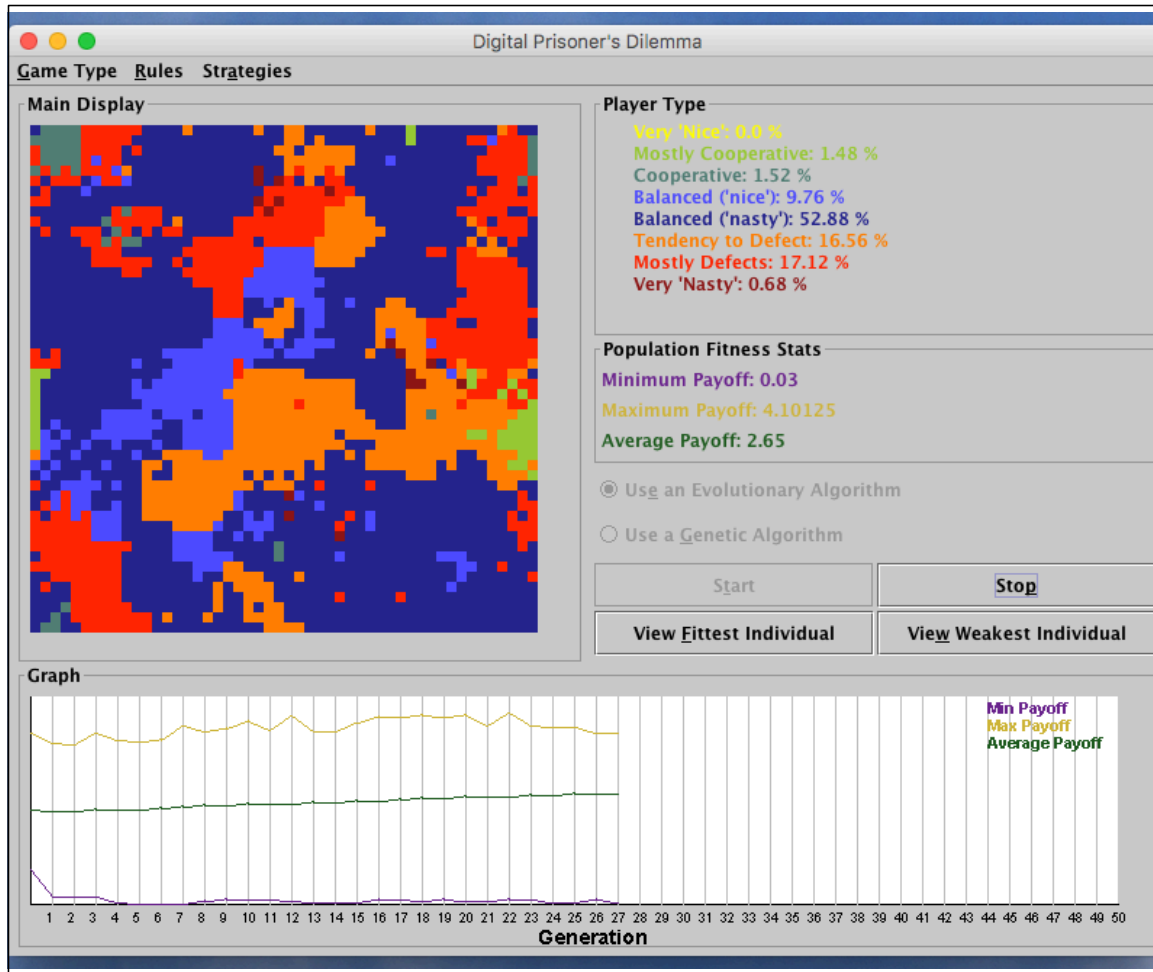
Agents choose from different strategies:

- Collaborate
- Defect
- During simulation: Agents predict next behavior of neighboring agents learned from observing past behavior.

Simulation observes tendency to maximize individual welfare instead of helping the group.

This type of simulation will be base to simulate more complex collaborations of autonomous organizations.

Research performed by Ameneh Deljoo, PhD candidate University of Amsterdam.

ABM Simulation

Ameneh Deljoo, Tom van Engers, Leon Gommans and Cees de Laat, "What is Going on: Utility-based Plan Selection in BDI Agents", KnowProS 2017 workshop, AAAI-17, San Francisco feb 5, 2017.

SARNET: Security Autonomous Response with programmable NETworks

Marc Lyonais, Leon Gommans, Rodney Wilson, Rob Meijer,
Frank Fransen Tom van Engers, Paola Grosso, Cees de Laat,
Ameneh Deljoo, Ralph Koning, Ben de Graaff, Gleb Polevoy, Stojan Travanovski.

UNIVERSITY OF AMSTERDAM

AIRFRANCE KLM

ciena   NWO Netherlands Organisation for Scientific Research   TNO   COMMIT/

Big Data: real time ICT for logistics
Data Logistics 4 Logistics Data (dl4ld)

Jan Burgmeijer, TNO, PI, Cees de Laat, UvA, Co-PI, Leon Gommans, KLM

simacan   TNO   AIR FRANCE KLM

Z   ORACLE   ciena   THALES

evofenedex   NWO Netherlands Organisation for Scientific Research   Gemeente Amsterdam   TRANSFIDES

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?

