Secure data sharing in the Responsible Internet.

Paola Grosso & Cees de Laat

University of Amsterdam

In this talk we present the architecture of cyber infrastructures that enable Secure Data Sharing.

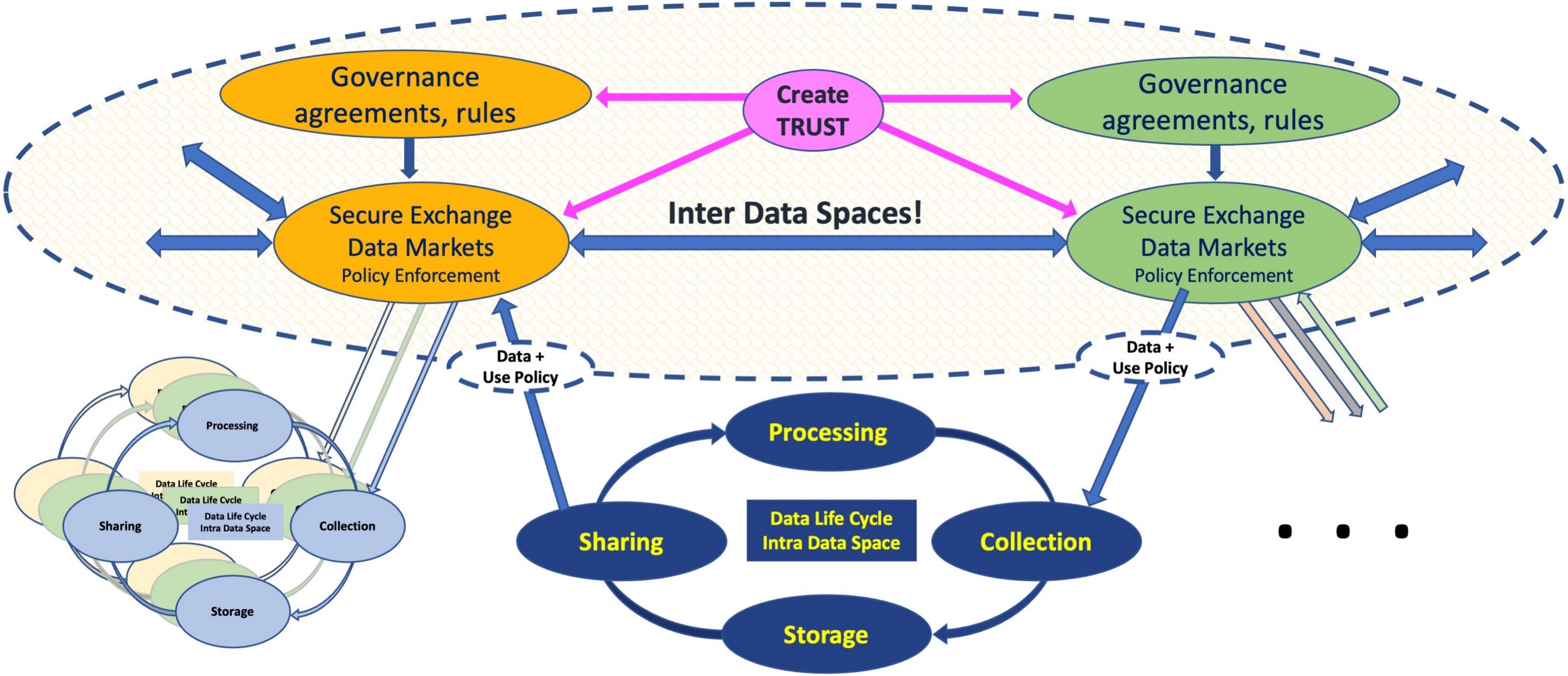
A core component of such architecture is the Responsible Internet, a programmable network that provides enhanced transparency and accountability to the end users.

Such transparent network is essential to be able to Enforce and Audit Policies in Digital Data Markets and Data Exchanges to to reduce risk of malicious data use and leakage.

The Roaring Twenties!

- In the 90's the Internet was running on top of the telco's
- We freed it in the 2000's with GLIF and the *Lights
- We developed the computer science for virtualization of CI
- Networking is (almost) not the problem anymore (DMC2022...)
- Data and algorithms & apps and services are now in the cloud
- Just a few large players emerge with an almost monopoly
- **Roaring 20's to free the Data with initiatives such as GRP!**

