1off.it Technical Presentation

MikroTik ISP Datacenter Architecture

06 mar 2025





About me Alessandro Campanella

I have been working in the telecommunications field for over 20 years as a producer, consultant, and trainer, both in Italy and abroad.

I was among the pioneers of wireless technology, conducting experiments on my network and gathering the experiences of the many operators and integrators with whom I have collaborated over the years.

MikroTik Evangelist Design, analysis, and consulting for ISPs MikroTik Training and Certification









3

MUM 2006

I was there (here)

Returning to the place of the first MUM after 18 years, reconnecting with friends and colleagues, evokes a mix of nostalgia and joy.

It's a powerful reminder of the journey undertaken and the enduring bonds formed.

This reunion is both a reflection on past experiences and a celebration of growth and enduring connections.

I was there Gandalf. I was there 3000 years ago...







 $\widehat{\mathbf{M}}$

4

Riga 2007 - 1° Train the Trainer







Module 1 Introduction





My Journey in ISP Network Design

My Story: From Simplicity to a Robust Architecture

When I started working with ISPs, MikroTik wasn't the first name that came to mind for building a high-performance data center.

The common belief was that only high-end vendors could provide reliable solutions for **scalability**, **redundancy**, and **performance**.

But over the years, I refined my approach, continuously testing and adapting my designs.

I saw MikroTik evolve rapidly, bringing powerful features that — when used correctly — transformed the way I built ISP infrastructures.



My Journey in ISP Network Design

Evolving ISP Networks with MikroTik

- From basic routing to full Layer 3 hardware offloading.
- From traditional STP-based failover to fast and resilient MLAG setups.
- From simple VLAN segmentation to scalable, **ISP-grade network** architectures.

MikroTik was no longer just an affordable alternative, it became a key enabler for ISP data centers.



My Journey in ISP Network Design

My Approach

The architecture I propose is the result of this experience: a **scalable**, **resilient**, and **high-performance** approach tailored to ISP needs.

A well-designed MikroTik-based network can rival traditional enterprise solutions... **if you know how to use its full potential.**







9

Module 2 Topology





MikroTik ISP Datacenter Architecture Designing and Managing an ISP Datacenter with MikroTik

10







11

Module 3 MLAG





MLAG: Redundancy Challenges

From Single Points of Failure to a Fully Redundant Network

One of the biggest challenges in ISP networks is ensuring redundancy without sacrificing performance.

Traditional solutions, like Spanning Tree Protocol (STP), block redundant links to prevent loops, but this limits bandwidth utilization and slows down failover in case of a failure.

Imagine a core switch failure in a traditional setup: traffic is disrupted, convergence takes time, and the impact on customers is immediate. ISPs can't afford this kind of downtime.



MLAG: Redundancy Challenges

From Single Points of Failure to a Fully Redundant Network

This is where Multi-Chassis Link Aggregation (MLAG) changes the game:

- **No More Blocked Links**. Unlike STP, MLAG allows all connections to be active, fully utilizing available bandwidth.
- **Instant Failover**. If a switch or link fails, the other switch takes over immediately, preventing service disruptions.
- **Better Load Balancing**. Traffic is dynamically distributed across multiple links, improving efficiency.
- **Seamless Scalability**. New devices can be integrated without major reconfiguration or downtime.



Creating a Unified, Redundant Switch Architecture

MLAG (Multi-Chassis Link Aggregation Group) allows two separate MikroTik switches to function as a single logical switch from the perspective of connected devices.

- Devices connected via LACP bonds believe they are linked to a single switch, ensuring uninterrupted operation even if one switch fails.
- Unlike STP, all links remain active, fully utilizing available bandwidth.
- The two switches continuously synchronize their MAC address tables to maintain seamless traffic forwarding.

14



MLAG: How It Works in RouterOS

How MLAG is Constructed in MikroTik

Two CRS3xx, CRS5xx, CCR2116, or CCR2216 devices form an MLAG pair using RouterOS v7.

The two switches communicate over a dedicated peer port using Inter Chassis Control Protocol (ICCP), sharing MAC tables and system states.

MLAG requires STP, RSTP, or MSTP enabled, and both switches must have the same STP priority and configuration.





MLAG: How It Works in RouterOS

How MLAG is Constructed in MikroTik

Peer communication should happen over a dedicated untagged VLAN, isolated from the rest of the network using VLAN filtering.

MLAG supports LACP bonding on peer ports, allowing connected devices to distribute traffic across both switches.

The MLAG is not compatible with L3 hardware offloading. When using MLAG, the L3 hardware offloading must be disabled... but only on switches!









Bonding 802.3ad



tagged vlan 101

untagged vlan 4000



/interface bonding
add mlag-id=10 mode=802.3ad name=client-bond slaves=sfp-sfpplus2

/interface bridge
add name=bridge1 vlan-filtering=yes

/interface bridge port
add bridge=bridge1 interface=sfp-sfpplus1-ICCP pvid=4000
add bridge=bridge1 interface=client-bond

/interface bridge vlan
add bridge=bridge1 tagged=sfp-sfpplus1-ICCP vlan-ids=1
add bridge=bridge1 tagged=sfp-sfpplus1-ICCP,client-bond vlan-ids=101

/interface bridge mlag
set bridge=bridge1 peer-port=sfp-sfpplus1-ICCP

/interface bonding
add mlag-id=10 mode=802.3ad name=client-bond slaves=sfp-sfpplus2

/interface bridge
add name=bridge1 vlan-filtering=yes

/interface bridge port
add bridge=bridge1 interface=sfp-sfpplus1-ICCP pvid=4000
add bridge=bridge1 interface=client-bond

/interface bridge vlan
add bridge=bridge1 tagged=sfp-sfpplus1-ICCP vlan-ids=1
add bridge=bridge1 tagged=sfp-sfpplus1-ICCP,client-bond vlan-ids=101

