AMTSO VPN Testing Guidelines Part 1 – Performance Assessment

The cybersecurity industry's
testing standard community

Notice and Disclaimer of Liability Concerning the Use of AMTSO Documents

This document is published with the understanding that AMTSO members are supplying this information for general educational purposes only. No professional engineering or any other professional services or advice is being offered hereby. Therefore, you must use your own skill and judgment when reviewing this document and not solely rely on the information provided herein.

AMTSO believes that the information in this document is accurate as of the date of publication although it has not verified its accuracy or determined if there are any errors. Further, such information is subject to change without notice and AMTSO is under no obligation to provide any updates or corrections.

You understand and agree that this document is provided to you exclusively on an as-is basis without any representations or warranties of any kind whether express, implied or statutory. Without limiting the foregoing, AMTSO expressly disclaims all warranties of merchantability, non-infringement, continuous operation, completeness, quality, accuracy and fitness for a particular purpose.

In no event shall AMTSO be liable for any damages or losses of any kind (including, without limitation, any lost profits, lost data or business interruption) arising directly or indirectly out of any use of this document including, without limitation, any direct, indirect, special, incidental, consequential, exemplary and punitive damages regardless of whether any person or entity was advised of the possibility of such damages.

This document is protected by AMTSO's intellectual property rights and may be additionally protected by the intellectual property rights of others.

AMTSO VPN Performance Testing Guidelines

Introduction

It is widely accepted that VPN services are an important tool to increase online privacy and security. However, choosing the right provider and package can be a challenge — even though various VPN protocols are standardized, the services and feature sets provided can be wildly different. Therefore, with this series of documents, we aim to draw clear guidelines on how VPN services should be tested for a fair vendor comparison. In this first paper we will focus on performance metrics and general VPN features; analysis of additional security features included in VPN packages will be covered in a separate document.

Table of Contents

General testing setup

- Only one VPN solution should be tested on one device at any given time. Some VPN providers may impact others, even if the app is not open or active, so it is recommended that testers prepare a “Master Copy” snapshot of a system fully prepared for testing but without any VPN solution installed; then, for each solution to be tested, to start with a clean “Master Copy” and install the solution. This process will create a set of systems with VPN products installed on exactly equivalent systems.

- Testers should disclose the platforms considered in scope and the reasoning behind the selection. If possible, a wide range of platforms should be covered, i.e. Windows, MacOS, Android, iOS, Linux, browser extensions. The same platforms should be compared if comparing VPN providers.

- Network configuration (Wi-Fi, cellular or wired connections) should be disclosed in the test scope and any associated reports, especially identifying available/guaranteed bandwidth or other restrictions.

VPN testing scope

1. Main VPN Features

These are core features common to most VPN solutions, with notes on how to measure them and potential issues testers should consider.

a. Launch on boot and auto-connection tests:

A VPN service should encrypt the traffic from the moment the device is powered up until it is shut down. Therefore, the launch-on-boot and auto-connect tests should show if a VPN provider is able to establish a VPN tunnel after boot without user interaction and that all inbound/outbound traffic is protected.

i. Testing steps: Enable auto-connect → enable start on boot → restart the device → observe whether the app starts, and the VPN tunnel is established automatically.

ii. It’s normal for the operating system to send some data outside of the encrypted tunnel while booting. It is important to ensure that as soon as the user or user-level applications are able to send and receive data, that data should immediately go through a VPN tunnel. After full booting, the user should open a browser and check the IP address. This test would confirm that when user starts browsing and sending information, this already goes through a VPN tunnel.

b. Internet kill switch test:

A common use case to protect device traffic from being exposed is to enable an automatic internet kill switch, which is activated when the VPN connection drops.

i. Test 1: Enable the kill switch to shut down the internet connection → drop the VPN connection → observe whether the access is lost → observe whether there are any data leaks. Testers should analyze all traffic data using a network protocol analyzer on the endpoint to spot data leaks, or analyze the traffic from the network.
ii. Test 2: (from previous state) Reconnect to VPN -> observe if internet access comes back.