Strategic Level

Tactical **Level**

Operational Level

Big Data Sharing use cases placed in airline context

Global Scale

National Scale

City / regional Scale

Campus / Enterprise Scale

Aircraft Component Health Monitoring (Big) Data
NWO **CIMPLO project**
4.5 FTE

Cargo Logistics Data
(C1) DaL4LoD
(C2) Secure scalable policy-enforced distributed data Processing
(using blockchain)

NLIP iShare project

iSHARE
powered by NLIP

Cybersecurity Big Data
NWO COMMIT/
SARNET project
3.5 FTE

System and Network Engineering

AIR FRANCE KLM

SAE Use Case envisaged **research** collaboration

Funding Agency	NSF — Big Data Hub / Spoke or Industry initiative funding		NWO — Netherlands Organisation for Scientific Research — Topsector Funding
International Networking	INTERNET2	ESnet — ENERGY SCIENCES NETWORK	GÉANT
Regional / National Networking	CENIC — Connecting California	SoX — Southern Crossroads	LEARN — LONESTAR EDUCATION AND RESEARCH NETWORK	SURFNET
Local University	Stanford	Georgia Tech	UT DALLAS	UNIVERSITEIT VAN AMSTERDAM
Aircraft MRO, OEM & Operators	BOEING	DELTA AIR LINES	Bell Helicopter	AIR FRANCE KLM
Industry Standards Body	SAE INTERNATIONAL	SAE AeroSpace Group — HM-1 working group — Use Case on aircraft sensor Big Data		

SE System and Network Engineering

AIR FRANCE KLM

Data Processing models

- Bring data to computing
- Bring computing to data
- Bring computing and data to (un)trusted third party
- A mix of all of the above
- Block chain to record what happened
- Block chain for data integrity
- Bring the owner of Data in control!
- Data owner policy + PEP technology