/interface bridge mlag
set bridge=bridge1 peer-port=sfp-sfpplus1-ICCP

MLAG

Switch 1



Switch 2

🕅 тікготік	Workspac	e: <none> > 🖯 4</none>	Q												😚 New WinBo	x road	🔊 Safe Mode 💈	; ≡
- 🎢 Quick Set	🛄 Interfa	ace List 👻 Interface	Interfac Ether	ne EolP Tur IP Tu	inni GRE Tur VL	AN VXLAN VRRP	VETH МАС	sec MACVL/	Bondinç LTE 💩 🗙	}-{ Bri	dge ~ Brid	ge Ports Port Exten	s VLANS MSTIS Port	MSTOV Filters NAT	Hosts MDB 🛛 🗙	}-{ MLAG		
🤶 WiFi	다 New	Enable 🕕 Disable	Remove				Q Find	Y Filter 💷	© Configuration	Ct Nev	v 🕟 Enable	🕕 Disable Remov		९ Find 🍸 Filter 🔲	© Configuration	Bridge	bridge	
🛄 Interfaces	Туре		~ Bonding					× Disable	Detect Internet	P	Bridge	∧ VLAN IDs	Current Tagged	Current Untagged ≡	MVRP Attributes	Deer Port		
👽 WireGuard		Name	^ Type	Actual MTU	L2 MTU Tx	F	8x			added	by pvid					Priority	129	
🕂 Bridge	S	bonding122	Bonding	9180	9570	0 bps		0 bps			bridge		1000	bridge, karsolink-02		FIGILY	120	
로 PPP	s	w bonding123	Bonding	9180	9570	O bps		0 bps		added	bridge		ICCP			Status	connected	
Switch >	> 📄 S	 bonding124 	Bonding	9180	9570	0 bps		0 bps		D	bridge	4094		ICCP		System Id	F4:1E:57:9A:29:7	3
🟷 Mesh	S	⇔ bonding131-Pr.	Bonding	9180	9570	0 bps		0 bps		CEPH						, Active Role	secondary	
v <u>4</u> IP →	, 🔲 RS	↔ bonding132-Pr.	Bonding	9180	9570	6.1 Mbps		2.3 Mbps			bridge	4090	ICCP, karsolink-02.net					
v v∮lPv6 >	, 🔲 RS	⇔ bonding133-Pr.	Bonding	9180	9570	5.6 Mbps		23.6 Mbps			bridge	3101, 3111, 3122,	ICCP, karsolink-02.net			Cancel	Apply	ОК
° ⊘ MPLS >	, 🔲 S	bonding134	Bonding	9180	9570	0 bps		0 bps			bridge	3201, 3211, 3212,	ICCP, karsolink-02.net				PD.	
	, s	bonding141	Bonding	9180	9570	0 bps		0 bps										
	S	♦ bonding142	Bonding	9180	9570	0 bps		0 bps									Model CRS520-4	XS-16XC
	Í S	♦ bonding143	Bonding	9180	9570	0 bps		0 bps								Serial N		MG10
	S	♦ bonding144	Bonding	9180	9570	0 bps		0 bps			all	VI ANs ta	agged on I					
😕 Dot1X	RS	(ICCP	Bonding	9180	9570	9.1 kbps		18.5 Mbps			un					Firmwar	e Type al64B	
Files	RS	 karsolink-02.ne 	Bonding	9180	9570	23.1 Mbps		9.0 Mbps								Factory Fire	mware 7.16.1	
E Log																Current Fir	mware 7.17.2	
🎦 New Terminal																Upgrade Fir	mware 7.17.2	
🗞 RADIUS								▼ 13 85						6				
Z Tools >	> 🕺 💾 MLAG	3-B 10.82.8.218 / arm64	4 / CRS520-4XS-1	16XQ / 7.17.2 (stat	ole) 🗰 CPU: 0	% 📑 Memory Free	e/Used/Tota	I: 3795.1 MiB	/ 300.9 MiB / 4096.0 M	liB 🛛 U	ptime: 05:37	7:47 🕗 Date: 20	25-02-27 17:24:27					

18





 \sim

19

Can be a bonding Starting from 7.17 active-backup supports HW offload (different speeds)

Priority 0x1000

Each switch pair believes it is communicating with a single switch





Priority 0x2000



MLAG: 100G with 25G ICCP failover









21









Why VLAN Design Matters in an MLAG Setup

In a MLAG environment, proper VLAN configuration is crucial for ensuring seamless communication between devices.

Unlike traditional setups, where VLAN traffic is confined to a single switch, MLAG allows VLANs to be distributed across both switches in a redundant and scalable way.

- Both MLAG switches must have the same VLAN settings to avoid inconsistencies.
- Ensures traffic is forwarded correctly based on VLAN membership.
- Separating device to device traffic improves security and performance.

M



MikroTik ISP Datacenter Architecture Designing and Managing an ISP Datacenter with MikroTik







Optimizing Network Architecture with VLAN Segmentation

In a network architecture involving **two BGP** routers (BGP1 and BGP2), three BRAS devices, and two firewalls, implementing five distinct VLANs ensures **efficient** routing, redundancy, and security.

iBGP is used to maintain **dynamic routing** updates, with BGP1 and BGP2 acting as route reflectors to optimize scalability and avoid a full-mesh iBGP configuration.





iBGP Synchronization Between BGP1 and BGP2

This VLAN establishes a direct Layer 2 connection between BGP1 and BGP2, facilitating internal BGP (iBGP) sessions for routing table synchronization.

Assign dedicated interfaces on both routers to this VLAN, configure IP addresses on these interfaces, and establish both IPv4 and IPv6 iBGP peering to ensure consistent routing information.





Connectivity from BGPs to BRAS Devices and Firewalls

This VLANs connect both BGP routers to all three BRAS devices and to the two Firewalls, allowing the BRAS and the Firewalls to establish iBGP sessions with BGP1 and BGP2 on their loopback addresses.

Since BGP1 and BGP2 are configured as route reflectors, the BRAS devices and Firewalls do not need a full-mesh iBGP connection.

OSPF is enabled on this VLANs but only to distribute loopback IP addresses (IPv4 and IPv6).

This configuration ensures that customer and Datacenter routes are propagated efficiently and BRASs and Firewalls obtain only default routes.