c. *Additional leak prevention control features:*

In addition to the commonly known "Internet kill switch", the VPN provider may offer other functionality for users to control the non-VPN traffic leak prevention. Some related questions include:

i. What are those features?

ii. How do they enable users to control non-VPN traffic leak prevention?

iii. Do they work as advertised?

d. Split tunneling test

Users often choose to have some applications bypass encryption while keeping the rest of the device's traffic routed through a VPN. This can be tested as follows:

i. Choose one application (e.g., browser A) that will be excluded from the VPN tunnel while leaving all other traffic (including browser B) encrypted → check the device's IP through browser A and compare it against the IP through browser B → check if applications that should send their traffic through a VPN tunnel are not leaking any data.

- Testers should analyze all traffic data using a network protocol analyzer on the endpoint, or analyze the traffic from the network.

ii. NOTE: Some Operating Systems do not support this feature, so VPN providers are unlikely to offer this feature for all platforms.

2. Speed/latency tests

One of the most important characteristics of a good VPN provider is its minimal impact on users' internet speed and latency. A common mistake is to launch a VPN app, check speed using a random online speed test tool, and write down the results. For a fair and accurate test, more sophisticated methods are required. Here are some of the most important factors to consider when performing speed tests.

a. *Environment and network factors:*

i. Device hardware should meet or exceed the minimum recommended hardware requirements for the latest OS version.

ii. The testing environment should be optimized.
  - For example, there should not be anti-malware solutions running full system scans, updates, or other resource-heavy processes running in parallel.

iii. Software running in the background (e.g., cloud storage or anti-malware programs) can have an even higher impact on network speeds than the browser. All unnecessary software should be switched off during the tests.

iv. Use the VPN app's recommended VPN server when testing for best possible performance.

v. The same device or multiple identical (in terms of software and hardware) devices should be used when testing different VPN providers, throughout the test.
  - When using cloud-hosted VMs, testers should ensure the same specifications are used throughout, and should be aware that differences in underlying hardware may affect system performance, so should always check and confirm the specifications are matching even within the same selected cloud VM profile.

vi. VPN performance highly depends on the device's CPU being fast enough to encrypt all network traffic. Therefore, all VPN speed tests should be performed in a controlled environment, where best efforts are made to ensure the device's varying resource usage cannot influence the results. In general CPU usage while running a speed test over VPN should be monitored.
  - Some VPNs are capable of maxing out the CPU or one of its cores when using a VPN protocol that does not support multithreading, such as OpenVPN. By contrast, other protocols may support and take advantage of multithreading to deliver better results using multiple CPU cores.

vii. Testers should ensure their chosen test environment (specific device-OS combination or specific VPS instance) does not have some edge-case incompatibility with any of the VPN provider-test case combinations (like CPU running at 100% and the throughput is lower than expected).
  - *Example: performing speed tests in a VM (Azure F2s_v2 (2vcpu, 4GB ram) resulted in some VPN providers underperforming and causing the VM CPU to throttle at 100%, while this issue was not reproduced in any other VM environment or consumer device.*

viii. ISP download and upload max throughput should be representative of, or higher than, throughputs to be expected in the use-case, and selection criteria should be disclosed.
    - Where limits are imposed testers should be aware that minor differences in throughput are considered to be of marginal interest.

ix. Consistent network conditions for testing must be assured.
    - Wired tests should be performed on high-throughput networks.
    - Public and crowded networks with significant interference should be avoided during wireless tests.

x. Test all VPN providers on several different ISPs if possible. If this is not viable, testers should disclose the ISP used and disclose the potential influence of the ISP selection on the VPN provider speed results.

b. Procedural and statistical factors:

i. Measure and compare download speed, upload speed and latency.

ii. Measure reference speed results without VPN each time all providers are tested to ensure equivalence in loading etc.
    - Where testers include reference measurements in results, they should emphasize results in the form of comparison to other solutions' results in similar testing; raw data for the "reference without VPN" result and absolute numbers may be provided for context, but be advised that direct comparison with reference data can be misleading in some cases and that comparison between throughput with solutions in place, or with industry averages, is more relevant.

