Uz efektīvām pakešu transformācijām balstīta datortīklu virtualizēšana

Promocijas darba aizstāvēšana

Leo Trukšāns
Rīga, 2015. gada 23. janvārī

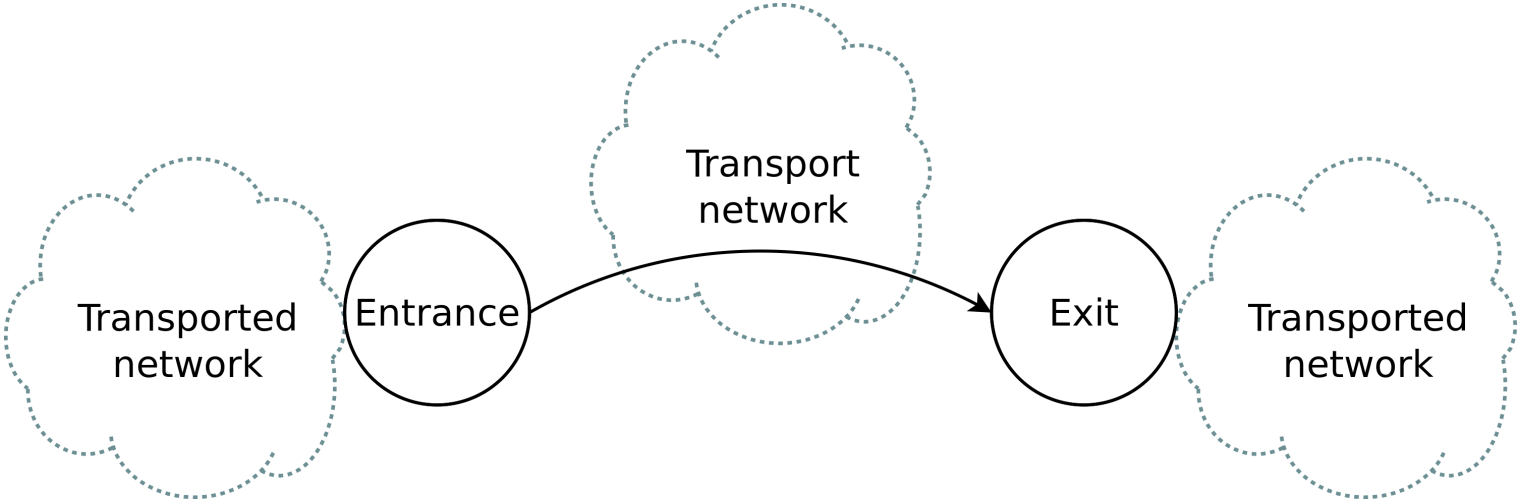
Darba galvenie rezultāti

- ZERO protokols
- Pakešu Transformāciju Valoda (PTL)

- LU MII mākoņa koncepcija un realizācija
- Astronomisko datu apstrāde mākonī
- Protokolu veiktspējas novērtējums meteoroloģisko datu pārraidei

ZERO protokols

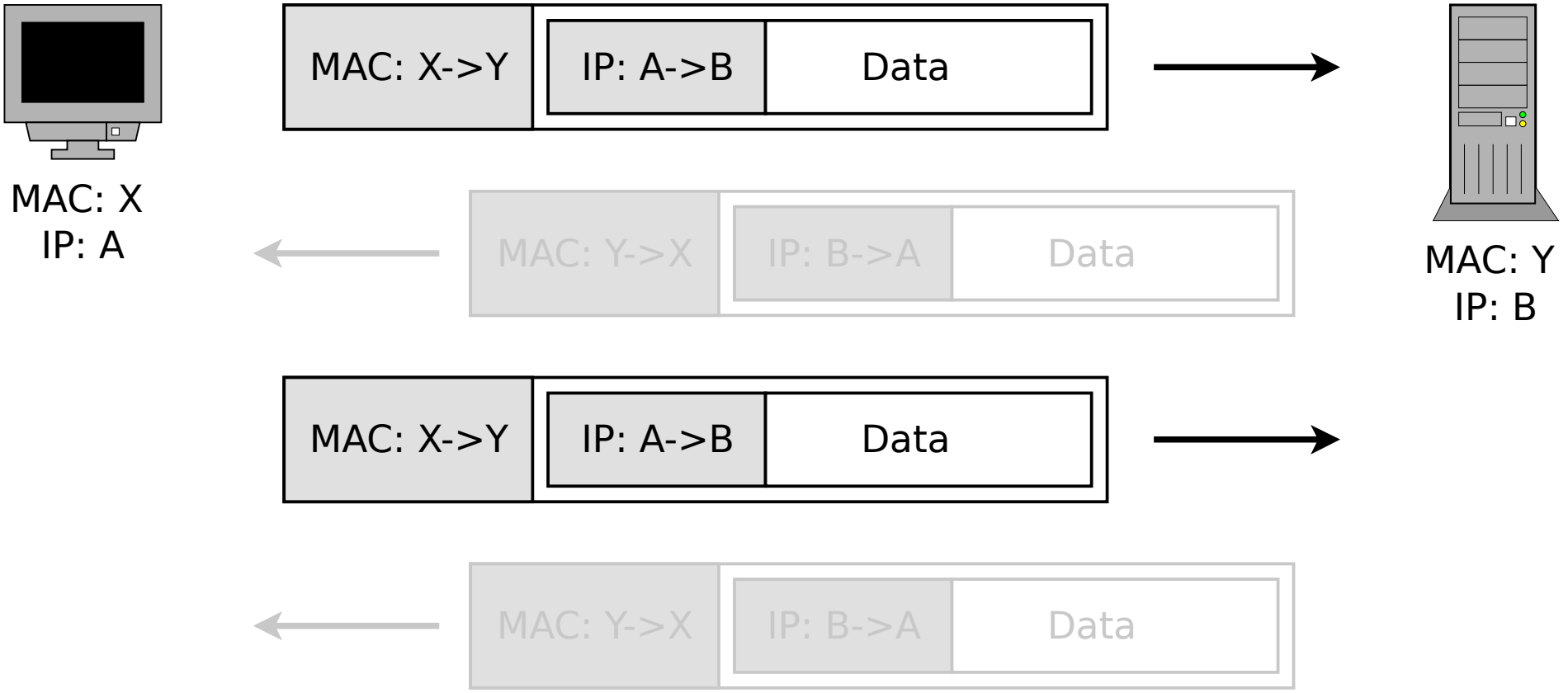
- Tunelējošs Ethernet-over-IP protokols



- Nepalielina datu struktūru – nefragmentē
- Simplekss
- Redundanta sinhronizācija

ZERO ideja

- Vienā sesijā visiem viena virziena kadriem MAC galva un gandrīz visa IP galva ir identiskas



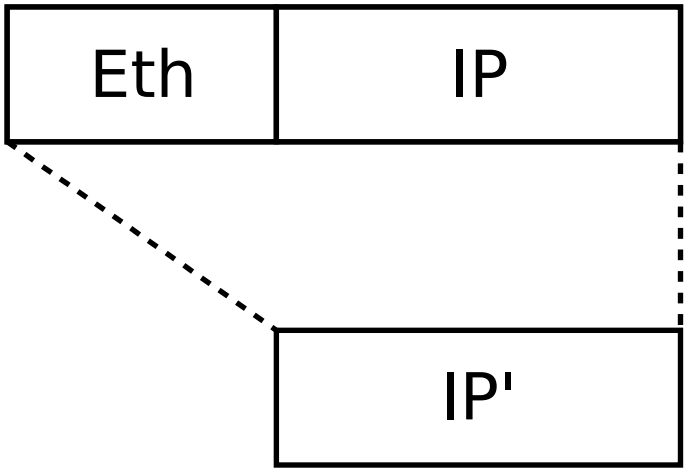
ZERO ideja

- Daži IP lauki praktiski **netiek lietoti**
  - *Fragment Offset* – lieto tikai, ja ceļā ir kanāls ar MTU<1500 (reti)
  - *Evil Bit* – netiek lietots saskaņā ar IP standartu

NICE kadri

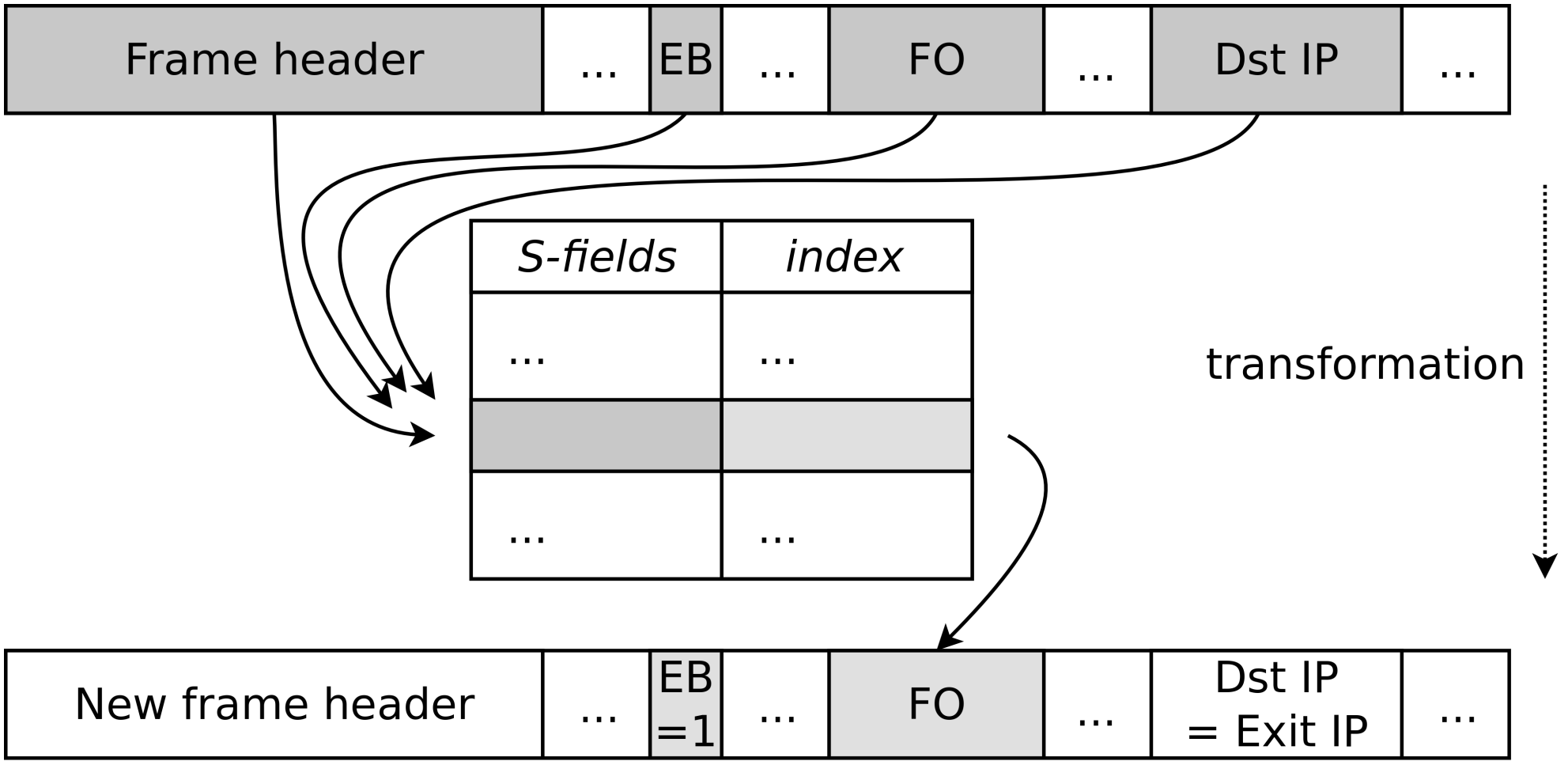
- **Definition 1.** *Ethernet frame is said to be NICE if it fulfills the following criteria:*
  - *it includes a legal IPv4 packet as the last part of Ethernet data field;*
  - *the header of the included IPv4 packet is 20 bytes long (has no options field such as Source Routing);*
  - *the TTL value of the included IPv4 packet is higher than certain minimum (ttl_min, described further);*
  - *the U-fields of the included IPv4 packet contain null data.*