The VM Turntable Demonstrator

Seattle

Toronto

Chicago

Starlight

Dynamic Lightpaths

Amsterdam

Netherlight

UvA

NYC

iGrid05
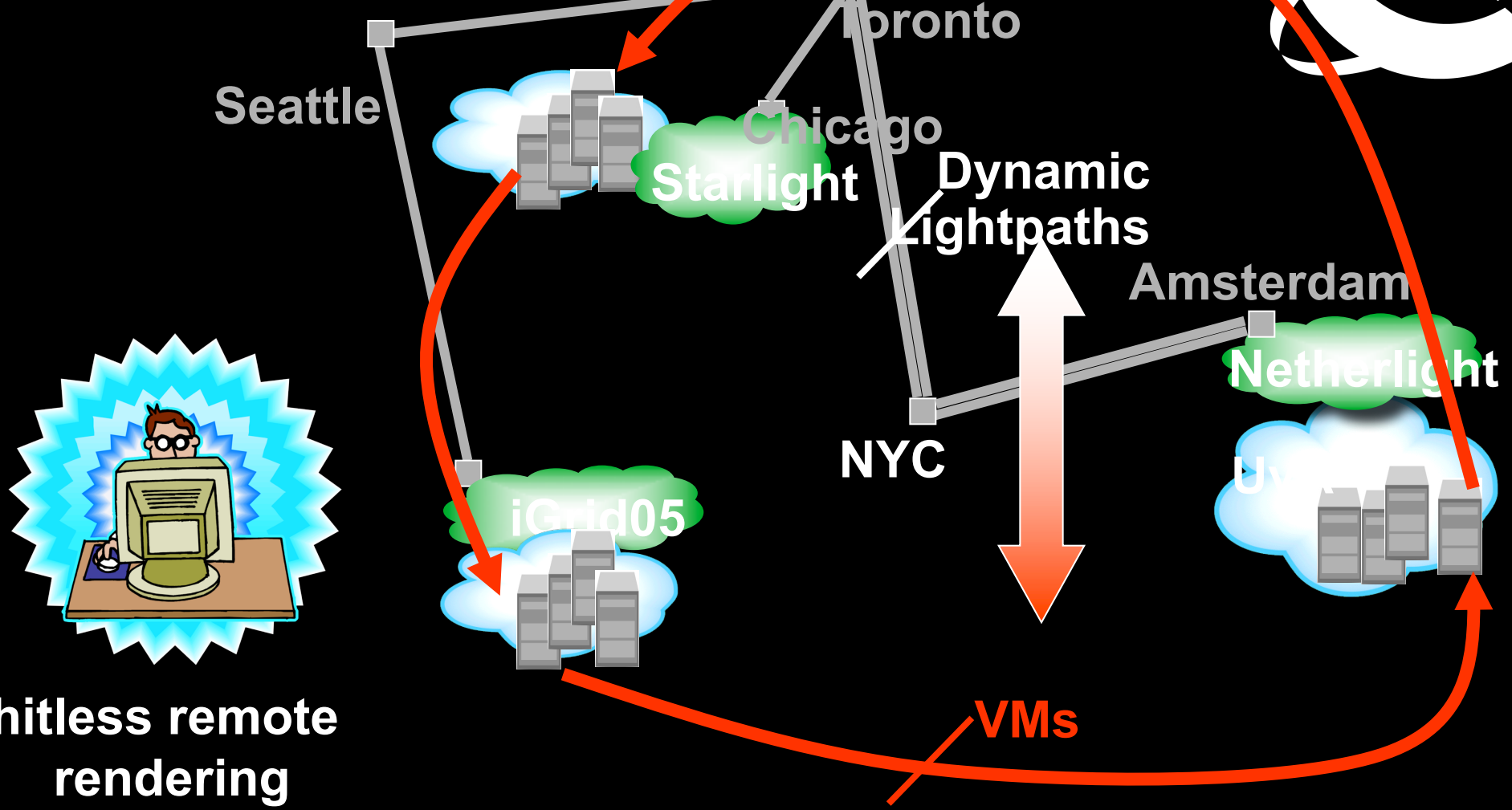
hitless remote rendering

VMs

The VMs that are live-migrated run an iterative search-refine-search workflow against data stored in different databases at the various locations. A user in San Diego gets hitless rendering of search progress as VMs spin around

Experiment outcomes
Note, this was in 2005 at SC and igrid2005!

We have demonstrated seamless, live migration of VMs over WAN

For this, we have realized a network service that

- Exhibits predictable behavior; tracks endpoints
- Flex bandwidth upon request by credited applications
- Doesn't require peak provisioning of network resources

Pipelining bounds the downtime in spite of high RTTs

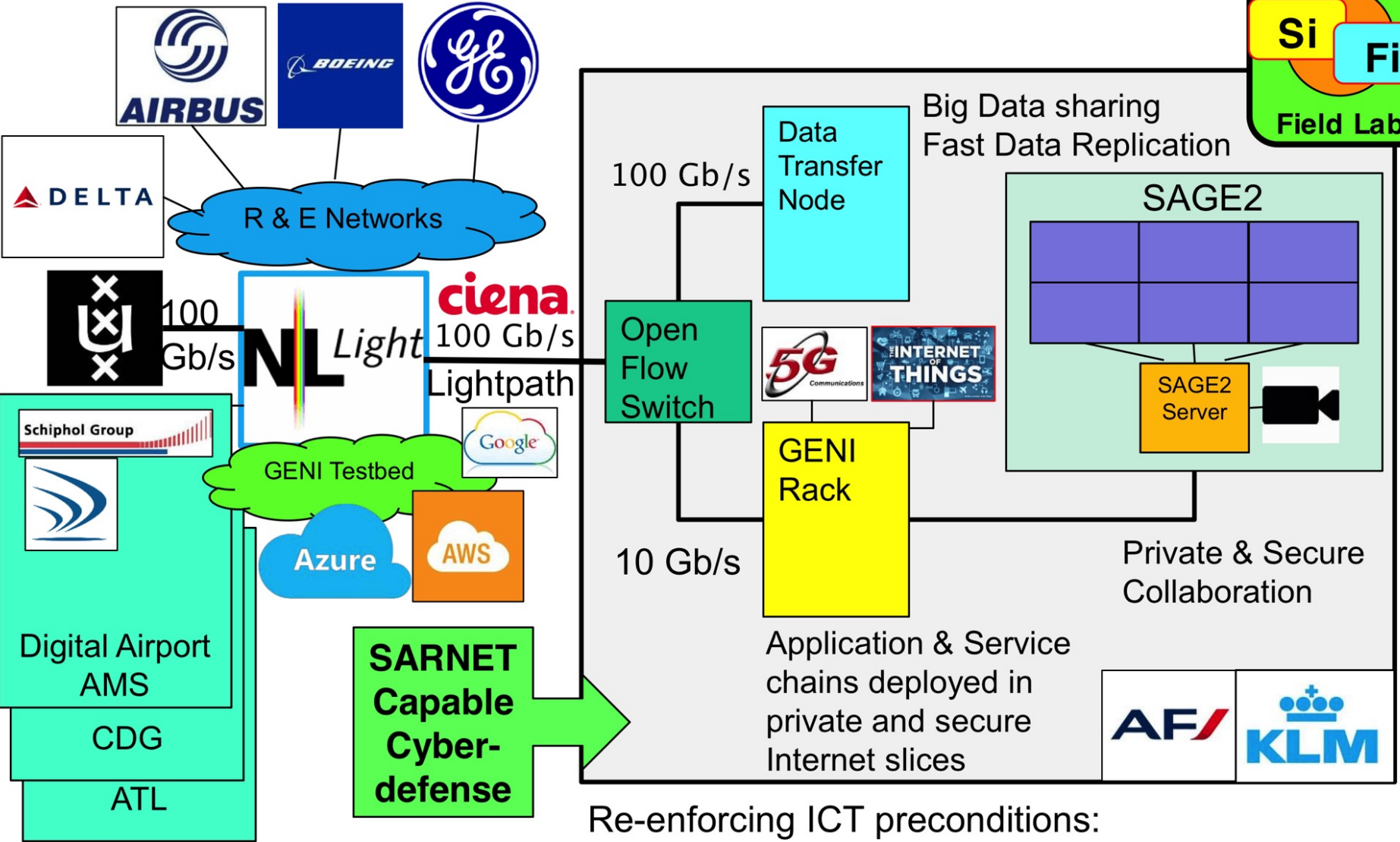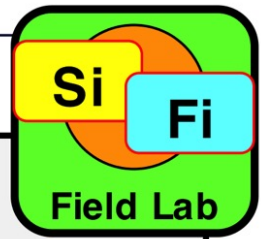- San Diego – Amsterdam, 1GE, RTT = 200 msec, downtime <= 1 sec
- Back to back, 1GE, RTT = 0.2-0.5 msec, downtime = ~0.2 sec*
- *Clark et al. NSDI 05 paper. Different workloads*