The Internet of Data



Application domain

Application domain

AMDEX

Data objects & methods

Data & Algorithms service

FAIR / USE

AMS-IX

Routers - Internet – ISP's - Cloud

IP packet service

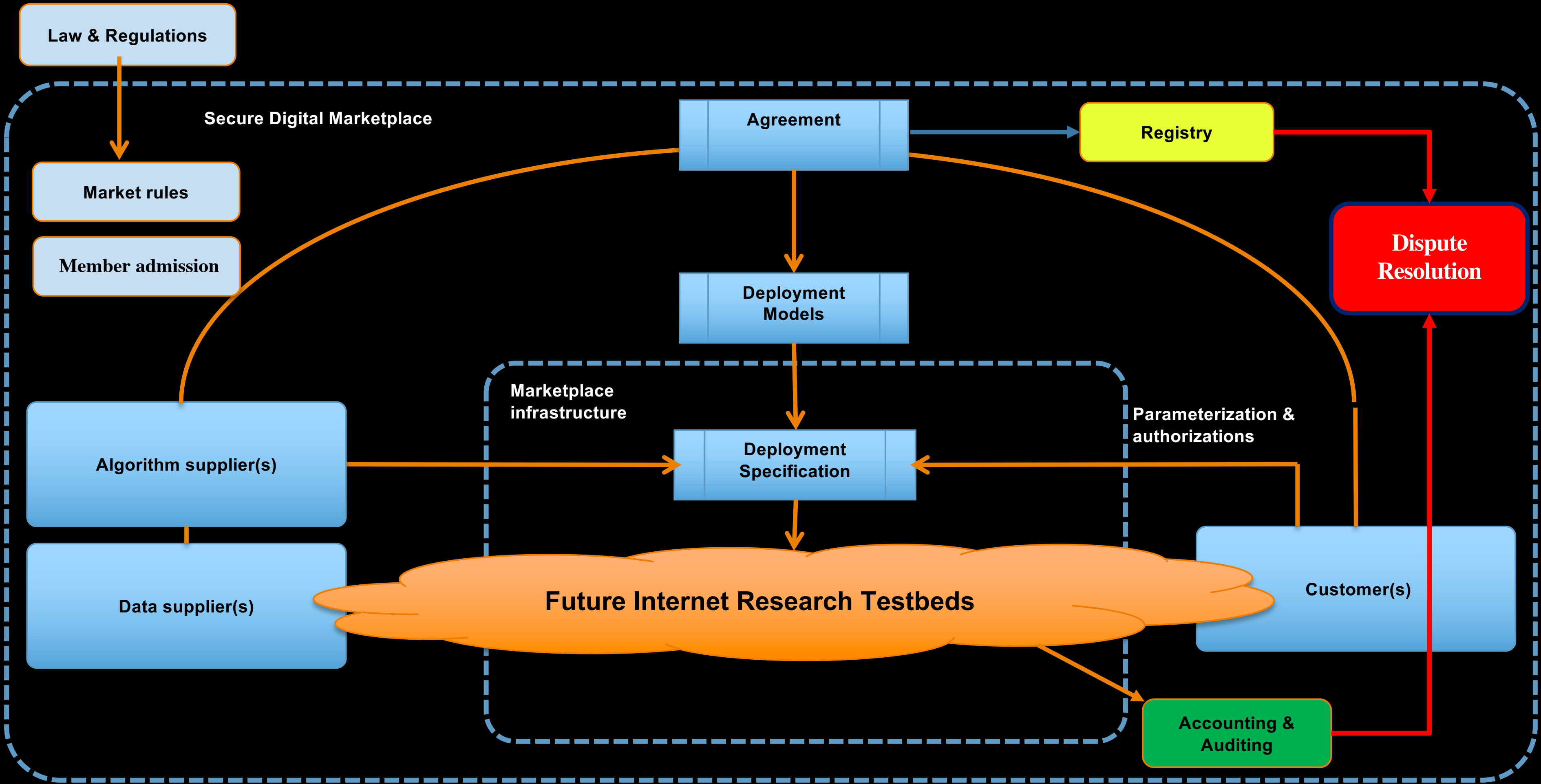
IP / BGP

Layer 2 exchange service

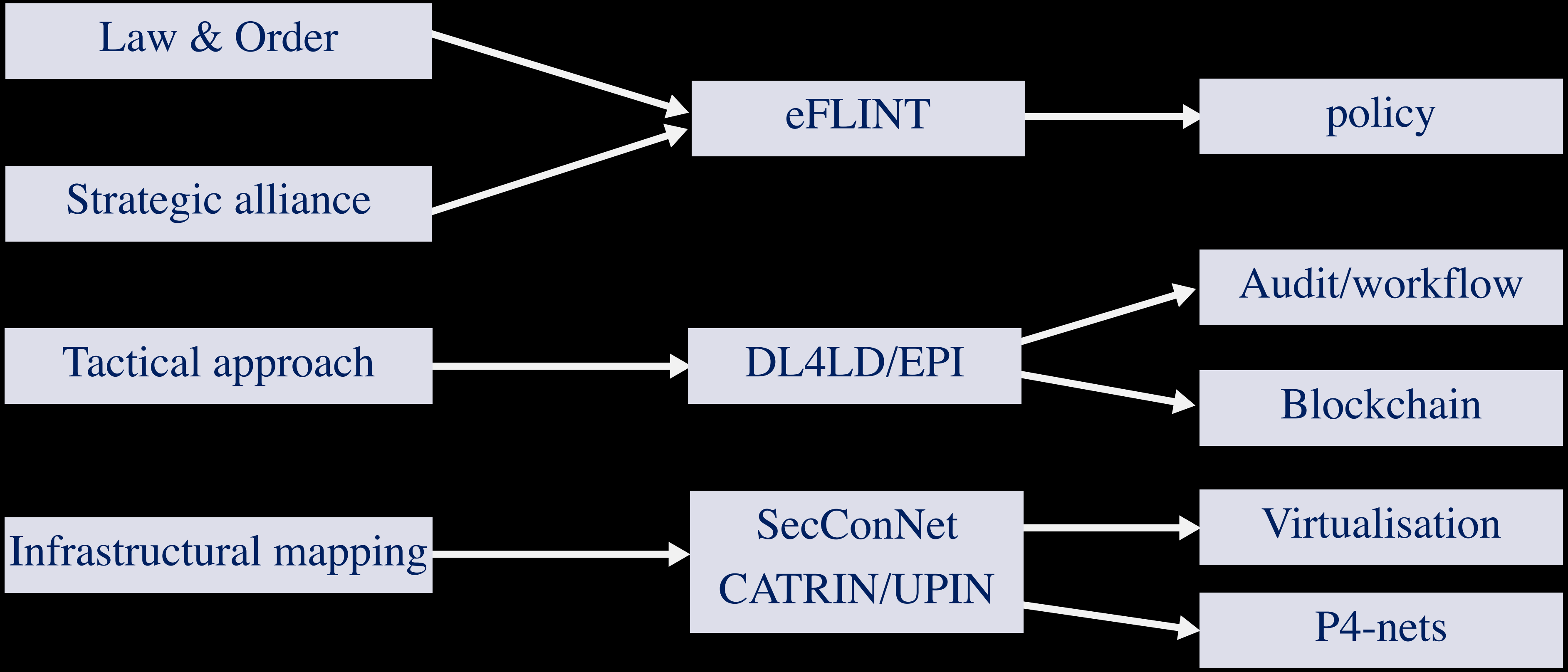
Ethernet frames

ETH / ST

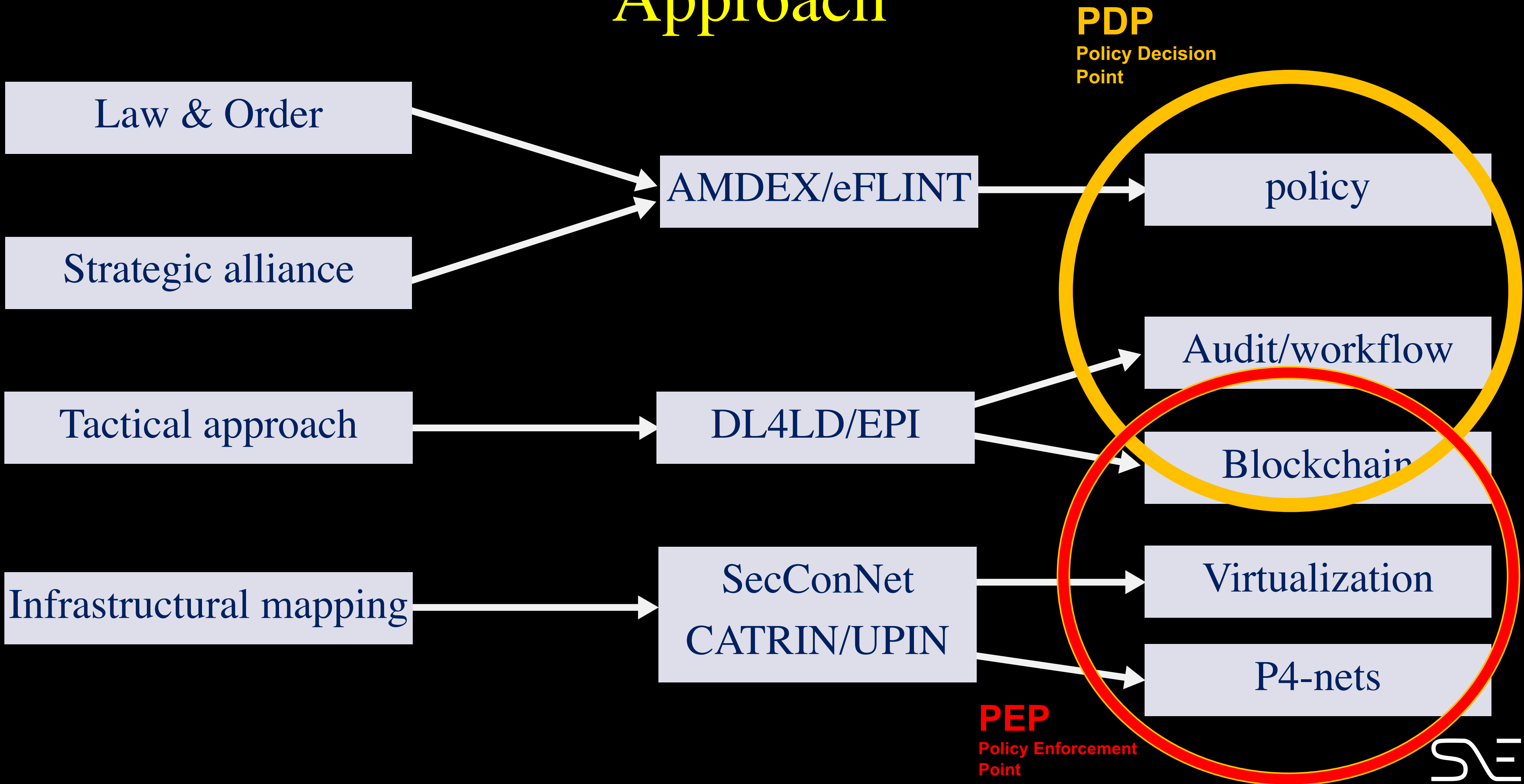
Secure Digital Market Place Research



Approach


Law & Order
Strategic alliance
Tactical approach
Infrastructural mapping
eFLINT
DL4LD/EPI
SecConNet
CATRIN/UPIN
policy
Audit/workflow
Blockchain
Virtualisation
P4-nets

Approach



PDP
Policy Decision Point
Law & Order
Strategic alliance
Tactical approach
Infrastructural mapping
AMDEX/eFLINT
DL4LD/EPI
SecConNet
CATRIN/UPIN
policy
Audit/workflow
Blockchain
Virtualization
P4-nets
PEP
Policy Enforcement Point

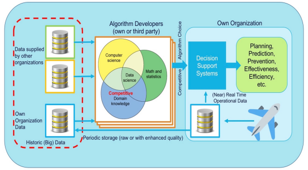
SC2018

https://delaat.net/sc

Training AI/ML models using Digital Data Marketplaces

Creating value and competition by enabling access to additional big data owned by multiple organizations in a trusted, fair and economic way

The more data - the better: an aircraft maintenance use-case



- AI/ML algorithm based Decision Support Systems create business value by supporting real-time complex decision taking such as **predicting the need for aircraft maintenance**.
- Algorithm quality increases with the availability of aircraft data.
- Multiple airlines operate the same type of aircraft.
- **Research Question:** *"How can AI/ML algorithm developers be enabled to access additional data from multiple airlines?"*
- **Approach:** *Applying Digital Data Marketplace concepts to facilitate trusted big data sharing for a particular purpose.*

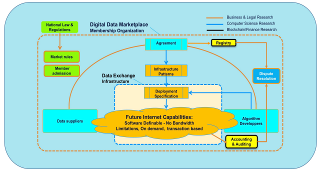
Digital Data Marketplace enabling data sharing and competition

A **Digital Data Marketplace** is a membership organization supporting a common goal: e.g. *enable data sharing to increase value and competitiveness of AI/ML algorithms.*

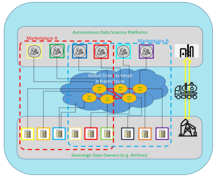
Membership organization is institutionalized to create, implement and enforce membership rules organizing **trust**.

