CLOUDSWITCHED

FREE RESOURCE — BEST-PRACTICES GUIDE

Network Performance Monitoring Guide

SNMP, NetFlow, alerting thresholds and baseline metrics for proactive network management — detect and resolve performance issues before they impact your business.

5	99.9%	5min	85%
MONITORING PILLARS	UPTIME TARGET	CRITICAL ALERT THRESHOLD	DOWNTIME REDUCTION

SNMP · NETFLOW / SFLOW · BASELINES · ALERTING · CAPACITY PLANNING

PREPARED FOR	PREPARED BY	VERSION	FORMAT
Cloudswitched Knowledge Library	Cloudswitched Ltd.	2026 Edition	Fillable PDF

01

Monitoring fundamentals

Proactive monitoring catches problems before users report them. Without it you are reactive — discovering issues only when staff complain. A well-monitored network reduces downtime by up to 85% and lets you plan capacity from real data rather than guesswork. The five pillars below are the foundation of any production-grade monitoring stack.

Core protocols and methods

SNMP (Simple Network Management Protocol). Industry standard for device monitoring. Collects CPU, memory, interface counters and uptime from routers, switches, firewalls and servers. Always use `SNMPv3` with authentication and encryption — v1/v2c are unauthenticated cleartext.

NetFlow / sFlow / IPFIX. Traffic-flow data showing which applications, users and protocols are consuming bandwidth. Essential for capacity planning and security analysis — you cannot tune what you cannot see.

ICMP (Ping). Basic reachability and latency monitoring. Simple, but the fastest way to detect a device or link failure and confirm round-trip time to internet hops.

Syslog. Centralised collection of device logs, events and errors. Crucial for troubleshooting and security investigations — without it, every device is a black box.

SNMP traps. Asynchronous, real-time alerts pushed from devices the instant something happens (interface down, power supply fault, fan failure). Faster than polling.

Protocol cheat sheet

PROTOCOL	WHAT IT MONITORS	BEST FOR	TYPICAL TOOLS
SNMP v3	Device health, interface stats, CPU, memory	Infrastructure monitoring	PRTG, Zabbix, LibreNMS
NetFlow / sFlow	Traffic flows, bandwidth by app/user	Capacity & security analysis	PRTG, ntopng, SolarWinds
ICMP Ping	Reachability, latency, packet loss	Basic uptime monitoring	Any monitoring tool
Syslog	Device logs, events, errors	Troubleshooting, security	Graylog, ELK, Splunk
SNMP traps	Real-time alerts pushed from devices	Immediate event notification	PRTG, Zabbix, Nagios

02

Baseline metrics & thresholds

You cannot detect anomalies without knowing what normal looks like. Collect baseline data for at least **2 weeks** (ideally **4 weeks** to capture month-end peaks) before setting alert thresholds. Capture both **average** and **95th-percentile** values — the 95th tells you what the link is really sustaining at peak, not just the smoothed mean.

Recommended baseline metrics

METRIC	BASELINE	NORMAL RANGE	WARNING	CRITICAL
WAN bandwidth utilisation	4 weeks	20–40%	70% / 10 min	90% / 5 min
Switch CPU utilisation	2 weeks	5–15%	60%	80%
Firewall CPU utilisation	2 weeks	10–30%	70%	85%
Internet latency (RTT)	2 weeks	5–20 ms	50 ms	100 ms
Packet loss	2 weeks	0–0.1%	0.5%	1%
WiFi client count per AP	4 weeks	Varies by model	80% of capacity	95% of capacity
DHCP pool utilisation	2 weeks	30–60%	80%	90%
Interface errors / discards	2 weeks	0 / hour	10 / hour	100 / hour

ALERTING BEST PRACTICE

Configure alerts at two levels — **Warning** (investigate when convenient) and **Critical** (investigate immediately). Avoid alert fatigue by tuning thresholds to your environment. If an alert fires more than once a day without any action being needed, the threshold is too sensitive: either raise it, add a sustained-time qualifier, or suppress it during known maintenance windows.

WHY SUSTAINED-TIME QUALIFIERS MATTER

A WAN link briefly touching 90% during a backup burst is not a problem. The same link *sustaining* 90% for 5 minutes is. Always pair a numeric threshold with a time window (e.g. `> 70% for 10 minutes`) — this is the single biggest knob for cutting noise without losing signal.

03

Alerting & escalation

Every critical alert should have a defined response — who is notified, what they check first, and when to escalate. Document this in a runbook alongside your monitoring configuration so the on-call engineer at 02:00 is not improvising. The matrix below is a starting point; tune notification channels and response times to your SLAs.