VM + Lightpaths across MAN/WAN are deemed a powerful and general alternative to RPC, GRAM approaches
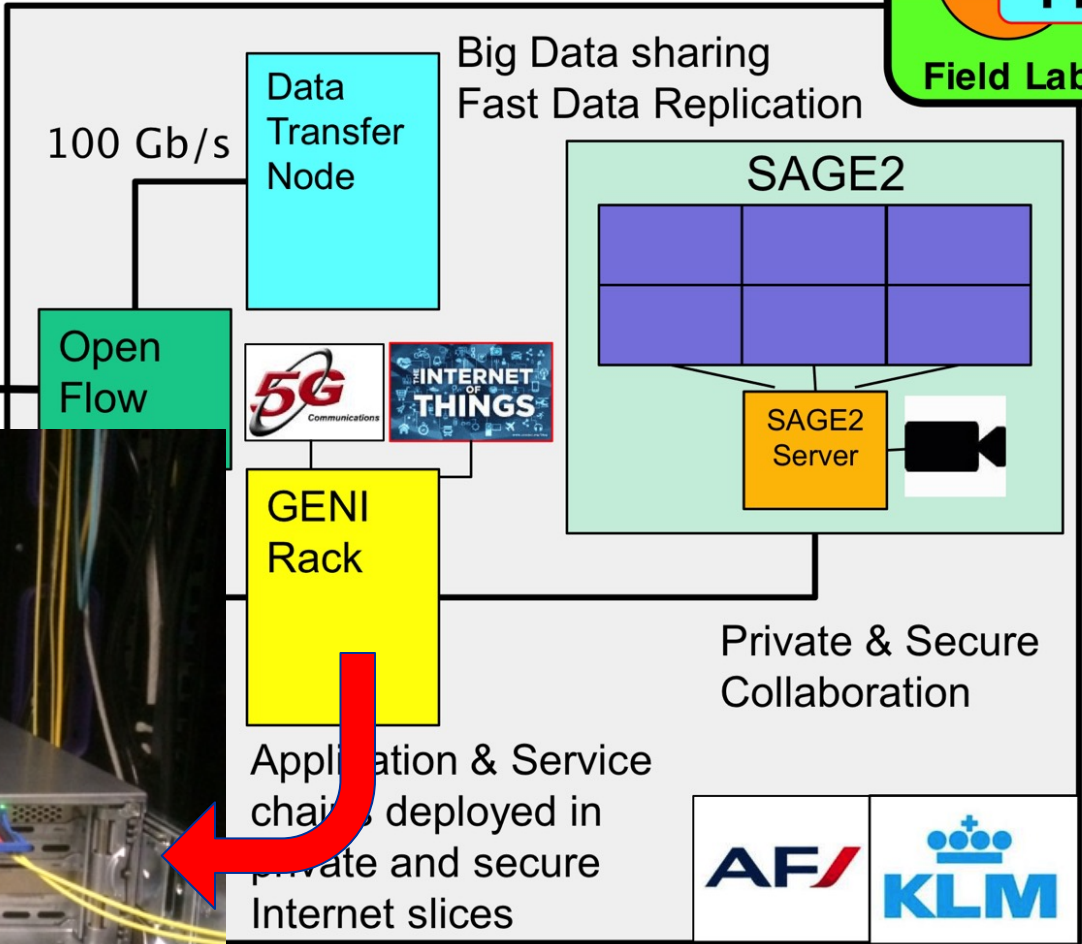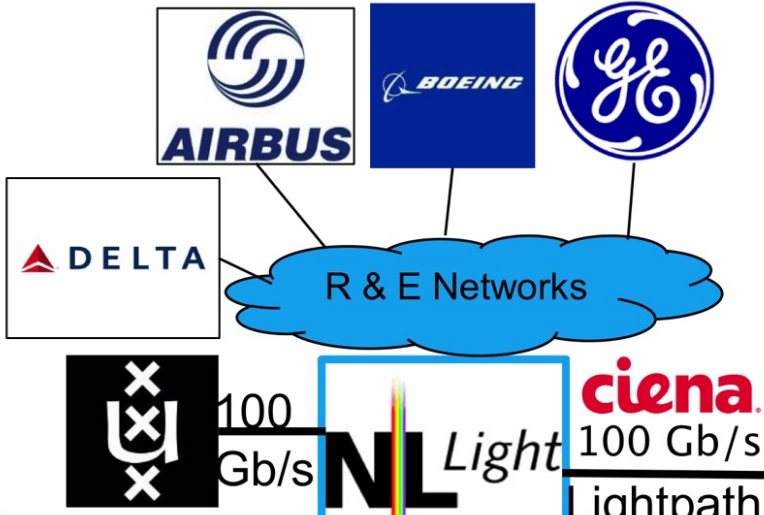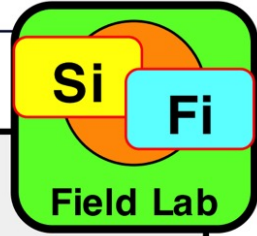
We believe it's a representative instance of active cpu+data+net orchestration