LACP

EDGE port, VLAN ingress filtering (admit only untagged), PVID of the Peering VLAN

/interface bonding

add mlag-id=20 mode=802.3ad name=peering-bond slaves=sfp-sfpplus3 add mlag-id=30 mode=802.3ad name=bgp1-bond slaves=sfp-sfpplus4 add mlag-id=40 mode=802.3ad name=bgp2-bond slaves=sfp-sfpplus5

/interface bridge port add bridge=bridge edge=ves frame-types=admit-only-untagged-and-priority-tagged $\$ interface=peering-bond pvid=101

add bridge=bridge interface=bgp1-bond add bridge=bridge interface=bgp2-bond

/interface bridge vlan add bridge=bridge1 tagged=sfp-sfpplus1-ICCP,peering-bond,bgp1-bond,bgp2-bond \ vlan-ids=101

Peering VLAN untagged on Switches EDGE ports tagged on ICCP and bonded interfaces

/interface bonding

add mlag-id=20 mode=802.3ad name=peering-bond slaves=sfp-sfpplus3 add mlag-id=30 mode=802.3ad name=bgp1-bond slaves=sfp-sfpplus4 add mlag-id=40 mode=802.3ad name=bgp2-bond slaves=sfp-sfpplus5

/interface bridge port

add bridge=bridge edge=yes frame-types=admit-only-untagged-and-priority-tagged \ interface=peering-bond pvid=101 add bridge=bridge interface=bgp1-bond add bridge=bridge interface=bgp2-bond

/interface bridge vlan add bridge=bridge1 tagged=sfp-sfpplus1-ICCP, peering-bond, bgp1-bond, bgp2-bond \ vlan-ids=101

Peering

BGP

MLAG



Configuring Peering Interfaces

In a peering setup, the configuration of physical interfaces plays a critical role in ensuring stability, security, and flexibility. Peering connections can be established in two main ways:

- Using LACP (Link Aggregation Control Protocol) When multiple links are available, LACP can be used to bundle them into a single logical interface, increasing bandwidth and providing failover capabilities.
- **Single Port Assignments** If different peering partners need to be assigned to specific switches, each connection can be terminated on separate interfaces, distributing the load across the MLAG pair.





Configuring Peering Ports in Edge Mode

To maintain stability and security, peering ports should be set as Edge Ports (also called PortFast in other networking vendors).

- **No STP Convergence Delays** Edge ports immediately transition to forwarding state without going through STP listening/learning phases, reducing downtime when links are flapped.
- No Risk of Loops Since peering ports connect directly to routers, they do not need to participate in STP calculations. If deemed appropriate, BPDU Guard can be enabled.





Filtering VLANs and Allowing Only Untagged Traffic

For security and operational consistency, ingress VLAN filtering should be applied, ensuring that only untagged traffic is allowed on peering interfaces.

This prevents unauthorized VLANs from being propagated across the network and ensures that each peering session is handled in its designated VLAN.

- Peering VLAN traffic should be untagged to avoid inconsistencies between our network and the peering partner.
- VLAN filtering prevents accidental trunking of VLANs that could interfere with traffic segmentation.











VL

VLAN Design

Configuring Interfaces for Wholesale Delivery Kits

In wholesale service delivery, the provider may hand over VLANs in two ways:

- **QinQ (802.1ad)** Where the Service VLAN (S-VLAN) encapsulates the Customer VLAN (C-VLAN), commonly used for wholesale access.
- **Double 802.1Q VLAN (802.1q)** Where only a single VLAN tag type is assigned, either for the service or the customer.

Our objective is to transport these VLANs transparently to the BRAS without altering their structure, allowing correct end-to-end VLAN mapping for the service provider and the customer.

To achieve this, we implement Tag-Stacking and accept only tagged packets.



Module 5 L3 Hardware Offloading



MikroTik ISP Datacenter Architecture Designing and Managing an ISP Datacenter with MikroTik R



33





L3 Hardware Offloading

Enhancing Routing Performance

Layer 3 Hardware Offloading (L3HW) enables MikroTik devices to **delegate routing tasks from the main CPU to a dedicated switch chip**.

This transition significantly boosts network performance, allowing for highspeed data transmission that would otherwise strain the CPU.

- Increased Throughput
- Reduced Latency
- Optimized CPU Utilization
- Improved Scalability





35

L3 Hardware Offloading

Prerequisites

L3HW is available on devices equipped with specific switch chips (CRS3xx) Series Switches, CRS5xx Series Switches, CCR2116 and CCR2216 Routers).

Some functionalities, such as **MLAG**, **VRF** and **VRRP**, are incompatible with L3HW and must be disabled to utilize hardware offloading.

These three features are essential, can L3HW still be used?



L3 Hardware Offloading

Тір

Visualize your router as comprising two components:

- **Switch Chipset**: Manages Layer 2 (L2) traffic, functioning similarly to a dedicated switch.
- **Router CPU**: Handles Layer 3 (L3) traffic, performing routing operations.

These elements are integrated within a single chassis, collaborating to provide both switching and routing functionalities.





L3 Hardware Offloading

L3HW inter-VLAN routing

- Configure a bridge to serve as the central point for traffic management.
- Add only the bonding interface to the bridge.
- Activate VLAN filtering on the bridge.
- Assign VLAN tags to both the bonding interface and the bridge
- Define VLAN interfaces on the bridge.
- Enable L3HW.





🕅 тікготік

🎾 Quick Set 훗 WiFi

🗔 Interfaces 🤨 WireGuard

F Bridge 🖳 PPP JII Switch 창 Mesh

부 IPv6 🖉 MPLS 🔀 Routing

🔅 System

🕢 Queues

💀 Dot1X

🖧 RADIUS

者 Tools 🔇 Partition

📋 Log New Terminal Workspace: <none> < 🗇 5 📿 🛄 Interface List 🖌 Interface

🗅 New 🕟 Enable 🕕 Disable

FP Tx/Rx Packet Rate 2

Tx/Rx Bytes 4

Tx/Rx Packets 3

Tx/Rx Drops 0 Tx Queue Drops

Tx/Rx Errors

🖌 R

}-{ ∃ef Bridge

Byte Graph

Packet Graph

🗇 Cancel

MMmmmmh

BGP1-WT-NM-IX 00:00:5E:00:

MikroTik ISP Datacenter Architecture Designing and Managing an ISP Datacenter with MikroTik