NICE transformācija



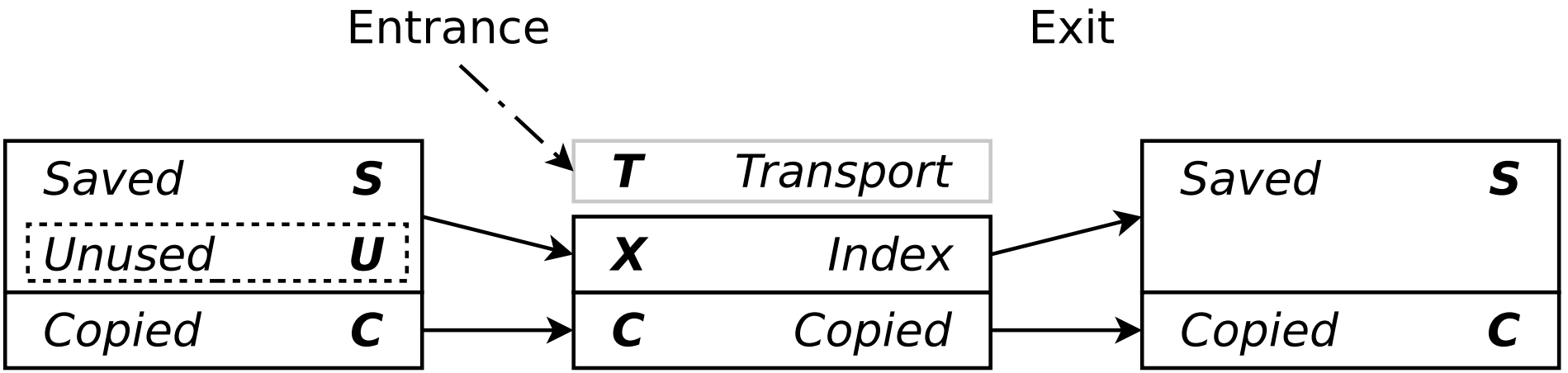
ZERO IP

NICE transformācija



ZERO IP

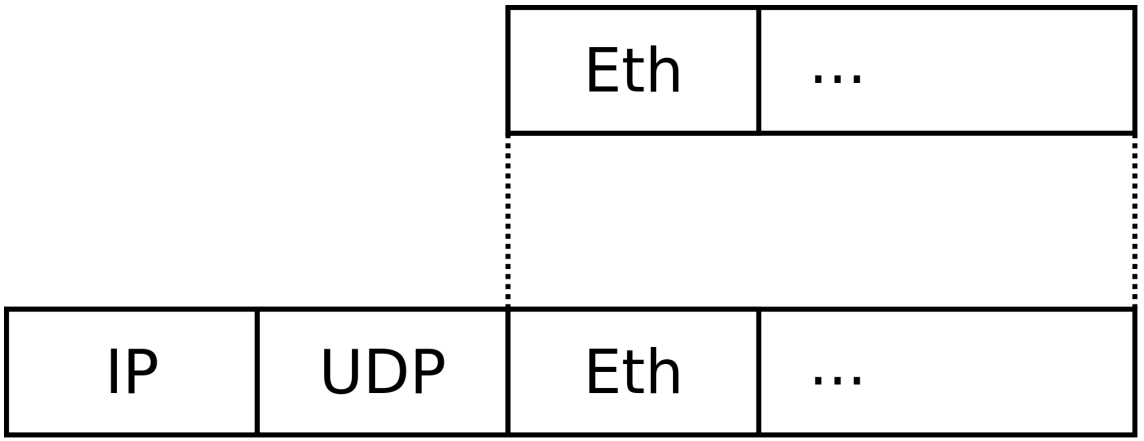
ZERO lauki


Entrance
Exit
Saved S
Unused U
Copied C
T Transport
X Index
C Copied
Saved S
Copied C

UGLY kadri

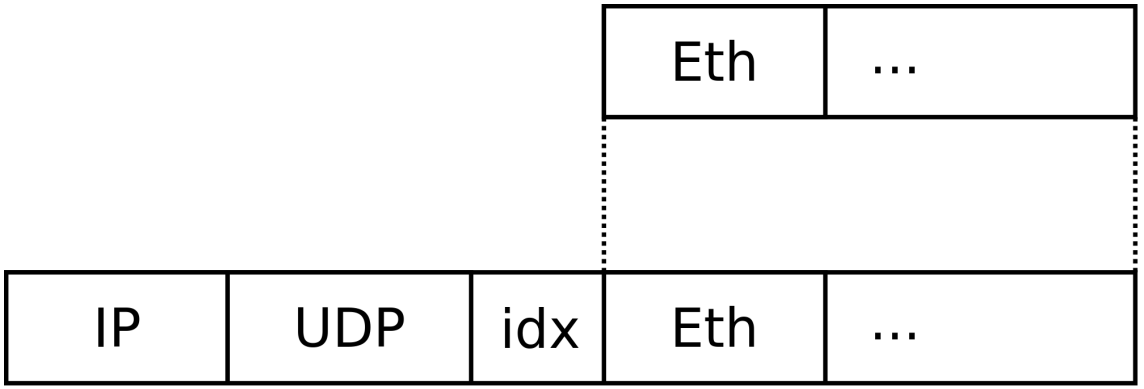
- **Definition 2.** *Ethernet frame is said to be UGLY if it is not NICE.*

UGLY transformācija



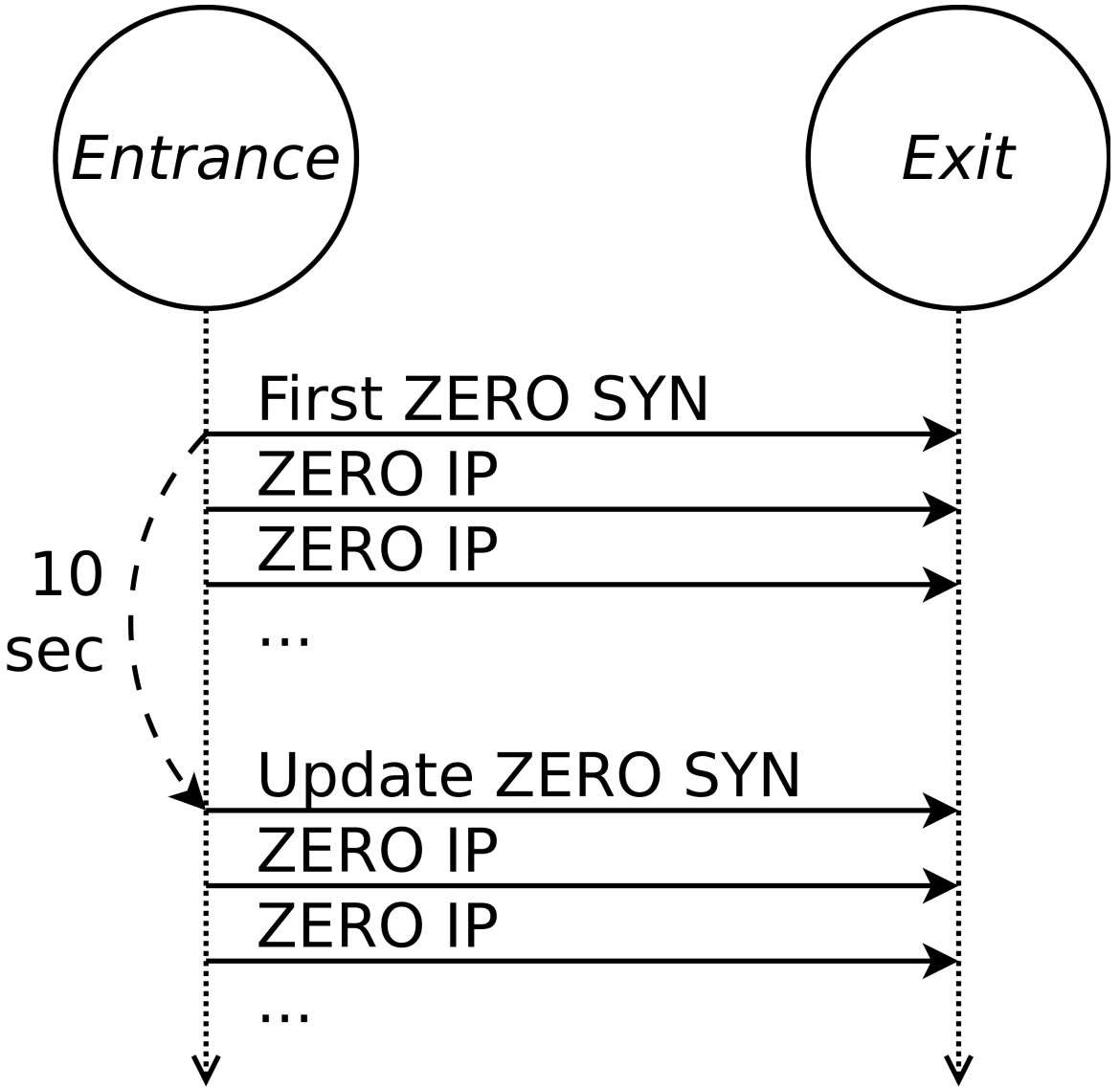
ZERO ENC

SYN transformācija



ZERO SYN

Kanālu redundance



ZERO IP korektums

- **Theorem:** *In converged state ZERO protocol will correctly tunnel NICE frames with zero-overhead through Transport infrastructures that:*
  - *do not fragment IPv4 packets with size<=1500;*
  - *do not filter IPv4 packets by Source IP address;*
  - *do not alter IP packet contents besides normal TTL and IP header checksum modification during forwarding.*