Market members arrange **digital agreements** to exchange data for a **particular purpose** under specific conditions.

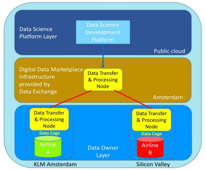
Agreements subsequently drive data science transactions creating processing infrastructures using infrastructure patterns offered by a Data Exchange as **Exchange Patterns**.



Researching Exchange Patterns to support Digital Data Marketplaces



Data Exchange Model



Research Infrastructure

Trust Modelling: What is the optimal infrastructure archetype, describing storage and processing locations and their relationships, which best suit member requirements when considering risk?

Processing Models: What are the implications of distributing data processing across membership organization owned infrastructures in terms of achievable model accuracy and processing performance using federated/distributed models vs centralized models

Marketplace Reference Architecture: What constitutes a marketplace? Researching needed functions, personas, flows, credentials, contracts & rules, conflict resolution, and much more ...

Research Elements

NWO · Commit2Data · TNO · ciena · DELL EMC · EQUINIX · amsterdam economic board · AMdEX · SAE ITC · Gemeente Amsterdam · SURF · CIMPL · ReMAP · AIR FRANCE KLM

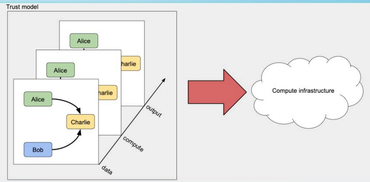
Leon Gommans, Anne Savelkoul, Wouter Kalfsbeek, David Langeveld, Erik IJzermans, Floris Freeman, Brend Dikkers, Cees de Laat, Tom van Engers, Wouter Los, Paola Grosso, Joseph Hill, Reggie Cushing, Giovanni Sileno, Lu Zhang, Ameneh Deljoo, Thomas Baeck, Willem Koeman, Laurie Strom, Axel Berg, Gerben van Malenstein, Kaladhar Voruganti, Rodney Wilson, Patricia Florissi

Dataharbours: computing archetypes for digital marketplaces

Reginald Cushing, Lu Zhang, Paola Grosso, Tim van Zalingen, Joseph Hill, Leon Gommans, Cees de Laat, Vijaay Doraiswamy, Purvish Purohit, Kaladhar Voruganti, Craig Waldrop, Rodney Wilson, Marc Lyonnais

The problem

How can competing parties share compute and data? The architecture of a digital marketplace is an active research field and has many components to it. Here we investigate a federated computing platform which is molded into different **archetypes** based on **trust** relationships between organizations.



The components

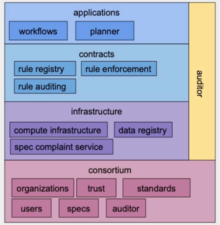
Consortium: is an initial document which brings together organizations that wish to collaborate. It defines static information such as keys to identify parties.

Infrastructure: A single domain organization infrastructure that securely hosts data, compute containers and, optionally, compute infrastructure. We dub this infrastructure a **data harbour**. A harbour implements a set of protocols that allows it to interact with other harbours.

Contracts: Are a set of rules that are shared amongst participating harbours which describe how objects (data, compute) can be traded between harbours and who can process data. In its simplest form is a 7-tuple which binds a user, data object, compute container, contract, consortium, harbour, and expiry date.

An application: Is a distributed pipeline which can make use of several contracts. The combination of application and contract defines the archetype of the computation i.e. how data and compute are moved to effect computation.

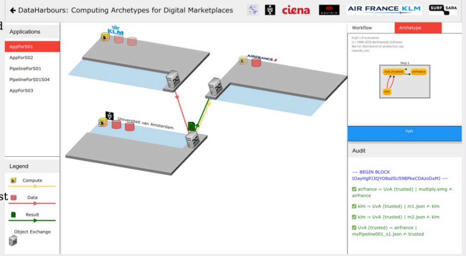
Auditor: A trusted entity that collects audit trails for use in litigation of policy violations.



In action

Federated computing on 3 distributed data harbours. Here we illustrate one archetype where KLM and Airfrance do not trust each other and employ a trusted 3rd party to send the data and compute for processing.

For the scenario to succeed the different harbours need to effect several transactions which are governed by contractual rules.

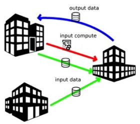


The transaction **protocol** involves first identifying both parties are who they say they are through pub/priv key challenges and secondly, that at least a **contract** rule is matched to allow the transaction. Important steps of the transactions are **audit** logged i.e. signed and published to and audit log collector.

ORACLE · TRANSFIDES · AIR FRANCE KLM · THALES · Gemeente Amsterdam · EQUINIX · TNO · SURF SARA · BiZdesign · evofenedex · TKI DINALOG · ciena · Commit2Data · NWO

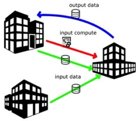
A secure network overlay for tracking and enforcement of data transaction rules.

Ralph Koning, Reginald Cushing Lu Zhang, Cees de Laat, Paola Grosso, University of Amsterdam



Competing companies can, together, generate value from collaborating on data and compute. Examples include airlines industry, ports, healthcare.

Clearly this poses a challenge of how to facilitate such collaborations through technology. Here we look at one piece of the puzzle i.e. setting up distributed multi-domain infrastructures between such parties to facilitate the running of applications.

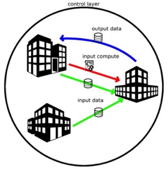


Motivation

- Multi-domain distributed applications need to share data and compute under different policies.

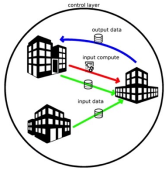
Challenges

- Map data sharing policies to infrastructure.
- Build an infrastructure that facilitates these applications.
- Control sharing of data and compute.
- Audit activity of the network.
- Minimize risk of policy/security breaches.



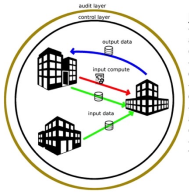
Overlay

- Nodes on the network are addressed using their public key.
- Nodes include: domain controllers, data buckets, auditors, application planners, users.
- Keys create chains of trust and verification through cryptographic signature trails.
- Applications are decomposed to a set of transactions.
- Transactions drive the overlay.



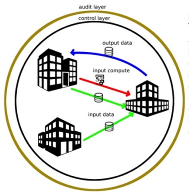
Control functions

- Securing bucket-to-bucket communication through transaction specific VPNs.
- Bucket node key address used as VPN keys.
- Opening connection endpoints on audit signatures.
- Network interfaces created on demand. Bucket containers have no network interface. Interfaces are only created and attached per signed transaction.



Network of Auditors

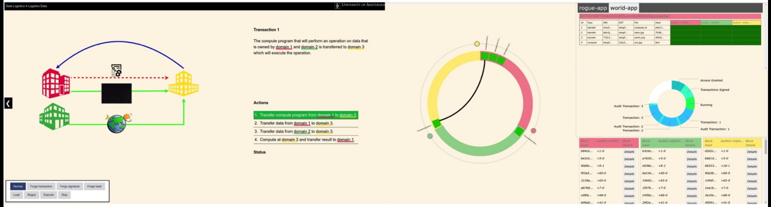
- Auditor nodes on the network provide a signing and verification layer that is checked by the control layer.
- Auditors sign network actions based on their internal policy.
- Auditors are independent of each other.
- The more number of signatures an action gets (e.g. transaction) the more confident the control layer is.
- Auditors cross-verify each other's logs to minimize log tempering.



In short..

- Overlay allows for a distributed infrastructure.
- Key-based addressing allows for node signature trails and trust chains.
- Network of auditors provide *rubber-stamping* of actions/transactions
- Control layer enforces security using inputs from auditors and minimizes attack vectors on data transfers.

Proof of Concept, see https://dl4ld.nl/



This research is funded by the Dutch Science Foundation in the Commit2Data program (grant no: 628.001.001) and by Equinix.

SC2019

https://delaat.net/sc

ICT-OPEN 2020-2021

Agent-Oriented Programming for Modern Cyber-Infrastructures

Mostafa Mohajeri Parizi, Giovanni Sileno and Tom van Engers. UvA, Complex Cyber Infrastructures (CCI) group

Introduction

- Importance of data in all domains of human activity has brought the requirement for more complex data-sharing Cyber-Infrastructures.
- These Infrastructures exhibit the double status of *computational* and *social* systems and regulating them requires higher level reasoning.
- Agent Oriented Programming (AOP) is extensively studied and used for modeling and simulation of social systems.
- The **A**gent**S**cript Cross-**C**ompiler (ASC) is built to bridge the modelling power of AOP with operational requirements of modern Complex Cyber-Infrastructures

Summary

This work introduces **A**gent**S**cript Cross-**C**ompiler (ASC):

- Provides a high level DSL agent programming language
- A Cross-Compiler to translate the Agent DSL into executable code.
- Allows use of modern development tools such as Testing, Debugging and Profiling.
- Enables seamless deployment into modern infrastructures with minimum runtime dependencies and transport-layer agnostic communication.