38

L3H Sim

	🕅 тікготік	Workspace: <none> ~</none>	∄ୁ ଦ							Ć	👌 New WinBox	∽ ~ ⊗ s	afe Mode 💈
	🔧 Quick Set	🛄 Interface List 🔽 🛛	nterface Interfac	ce List Ethernet	EolP Tunnel	IP Tunnel GRE 1	Funnel VLAN V	KLAN VRRP VETH MACsec	MACVLAN Bonding	AMT LTE			
	😤 WiFi	🗅 New 🕞 Enable 🕕	Disable 🔀 Remo	ove							Q FI	ind 🍸 Filter 🛄	
	Interfaces	■ P Name	^ Туре		Actual MTU	L2 MTU Tx		TX Pa	acket Rx Pack	et FP Tx	FP F	Rx ≡	
	WireGuard	🔲 R 🛛 — 💥 bric	ge Bridge		1500	1584	9.1 Gbps	9.1 Gbps	755 760	755 943	9.1 Gbps	9.1	Detect interne
		🔽 R 🗆 🗤	lan10-ibgp VLAN			1580			755 756		9.1 Gbps		
	토 PPP	R (∞) \	lan20-FW VLAN		1500	1580	1216 bps	1024 bps			0 bps		
	釽 Switch >	R (@)\	lan30-PP VLAN		1500	1580	784 bps	9.0 Gbps		747 968	0 bps	9.0	
	🐎 Mesh												
	₩ IP >	- F Bridge		bridgo	ß					<u>љ</u>	Duritati	euitek 1	
1 . •	V [®] IPv6 >	-11 bridge		C RUNNING SLAVE PAS	SSTHROUGH INACTI			DISABLED DYNA		ROUGH INACTIVE		INVALID	
Ilation	X Routing	ED Ty /Dy Dool	ot Data 755 7/6 r			General	ED	Tx/Dx Dacket Date 755 7/8 n		General		Name Quite	b1
ulation	System >			uis 1 75560	02 p/s	STP				Loop Protec	1	Type Marve	ell 98DX3255
	Queues	Tx/R	x Bytes 95.6 GiB	/ 79.7 G	ЭiВ	VLAN		Tx/Rx Bytes 407.1 GiB	/ 377.3 MiB	Status	Mirro	or Target +	UI 300X3233
	- Dot1X	Tx/Rx F	Packets 68 2 39 01	10 / 57626	6 405	Status		Tx/Rx Packets 288 624 2	91 / 1053196	Traffic			
	Files	Tx/R	x Drops 0			Traffic		Tx/Rx Drops 0		4 Actions	RSPAN Ingres	s Vian ID	
				KT N	lew WinBox		fe Mode 22 =	Tx Queue Drops 0		7 Actions	RSPAN Egres	s Vian ID 1	
								Tx/Rx Errors 0		Reset Traffi			
errace List Ethernet Eolp lunnel ip lunnel GRE lunnel	VLAN VXLAN VRRP VETH	I MACSEC MACVLAN I	Sonding AMT LI					× Jraph		Neset Ham	Switch	All Ports	
Remove					Q Fi	nd Y Filter 🛄	Onfiguration	n II.	[^{50.0} G	bps	Cpu Flow	Control +	
ype Actual MTU L2 MTU Tx	KX OF F	Tx Packet	Rx Packet	FP Tx	FP R	x =	Detect Internet		-40.0 G	bps	L3 Hw Of	floading	
ridge 1500 1584	12.8 KDps 95.5	KOPS 23		.51	1440 bps	30.7			30.0 G	bps	QoS Hw Of	floadir.y	
LAN 1500 1580	3.0 kbps 2.9	kbps 4		4	0 bps	10.2				ops	Cancol		Apply 0
LAN 1500 1580	6.8 kbps 9.7 (Gbps 9	802 7	28	0 bps				0 bps	ops	Calicer		
								2025-02-22 16:43:22 2025-02-2	2 16:44:222025-02-22 16:45	:37 📥 Profi	le		
								TX R		CPU	total		
bridge 🗅 c		vlan10-ib	ap (C)	2	l Switch	switch1							
YNAMIC RUNNING SLAVE PASSTHROUGH INACTIVE		ISABLED DYNAMIC RUNNING SLA	VE PASSTHROUGH INAC					t Graph					Q Find Y
General		0.010		General		No			[^{5 000 000}	p/s Nam	e U	Jsage	16.0
S 7 46 p/S STP	FF 1X/RX Facket Rate	0p/s / 8p		Loop Protect		Name Switch			3 000 000	p/s netv	vorking		10.5
GIB / 25.8 GIB VLAN	Tx/Rx Bytes	341.6 GiB / 37	7.2 MiB	Status		Type Marvel	198083255		-2 000 000	p/s ethe	ernet		5.4
24 379 / 19 503 093 Status	Tx/Rx Packets	242 212 205 / 1 C	52875	Traffic	MILLO	Poper			1 000 000	p/s bridg	jing		2.2
/ 0 Traffic	Tx/Rx Drops			7	DODANUS					kern	el		2.0
77. Anthony	Tx Queue Drops			<pre></pre>	RSPAN Ingres			025-02-22 16:45:2 2 025-02-22 .	16:44:2 2 025-02-22 16:45:57	pres	tera_dx_mac		0.5
/ 0	Tx/Rx Errors			Torch	RSPAN Egres:			Tx Packet R	x Packet	rout	na		0.4
Iorch Beast Traffia Countr	Byte Graph			Reset Tramic	Switch /	All Ports							
5.0 Mbps	JIS		f50.0 Gbps		Cpu Flow	Control +		ancel		Cance			
-4.0 Mbps			-40.0 Gbps		L3 Hw Off	floading 🔽 🛛 🚽		ory Free/Used/Total: 15.0	GiB / 1020.6 MiB / 16.0 Gi	B 🛛 Uptime: 01	32:51 () Da	e: 2025-02-22	16:45:38
-3.0 Mbps			-30.0 Gbps		QoS Hw Off	floading							
-2.0 Mbps			-20.0 Gbps		01		Annual OK						
		^7	O hos		Jancel		Арріу						
02-22 16:43:362025-02-22 16:44:36	2025-02-22 16:42:3	\$62025-02-2216:43:362025-0	2-22 16:44:36	📥 Profile				×					
Rx A		Tx Rx		CPU tota	al						L3HW	/ disa	bled
	,												
	Packet Graph						Q Find Y Filt	er	\		91 Gł	ามร มาย	n CPL
500 p/s			-5 000 000 p/s	Name	U	Isage	0.5		\mathbf{N}		J.1 G.	·p5 01	
		ll i i	-4 000 000 p/s	routing			0.5		\mathbf{N}		16 9%	CPII	load
-300 p/s			-2 000 000 p/s	romon			0.0				10.570		1000
Man Man 100 p/s			-1 000 000 p/s	prestera	_dx_mac		0.0	L3H	W enab	led			
m. M. M. M. Op/s			0 p/s	ipv6			0.0						
25-02-22 16:43:36 2025-02-22 16:44:36	2025-02-22 16:42:30	2025-02-2216:43:32025-02-2	2 16:44:36	unix			0.0	95 5	Khns o	n CPU			
Rx Packet	KT =	< Packet 🔳 Rx Packet		kernel			0.0						
									6 CPIIIA	ad			
	Cancel			Cancel			II Stop	0.57		uu			
L:C9 / arm64 / CCR2116-12G-4S+ / 7.18rc2 (testing)	0% T Memory Free/Used/J	Total: 15.0 GiB / 1035.6 Mi	B/160GiB	S Uptime: 01:31:5	50 (J) Da	te: 2025-02-22.1	6:44:37						
		2010 0101 200010 1011		01101.0	U								