ZERO realizācijas

- Lietotnes līmenī
  - Sākotnējie rezultāti
- Linux kodola līmenī
  - Stabila, augstas veiktspējas realizācija
  - Ielādējama kodola moduļa veidā
  - Stendā tunelēšana pievienoja 300-400 mikrosekundes RTT laikam (bez tās – 200us)
  - Eksperimenti dažādos scenārijos
- Realizācija ar PTL

ZERO diskusija

- Daudzpunktu tunelēšanas topoloģija
- Par drošības sekām
- IPv6 iepakošana IPv4 tunelī nefragmentējot
- IP *Ident* lauka lietošana indeksam
- Alternatīva TTL lauka kompensēšana
- ZERO noturība pret NAT/PAT
- MEPS (*Maximum encapsulated packet size*) jēdziens

PTL

- Pakešu Transformāciju Valoda (*Packet Transformation Language, PTL*)
- PTL ir projektēts ar sekojošām spējām:
  - 1. transformēt dotu OSI līmeņa PDU uz citu OSI līmeņa PDU;
  - 2. nodrošināt universālu valodu PDU transformāciju definēšanai;
  - 3. nodrošināt PDU transformāciju atgriezeniskumu (*reversibility*).

PTL

- Atgriezeniskuma princips (*Reversibility principle*):
  - Visām tiešajām transformācijām (*FT*) *Sūtītājā*, jāatbilst funkcionāli pretējām transformācijām (*RT*) *Saņēmējā*

$$FT(P) = P' \quad \text{un} \quad RT(P') = P$$

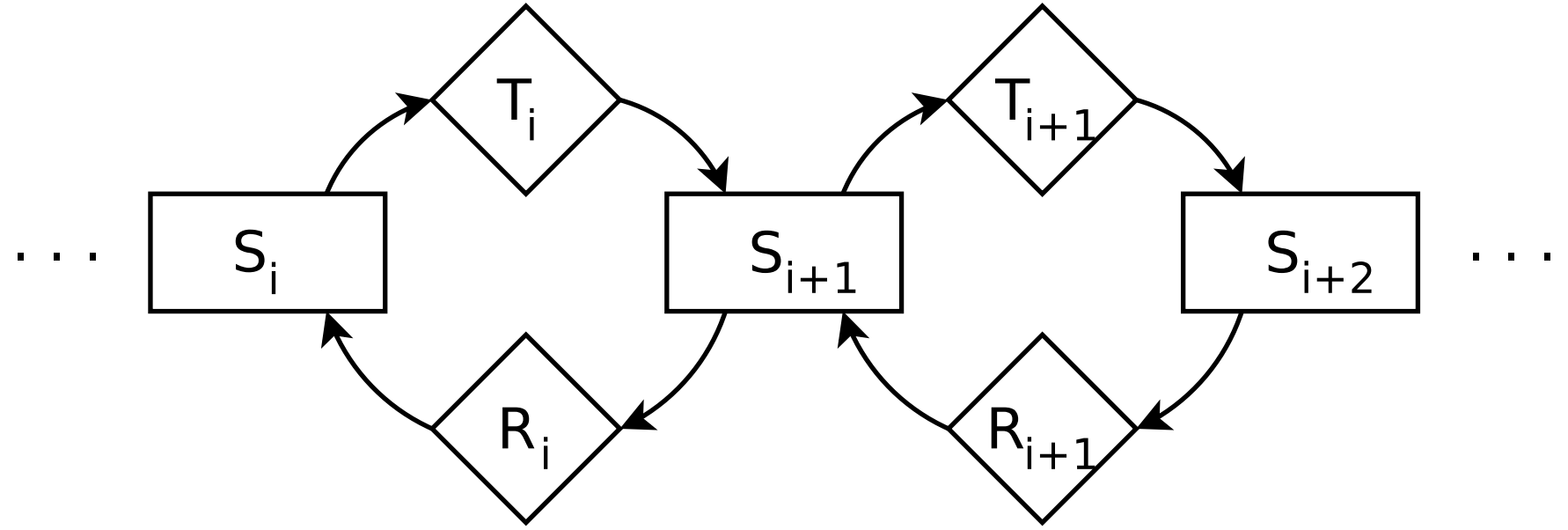
$$\text{tātad,} \quad RT(FT(P)) = P$$

PTL

- Tieša transformācija (*FT*) var neizdoties, ja kāda tās funkcija neizdodas
  - Ja visas funkcijas izdodas, transformācija izdodas
- Var paredzēt transformāciju ietvaru ar optimālām FT un “atkāpšanās” FT
  - Katram kanālam – pieņemamākā transformācija