AgentScript's Compile, Build and Deploy Process

Agent's Script

High Level Logic-Based DSL motivated by *AgentSpeak(L)*

Off-the-shelf development tools
- Test tools: JUnit, ScalaTest, etc.
- Debuggers
- Profilers: Flight Recorder, etc.

- Intuitive modelling of social agents
- *Readable* and Verifiable DSL
- Enables logical reasoning

Scala/Java Presentation of the Agent's Script

- Off-the-shelf build tools:
  - sbt, maven
- Standard CI/CD operations

- Stand-alone Application
- Only Requires a JVM to execute
- Can use containerization tools for seamless deployment: docker, k8s

JVM-Based ByteCode Presentation of the Agent's Script

Verifiable via ByteCode verification tools: JPF

- Agent's communications are Transport layer agnostic
- Enables plugins for interoperability
  - rest, amqp, kafka, etc.

Acknowledgments

This work results from work done within Data Logistics for Logistics Data project (DL4LD, www.dl4ld.net). The DL4LD is funded by the Dutch Science Foundation in the Commit2Data program (grant no: 628.001.001).

Contact info:
1: m.mohajeriparizi@uva.nl
2: g.sileno@uva.nl
3: tom.vanengers@tno.com

Digital Enforceable Contracts (DEC): Making Smart Contracts Smarter

Lu-Chi Liu, Giovanni Sileno, Tom van Engers

Complex Cyber Infrastructure Group, Informatics Institute, University of Amsterdam

Background

- Current smart contracts have limited capabilities of normative representations, making them distant from actual contracts.
- Normative contents (duty and power) can be modeled into logic-based representation.
- DEC provides a general architecture where various enforcement mechanisms are enabled by normative reasoning. For example, to check whether an action will lead to a duty.

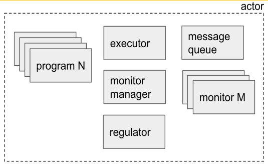
```
// written in eFLINT
Act request to modify consent
  Actor subject
  Recipient controller
  Related to consent, other purpose
  Conditioned by
    consent && consent.purpose != other purpose
  Creates duty to modify consent()

 Duty duty to modify consent
  Holder controller
  Claimant subject
  Related to consent, other purpose
```

Norms related to GDPR

Actor-based Modular Architecture

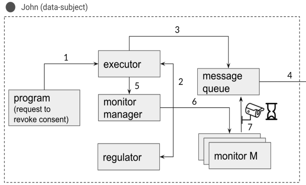
The architectural model is composed of a selected set of modules providing the functionality to run enforcement constructs.

Actor (the minimal unity of agency):
Program - plan to achieve a given design goal
Executor - internal control of the actor
Message queue - communication channel
Monitor - listeners that hook to events or facts
Monitor manager - handle monitors
Regulator - normative reasoning

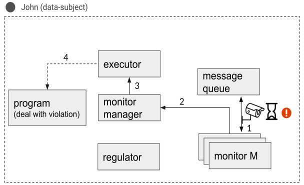


Prototype being developed using Akka-typed actor-oriented programming framework

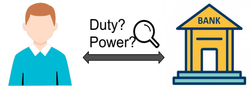
Example: A Data-sharing Scenario with GDPR



1) John (data-subject) attempts to revoke his consent of using his data from Bank (data-controller).
2) The executor sends query to the regulator to check related permissions and duties. (According to GDPR, Bank, as data-controller, has the duty to fulfill this request.)
3) The executor sends this request to the queue.
4) The request is then sent to Bank.
5) The executor asks monitor manager to create a monitor to check for violation.
6) A monitor is created.
7) The monitor checks messages from Bank with a timeout mechanism.



1) When the duty is due and not fulfilled, the monitor will be aware of this violation.
2) The monitor reports the violation.
3) Monitor manager notifies the executor of the violation.
4) The executor takes actions to deal with the violation.



Acknowledgements
This research is funded by the Dutch Organization for Scientific Research (NWO) under contracts 628.009.014 (SSPDDP project) and 628.001.001 (DL4LD project).

References
[1] Lu-Chi Liu, Giovanni Sileno, and Tom van Engers. Digital enforceable contracts: making smart contracts smarter. JURIX. 2020.
[2] L. Thomas van Binsbergen, Lu-Chi Liu, Robert van Doesburg, and Tom van Engers. eFLINT: a domain-specific language for executable norm specifications. In Proceedings of GPCE '20. ACM. 2020.

Policy Enforcement for Secure and Trustworthy Data Sharing in Multi-domain Infrastructures

Xin Zhou
University of Amsterdam
Amsterdam, 1098XH
x.zhou@uva.nl

Reginald Cushing
University of Amsterdam
Amsterdam, 1098XH
r.s.cushing@uva.nl

Adam Belloum
University of Amsterdam
Amsterdam, 1098XH
a.s.z.belloum@uva.nl

Tom van Engers
University of Amsterdam
Amsterdam, 1098XH
T.M.vanEngers@uva.nl

Sander Klous
University of Amsterdam
Amsterdam, 1098XH
klous.sander@kpmg.nl

Cees de Laat
University of Amsterdam
Amsterdam, 1098XH
delaat@uva.nl

January 31, 2021

1 Abstract

The push for data sharing and data processing across organisational boundaries creates challenges at many levels of the software stack. Data sharing and processing rely on the participating parties agreeing on the permitted operations and expressing them into actionable contracts and policies. Converting these contracts and policies into an operational infrastructure is still a matter of research. In this paper, we investigate the architecture of a multi-domain distributed architecture for policy driven application. The architecture spans components from auditing policies to managing network connections.

The architecture is based on an auditable secure network overlays[3] proposed by Cushing et al. in 2020, the overlays have already introduced an audit layer and a control layer. The audit layer aims at checking if a certain data process is compliant, only those compliant ones can collect signatures, and forwarded to the control layer for further processing, such a mechanism ensures that all operations are audited before execution. This process is shown as fig 1: [1]

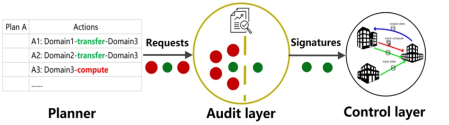
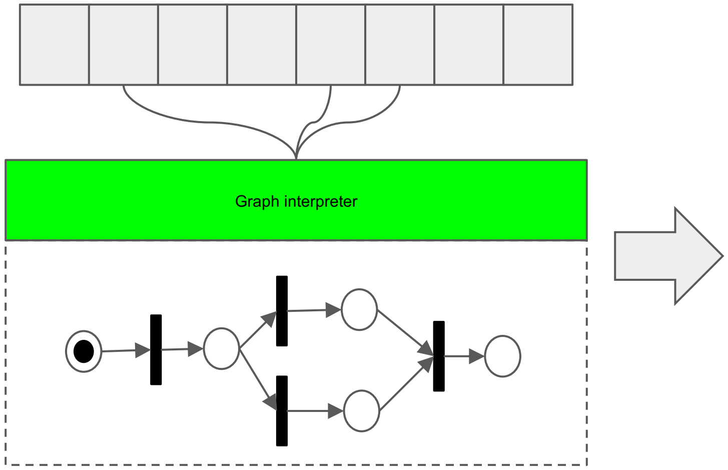


Figure 1: Auditable network overlays: the audit layer aims at checking the requests sent by a planner, only those compliant requests can receive signatures, and then being further executed in the control layer

