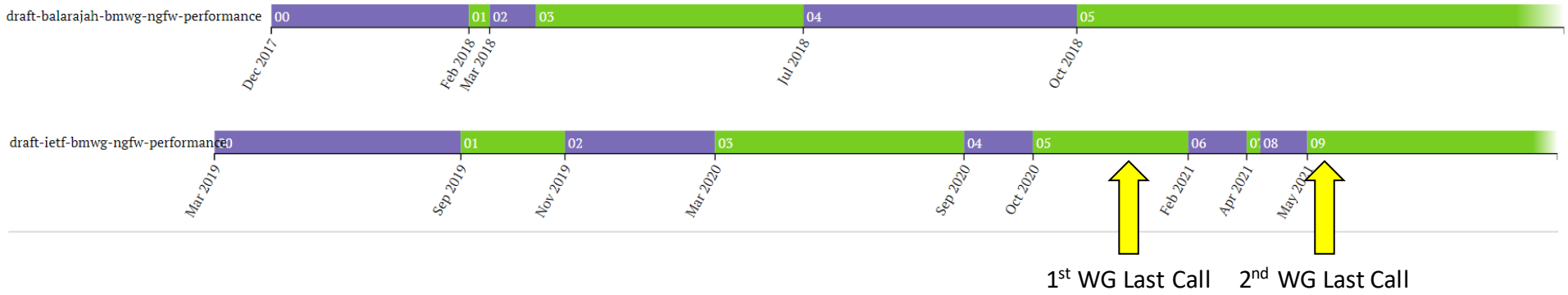
Benchmarking Methodology for Network Security Device Performance

draft-ietf-bmwg-ngfw-performance-09
IETF 111, July 26, 2021

Agenda

- Timeline
- Draft status
- Review of open comments
- Next Steps

Timeline



- 16 draft versions over 3.5 years, discussed in ten IETF Meetings and two interim meetings
- Contributors: Alex Samonte, Amritam Putatunda, Anand Vijayan, Aria Eslambolchizadeh, Baski Mohan, Chao Guo, Chris Brown, Chris Marshall, David DeSanto, Jay Lindenauer, Jurrie Van Den Breekel, Michael Shannon, Mike Deichman, Ray Vinson, Ryan Liles, Ryan Riese, Samaresh Nair, Stephen Goudreault, Tim Carlin, Tim Otto, and Toulnay Orkun
- 1st WG last call in December 2020, 2nd WG last call closed on May 21, 2021
- Received more than 150 comments/suggestions from 12 contributors during the last call(s)
- All comments resolved, except the comments received from Sarah Banks on May 20 and July 12

Review of open comments

Around 30 comments/suggestions were posted to the BMWG mailing list by Sarah Banks on May 20, 2021. Eleven of them are remaining to be resolved, since the authors' initial responses did not resolve the contributor's concerns:

- Scope definition regarding NGFW, IPS, passive devices (#1, #2, #3, #8)
- Test bed environment (#4, #5, #6, #7, #9, #10)
- Traffic mix (#11)

NOTE that none of the remaining open comments relates to the core sections of the draft – i.e. specification of benchmarking methodology

Review of open comments (1/11)

Initial comment (Banks)	Author's response	2nd Reply to Authors (by Banks)
The draft aims to replace RFC3511, but expands scope past Firewalls, to "next generation security devices". I'm not finding a definition of what a "next generation security device is", nor an exhaustive list of the devices covered in this draft. A list that includes is nice, but IMO not enough to cover what would be benchmarked here - I'd prefer to see a definition and an exhaustive list.	**1) Scope definition is desired** **2) "Next-Generation security device" is a well-known term in the industry (-> Google search analysis)** **3) We avoid limiting the draft by explicitly adding a list of NG security devices currently available in the market only. In the future, there may be more and more new types of NG security devices that will appear on the market.**	I think there are 2 types of devices called out; I'm not seeing a definition of what a "NG security device" is, and I'm not comfortable with a draft that has a blanket to encompass what would come later. Who knows what new characteristics would arrive with that new device? I think the scope here is best suited for the devices we know about today and can point to and say we're applying knowledgeable benchmarking tests against

Authors' Proposal: Define Next-generation firewall (NGFW) as: "This term is widely used for the modern, state-of-the-art technology firewalls (as of 2021) that can do application-level traffic inspection including several, sometimes optional features." (Perhaps a few of them could be listed here.)