Ambition to put capabilities into fieldlab

Ambition to put capabilities into fieldlab

AIRBUS

BOEING

GE

DELTA

R & E Networks

100 Gb/s

ciena

NL Light

100 Gb/s
lightpath

100 Gb/s

Data Transfer Node

Big Data sharing
Fast Data Replication

Si Fi
Field Lab

SAGE2

Open Flow

5G Communications

THE INTERNET OF THINGS

SAGE2 Server

GENI Rack

Private & Secure Collaboration

Application & Service chains deployed in private and secure Internet slices

AF/ KLM

ing ICT preconditions:
saged site has similar elements

AIRFRANCE / KLM

AF/KLM FieldLab

Ambition to put [o]

AIRBUS
BOEING
GE

DELTA

100 G[...]

R & E Networks

ciena

100 Gb/s
NL Light
100 Gb/s
lightpath

Open
Flow

GENI Rack

Private & Secure Collaboration

Application & Service chains deployed in private and secure Internet slices
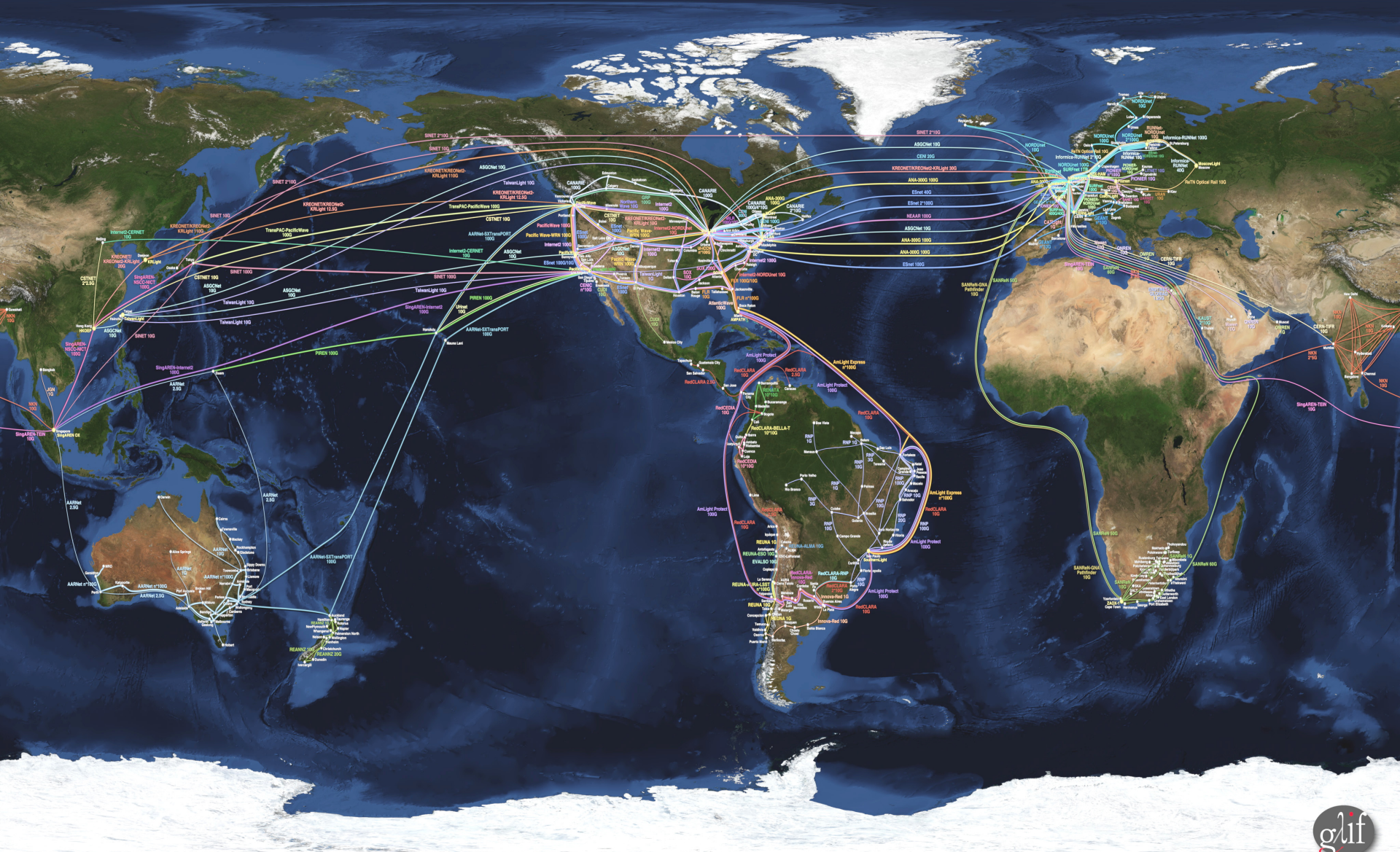
AF/ KLM
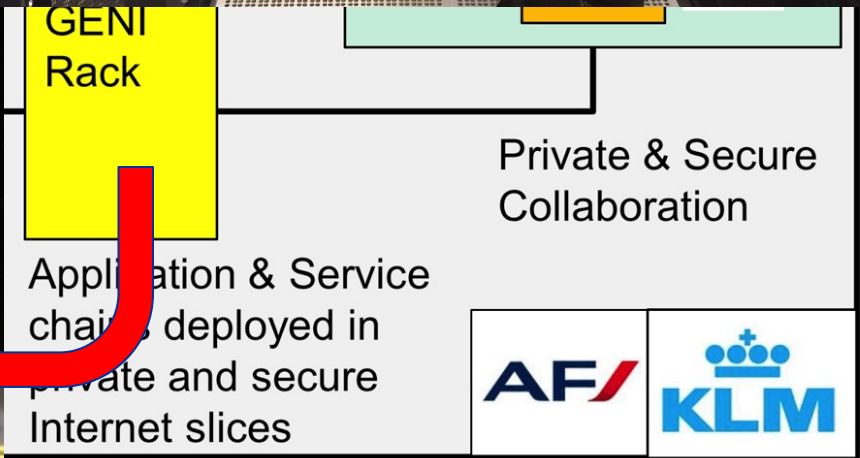
[...]ing ICT preconditions:
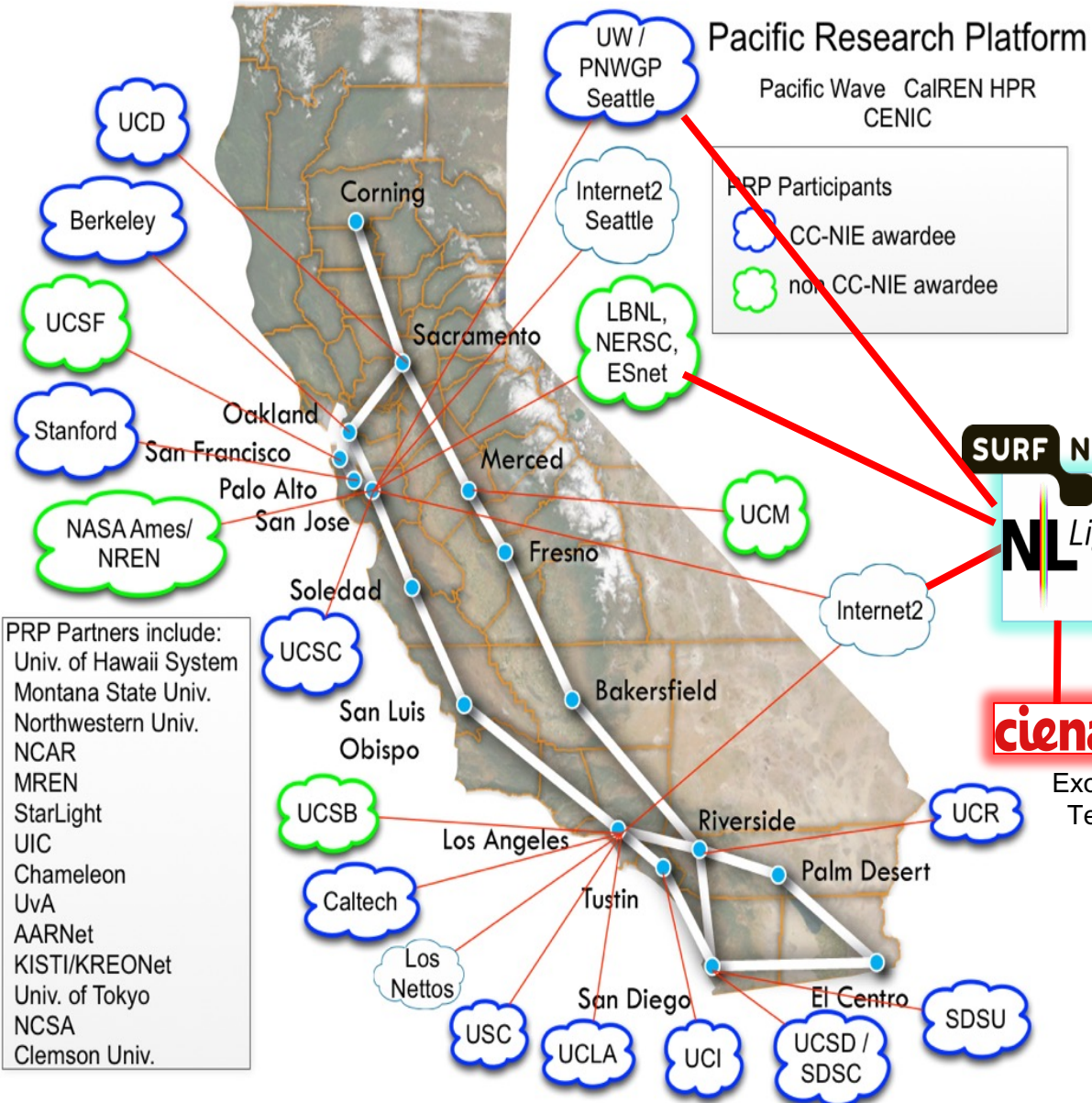[...]saged site has similar elements

AIRFRANCE / KLM

The GLIF – LightPaths around the World

F Dijkstra, J van der Ham, P Grosso, C de Laat, "A path finding implementation for multi-layer networks",
Future Generation Computer Systems 25 (2), 142-146.