MikroT



BGP1 BGP2

/interface bridge
add name=bridge protocol-mode=mstp vlan-filtering=yes

/interface bonding
add lacp-rate=1sec mode=802.3ad name=MLAG slaves=sfp-sfpplus1,sfp-sfpplus2

/interface bridge port
add bridge=bridge interface=MLAG

/interface vlan

add interface=bridge name=vlan10-ibgp vlan-id=10 add interface=bridge name=vlan20-Firewall vlan-id=20 add interface=bridge name=vlan30-BRAS vlan-id=30 add interface=bridge name=vlan101-Peering1 vlan-id=101 add interface=bridge name=vlan102-Peering2 vlan-id=102

/interface bridge vlan

add bridge=bridge comment=ibgp tagged=bridge,MLAG vlan-ids=10 add bridge=bridge comment=Firewall tagged=bridge,MLAG vlan-ids=20 add bridge=bridge comment=BRAS tagged=MLAG,bridge vlan-ids=30 add bridge=bridge comment=Peering1 tagged=MLAG,bridge vlan-ids=101 add bridge=bridge comment=Peering2 tagged=bridge,MLAG vlan-ids=102

/interface ethernet switch l3hw-settings set autorestart=yes ipv6-hw=yes

/interface ethernet switch set 0 l3-hw-offloading=yes

Same configuration on BGP2

Peering BGP

MLAG





L3 Hardware Offloading

iBGP Through Direct Connection Interface

L3HW requires the connection interface to be in the bridge to enable hardware-accelerated routing for iBGP sessions.

However, adding the interface to the bridge causes STP to block the port, preventing VLAN traffic from passing through.

Multiple Spanning Tree Protocol (MSTP) is required to allow the iBGP VLAN to traverse the interconnection interface while maintaining loop prevention and efficient traffic routing.





 $\overline{\mathbb{O}}$

41



/interface bonding add mode=802.3ad name=MLAG slaves=sfp-sfpplus1,sfp-sfpplus2 add mode=802.3ad name=BGP1-BGP2 slaves=sfp-sfpplus3,sfp-sfpplus4

/interface bridge

add name=bridge1 protocol-mode=mstp vlan-filtering=yes \
 region-name=MLAG region-revision=1

/interface bridge port
add bridge=bridge1 interface=MLAG
add bridge=bridge1 interface=BGP1-BGP2

/interface vlan

add interface=bridge1 name=vlan101 vlan-id=101 add interface=bridge1 name=vlan201 vlan-id=201 add interface=bridge1 name=vlan301 vlan-id=301 add interface=bridge1 name=vlan10-ibgp vlan-id=10

/interface bridge vlan

add bridge=bridge1 tagged=MLAG,bridge1 vlan-ids=101 add bridge=bridge1 tagged=MLAG,bridge1 vlan-ids=201 add bridge=bridge1 tagged=MLAG,bridge1 vlan-ids=301 add bridge=bridge1 tagged=BGP1-BGP2,bridge1 vlan-ids=10

/interface bridge msti

add bridge=bridge1 identifier=1 priority=0x1000 vlan-mapping=10 add bridge=bridge1 identifier=1 priority=0x8000 vlan-mapping=101,201,301

> MSTP is essential otherwise VLAN 10 must pass through the switches







42

L3 Hardware Offloading

L3HW does not support VRRP, but...

Establish VRRP using IPv6 link-local addresses.

Assign the IPv6 peering address to the VRRP interfaces on both Master and Backup routers.

On Master, configure the IPv4 address directly on the VLAN interface associated with VRRP.

On Backup, assign the IPv4 address to the VRRP interface.

This setup allows all IPv4 routes on the Master to be offloaded to hardware, optimizing performance.