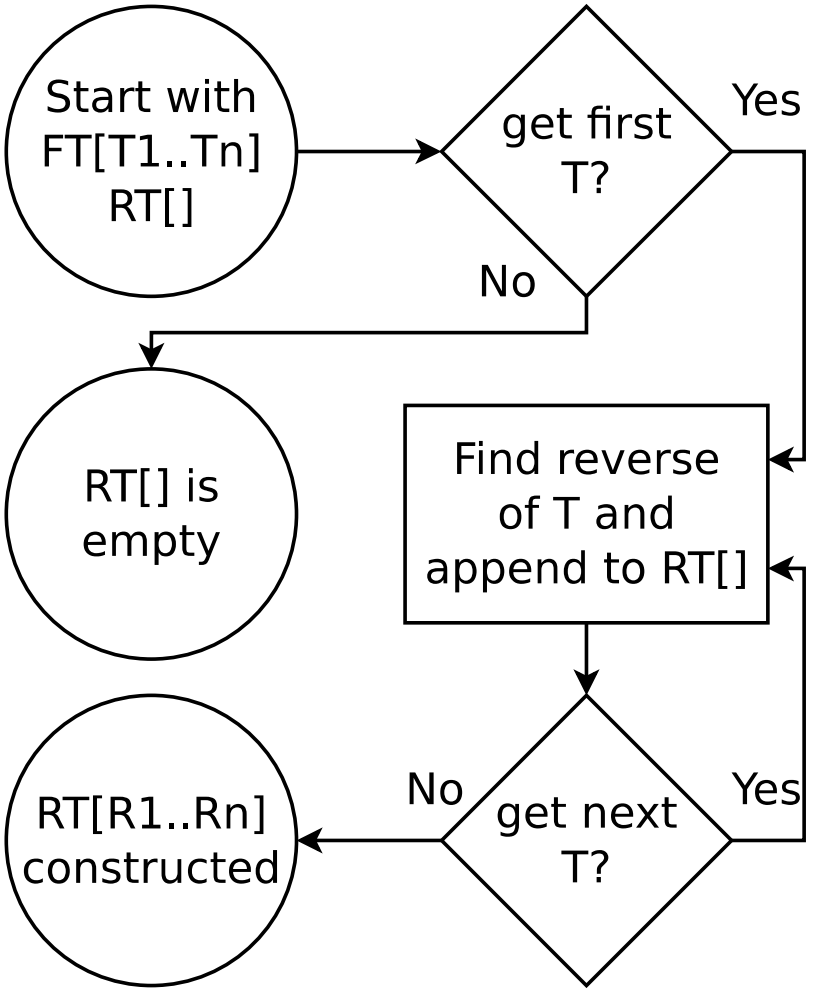
PTL

- FT un RT funkciju ķēdes transformē datus
- Lietotājs uzdod tikai FT



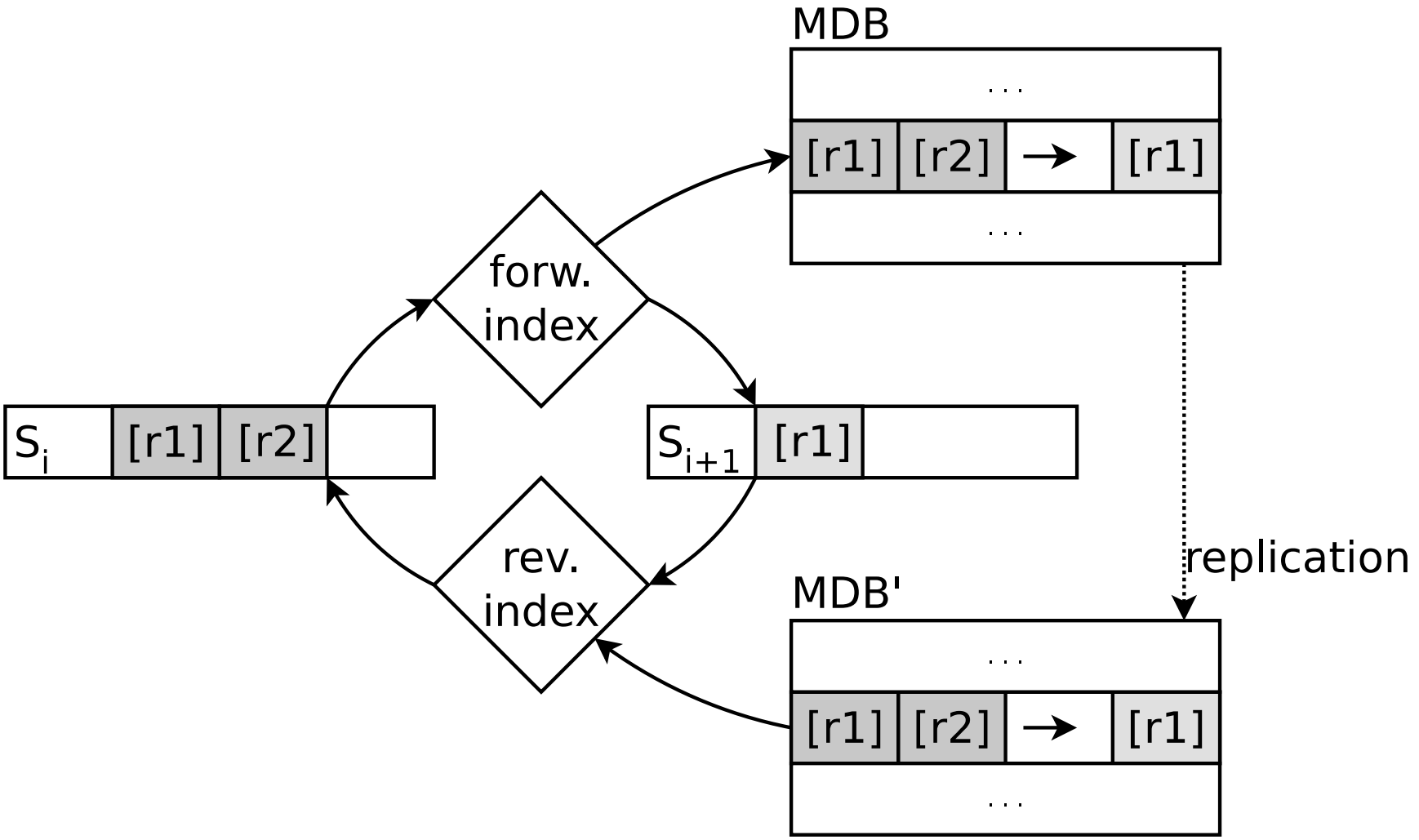
PTL

- Reverso Transformāciju Ģeneratora *(RTG)* algoritms



PTL

- `index` funkcija



PTL sākotnējās funkcijas

- Izvērtēšanas funkcijas

`equal` – salīdzina divu parametru vērtības

`exists` – pārbauda parametra eksistenci apskatāmajā PDU

PTL sākotnējās funkcijas

- Satura maiņas funkcijas

`swap` – maina vietām abu parametru saturu

`insert` – saturu no p2 ievieto pirms p1

`encapsulate` – iepako p1 jaunā p2 struktūrā

`index` – kartē sr saturu indeksa veidā dr apgabalā

`DES` – šifrē p1 ar atslēgu key

`flag` – iezīmē p1 ar saturu no p2

ZERO pieraksts PTL valodā

- ZERO *DC* transformācija:

```
equal(([P0-P7],[P50-P63]), (0x45, 0))
flag(P48, 0, 1)
index([P51-P63], ([F0-P0), [P51-P63], [P128-P159]), 1, 4096)
```

ZERO pieraksts PTL valodā

- ZERO *IC* transformācija:

```
equal(([P0-P7],[P50-P63]), (0x45, 0))
flag(P48, 0, 1)
swap([P96-P127], [P128-P159])
index([P51-P63], ([F0-P0), [P51-P63], [P128-P159]), 1, 4096)
```

Aprobācija un citi rezultāti

- Linux kodola ZERO realizācija
- LU MII mākoņa koncepcija un realizācija
- Astronomisko datu apstrāde mākonī
- Protokolu veiktspējas novērtējums meteoroloģisko datu pārraidei

EUMETSAT

ZERO aprobācija

- Testu rezultāti apstiprina ZERO protokola gandrīz nulles virstēriņu efektivitāti – **99,94%** paketēm nav virstēriņu
- Kodola realizācija uzrāda augstu veiktspēju un mazas aiztures
- Aprobēta mājas apstākļos un augstās slodzēs
- Praksē atrasti jauni aspekti
  - “Mazo sinhronizāciju faktors”
  - Reālu transporta tīklu ierobežojumi

Pirmās paaudzes mākonis – E-spiets

- Sākts 2008. gadā, bez esoša pieprasījuma
- Vairāku tipu pakalpojumi:
  - IaaS, PaaS, datnes.lv, Lieli Dati
- Mūsdienīgi tīklošanas risinājumi
  - VPN tuneļi, VLANi apakštīkliem, plūsmu filtri
- Pastāvīgi projektēts, ieviests, attīstīts
  - Tālredzīga izvietošana (*provisioning*)
  - Reālā laika pārraudzība (*monitoring*)
  - Pārslodzes paredzētājs un jaudas mērogotājs

Pirmās paaudzes mākonis – E-spiets

Pirmās paaudzes mākonis – E-spiets

- SAN datu tīkls
  - Augstas pieejamības arhitektūra
  - 6 kontrolieri, 624 SATA + 16 FC diski
  - 468 TB telpa
- 384 GB RAM (8*48)
- 64 Xeon kodoli (8*8)

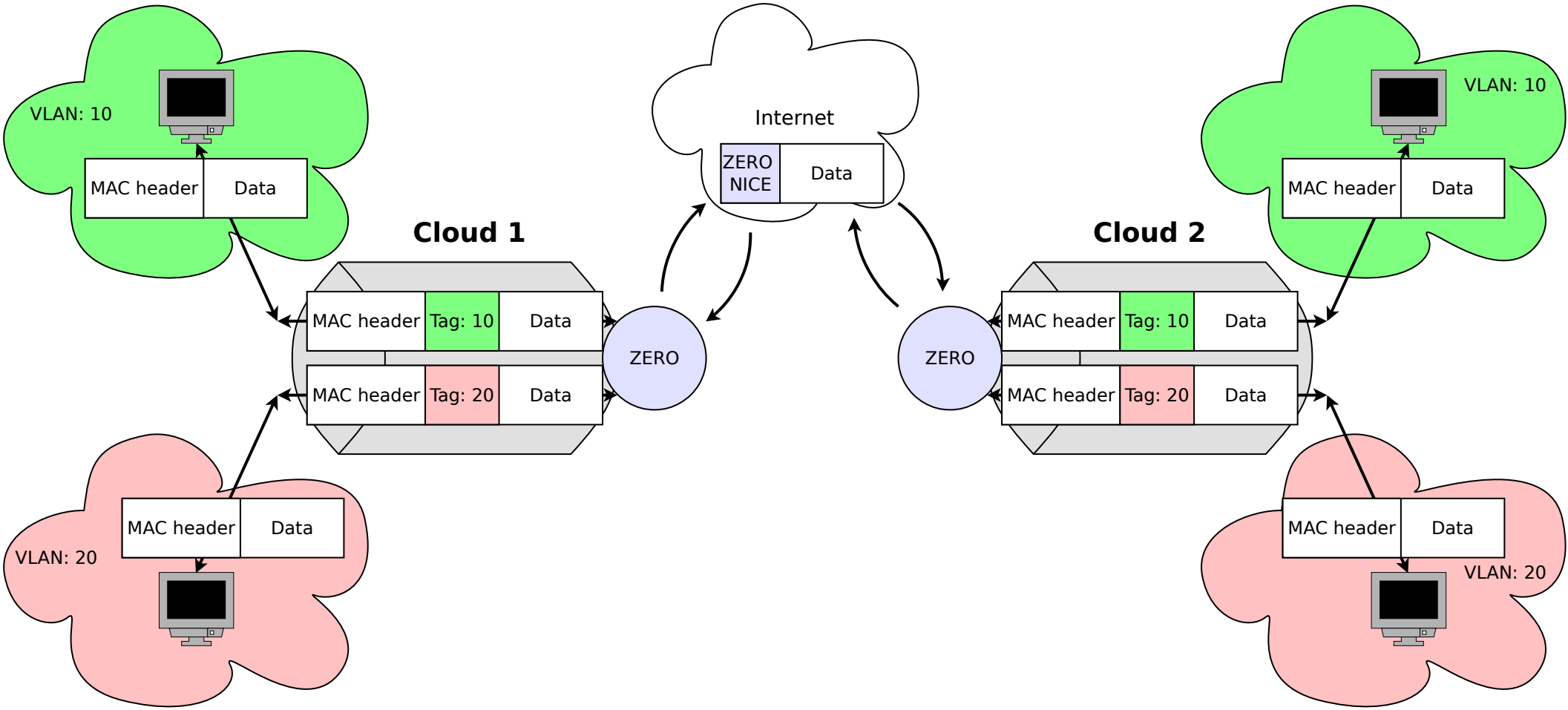
Pirmās paaudzes mākonis – E-spiets

- Jau 2011. gadā pirmais resursu deficīts
- 2015. gadā – praktiski pilns
  - CPU: 92,97%
  - RAM: 91,56%
  - SAN: 84,64%
- 179 VM, 274 virtuālie diski, 22 tīkli
- 6 gadu laikā – neviena būtiska pārtraukuma!

Nākamās paaudzes mākonis

- Plašs pakalpojumu klāsts 3 līmeņos:
  - IaaS, PaaS, SaaS
- Lietotnes var izvietot dažādos līmeņos
  - MaaS, High Performance Computing (HPC), High Throughput Computing (HTC), GPU
- Zaļā skaitļošana – nelietotus resursus slēdz ārā
- Var savienot mākoņus federatīvās sistēmās
  - ZERO protokolu iespējams lietot mākoņu kodolu tīklu loģiskai apvienošanai bez fragmentācijas

ZERO “Intercloud”



Laiks jautājumiem

Layer2 paplašināšana

- Ir vēlme paplašināt Layer2 kanālus
  - Nodalāmas lietotāju grupas sadalītā tīklā
  - Virtuālo mašīnu grupas mākoņos (*IaaS clouds*)
  - Lietu Internets (*Internet of Things, IoT*)
  - Attālināti uzņēmuma biroji
  - Attāli lietotāji

Layer2 paplašināšana

- Kā to darīt?
  - Kā konceptuāli atdalīt Layer2 kanālu no tam izdalītas fiziskās infrastruktūras?