Review of open comments (2/11)

Initial comment (Banks)	Authors' response	2nd Reply to Authors (by Banks)
The draft aims to replace RFC3511, but expands scope past Firewalls, to "next generation security devices". I'm not finding a definition of what a "next generation security device is", nor an exhaustive list of the devices covered in this draft. A list that includes is nice, but IMO not enough to cover what would be benchmarked here - I'd prefer to see a definition and an exhaustive list.	**This draft includes a list of security features that the security device can have (RFC 3511 doesn't have such a list). Also, we describe in the draft that the security devices must be configured "in-line" mode.** **We believe these two points qualifying the definition of next generation security.**	I strongly disagree. Well, I mean OK, for active inline devices maybe this is OK, but to say that the only way a device can be "NG" is to be active/inline, I disagree with. And if there is, have we gathered all of the items we'd want to actively test for in that case? For example, what about their abilities to handle traffic when a failure occurs? (fail open/closed). What about alerts and detections and the whole federation of tests around positives/false positives/false negatives, etc? I'm onboard with expanding the scope, but then we have to do the devices benchmarking justice, and I feel we're missing a lot here.

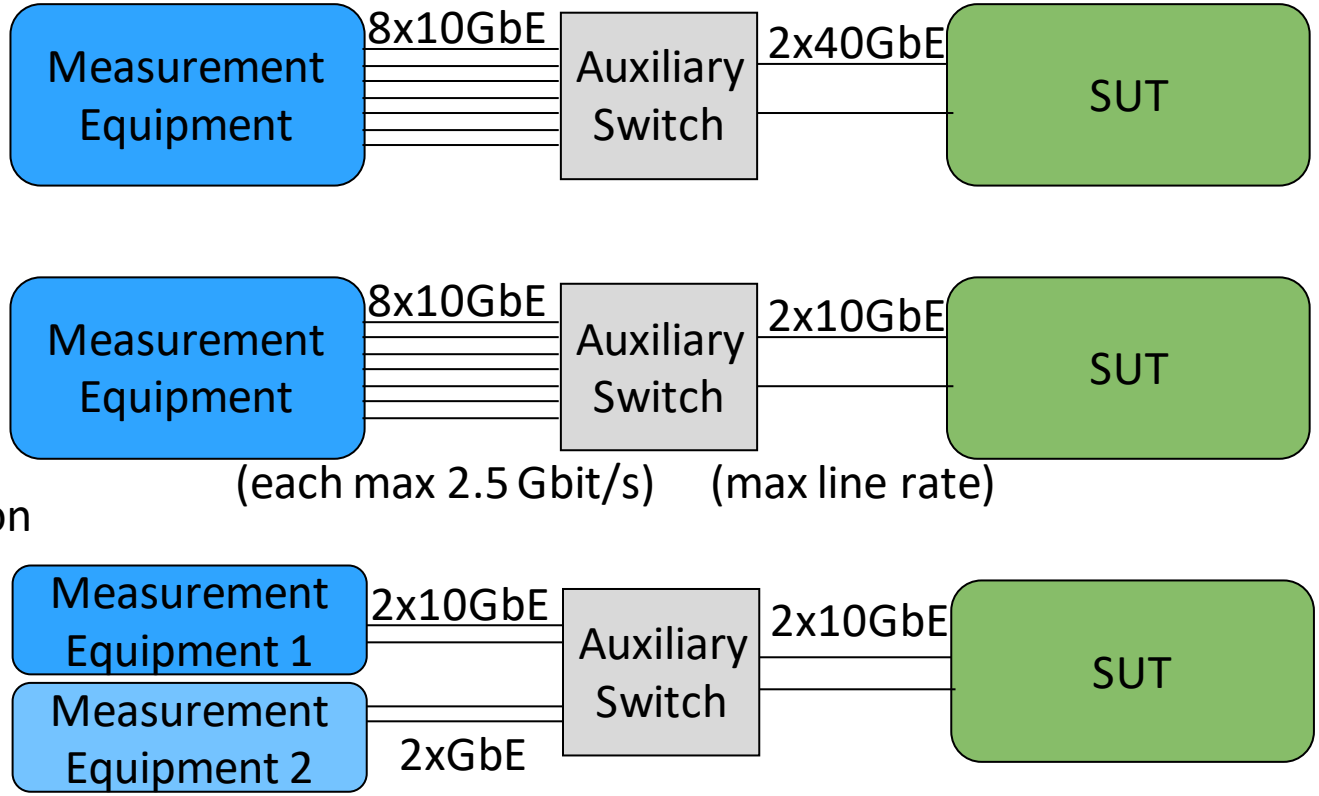
Authors' Proposal: Update list of security features that security devices can have; describe that security devices must be configured in in-line mode