🥍 Quick Set	^{火식} Route Li	st				2	III L3 Hw Settings	x ت	
훗 WiFi	🗅 New 🖻	Enable 🕕 Disable	× Remove				Error Code 0	General	Q Find Y Filter
🛄 Interfaces	Active	∽ is ∽ ye	s				State OK	Advanced	
WireGuard		Dst. Address	^ Distance	Routing Table	AS Path			Monitor	
} ∮ Bridge	DAbH	₽ 1.6.146.0/24	20	main	41327,6453,4755,9583		IPv4 Routes Total 971263		
토 PPP	DAbH	₽ 1.6.147.0/24	20	main	41327,9583		IPv4 Routes HW 28614		
釽 Switch >	DAbH	₽ 1.6.148.0/22	20	main	41327,9583		IPv4 Routes CPU 942736		
⋩ Mesh	DAbH	₽ 1.6.152.0/22	20	main	41327,9583		IPv4 Shortest HW Prefix 24	actual	
<u>v</u> ч ₽ >	DAbH	₽ 1.6.156.0/22	20	main	41327,9583		IPv4 Hosts 175		
v≞ IPv6 >	DAbH	₽ 1.6.156.0/24	20	main	41327,9583			offloaded	
MPLS >	DAbH	₽ 1.6.157.0/24	20	main	41327,9583		IPv6 Routes Total 12128	nrefives	
X: Routing	DAbH	〒 1.6.158.0/24	20	main	41327,9583		IPv6 Routes HW 322	prenxes	
System	DAbH	➡ 1.6.160.0/22	20	main	41327,9583		IPv6 Routes CPU 11805		
	DAbH		20	main	41327,1299,9583		IPv6 Shortest HW Prefix 48		
	DAbH	₽ 1.6.163.0/24	20	main	41327,3320,9583	43	IPv6 Hosts 114		
	DAbh	₽ 1.6.164.0/22	20	main	41327,9583				
	DADH		20	main	41327,3320,9583		Route Queue Size 0		
E Log	DADH	P 1.6.166.0/24	20	main	41327,6453,4755,9583		Route Queue Rate 2		
▶→ New Terminal	DADH	P 1.6.167.0/24	20	main	41327,9583		Route Process Rate 2		
🔗 RADIUS	DADH	¬ 1.6.168.0/22 ¬ 1.6.168.0/22	20	main	41327,9583				
🛃 Tools 🛛 🔪 👌	DADH	- 1.6.168.0/24	200	main	28716,6453,4755,9585		Nexthop Cap 8192		
🔇 Partition		¬ 1.6.169.0/24 ¬	200	main	28/10,0403,4/50,9585		Nexthop Usage 310		
Make Supout.rif	ПАРИ	-1.6.172.0/22	20	main	41327,1299,9900				
		- 1.6.172.0/24	20	main	41327,2914,9965		VXLAN MTU Drop Packet 0		
	ПАРН	■ 1.6.177.0/24	20	main	41327,9303				
	ПАРН		20	main	41327,1239,9303		FastTrack IPv4 Conns 0		
	ПАРН	P 1 6 183 0/24	20	main	28716 6453 4755 9583		FastTrack Queue Size 0		
	DAbH	P 1 6 184 0/22	200	main	41327 9583		FastTrack Queue Rate 0		
	DAbH		20	main	41327,3320,9583		FastTrack Process Rate 0		
	DADH	₽ 1.6.188.0/22	20	main	41327,9583		FastTrack HW Min Speed 0		
	DAbH	₽ 1.6.188.0/24	20	main	41327,9583		FastTrack HW Offloaded 0		
L3HW	DAbH	₽ 1.6.191.0/24	20	main	41327,9583		FastTrack HW Unloaded 0		
	DAbH	₽ 1.6.192.0/22	20	main	41327,9583				
	DAbH	₽ 1.6.194.0/24	20	main	41327,9583		LPM Cap 6720		
	DAbH	₽ 1.6.195.0/24	20	main	41327,9583		LPM Usage 6535		
	DAbH	₽ 1.6.196.0/22	20	main	41327,9583		LPM Bank Cap 336		1.4M rou
	DAbH	₽ 1.6.197.0/24	20	main	41327,1299,9583		LPM Bank Usage 336		
	DAbH	₽ 1.6.198.0/24	20	main	41327,1299,9583		336		
	ПАЬЦ		20		11707 6/67 / 766 0607				

BGP1-WT-NM-IX 00:00:5E:00:01:C9 / arm64 / CCR2116-12G-4S+ / 7.18rc2 (testing) 🛱 CPU: 1% 🔂 Memory Free/Used/Total: 15.0 GiB / 1041.6 MiB / 16.0 GiB 🛛 🖉 Uptime: 01:03:49 🕑 Date: 2025-02-22 16:16:31



L3 Hardware Offloading

shortest-hw-prefix

If the entire routing table does not fit into the hardware memory, partial offloading is applied.

In this scenario, the longest prefixes are hardware-offloaded, while shorter ones are redirected to the CPU. (/24 is shorter than /16)

This parameter shows the shortest route prefix (/x) that is offloaded to the hardware memory, all prefixes shorter than this are processed by the CPU.

An "ipv4-shortest-hw-prefix=0" value indicates that the entire routing table is offloaded to the hardware memory.



L3 Hardware Offloading

Efficient Management of IPv4 and IPv6 Prefixes

In high-performance networking, the ability of a router to manage a substantial number of IPv4 and IPv6 prefixes is crucial.

CCR2216 and CCR2116 models both offer Layer 3 Hardware Offloading (L3HW) capabilities to optimize routing. However, they differ significantly in their prefix handling capacities.

Model	Offloadable IPv4 Prefixes	Offloadable IPv6 Prefixes
CCR2116-12G-4S+	16,000 - 36,000	4,000 - 6,000
CCR2216-1G-12XS-2XQ	60,000 - 120,000	15,000 - 20,000

1	
(1)	Π
\bigcirc	11



0.00	18:00	16	06:00	12:00	18:00 2025-02-16 21:50:00	17 06:00	12:00
	SRC AS			м	IN MAX	AVERAGE	
	204471: Karsolink			222.01Mb	ps 5.71Gbps	1.26Gbps	
	15169: Google			150.45Mb	os 3.26Gbps	1.24Gbps	
	32934: Meta			53.31Mb	os 1.71Gbps	859.45Mbps	
	16509: Amazon.com			19.64Mb	os 3.56Gbps	733.83Mbps	
	2593: VSIX Network			4.37Mb	os 2.11Gbps	707.31Mbps	
	41327: Fiber Telecom (FT)			1.91Mb	os 1.74Gbps	457.80Mbps	
	16276: OVHcloud			11.15Mb	os 1.95Gbps	394.57Mbps	
	20940: Akamai Technologies			5.44Mb	os 1.65Gbps	313.99Mbps	
	54113: Fastly			1.91Mb	os 2.45Gbps	294.41Mbps	
	51185: MainStreaming			31.46Mb	os 1.35Gbps	261.68Mbps	
	60530: INTERCONN			5.00b	os 10.15Gbps	257.91Mbps	
	32590: Valve Corporation			 117.00b	os 5.28Gbps	206.87Mbps	
	396986: ByteDance			47.75Kb	os 433.28Mbps	146.22Mbps	



46



L3 Hardware Offloading

Efficient Management of IPv4 and IPv6 Prefixes

Identify High-Traffic Prefixes

- Monitor network traffic to pinpoint prefixes that generate significant load.
- Focus on Autonomous Systems (AS) or specific IP ranges that are critical to your operations.