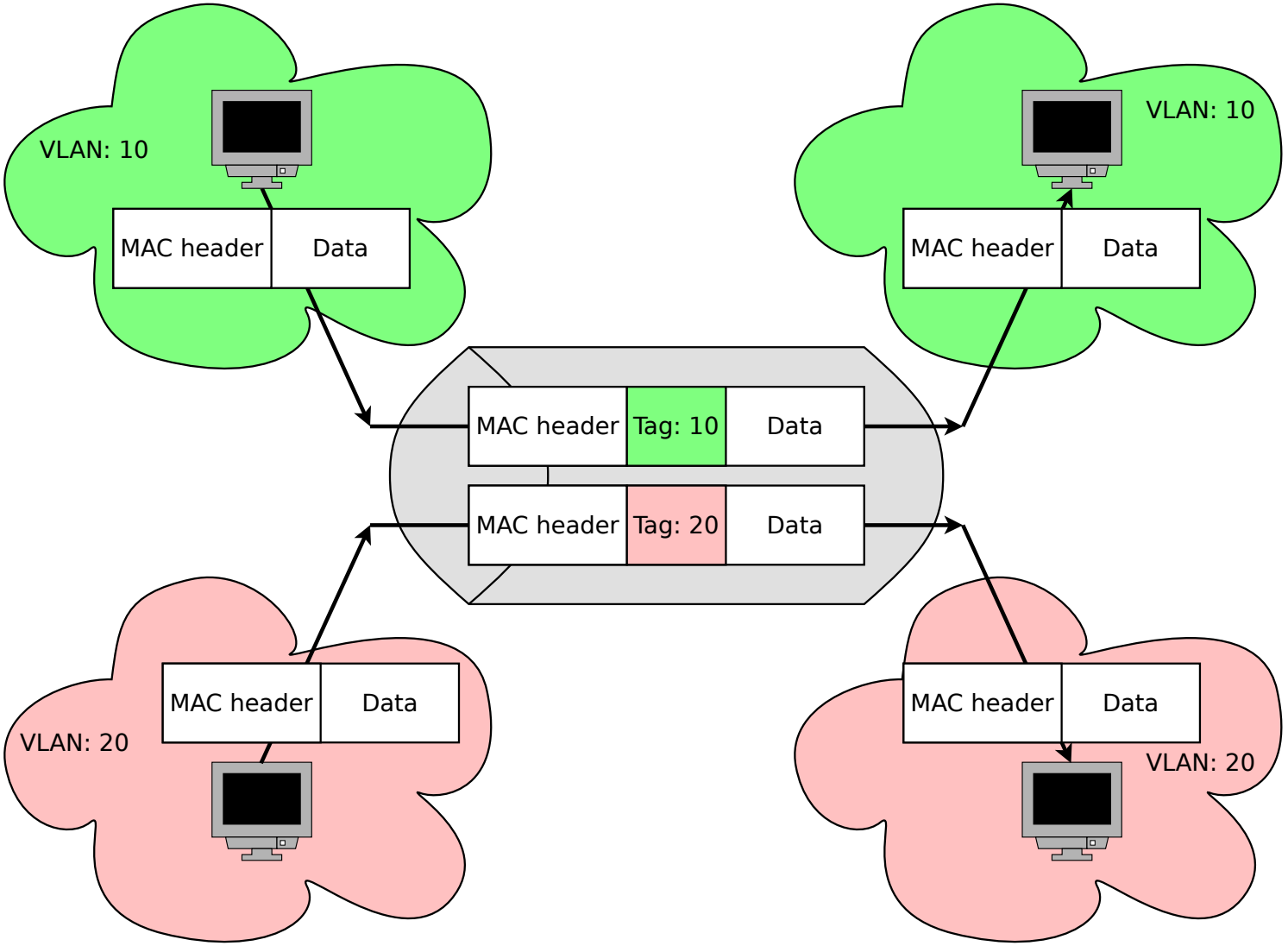
Layer2 paplašināšana

- **Virtualizēt Layer2 kanālu** kā loģisku vienību – uz pieejamās Layer2/3/4 infrastruktūras (transporta tīkla)
- Koplietot transporta tīklu vairākiem šķietami nodalītiem loģiskiem Layer2 kanāliem
- Ļaut radīt jaunus [loģiskus] tīklus, neradot jaunu infrastruktūru

- **Datoru tīklu virtualizēšana** ir fundamentāla iespēja mūsdienu datoru tīklos

Tīklu virtualizēšanas metodes

- **Plūsmu iezīmēšana** (*labeling*)



Tīklu virtualizēšanas metodes

- **Plūsmu iezīmēšana** (*labeling*)
  - VLAN (802.1q)
  - Ethernet-over-MPLS
  - VPLS
  - utt.

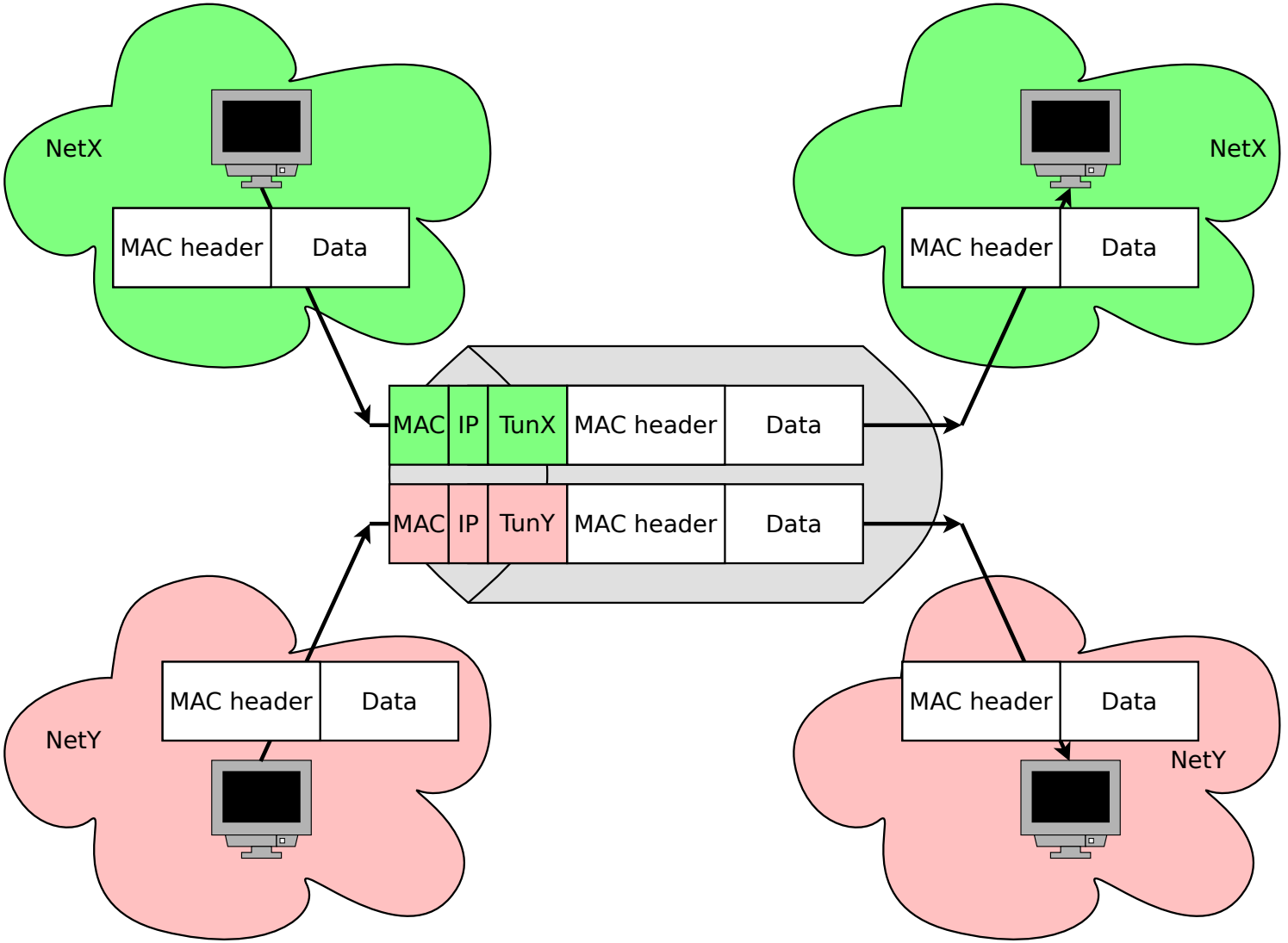
Tīklu virtualizēšanas metodes

- **Plūsmu iezīmēšana** (*labeling*)

- Nedaudz palielina kadru ar papildus lauku
- Nepieciešama īpaša, protokolu atbalstoša **infrastruktūra** ar lielāku (>1500) MTU
  - **Zaudē** Layer2 informāciju, pārejot citā Layer2 protokolā (aiz maršrutētāja)