iii. Perform a single test run per connection and gather statistically significant speed data from that run.
    - This can be automated to perform multiple speed measurements per VPN connection.

iv. Multiple speed test runs should be performed throughout the day and week to account for network load variation.
    - Tests should cover multiple workdays and weekend days.
    - Multiple runs should be done each day for each vendor to cover different parts of the day (e.g. morning run, noon run, evening run).

v. Different providers should be tested in close succession to each other (as close as possible in order to experience the same conditions).

vi. Due to spikes and anomalies in speed data, drastically higher or lower outliers can strongly impact the average on smaller data samples.
    - To mitigate this issue testers should remove extreme values or use a median value to evaluate the results fairly.

vii. For transparency and reproducibility of test methods, industry-standard speed test tools such as OOKLA, speed.cloudflare.com, nperf.com, etc. are recommended.

c. *Choosing the VPN server for testing:*

i. When measuring the network throughput of a VPN, the same location should be selected for all VPN apps. Ideally, the same city or at least the same country should be tested. If that's not possible, then the tester should explain in the report the steps they took to select the location and the decision behind choosing a specific location.
It is common for various speed testing tools to suggest the best server based on the lowest ping. However, this method of choosing the fastest server is not accurate and can eventually lead to suboptimal performance results due to the selection algorithm of the target speed server. For this reason, it is recommended that testers perform several speed tests with no VPN before the actual testing session with different target speed servers to determine the fastest ones (at least 2 different servers) for a short period of time (max for 1 day) and perform all VPN provider speed tests to those servers. The selection and selection process should be recorded as part of the test report.

ii. The most common internet connection highways should be tested to reflect the experience of the majority of users. The selection method and reasoning behind it should be disclosed in the test.
   - Same country to the same country (for example: US -> US, UK -> UK).
   - Different country highways (for example: US -> UK, UK -> US).

d. *Feature sets and configuration:*

i. When testing different VPN providers, the appropriate comparable feature sets should be identified, tested, and compared separately.

ii. For tests aiming to measure the fastest speeds available, security features unrelated to VPN tunnelling (e.g. anti-malware tools) should be disabled. In such cases, the tester should include a clear disclaimer in their report informing users that enabling additional security features is likely to significantly affect the performance of the solution.

iii. Where some solutions include additional features not present in all solutions under test, and the test compares or otherwise measures those additional features, the tester should highlight any differences observed in speeds recorded with and without additional features active.

iv. Reporting only speed results obtained with features disabled, but security efficacy measured with them enabled, can produce misleading data. In such cases the tester should include appropriate disclaimers making clear the differences in configuration and the potential impact on how the results should be understood.

e. *Use of tools:*

   i. Tools including open-source options are available to assist with automating speed measurements.
      - Examples include https://github.com/NordSecurity/VPN-Speed-Comparison-Tool

   ii. Some tools may not work for some products, or may work in different ways for different solutions. Testers should ensure any tools used are compatible with the solutions under test.

   iii. Use of third-party tools, including open-source options, should be disclosed in the report.

3. Website accessibility

Some websites block VPN users, or use bot protection that affects VPN users, so it is important to test providers' ability to ensure access and uncensored content.

a. Test prerequisites:

   i. Using incognito mode in browser.
   ii. Testing accessibility for several days.
   iii. Re-connecting to the VPN server for each test iteration.

b. Testing steps:

   i. Select locations.
   ii. Prepare a list of websites to target.
      - For example, the top 100 most visited websites globally – this can be obtained from publicly available listings, Testers should disclose their source and the reasoning behind their selection.

   iii. Connect to VPN → observe whether all websites are accessible → observe whether presentation rate of CAPTCHAs increases compared with direct (without VPN) connections.