Pacific Research Platform testbed involvement

Research goal:
Explore value of academic network research capabilities that enable innovative ways & models to share big data assets

ExoGENI Testbed

PRP Partners include:
Univ. of Hawaii System
Montana State Univ.
Northwestern Univ.
NCAR
MREN
StarLight
UIC
Chameleon
UvA
AARNet
KISTI/KREONet
Univ. of Tokyo
NCSA
Clemson Univ.

Note: this diagram represents a subset of sites and connections.

System and Network Engineering

AIR FRANCE KLM

Approach

- Strategic:
  - Translate legislation into machine readable policy
  - Define data use policy
  - Trust evaluation models & metrics
- Tactical:
  - Map app given rules & policy & data and resources
  - Bring computing and data to (un)trusted third party
  - Resilience
- Operational:
  - TPM & Encryption schemes to protect & sign
  - Policy evaluation & docker implementations
  - Use VM and SDI/SDN technology to enforce
  - Block chain to record what happened (after the fact!)

Strategic Level
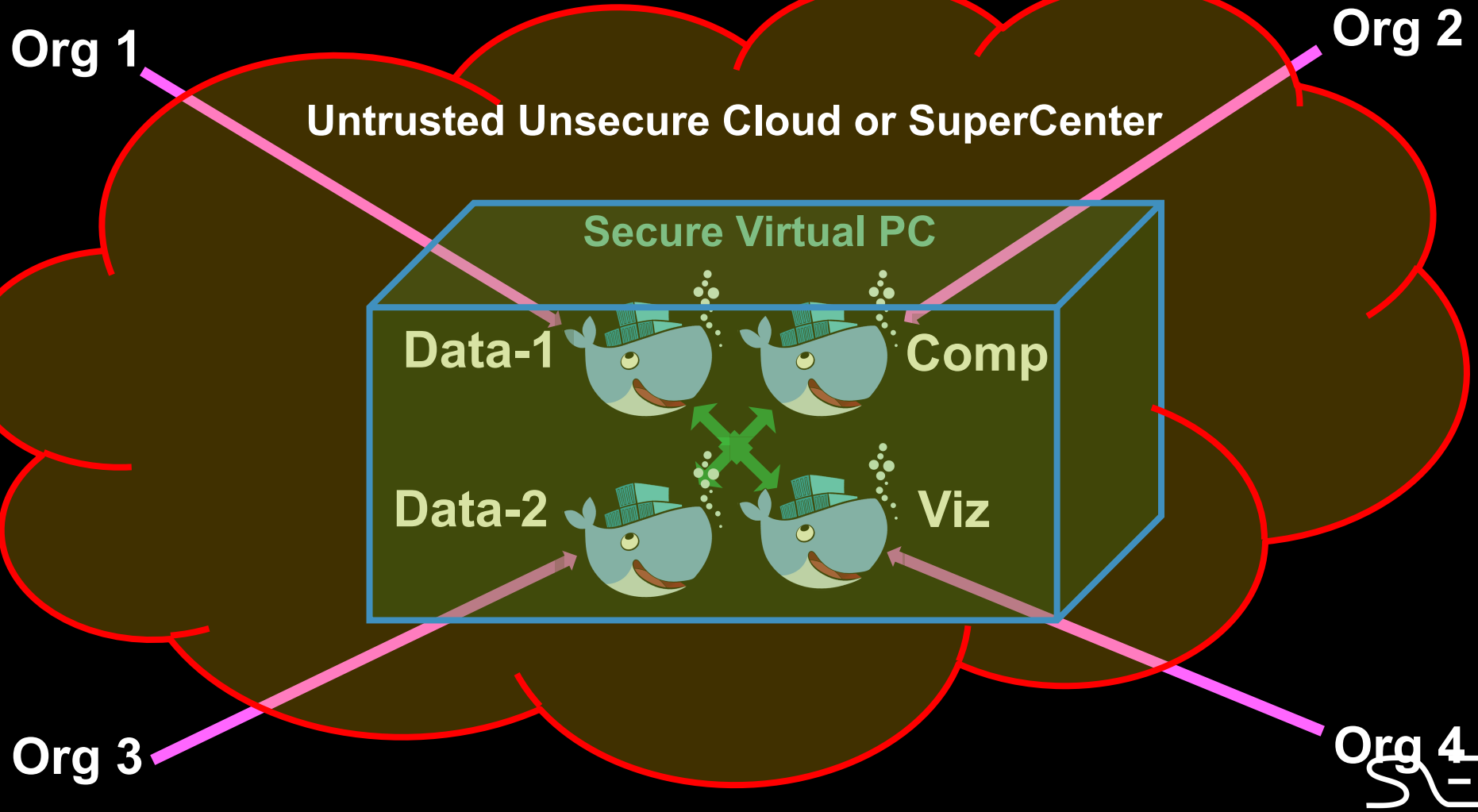
Tactical Level

Operational Level
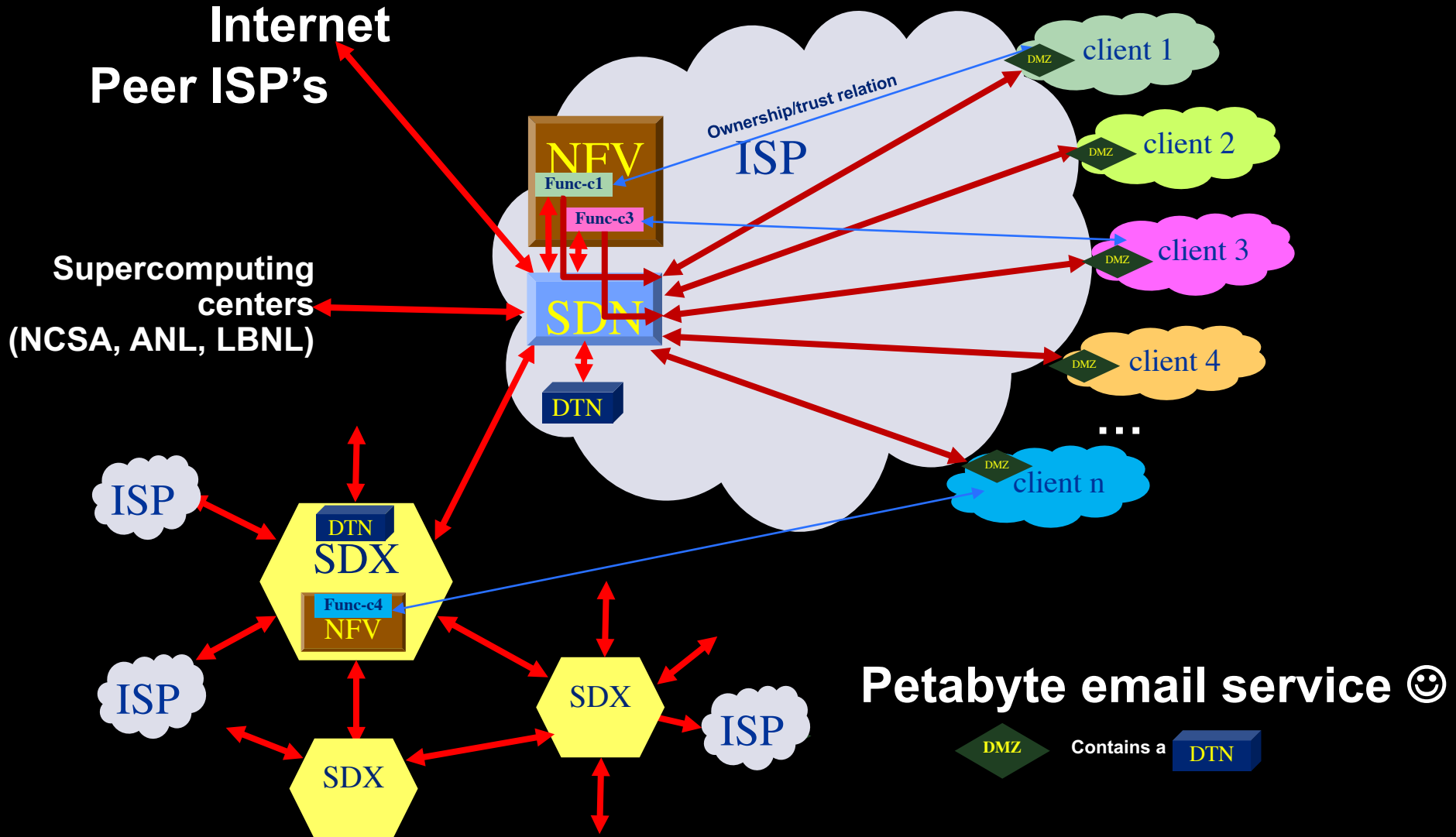
Secure Policy Enforced Data Processing
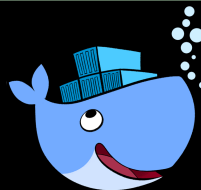
- Bringing data and processing software from competing organisations together for common goal