Tīklu virtualizēšanas metodes

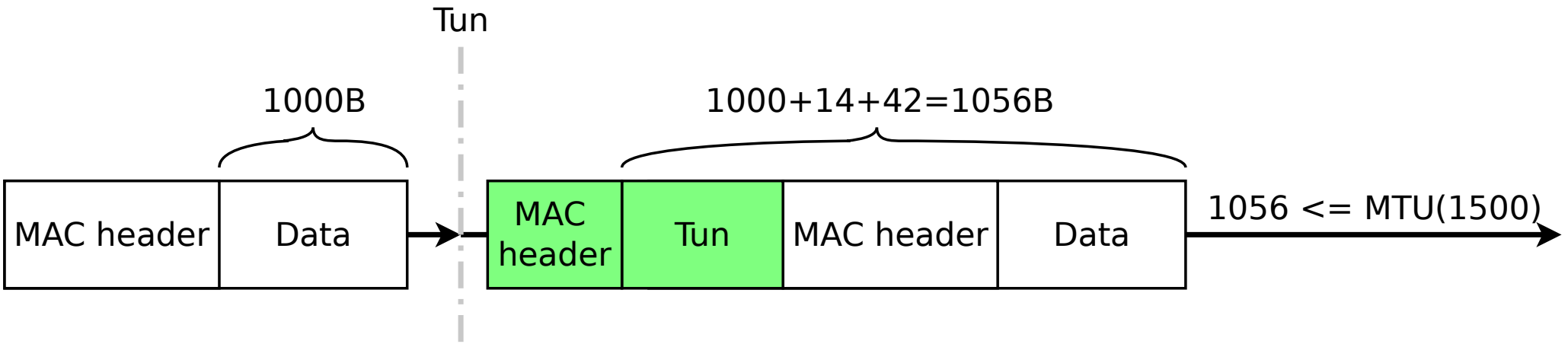
- **Tunelēšana**



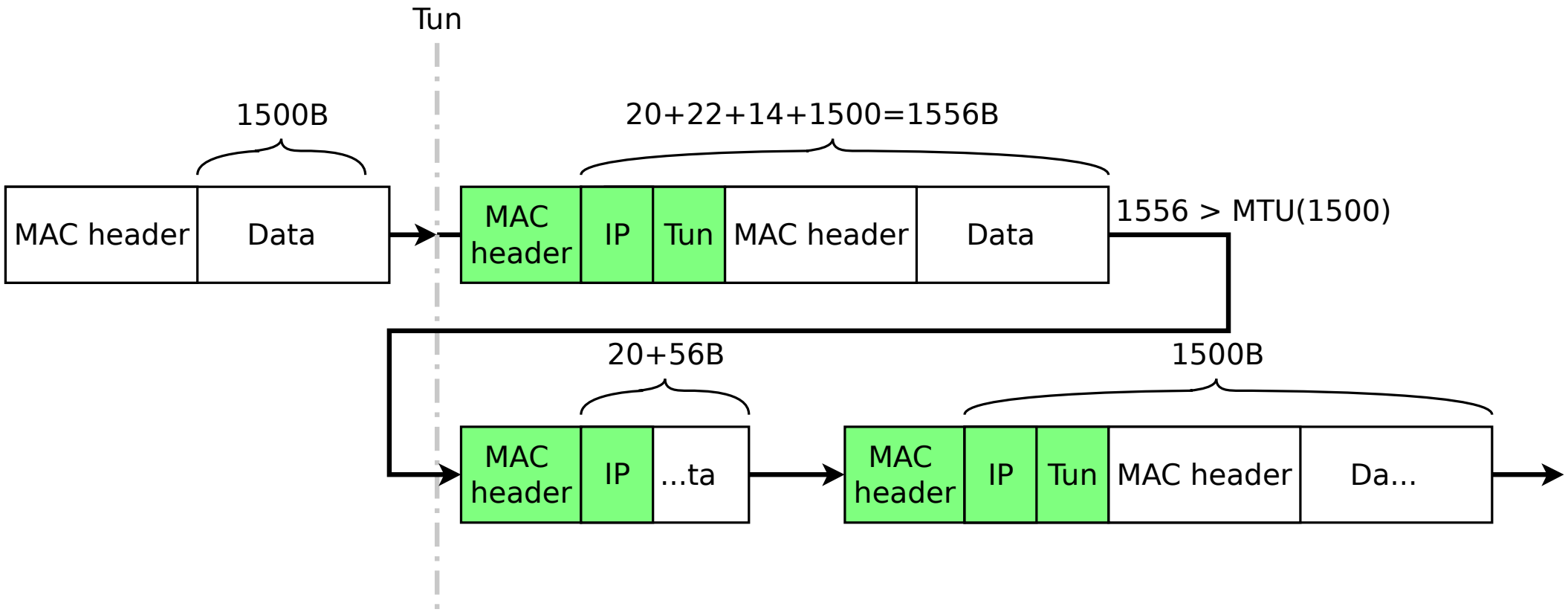
Tīklu virtualizēšanas metodes

- **Tunelēšana**
- Layer3 plūsmas
  - IP-IP
  - GRE
  - IPsec
- Layer2 plūsmas
  - L2TP
  - EtherIP
  - OpenVPN
  - VXLAN
  - NVGRE
  - STT

Fragmentēšana tuneļos



Fragmentēšana tuneļos



- Virstēriņi, mainīgas aiztures

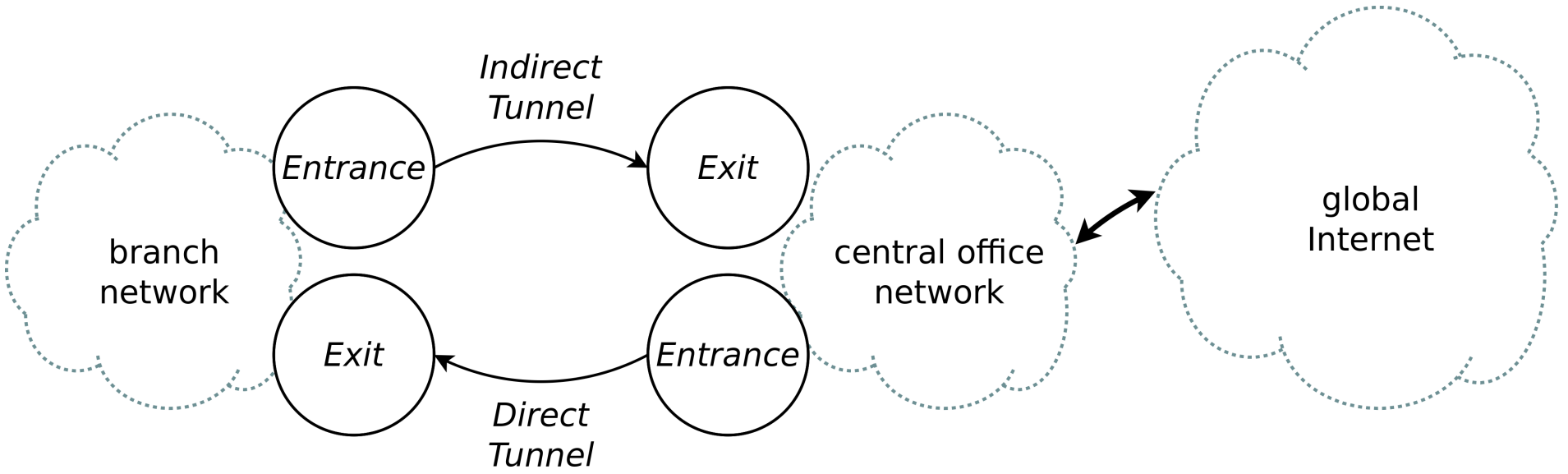
Tīklu virtualizēšanas metodes

- **Tunelēšana**

- Lietojama Internetā u.c. transporta tīklos
- Vienmēr **iepako** kadru/paketi transporta tīklā lietotā struktūrā
  - **Saglabā** visu iepakoto datu struktūru
  - Būtiski **palielina** pārsūtāmo tilpumu
- Rada fragmentācijas iespēju, mainīgas aiztures

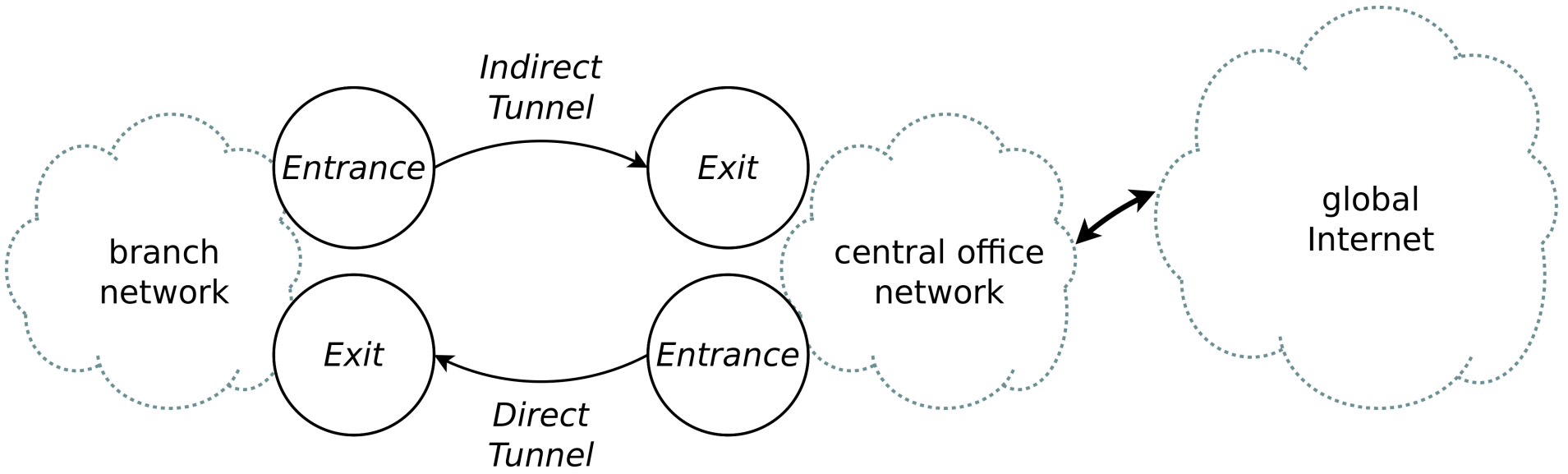
Tiešie un netiešie kanāli

- "*Daži -> daudzi*" problēma
- *IC NICE* transformācijā: *SrcIP* <-> *DstIP*
- *SrcIP* kļūst par *S-field*



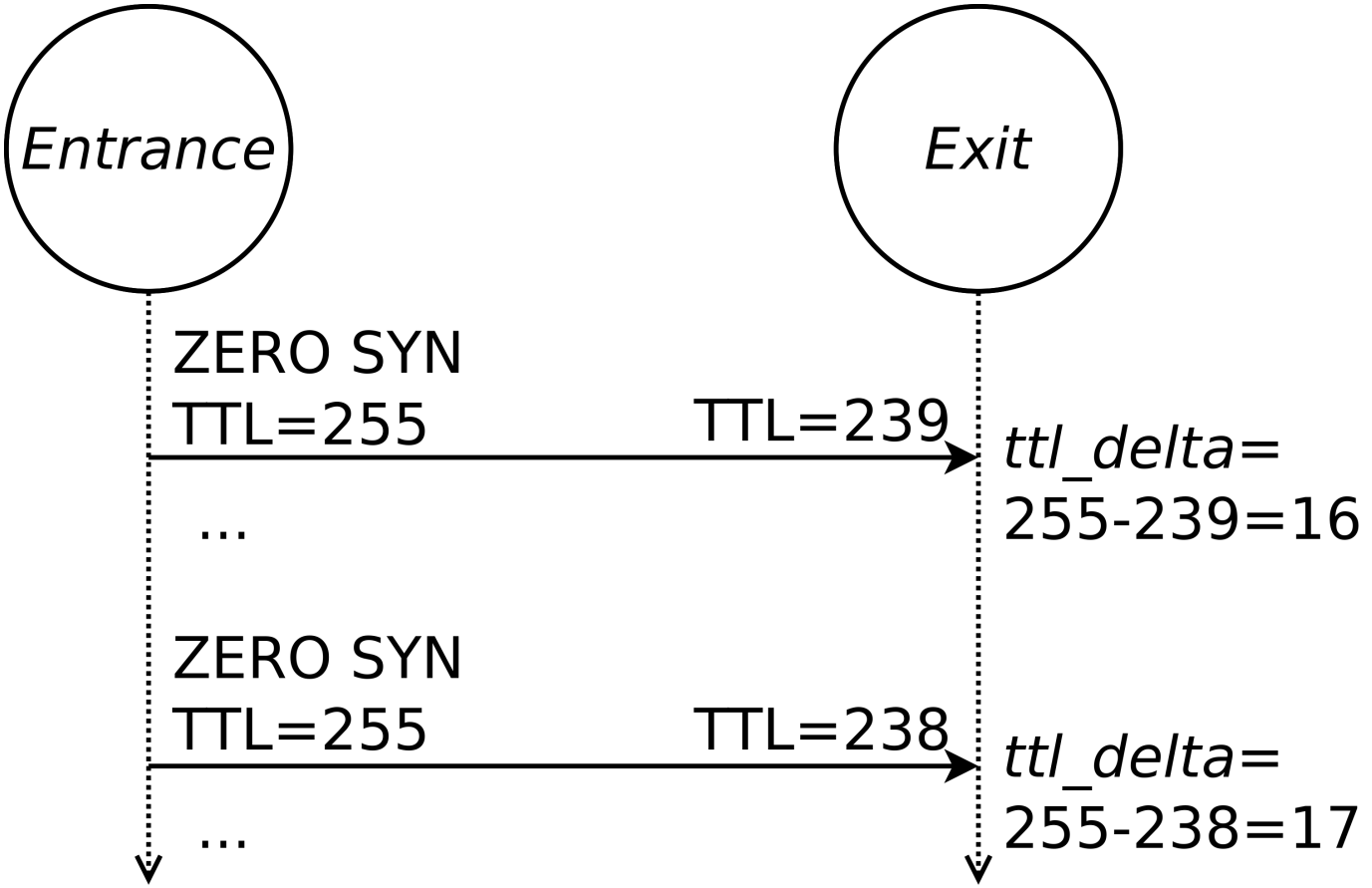
Tiešie un netiešie kanāli

- **Theorem 1:** *Every branch computer requires only one Channel in each direction through ZERO tunnel to communicate with all global Internet hosts connected behind the central office network.*