4. Connection success and time to connect

VPN providers are expected to ensure a fast and reliable service. This includes rapid and dependable establishment of connections.

a. If a connection success metric is being measured, it should be measured through multiple tests comparing successful connections vs. all attempts to connect.

b. Time-to-connect, where recorded, should be the average time from multiple tests (for each OS separately), measured from the initiation of the attempt to connect (button clicked) to a successfully-established connection.

c. The same user-server country combinations should be tested with all VPN providers for comparison.

   i. Testing should be performed for several connection location combinations for accurate results.
   ii. Testers should explain in the report how the user-server country combination(s) was(were) chosen.
   iii. Inappropriate selection of country combinations risks giving advantages to some solutions over others.

d. Testers should be aware of the possibility that UI elements may not always represent the actual state of the connection (for example, the VPN app UI may display "Connected" when in reality the connection is still in process of being established).

   i. The goal here is to evaluate whether there is a significant difference between providers (for example, 3 seconds vs 15 seconds could be considered a significant difference, whilst the difference between 3.2 seconds and 3.3 seconds should not be considered significant).
   ii. Testers should consider the best ways this could be tested, be transparent in the methodology used, and ensure results are presented in a fair and balanced manner.

5. Resource consumption

Different VPN providers consume different amounts of the device's resources. Users expect VPN providers to ensure that resource consumption is as low as possible.

a. Resources to measure and compare include:

   i. RAM usage.
   ii. CPU usage.

b. Measure and compare full system resources (not specific apps).

    i. When measuring full-system resource usage, efforts should be made to avoid or minimize the impact of unrelated changes in system activity, such as updates or other background OS tasks.

c. All tests should be run for at least a few hours with simulated normal usage (browsing, streaming, changing VPN server).

d. To produce meaningful results, the VPN services should be tested in identical environments for all VPN vendors.

    i. Resources allocated for the testing system (either physical or virtual) should be at least equal to the recommended specifications for the tested Operating System.
    ii. Full environment and networking setup advice is included in the speed testing section above.

e. The default feature set should be used when testing performance, with some exceptions:

    i. In cases where some VPN products have anti-malware and/or other additional security features turned ON by default, and the other VPN products do not provide similar features, then the additional features should be turned OFF in all the compared VPN products to provide a fair comparison. The tester should include a clear disclaimer in their report informing users that enabling additional security features is likely to significantly affect the performance of the solution.
    ii. In cases where all products support additional security features but have different default configurations, the tester may choose either to enable those features in all VPN products, or disable them all. Testers should disclose the reasoning behind the choice.

f. All VPN product changes from the default configuration made by the tester, and also all default configuration differences between VPN products left unchanged, should be disclosed in the results.

g. Testers should disclose the methods used to measure resource consumption and changes in resource consumption; the most direct method is to track all consumption across the entire system, rather than trying to measure only specific areas.

6. Leaks (advanced test)

A VPN is intended to ensure privacy, but not all VPN providers provide additional measures to prevent data leaking from the VPN tunnel. Most common are DNS, IP, and packet leaks that happen in all connection states and network environments and between them.

a. Various types of data leak should be tested for all vendors. Types of leak test include:

   i. Test for leaks while connected to a VPN server.
      - No data packets should exit the device unencrypted (via any other network interface except the VPN interface).

   ii. Test for leaks while switching VPN servers.
      - While connected to the VPN server and upon choosing a different VPN server there will be a moment when your device will disconnect from the current and start the connection attempt to the new VPN server. During that time period while the new connection is not yet established and the previous connection is already terminated - no data should exit the device.

   iii. Test for leaks while switching between Wi-Fi hotspots.
      - When connected to a VPN server and while the device is switching Wi-Fi access points no data packets should exit the device unencrypted (via any other network interface except the VPN interface).

   iv. Test for leaks while switching from Wi-Fi to mobile data and vice versa.
      - When connected to a VPN server and while the device is switching connection interface from Wi-Fi to mobile, no data packets should exit the device unencrypted (via any other network interface except the VPN interface).

   v. Test for WebRTC leaks.
      - While connected to VPN visit https://browserleaks.com/ and perform a WebRTC test to see if the IP you see there is your local IP or the IP of a VPN server.

   vi. Test for Ipv6 leaks.
      - When connected to IPv4 VPN connection no IPv6 traffic leaves the device since it would be unencrypted.