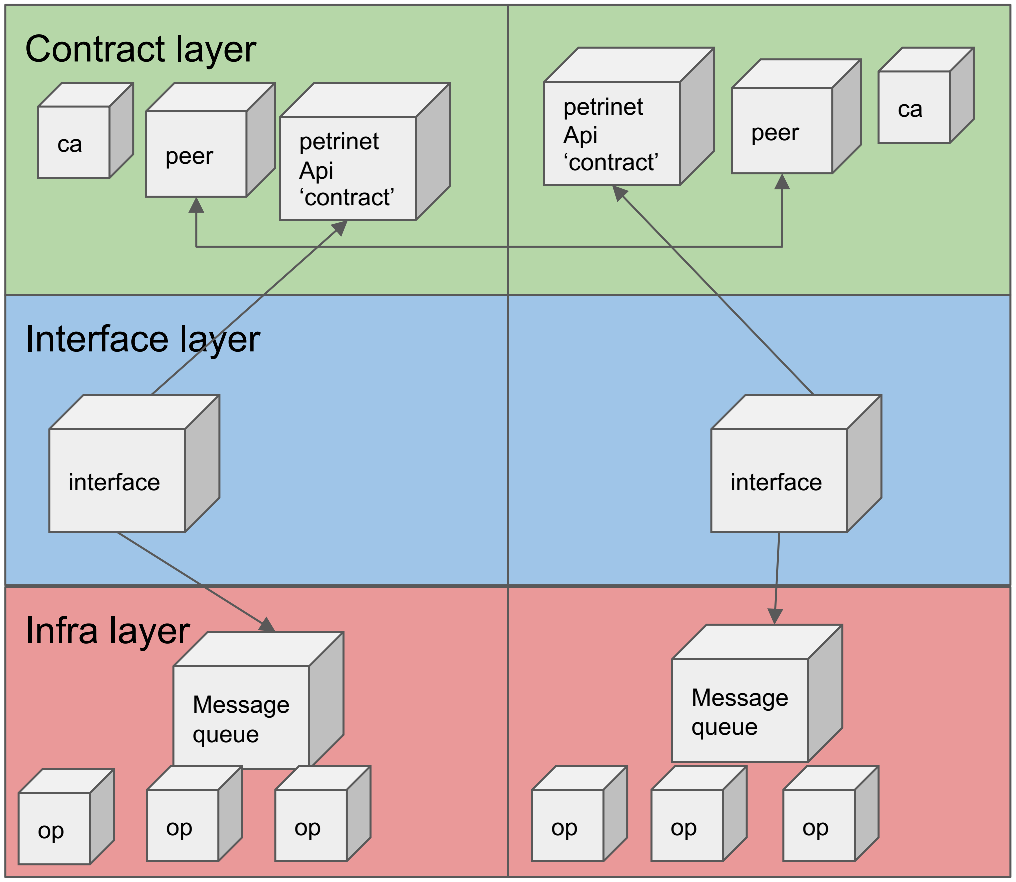
To enforce the policies by the audit overlay, the unstructured or semi-structured policies expressed in natural language need to be structured and formalized first, before it can be used as input to the audit overlay and combined with the environment conditions (such as region, risk level, etc.) that clarify which policies should be applied. Fig 2 presents the conceptual view of the policy which contain authorisations, obligations, and environmental conditions [4, 2].

[1]This research is funded by the Dutch Science Foundation in Commit2Data program (grant no: 628.001.001).

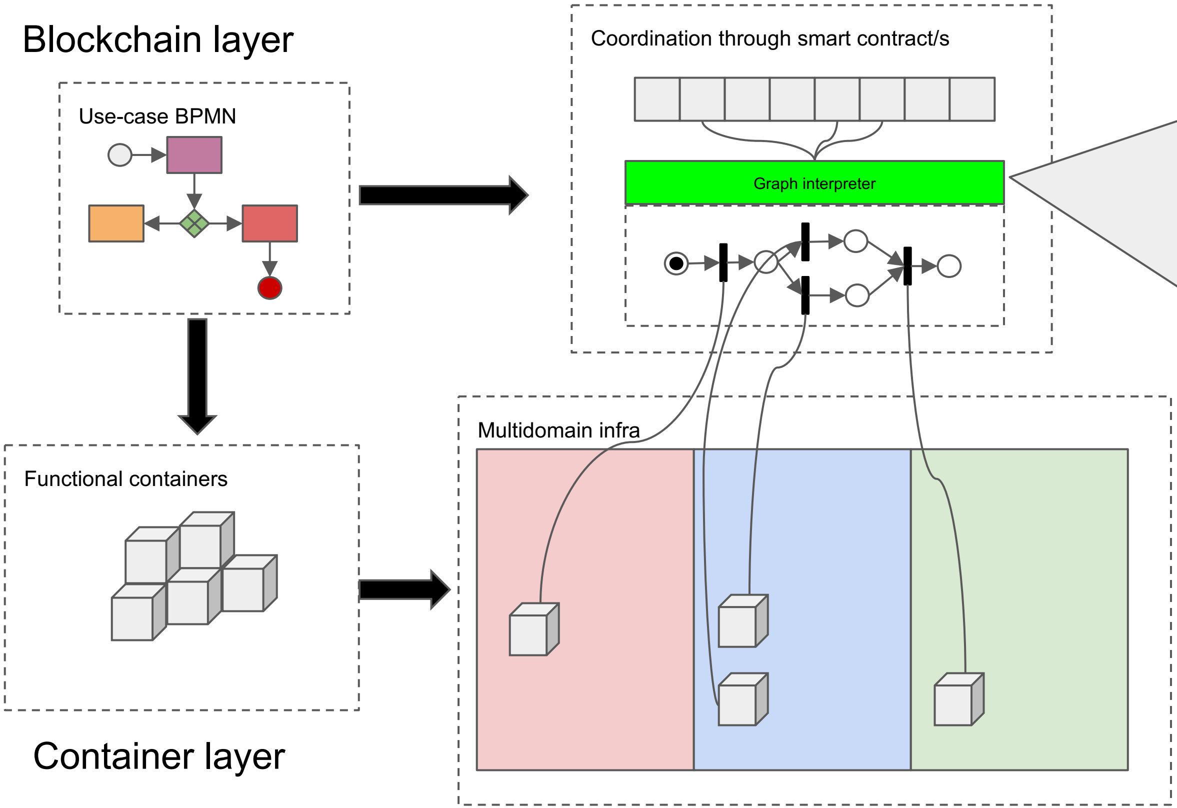
In this work we propose to encode the application agreement as a smart contract using Petrinet as a model to track state changes.



Architecture



Process model to infrastructure



- **Generic** dataflow/petrinet executor running on a blockchain i.e. every peer is running the executor.
- Domains/actors are assigned a set of tokens.
- Actors define functions as a task with token input, token outputs and webhooks to interact with the outside world.
- So actors own tokens and tasks
- A task needs certain amount of tokens to **fire**
- Blockchain transactions copy tokens between actors.
- When a task has enough input tokens it will **fire** which in turn generates blockchain events.
- Containers monitor the ledger to trigger a process inside a container (the task).
- The container will make blockchain transactions to signal the task is completed and move the state machine.

Universiteit van Amsterdam

Multiscale Networked Systems

Under the hood:
The responsible Internet

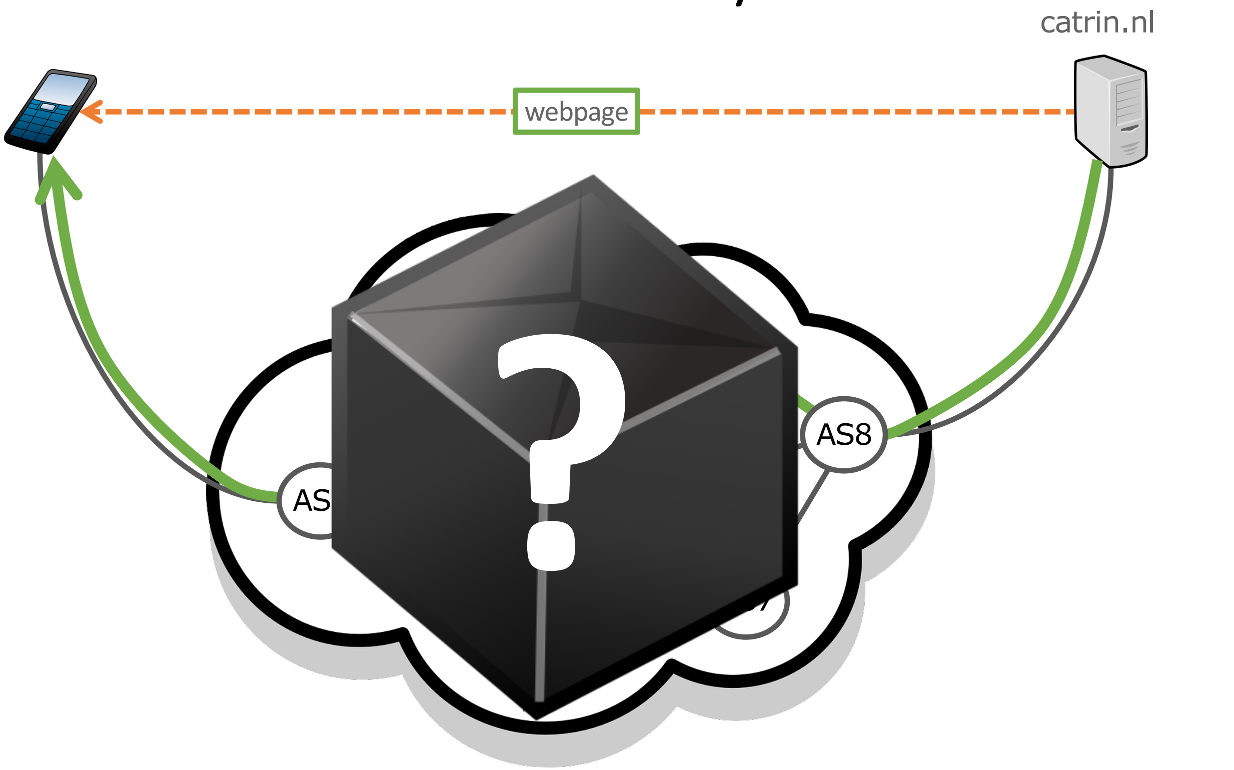
Paola Grosso

Multiscale Networked Systems research group
University of Amsterdam

Perception of the Internet vs. reality

User's perception
"It gets there"
Organizations
Individuals
services

Infrastructure-level
network operators
DNS operators
DDoS scrubbing centers
content distribution networks
names
addresses
routes