Alert response matrix

ALERT CATEGORY	NOTIFICATION METHOD	RESPONSE TIME	ESCALATION PATH
Device down (critical)	SMS + email + Teams	5 minutes	Escalate to vendor / ISP after 15 min
Interface down	Email + Teams	15 minutes	Check physical layer, escalate if unresolved in 30 min
High bandwidth sustained	Email	30 minutes	Investigate top talker, escalate if sustained > 1 hour
High latency / packet loss	Email + Teams	15 minutes	Trace hop-by-hop, open ISP ticket if external
Security event	SMS + email	Immediate	Follow incident response plan; isolate if confirmed
Backup failure	Email	1 hour	Investigate and retry, escalate if a second failure follows

Notification design principles

Match channel to urgency. SMS / phone for critical only. Email and Teams for everything else. If everything is critical, nothing is.

Group correlated alerts. A WAN failure can fire 50 dependent alerts in seconds. Use dependency-aware suppression (PRTG dependencies, Zabbix triggers) so the on-call gets one parent alert, not 50 noisy children.

Schedule maintenance windows. Suppress non-critical alerts during planned change windows. Always re-enable them on the way out — a forgotten suppression is how outages get missed.

Auto-resolve. Every alert that fires must also clear automatically when the condition recovers. Manual acknowledgement-only workflows accumulate stale alerts no one trusts.

04

Capacity planning & reporting

Monitoring data is only useful if it drives decisions. A monthly network report turns raw metrics into board-ready evidence — uptime, peak utilisation, top consumers, alert summary, and the upgrades you need budget for. Use trend analysis to predict when upgrades will be required; do not wait until you hit capacity, by which point you are already buying in a hurry and paying expedite premiums.

What a monthly network report should include

- **Uptime percentage** per critical service, against your SLA target (typically 99.9% = 43 min / month tolerance).
- **Bandwidth utilisation** — average, peak, and 95th-percentile per WAN circuit.
- **Top bandwidth consumers** — top 10 talkers and top 10 applications from NetFlow data.
- **Alert summary** — count by severity, mean time to acknowledge, mean time to resolve.
- **Trend analysis** — growth rates that drive next quarter's capacity decisions.

Trend signals to act on

Bandwidth trends. If WAN utilisation is growing at 10% per quarter, plan an upgrade two quarters *before* you hit 80% — ISP delivery lead-times of 60–90 days are routine for higher-tier circuits.

WiFi density. As headcount and BYOD client counts grow, monitor clients per AP. Plan additional APs before users start complaining — once you are above 95% capacity, performance is already degraded.

Switch port availability. Track used vs. available ports across the stack. Order additional switches before you run out, not the day someone needs a new desk drop.

ISP SLA validation. Use monitoring data to verify your ISP is meeting contracted bandwidth, latency and uptime commitments. This is the evidence you need to claim service credits when they miss.

Need network monitoring set up?

Cloudswitched deploys and manages monitoring platforms that give you complete visibility — SNMPv3, NetFlow, syslog, dashboards and on-call alerting, with monthly reports your board can actually read.

info@cloudswitched.com

cloudswitched.com/solutions/cloud-networking

05

Monitoring setup checklist

Work through this checklist when you stand up a new monitoring platform or audit an existing one. Tick each item as it is configured and tested. The five threshold lines at the bottom capture the values you settled on after baselining — keep them with the runbook.

Platform & protocols

- [ ] **Monitoring platform deployed** on a dedicated host or VM with sufficient storage for 12 months of polling history.
- [ ] **SNMPv3 enabled** on every router, switch, firewall and server — v1/v2c disabled.
- [ ] **NetFlow / sFlow exporters** configured on every WAN-facing interface and core switch.
- [ ] **Syslog collector** receiving from every infrastructure device, retention configured to meet policy.
- [ ] **ICMP probes** running against critical external endpoints and SaaS services.

Baselines, thresholds & alerting

- [ ] **Baseline captured** for at least 2 weeks across all critical metrics before alert thresholds enabled.
- [ ] **Warning and Critical thresholds** set per device class, with sustained-time qualifiers attached.
- [ ] **Notification channels** tested end-to-end (SMS, email, Teams) and on-call rota documented.
- [ ] **Dependencies configured** so a single WAN failure does not fire 50 child alerts.
- [ ] **Runbook linked** from every critical alert; auto-resolve enabled when the condition clears.
- [ ] **Monthly report** templated and scheduled, distribution list confirmed with stakeholders.

Threshold log & sign-off (post-baseline)

WAN bandwidth (warn / crit): _____

Latency / packet loss (warn / crit): _____

CPU — switch / firewall (warn / crit): _____

PREPARED BY	DATE	APPROVED BY

CLOUDSWITCHED

CLOUD NETWORKING & MANAGED MONITORING

info@cloudswitched.com

New London House, 6 London St
London EC3R 7LP · United Kingdom