TTL kompensēšana

- *ttl_delta = 255 - syn_ttl*



ZERO realizācijas

- Sākotnējais novērtējums

(R1)	ZERO	OpenVPN
Packet count	61289	61309
Total volume, bytes	4430144	7001420

No lietotāja iegūst tilpumā: **63,27%** no OpenVPN

(R2)	ZERO	OpenVPN
Packet count	125491	248473
Total volume, bytes	187155054	197586861

Uz lietotāju iegūst skaitā: **50,5%** no OpenVPN

ZERO realizācija mājas apstākļos

- Lietota gada garumā
- Cauri vismaz 2 IPS, tuvākajā ir DHCP klients
- Ierasti mājas lietojumi:
  - Tīmeklis, E-pasts, Skype, Youtube

ZERO realizācija mājas apstākļos

- Statistika par 31 dienas periodu:

Premises	Packets	NICE	NICE, SYN	UGLY
SOHO	Entrance packet counts	3286143, **96,5%**	95852, **2,8%**	22840, **0,7%**
SOHO	Entrance byte counts	241715177, **93,0%**	16750197, **6,4%**	1417140, **0,5%**
Remote	Entrance packet counts	6807934, **98,4%**	87025, **1,3%**	23510, **0,3%**
Remote	Entrance byte counts	9694042054, **99,8%**	16613563, **0,2%**	1363580, **0,0%**

- NICE/UGLY, SOHO->Remote = **99,3%** / 0,7%
- NICE/UGLY, Remote->SOHO = **99,7%** / 0,3%

ZERO realizācija mājas apstākļos

- ZERO SYN ietekme uz fragmentēšanu

Scenario	Packets	Bytes	SYNs	Fragd SYNs
Long TCP sessions, tunn. to client	1412722	2118545373	175	169
Long TCP sessions, tunn. to server	690894	35976562	175	0
Short TCP sessions, tunn. to client	37889	56403591	54	1
Short TCP sessions, tunn. to server	12161	643252	54	0

- Garās TCP sesijās ZERO SYN fragmentē bieži (169 no 175)
- Īsās TCP sesijās ZERO SYN fragmentē reti (1 no 54)

ZERO realizācija LU MII mākonī

- Aktīvs Tīmekļa serveris – aiz lokāla tuneļa
  - Transporta tīkls – OS tilts (*bridge*)
- Lieto nelielu (1-10MB) statisku objektu lejupielādei
- Parasti – īslaicīgas HTTP sesijas

ZERO realizācija LU MII mākonī

- Statistika par 1 stundu:

Scenario	Sessions	Total sessn. time,s	Packets	Bytes	SYNs	Fragd SYNs	Frags/ packets	Frags/ SYNs
Tunn. to client	3164	53331	9852 k	14317 MB	6666	390	0,0040%	**5,85%**
Tunn. to server	3364	57416	4599 k	284 MB	7107	166	0,0036%	**2,33%**

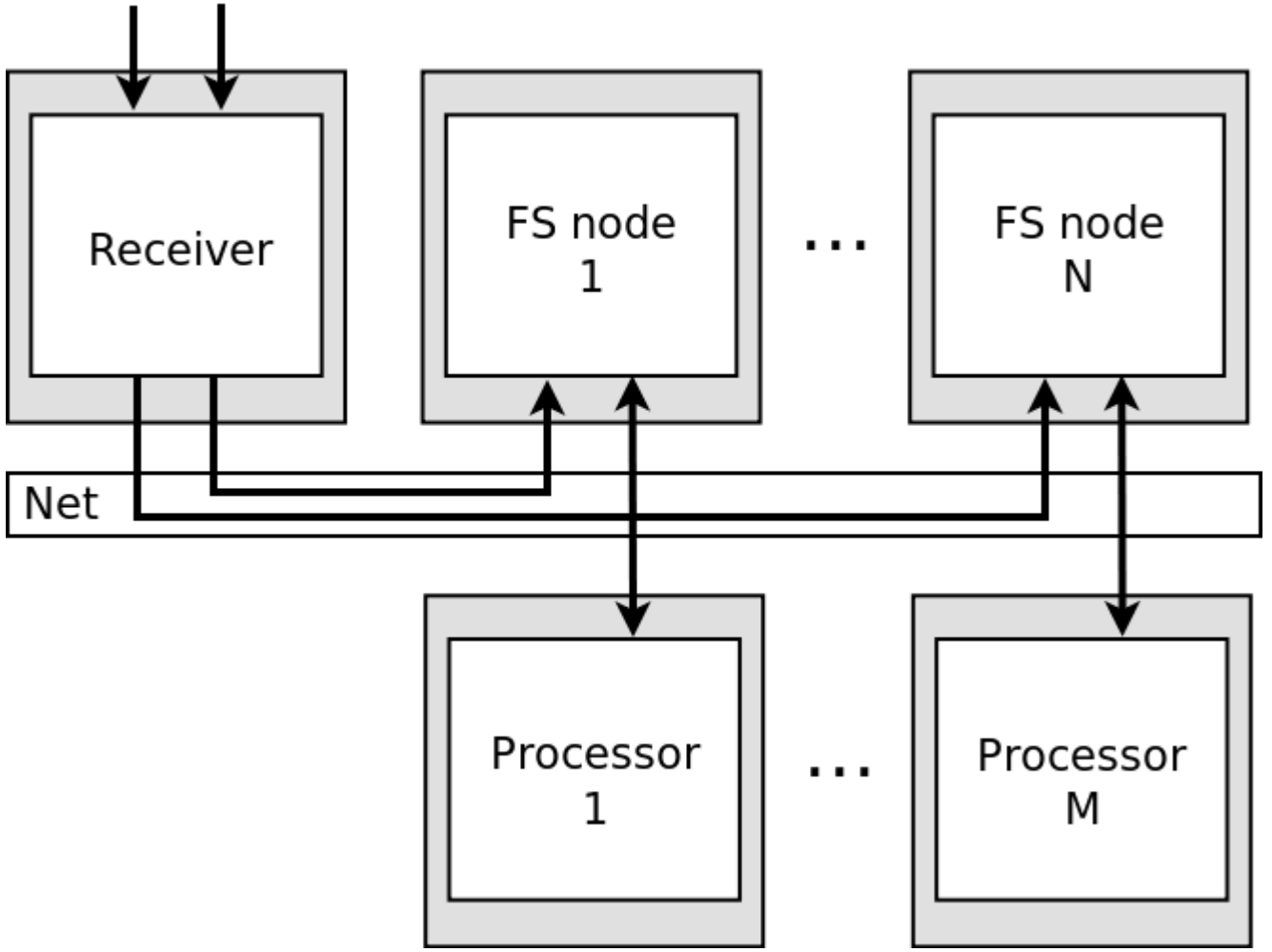
- Pat īpašās SYN paketes tiek fragmentētas **2-6%**
- No kopējā NICE pakešu skaita tie ir ~**$4e^{-5}$**

Dinamiska astronomisko datu apstrāde

- Hibrīda sistēma:
  - jēlie dati tiek sinhroni glabāti pagaidu atmiņā,
  - apstrādi veic asinhroni
- Apstrādei “strādniekus” piekārto dinamiski pēc jaudas nepieciešamības
  - (pagaidu atmiņas “uzpildīšanās”)
- Izpētītas piemērotas sadalītās datņu sistēmas
  - Lustre ir piemērotākā – ātra, pieejama, neprasa papildus mehānismus (SAN, *fencing*)