- Docker with encryption, policy engine, certs/keys, blockchain and secure networking
- Data Docker (virtual encryped hard drive)
- Compute Docker (protected application, signed algorithms)
- Visualization Docker (to visualize output)

Org 1

Org 2

Untrusted Unsecure Cloud or SuperCenter

Secure Virtual PC

Data-1

Comp

Data-2

Viz

Org 3

Org 4

Q&A

- More information:
  - http://delaat.net/sarnet
  - http://delaat.net/dl4ld

Networks of ScienceDMZ's & SDX's

Data Hub System Applicability

Industry

- Cross Cutting Field lab
- Innovation with SURF

Science

- European Open Science Cloud
- FAIR model
  - Findable – Accessible – Interpretable - Reusable

Society

- Smart Cities & Arena
- Streaming Data Decision Support

Validation Fieldlab and Dissemination

UVA - OpenLab

KLM
NetherLight
GENI
Fed4Fire
Cloud
SURFSARA
...

TNO - Intrepid

Smart Data
Factory
Innovations
Smart Rail
To-Grip
...

DataHub
DTN

DataHub
DTN

DataHub
DTN

DataHub
DTN

ams*ix*
amsterdam
internet exchange

NL *Light*

C2D – Big DataHubs

Arena
KAVE
AZURE
Use Cases
...

- Experimental facilities from day one!
- Proof of concepts demonstrating secure data sharing
- Blueprint, roadmap and standards where applicable
- Model for FAIR EOSC Infrastructure