Implement Routing Filters

- Utilize routing filters to control which routes are offloaded to hardware.
- Apply filters to selectively enable or disable hardware offloading based on criteria such as AS numbers or prefix lengths.



🕅 miкrотiк ^{Workspace:} «

🚀 Quick Set

i WireGuard

♥♥ IP ♥● IPv6 Ø MPLS C Routing

🕹 Tools

Mik	Desig
6	D

48

L3HW Real life	Write New Find vertice 0 Find vertice 10	till dung(C-z. pause]	V Filter 20 1 Stop 2 × vration prmet
Workspace manes ~ ⊖ 5 Q X BOP ~ Connection Templates Sessions VPN C × ⊡ Terminal	ල් New WinBox දෙ බ ද x සි Profile	Safe Mode 72 = 1260 3000 6.4 KMp3 1355 KMp3 4400 13293 2 X ↓ VLAN 9000 9556 0 0bps 1752 bps 0 3	
Cì New Enable () Disable Remove Q. Find Y. Filter [1off@:dge-01.core.karsolin	<pre>c.com) > interface/ethernet/switch/l3hw-settings/monitor</pre>	v 1 32	2 X
P Name AFI Remote AS Local Role Multih Routing T. VRF = ipv4-routes-total: ipv4-routes-total: ipv4-routes-total: ipv4-routes-total: ipv4-routes-total: ipv4-routes-total:			
catchpointCollector6 pp/6 397601 ebgppeer yes main main ipv4-shortest-ha-prefix: ipv4-shortest-ipv4-nosts: ipv4-shortest-ipv4-nosts: ipv6-shortest-ipv6-s	24 235 Viane Usage		
Sommologoons Arver ip 00332 elights link yes main main 1pv5-routes-total: \$\vert vert for the source of			
r CymruBagoons-B-v4 ip 65332 ebgprsclient yes main ipv6-shortest-hw-prefix: r CymruBagoons-B-v6 ipv6 65332 ebgorsclient yes main main			
s edge-02.core-v4 ip 204471 ibgp yes main main Fasttrack-ipv4-conns:		Traffic	
status edge-02.core-v6 ipv6 204471 ibgp yes main main tastTrack-nw=n1n-speed: status status edge-03.core-v4 ip 204471 ibgp yes main main nexthop-cap:	8192 286	20 70 00 Mbps 97 Actions	
et edge-03.core-v6 ipv6 204471 ibgp yes main main - [Q quit]D dump[c-z pause		III Stop Object Torch 2025-02-221 / 26-40 Reset Traffic Co-	unters
	Interface List Ethernet EolP Tunnel IP Tunnel GRE Tunnel VLAN VXLAN VRRP VETH MACsec MACVLAN Bonding LTE		
PiberTelecom-IPv4 ip 41327 ebgppeer main main PiberTelecom-IPv6 inv6 41327 ebgppeer main main P Name P Name	C Find Y Hitter C Find Y Hitter C Type Actual MTU L2 MTP A Rx Packet Rx Packet I	Configuration	
GOLINE-MINAP-v4 ip 202032 ebgp.peer main main	Bridge 9180 9. 56 9 Mbps 56.8 Mbps 19 373 19 388	-1000as	
s* COLINE-MINAP-v6 ipv6 202032 ebgp peer main R ew VLAN100-pr s* Google-MINAP-v4 ip 15169 ebgp peer main main R ew VLAN1001-r	er-FT VLAN 9180 9566 2015 Minor 19972 28429 dge-02.core-A VLAN 9180 9566 212.5 Mips 39.0 Mips 30.078 9250		
s Google-MINAP-v6 ipv6 15169 ebgppeer main main R (* VLAN1002-6	dge-02.core-B VLAN 9180 9566 1232 bps 6.5 kbps 1 4		
s* HE-MINAP-vid. ip 69339 ebgp peer main main R <ev th="" vlan1051-r<=""> s* HE-MINAP-vid ipv6 69339 ebgp peer main main R <ev td="" vlan1051-r<=""></ev></ev>	rosek-ul.come vlan 9178 9566 143.1 Mbps 17.4 Mbps 25.780 13.050 IINAP Vlan 1500 9566 27.9 Mbps 103.3 Mbps 9124 14.053	July Arm AM And 4000 bis	
St Messagenet-v4 ip 58082 ebgp peer main main main	C101-A VLAN 9000 9566 0 bps 7.3 kbps 0 9	-20 000 pis	
s* META-MINAP-v4-A ip 32934 ebgp peer main main s* META-MINAP-v4-B ip 32934 ebgp peer main main		32 Opis 2025-02.22.17.28.40	
	BRIDGE_01	ک × Apply	OK
c* META-MINAP-v6-B ipv6 32934 ebgp poer main main c* MINAP_RS1v4 ip 43369 ebgp rs client no main Byte Graph	DISABLED DYNAMIC RUMMIND BLAVE PASSTHROUGH INACTIVE	38/1805.1 MB/16.0 G/B SUptime: 8d11:17:46 () Date: 2025 12-22 17:28:41	
s MINAP_RS1v6 ipv6 43369 ebgprsclient no main main		STP	
s* MINAP_RS2v4 lp 43369 ebgp rs client no main s* MINAP_RS2v6 lpv6 43369 ebgp rs client no main	-600.0 Mops	VLAN	
st Prosek-01v4 ip12v 204471 ibgp yes main main		Status Traffic	
s* Prosek-01v6 ipv6 204471 ibgp yes main main s* SKYIT-v4 ip 210278 ebgppeer main main	2000.0kps	I 3H/W disabled	
st SKYT-v6 ipv6 210278 ebgppeer main main	Manun Martin Martin Martin Martin Martin 1000 Mages		
\$\phi\$ Trst-01v4 ip12v 204471 ibgp yes main main \$\phi\$ Trst-01v6 ipv6 204471 ibgp yes main 2025-02-14.06.10.55	00ps 2025-02-221715-02 2025-02-221716-02 2025-02-221717-02 2025-02-221718-02 2025-02-221718-02	Reset Traffic Counters 128 Mhnc on CD	ЛТ
s UTRS4-Karsolink ip 204471 ibgp yes main main			0
		A 2% CPI load	
s UTRSv4-B ip 64496 ebgprsclient yes main main		4.270 CF O IUdu	
PUTRSv6-A Ipv6 64496 ebgprsclient yes main main VUTRSv6-B Ipv6 64496 ebgprsclient yes main main	-40 000 p/s	L3HW enabled	
	1 Mar Mar Mar Mar Mar Mar Mar A Mar		
- Mug	Manual Market and and a second of the and the and the second seco	56.8 Mbps on CPU	
41 active BGP peers			
2025-02-14/06:13:10	2025-02-2217-15-02 2025-02-2217-16-02 2025-02-2217-17-02 2025-02-2217-18-02 2025-02-2217-18-02 2025-02-2217-19-02	U.8% CPU load	
41 D Cancel		Apply OK	
dge-01.core.karsolink.com 10.255.255.201/am64/CCR2216-1G-12XS-2X0/7.17.2 (stable) 🔅 CPU: 1% 🕕 Memory Free/Used/Total:	14.0 GB/2091.6 MB/16.0 GB SUptime: 8d 11:08 08 (2) Date: 2025-02-22 17:19:03		