Slide from CATRIN kick-off (Ralph Holz)

Why we care: digital autonomy on the decline

- Increasing dependency on digital services in all societies
  - "Can we rely on the Internet as a neutral, trustworthy infrastructure?"
  - Limited insight in/control over dependencies, mesh of systems/operators

- Concerns world-wide about integrity of digital systems

- Dominance of few, large, powerful companies

Slide from CATRIN kick-off (Ralph Holz)

UNIVERSITEIT VAN AMSTERDAM

Multiscale Networked Systems

SpringerLink

Open Access | Published: 07 September 2020

A Responsible Internet to Increase Trust in the Digital World

Cristian Hesselman, Paola Grosso, Ralph Holz, Fernando Kuipers, Janet Hui Xue, Mattijs Jonker, Joeri de Ruiter, Anna Sperotto, Roland van Rijswijk-Deij, Giovane C. M. Moura, Aiko Pras & Cees de Laat

Journal of Network and Systems Management **28**, 882–922(2020) | Cite this article

Abstract

Policy makers in regions such as Europe are increasingly concerned about the trustworthiness and sovereignty of the foundations of their digital economy, because it often depends on systems operated or manufactured elsewhere. To help curb this problem, we propose the novel notion of a responsible Internet, which provides higher degrees of trust and sovereignty for critical service providers (e.g., power grids) and all kinds of other users by improving the transparency, accountability, and controllability of the Internet at the network-level. A responsible Internet accomplishes this through two new distributed and decentralized systems. The first is the Network Inspection Plane (NIP), which enables users to request measurement-based descriptions of the chains of network operators (e.g., ISPs and DNS and cloud providers) that handle their data flows or could potentially handle them, including the relationships between them and the properties of these operators. The second is the Network Control Plane (NCP), which allows users to specify how they expect the Internet infrastructure to handle their data (e.g., in terms of the security attributes that they expect chains of network operators to have) based on the insights they gained from the NIP. We discuss research

Challenges: transparency, acccountability and controllability

Two arguments

1. In the current effort to create ‘responsible’ practices the infrastructure view is negleted: the black box approach
2. Digital sovereignity is desirable but hard to achieve: critical infrastructure dependency on ‘foreign’/external actors

How can we provide transparency, accountable and controllability in the networks of the Future?

Enter programmability

Per packet processing in the dataplane provides advantages compared to out-of-band approaches for fine grained telemetry and for more granular control.

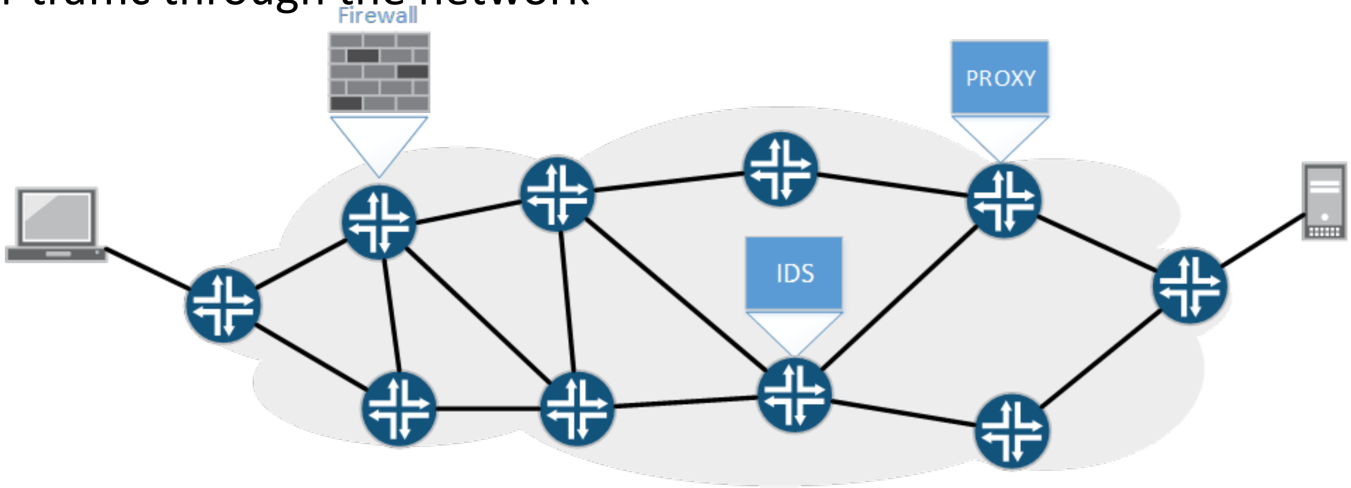
- Transparency:
  - From telemetry we acquire insights in what is happening in the network, eg the path taken by flows.
- Accountability goal:
  - From telemetry follows the possibility to identify attacks and feed intrusion detection systems.
- Controllability goal:
  - Users can select functionalities that better suit their intended network usage.

Enter Virtual network functions

Network Function Virtualization serves to more dynamically deploy network functions

- Moving Functions
- Creating Service Function Chains

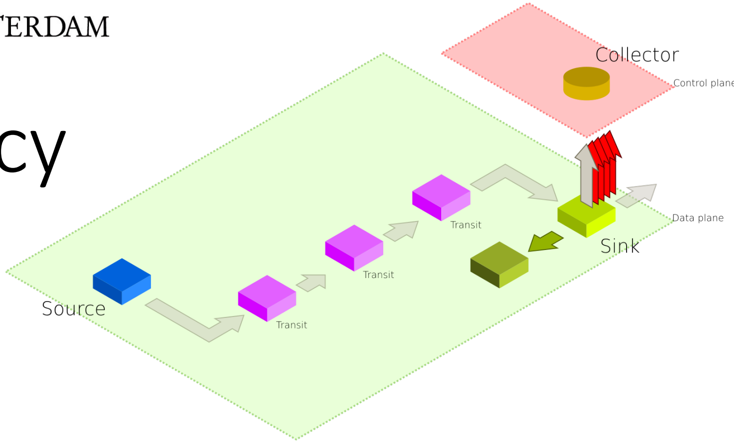
Steer traffic through the network



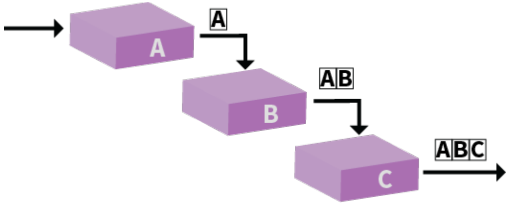
UNIVERSITEIT VAN AMSTERDAM

Multiscale Networked Systems

Transparency

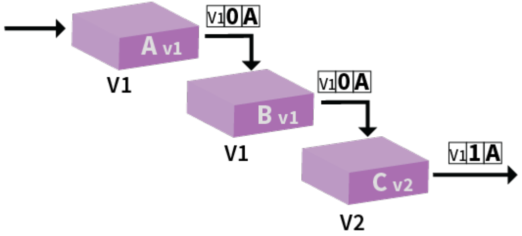


Knossen, S., Hill, J. and Grosso, P., 2019, November. *Hop recording and forwarding state logging: Two implementations for path tracking in p4. In 2019 IEEE/ACM Innovating the Network for Data-Intensive Science (INDIS) (pp. 36-47). IEEE.*



Next Header : next header type	Header Extension Length: 0x02	Padding: 0x00 * 6
Option Type: 0x3F	Option Data Length: 0x06	Option Data: Node Identifier
Option Type: 0x3F	Option Data Length: 0x06	Option Data: Node Identifier

8 bytes



Next Header : next header type	Header Extension Length: 0x01	Padding: 0x00 * 6
Option Type: 0x3F	Option Data Length: 0x06	Option Data

version	trackable	Node Identifier

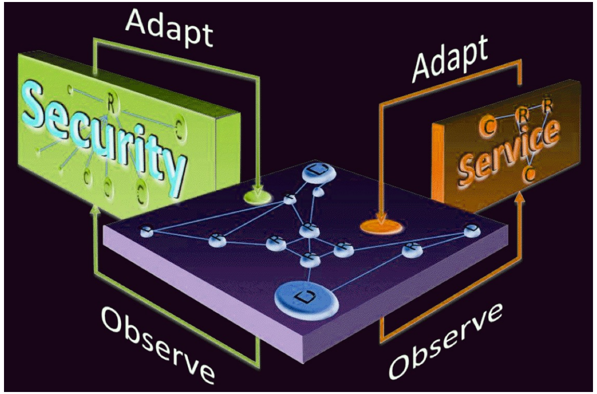
8 bytes

Beltman, R., Knossen, S., Hill, J. and Grosso, P., 2020, November. Using P4 and RDMA to collect telemetry data. In 2020 IEEE/ACM Innovating the Network for Data-Intensive Science (INDIS) (pp. 1-9). IEEE.