   vii. Test for leaks tied to known cases.
      - Examples include the TunnelVision configuration – details can be found at https://www.tunnelvisionbug.com/.

   viii. Test for leaks when split tunneling features unexpectedly exclude traffic from a browser that was supposed to remain in the VPN tunnel.
      - App split tunneling splits app traffic to direct via the regular or virtual network adapter. For example, in Windows the OS provides several ways to implement DNS requesting and on top of that the application itself can implement its own implementation for that purpose. It is important

to make sure all the traffic of the app is split when implementing the split tunneling feature and there are no unexpected leaks. So, in cases where a browser that is included in the split tunneling list of apps that should exit via VPN, when the rest are allowed to exit the physical network adapter, no traffic from the browser should be leaked though the physical adapter, and that can be tested. The same applies to other apps too, but the browser case is known as having this issue in Windows.

b. When selecting a tool to test DNS leaks, the tester should provide the reasoning for their choice. Some tools may not be appropriate for all tested products.
c. Testers should avoid using tools provided by one of the vendors being tested to evaluate other vendors.
d. Best efforts should be made to include all VPN protocols across the platforms tested.

Additional VPN Evaluation topics

Some additional aspects of VPN solutions are important to evaluate, but may not be simple or indeed possible to accurately measure. In cases where several vendors are being compared, these additional evaluation topics could be taken into scope. Where no direct test or measurement is possible, supporting information for features claimed by the vendors should be investigated.

- Does the vendor provide multi-platform support?
    - *If so, which platforms are supported (Windows, MacOS (both CPU architectures), iOS, Android, Linux, Browser Extensions, Android TV, Apple TV)?*
- How many server countries, and how many server locations within those countries, are provided?
    - *The selection of available countries may also be of interest.*
    - *Note that server availability may be subject to significant change over time.*
- Does the VPN app provide a "Quick Connect" feature?
    - *A one-click option where the best server is automatically picked.*
- Is there support for "Onion over VPN"?
    - *Web traffic is sent through a network of volunteer-operated servers. Instead of taking a direct route, internet data enters the Onion network via a random relay anywhere around the globe.*
- Does the vendor provide the option of a dedicated IP address?
    - *The user can obtain a dedicated IP address which remains unchanged over time.*
- How many concurrent devices are allowed under one account?
    - *Part of "total cost of ownership", which testers may wish to report on.*
- Does the vendor support "Multi-Hop" connections (double VPN or more)?
    - *If so, testers should investigate and report on how this is implemented, and possibly also the impact on speed metrics.*

- Does vendor support "RAM only" servers?
    - *Use of such technology can improve privacy by reducing data retention.*
- Which protocols and encryption methods are supported?
- What is the vendor's logging and audit policy?
    - *For example, does the vendor promise to keep no logs of user activity, and run no security audits?*
- Does the vendor produce a regular "Transparency report"?
    - *Such reports provide useful information on potential issues such as data breaches and requests for user data from governments/law enforcement bodies.*
- Under what jurisdiction does the vendor operate?
    - *This impacts the potential for private data to be accessed by governments, law enforcement, and other third parties.*
- What is covered in the vendor's privacy policy?
    - *This also impacts the retention and availability of private data.*
- What additional security features are included in the VPN package?
    - *VPN providers often provide additional security features as a part of the VPN application, including web filtering and detection of malicious files or vulnerable software. For a full-spectrum test of VPN packages these features should be tested in depth for efficacy and performance impact. Detailed guidance for such testing will be provided by AMTSO in a separate guideline paper.*
    - *Testers focusing on core VPN features only should carefully consider their approach to enabling or disabling additional features, bearing in mind the caveats provided at various points above (see "VPN Testing Scope" sections 2-d and 5-e for examples). Testers should ensure their approach to this is applied equally across different product types, and should provide a clear statement of their policy, why it was chosen, and its potential impact on how results should be interpreted, alongside or within their test reports.*
    - *Where additional features are discussed but not tested, it should be made clear in the test report that no judgement on the quality of those features can be derived from the test results and any attempt to use the additional features could lead to changes in the measured performance.*

This document was adopted by AMTSO on DATE