Review of open comments (3/11)

Initial comment (Banks)	Authors' response	2nd Reply to Authors (by Banks)
I still have the concern I shared at the last IETF meeting, where here, we're putting active inline security devices in the same category as passive devices. On one hand, I'm not sure I'd lump these three together in the first place; on the other, active inline devices typically include additional functions to allow administrators to control what happens to packets in the case of failure, and I don't see those test cases included here.	**This draft focuses on "in-line" mode security devices only. We describe this in section 4 in more detail.** **Additionally, the draft focuses mainly on performance tests. The DUT must be configured in "fail close" mode. We will describe this under section 4. Any failure scenarios like "fail open" mode is out of scope.**	OK, but I think an RFC that is going to encompass this device under the "NG security devices" classification is missing out on large portions of what customers will want to test. It'll also beg for another draft to cover them, and then I'm not sure we're serving the industry as well as we could.
Authors' Proposal: Clarify that passive security devices are out of scope; explain more clearly that devices must be configured in „fail close" mode		

Why Add Ancillary Routers/Switches?

- Test equipment physical interfaces may not match SUT interface types
- Test equipment interfaces may be less performant than SUT interfaces, requiring link aggregation
- Test equipment roles may be distributed so that two emulators connect to one SUT



Review of open comments (4/11)

Initial comment (Banks)	Authors' response	2nd Reply to Authors (by Banks)
Section 4.1 - it reads as if ANY device in the test setup cannot contribute to network latency or throughput issues, including the DUTs - is that what you intended?	**Our intention is, if the external devices (routers and switches) are used in the test bed, they should not negatively impact DUT/SUT performance. To address this, we added a section (section 5 "Test Bed Considerations") which recommends a pre-test. We can rename this as reference test or baseline test.**	I think there's still a concern there. Who defines what "negative impact" is? You're traversing at least another L2 or L3 step in the network with each bump, which contributes some amount of latency. If they don't serve in control plane decisions and are passively passing data on, then we could consider removing them from the setup and removing the potential skew on results.
Authors' Proposal: No change to draft		

Review of open comments (5/11)

Initial comment (Banks)	Authors' response	2nd Reply to Authors (by Banks)
Section 4.1: Option 1: It'd be nice to see a specific, clean, recommended test bed. There are options for multiple emulated routers. As a tester, I expect to see a specific, proscribed test bed that I should configure and test against.	**The draft describes that Option 1 is the recommended test setup. However. We added emulated routers as optional in option 1. The reason for that: Some type of security devices for some deployment scenarios requires routers between test client/server and the DUT (e.g., NGFW) and some DUT/SUT doesn't need router (e.g. NGIPS)**	Maybe I'm missing something here - a device can't perform a function for free, right? Even if it's impact is negligible, it's an impact of some sort. We're saying the emulated router is doing the routing - OK - but I think the same thing applies to the physical router - how do you know what else the emulated router is doing? if the test gear can call out the latency, I'd like to see clarification around how it's doing that and distinguishing the latency introduced by Device A, versus Device B, versus the DUT, etc.
Authors' Proposal: No change to draft.		