Universiteit van Amsterdam

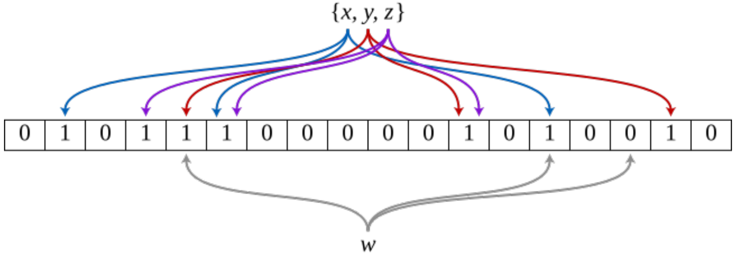
Multiscale Networked Systems

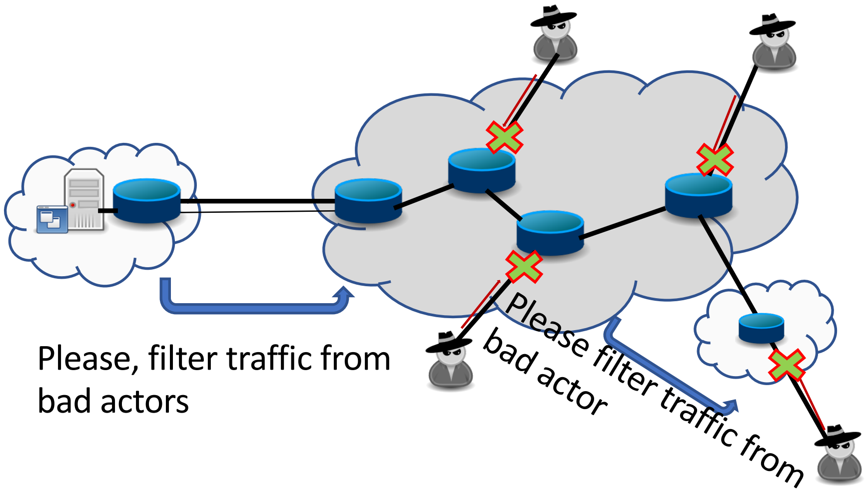
Controllability



Adapting for autonomous response (ML learning)

Bloom filters in P4

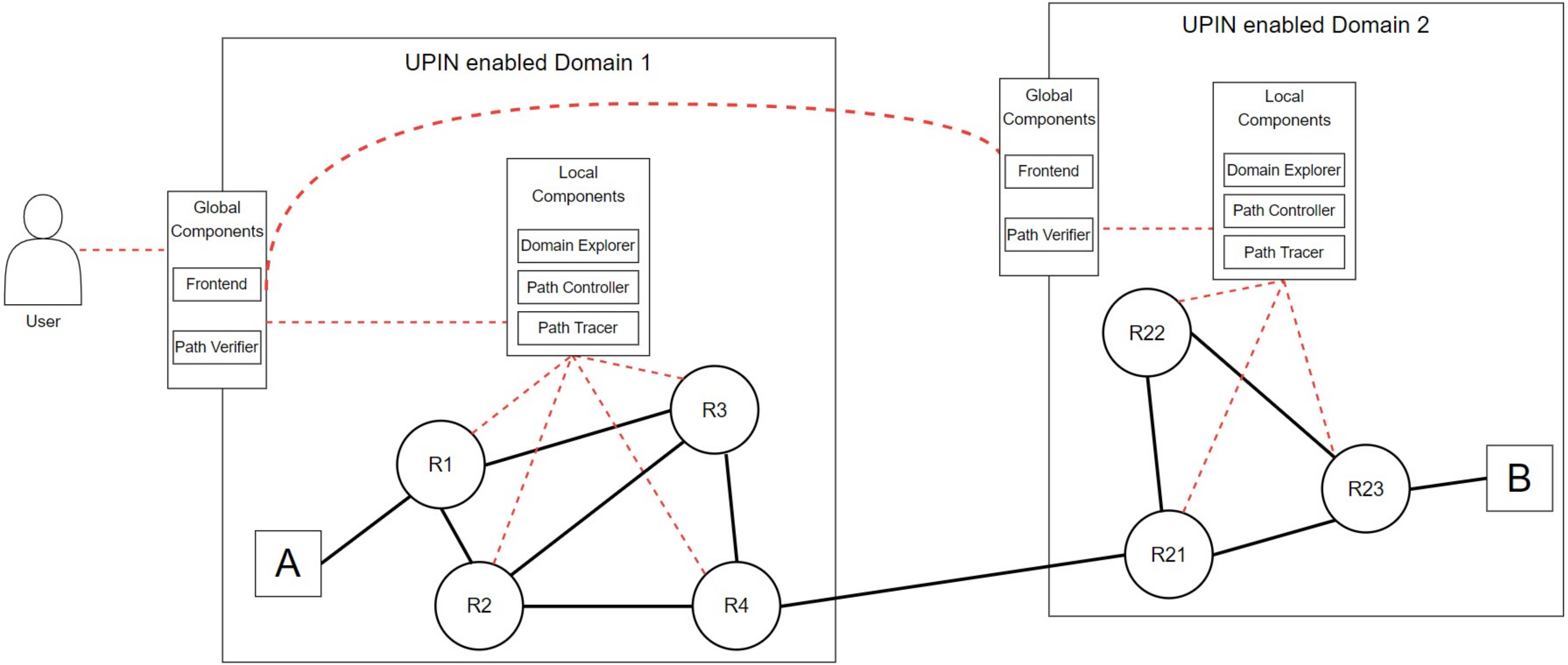




Hill, J., Aloserij, M. and Grosso, P., 2018, November. Tracking network flows with P4. In 2018 IEEE/ACM Innovating the Network for Data-Intensive Science (INDIS) (pp. 23-32). IEEE.

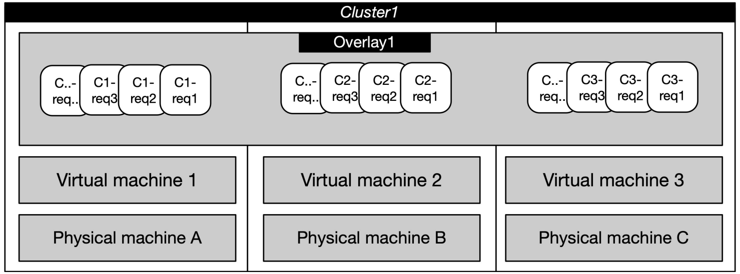
Koning, R., Deljoo, A., Meijer, L., de Laat, C. and Grosso, P., 2019, October. Trust-based collaborative defences in multi network alliances. In 2019 3rd Cyber Security in Networking Conference (CSNet) (pp. 42-49). IEEE.

Controllability in the UPIN model

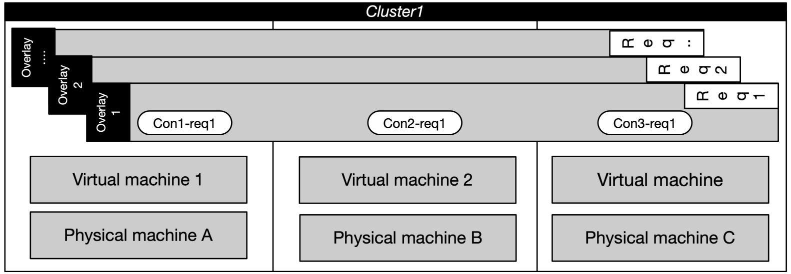


Bazo, R., Boldrini, L., Hesselman, C. and Grosso, P., 2021, August. Increasing the Transparency, Accountability and Controllability of multi-domain networks with the UPIN framework.
In *Proceedings of the ACM SIGCOMM 2021 Workshop on Technologies, Applications, and Uses of a Responsible Internet* (pp. 8-13).