Dinamiska astronomisko datu apstrāde

- Sistēmas arhitektūra



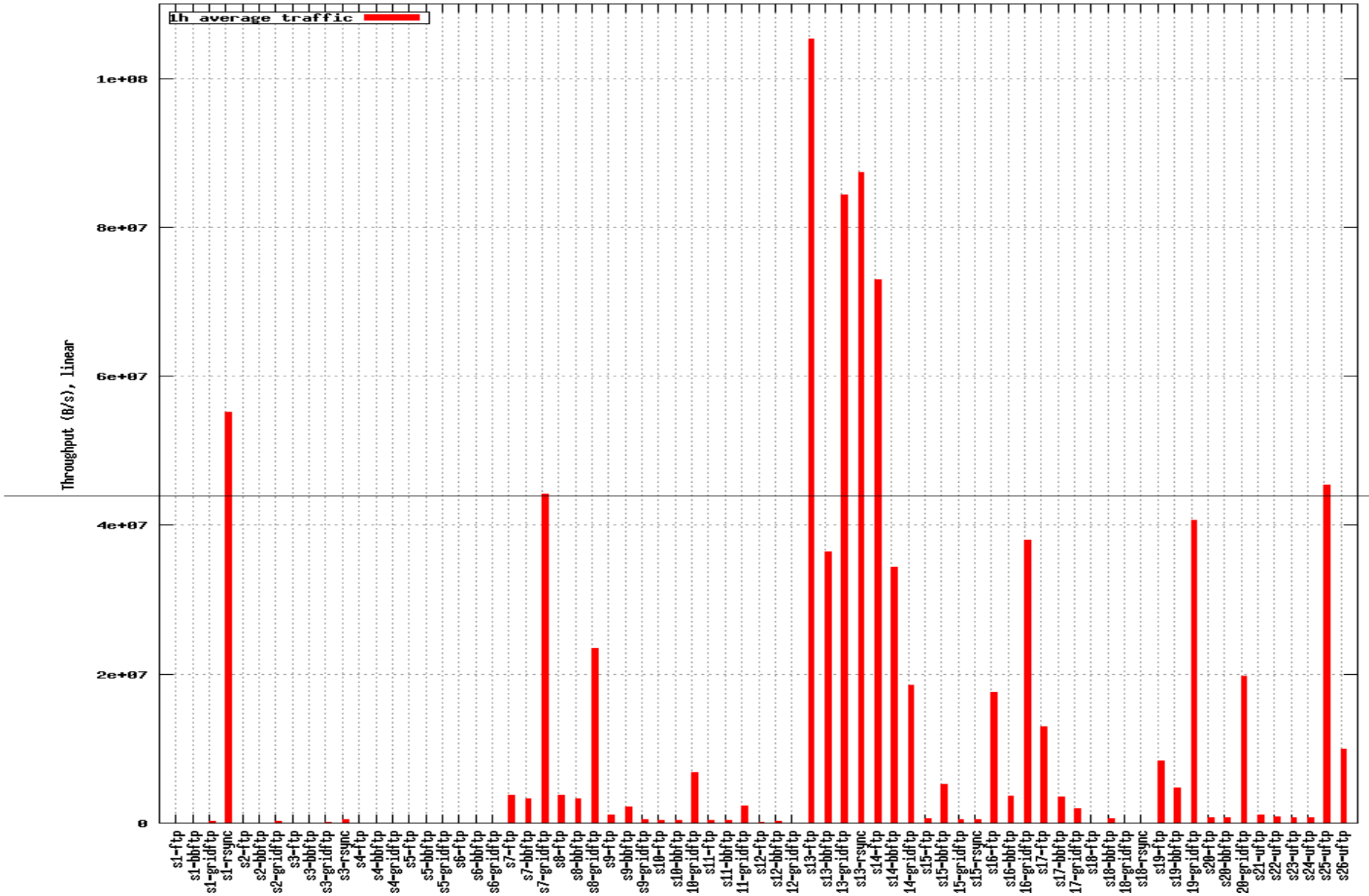
Protokolu veiktspēja meteoroloģijas datiem

- EUMETSAT trešās paaudzes programmā nepieciešama 300-400Mb/s pastāvīga datu plūsma ģeogrāfiski lielos attālumos
- **Jautājums**: vai datu pārraides protokoli nebūs šķērslis?
- Vairāki scenāriju apstākļi:
  - Aiztures, pakešu zudumi, failu kopas
- **Atbilde**: virszemes (70ms RTT) pārraidei labākais ir RSYNC
  - pat ar mazām datnēm ir 440Mbps

Protokolu veiktspēja meteoroloģijas datiem

Scenarios	***Category: file sizes***	***RTT***	***Packet loss rate***
1,2,3,4,5,6	Cat 1: 10kB	1,2,3: 70ms 4,5,6: 700ms	1,4: 0 2,5: 10^{-6} 3,6: 10^{-3}
7,8,9,10,11,12	Cat 2: 5MB	7,8,9: 70ms 10,11,12: 700ms	7,10: 0 8,11: 10^{-6} 9,12: 10^{-3}
13,14,15,16,17,18	Cat 3: 2GB	13,14,15: 70ms 16,17,18: 700ms	13,16: 0 14,17: 10^{-6} 15,18: 10^{-3}
19,20	Cat 4: mix of 10kB, 512kB, 5MB, 50MB, 2GB	19: 70ms 20: 700ms	0
21,22,23,24	Cat 5: mix of 10kB, 512kB, 5MB, 50MB, 2GB	21,22: 70ms 23,24: 700ms	21,23: 10^{-6} 22,24: 10^{-3}
25,26	Cat 6: 2GB	25: 70ms 26: 700ms	10^{-3}

Protokolu veiktspēja meteoroloģijas datiem

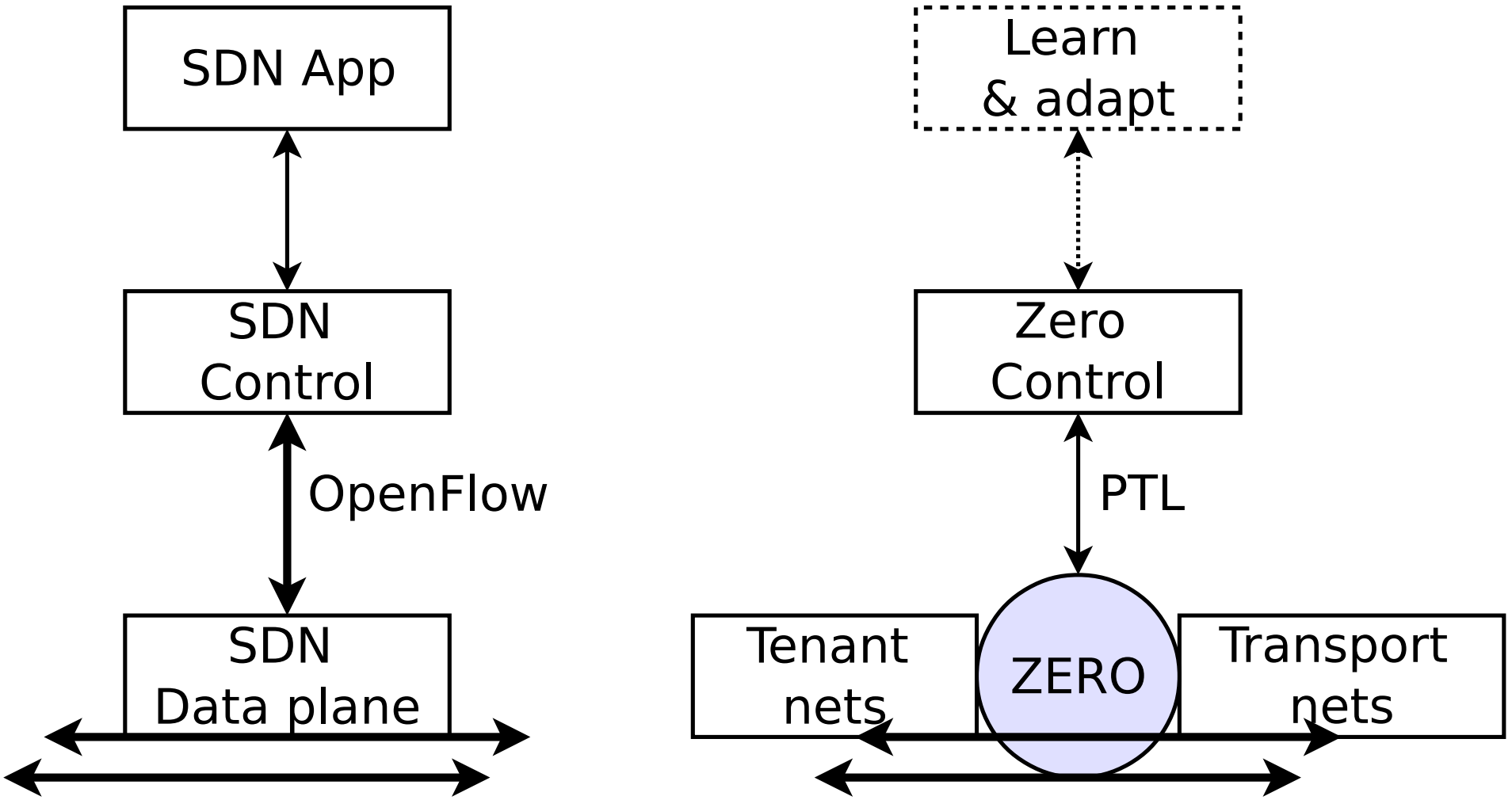


43MB/s

Ievērojami rezultāti

- Highest performance – FTP in scenario 13: 105MBps (**840Mbps**);
- For small files (10kB) only RSYNC reached target (scenario 1: 55MBps (**440Mbps**));
- At worst packet loss (10^{-3}) only UFTP reached target (scenario 25: 45MBps (**360Mbps**));
- At 700ms RTT only GridFTP came close to target (best case – GridFTP in scenario 16: 38MBps (**304Mbps**)).

ZERO SDN kontekstā



- IEEE EDUCON - Global Engineering Education 2015

Remote Security Labs in The Cloud

ReSeLa

Anders Carlsson
BTH
Sweden
aca@bth.se

Leo Truksans
IMCS UL
Latvia
leo.truksans@lumii.lv

Rune Gustavsson
BTH
Sweden
rgu@bth.se

Martens Balodis
IMCS UL
Latvia
martins.balodis@lumii.lv

***Abstract*—The paper describes on-going work on a configurable network based experimental platform ReSeLa. The platform is a key component of the ongoing EU funded TEMPUS project ENGENSEC. The project is aiming at providing courses and training material to educate future generation of Cyber security experts. Our project is based on the educational Framework; Conceive, Design, Implement, Operate (CDIO[1]). The paper outlines the background of earlier efforts on similar platforms in Sweden and Latvia. In the paper we also compares our approach with lessons learned in international projects such as PlanetLab, EmuLab, and GENI. Our approach of Cloud based environments is based on recent advanced in distributed computing such as OpenStack. Some identified and addressed challenges care presented. We also present the architecture of ReSeLa as well as its basic functionalities. The paper ends with a section on Future work and a short list of references**

Keywords— Cloud Based Security Engineering, Open Stack, learning and training, international cooperation

I. BACKGROUND

We are partners in the on going EU Project ENGENSEC[2] *Educating the Next generation experts in Cyber Security* (Engensec.eu). The main objective of the ENGENSEC project EU universities, ECTS grading and mutual degree recognition.

A *crucial component* in the project is to design, develop and implement sustainable experimental environments. That is, our Cloud based Remote Security Labs (ReSe-La).

The ReSeLA architecture is a result of earlier efforts by the partners, for example at Blekinge Institute of Technology (BTH) and Institute of Mathematics and Computer Science of University of Latvia (IMCS LU).

In 2006, Blekinge Institute of Technology (BTH) deci-ded to create a remotely accessible laboratory supporting advanced security exercises and training (*OpenLabs Security laboratory*). In the laboratory, students were able to experiment with insecure protocols, software vulnerabilities and other harmful software, e.g. viruses, in a safe and isolated environment.

This remotely accessible security laboratory was the