Review of open comments (6/11)

Initial comment (Banks)	Authors' response	2nd Reply to Authors (by Banks)
Follow on: I'm curious as to the choice of emulated routers here. The previous test suggests you avoid routers and switches in the topo, but then there are emulated ones here. I'm curious as to what advantages you think these bring over the real deal, and, why they aren't subject to the same limitations previously described?	**Comparing with real routers, the emulated router provides more advantages for L7 testing. Emulated routers do not add unknown latency. Even if there is any added delay due to the routing process, the test equipment can report the added latency, or it can consider this for the latency measurement.**	See reply to comment #5
Authors' Proposal: No change to draft.		

Review of open comments (7/11)

Initial comment (Banks)	Authors' response	2nd Reply to Authors (by Banks)
Follow on: I'm curious as to the choice of emulated routers here. The previous test suggests you avoid routers and switches in the topo, but then there are emulated ones here. I'm curious as to what advantages you think these bring over the real deal, and, why they aren't subject to the same limitations previously described?	**Question regarding the need for routers:** **- We avoid impacting the DUT/SUT performance due to ARP or ND process** **- Represent realistic scenario (In the production environment the security devices will not be directly connected with the clients.)** **- Routing (L3 mode) is commonly used in the NG security devices.**	See reply to comment #5
Authors' Proposal: No change to draft.		

Review of open comments (8/11)

Initial comment (Banks)	Authors' response	2nd Reply to Authors (by Banks)
Table 2: With the assumption that NGIPS/IDS are required to have the features under "recommended", I disagree with this list. For example, some customers break and inspect at the tap/agg layer of the network - in this case, the feed into the NGIDS might be decrypted, and there's no need to enable SSL inspection, for example. Table 3: I disagree that an NGIDS IS REQUIRED to decrypt SSL. This behaviour might be suitable for an NGIPS, but the NGIDS is not a bump on the wire, and often isn't decrypting and re-encrypting the traffic.	**IDS is being removed.**	I'm not sure this addresses the feedback though :) A NGFW for sure will do break/inspect as well, right?
Authors' Proposal: Explicitly remove IDS from scope.		

Review of open comments (9/11)

Initial comment (Banks)	Authors' response	2nd Reply to Authors (by Banks)
4.3.1.1 - As a tester with long time experience with major test equipment manufacturers, I can't possibly begin to guess which ones of them would conform to this - or even if they'd answer these questions. How helpful is this section to the non test houses? I suggest expansion here, ideally with either covering the scope of what you expect to cover, or hopefully which (open source/generally available) test tools or emulators could be considered for use as examples.	**We extensively discussed with Ixia and Spirent about this section. This section was developed with significant input from these test tools vendors in addition to others.**	OK, that's really good to know, but there are plenty of us working with and looking for more cost effective options to Ixia and Spirent. :) I think the expansion would be good here.
Authors' Proposal: No change to draft.		

Review of open comments (10/11)

Initial comment (Banks)	Authors' response	2nd Reply to Authors (by Banks)
6.1 - I would suggest that the test report include the configuration of ancillary devices on both client/server side as well	**We believe that adding configuration of the ancillary devices doesn't add more value in the report. Instead, we recommend documenting the configuration of the ancillary devices by conducting reference tests.**	I think including them assists greatly in the repeatability of the testing, for what it's worth.
Authors' Proposal: No change to draft.		

Review of open comments (11/11)

Initial comment (Banks)	Authors' response	2nd Reply to Authors (by Banks)
7.1.3.3 - what is a "relevant application traffic mix" profile?	**This is described in section 7.1.1 (2nd paragraph). We added the word "relevant" in the 1st sentence of the 2nd paragraph.**	A set of example(s) could be helpful. Not required, just helpful.
Authors' Proposal: Modify 7.1.1, 2nd paragraph, 1st sentence: "Based on customer use case, users can choose the relevant application traffic mix for this test. The details about the traffic mix MUST be documented in the report. At least the following traffic mix details MUST be documented and reported together with the test results.		

Next Steps

- Authors to update the final draft version 10 with results of today's meeting
- Agree whether a third WG Last Call is required, and if so, how it will be scheduled

It is important to note that a Last-Call is intended as a brief, final check with the Internet community, to make sure that no important concerns have been missed or misunderstood. The Last-Call should not serve as a more general, in-depth review.

(from RFC2418 section 8)