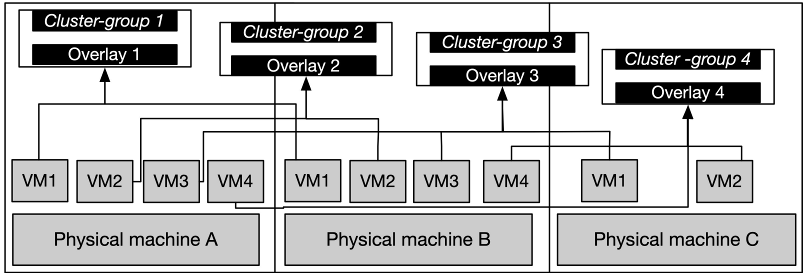
Intra domain connectivity



Overlay per DDM (Kubernetes and Calico)



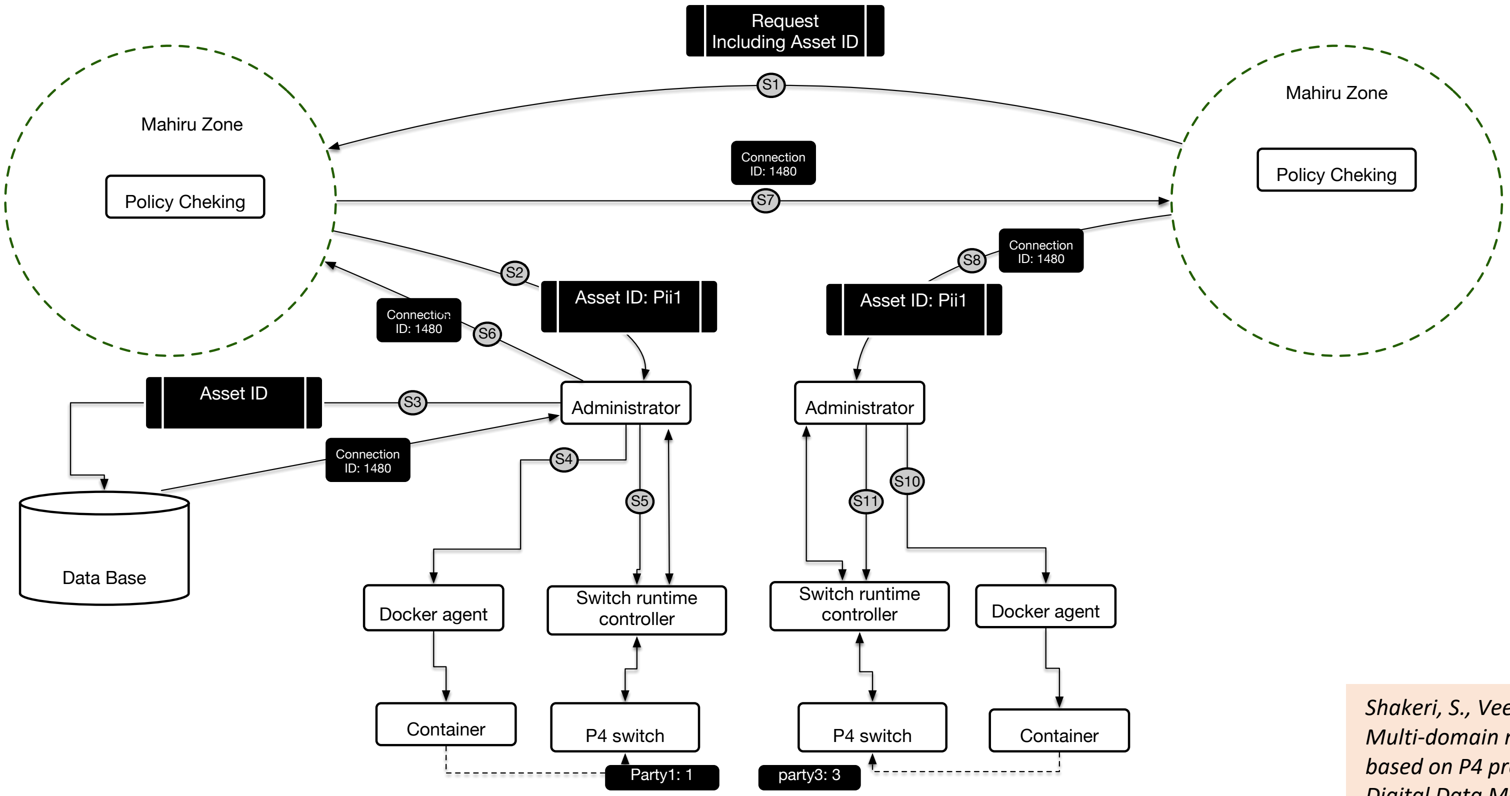
Overlay per Request (Swarm)



Overlay per Group (Kubernetes and Calico)

Shakeri, S., Veen, L. and Grosso, P., 2020, November. Evaluation of container overlays for secure data sharing. In 2020 IEEE 45th LCN Symposium on Emerging Topics in Networking (LCN Symposium) (pp. 99-108). IEEE.

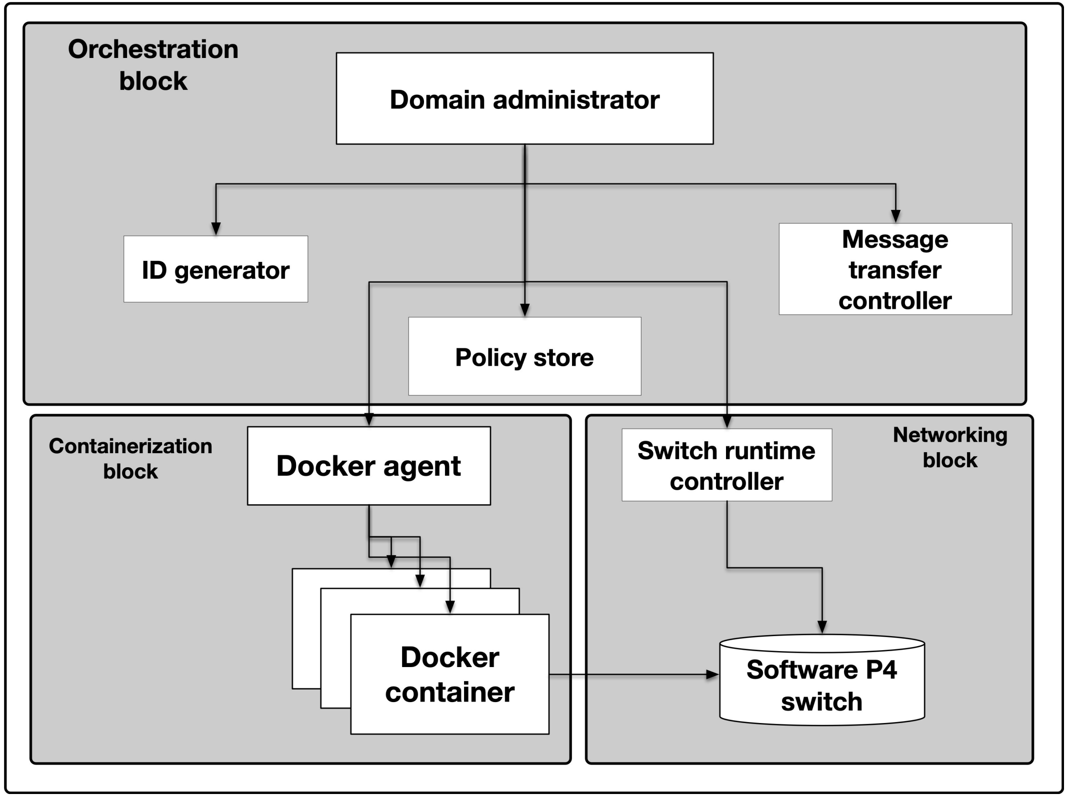
Multi-domain connectivity



Shakeri, S., Veen, L. and Grosso, P., 2022. Multi-domain network infrastructure based on P4 programmable devices for Digital Data Marketplaces. Cluster Computing, pp.1-14.

Putting it all together

All these networking technologies are at the basis of secure data sharing platforms!



Conclusions, Info, Acknowledgements, Q&A

- Data hindered by risk of unexpected use, lack of trust
- Using market principles, enforcement and determining incentives and value in the data life cycle to make data flow
- More information & published papers:
  - http://delaat.net/dl4ld http://delaat.net/epi http://delaat.net/sc
  - https://www.esciencecenter.nl/project/secconnet
  - https://towardsamdex.org https://upin-project.nl/
  - Slides with help from: Reggie Cushing, Sara Shakeri, Lu Zhang, Leon Gommans, Xin Zhou, Thomas van Binsbergen and many others.