Safe Mode 🐉

🕅 ШІКГОТІК Workspace: «none» 🗸 🗇 5 Q



L3 Hardware Offloading

L3 Hardware Offloading and VRF Compatibility

L3HW is not compatible with VRF, which means all user traffic must pass through the main VRF.

Despite this limitation, it is possible to maintain security and network segmentation by using a separate VRF for management traffic.

By creating a dedicated VRF for the management interface and moving all access services (like SSH, WinBox, and SNMP) to this VRF, management traffic remains isolated from user data.



50

L3 Hardware Offloading

Leveraging Layer 3 Hardware Offloading with NAT

NAT operations are typically processed by the CPU.

To offload NAT processing to hardware, connections must be eligible for FastTrack. Without FastTrack, NAT traffic remains CPU-bound.

When the HW limit of NAT entries is reached, other connections will fall back to the CPU. MikroTik's smart connection offload algorithm ensures that the connections with the most traffic are offloaded to the hardware.

CCR2116 supports up to 2.25K NAT entries, while CCR2216 supports up to 4K.

Starting from version 7.18, VXLAN can be offloaded. However, VTEPs are not supported on bonded or VLAN interfaces making it challenging to use MLAG.





 Image: Second state of the second s

be used on BRAS

because PPPoE interfaces cannot be offloaded /interface bridge port
add bridge=bridge1 interface=sfp-plus1 pvid=1000
add bridge=bridge1 interface=sfp-plus2 pvid=2000

/interface bridge vlan
add bridge=bridge1 tagged=bridge1 vlan-ids=1000
add bridge=bridge1 tagged=bridge1 vlan-ids=2000

/interface vlan
add interface=bridge1 name=vlan1000 vlan-id=1000
add interface=bridge1 name=vlan2000 vlan-id=2000

/ip firewall nat add action=masquerade chain=srcnat out-interface=vlan2000

/ip firewall filter add action=fasttrack-connection chain=forward

/interface/ethernet/switch set switch1 l3-hw-offloading=yes













HW QoS

Optimizing Traffic Flow and Performance

In ISP datacenters, Quality of Service (QoS) is essential for managing network traffic efficiently.

It ensures that critical applications like voice, video, and real-time services receive priority and low latency, while other traffic types are managed without impacting overall performance.

MikroTik implements QoS using Marvell Prestera DX switch chips, supporting advanced features from RouterOS v7.15+.



HW QoS

Key Features

- **QoS Marking**: Classifies traffic at ingress, altering headers to prioritize packets.
- **QoS Enforcement**: Manages congestion with traffic shaping and rate limiting.
- Active Queue Management (AQM): Utilizes WRED and ECN to prevent buffer overflow and maintain low latency.
- **Priority Flow Control (PFC)**: Allocates bandwidth for high-priority traffic, minimizing packet loss.





Module 7 Conclusions





56

Conclusions

MLAG and Resilience

MLAG provides high availability and redundancy by enabling dual-homing without STP blocking.

Precise VLAN and MSTP configurations are essential to avoid loops and ensure efficient traffic flow.



57

Conclusions

L3 Hardware Offloading (L3HW)

L3HW significantly enhances routing performance, but compatibility limitations exist with VRF, VRRP, and PPPoE.

Careful design is needed to leverage its benefits without compromising functionality.

Selective route offloading optimizes hardware utilization, particularly on CCR2116.



NAT and L3HW Integration

NAT offloading is supported but limited to 2.25K entries on CCR2116 and 4K on CCR2216.

A strategic approach is needed to balance FastTrack and non-FastTrack traffic, ensuring efficient resource utilization.





QoS and Traffic Prioritization

MikroTik allows the creation of advanced QoS systems suitable for complex ISP environments.

However, due to the variety of QoS mechanisms, it is crucial to evaluate and test different QoS models to find the most effective solution for specific network needs.

Deep knowledge and continuous tuning are required to balance latency, throughput, and resource usage effectively.



VXLAN Support and Future Evolution

VXLAN support has just been released in RouterOS 7.18, marking a significant step forward for network virtualization.

However, VTEPs are not supported on bonded or VLAN interfaces, complicating MLAG implementations.

This release paves the way for EVPN, indicating MikroTik's commitment to evolving towards next-generation data center networking.



Final Note: A Proven Yet Personal Approach

This network architecture is the result of my personal approach to designing ISP datacenters using MikroTik's advanced capabilities. It is built on years of hands-on experience, continuous optimization, and adapting to MikroTik's evolving features.

Although this method reflects my unique perspective.

This approach is not just theoretical, it's battle-tested and continues to evolve with each deployment, ensuring it meets the demanding requirements of ISP environments. **1off.it Technical Presentation**

Thank you Q&A



Alessandro Campanella

a.campanella@1off.it

🥑 @alessandrocampanella

ðff