ARTIFACT EVALUATED usenix ASSOCIATION AVAILABLE

ARTIFACT EVALUATED usenix ASSOCIATION FUNCTIONAL

ARTIFACT EVALUATED usenix ASSOCIATION REPRODUCED

USENIX Security ’24 Artifact Appendix: SPF Beyond the Standard: Management and Operational Challenges in Practice and Practical Recommendations

Md. Ishtiaq Ashiq
Virginia Tech

Weitong Li
Virginia Tech

Tobias Fiebig
Max-Planck-Institut für Informatik

Taejoong Chung
Virginia Tech

A Artifact Appendix

This artifact appendix is meant to be a self-contained document which describes a roadmap for the evaluation of our artifact. It includes a clear description of the hardware, software, and configuration requirements. It also includes the major claims made by our paper and instructions on how to reproduce each claim.

A.1 Abstract

[Mandatory]

We present data and source artifacts for reproducing Table 1, 2, and 3 and Fig. 3, 4, 7, and 9 in our final version. Data artifacts required to generate other figures require huge disk space. Although they may be available upon request, we can not make them publicly available and hence out of the scope of this artifact evaluation.

A.2 Description & Requirements

[Mandatory] We are providing our own environment with everything set up for the reviewers to run our scripts and reproduce our results.

A.2.1 Security, privacy, and ethical concerns

[Mandatory] N/A

A.2.2 How to access

[Mandatory] We plan to host the datasets in a publicly accessible S3 bucket. Links to the dataset and scripts are available in this GitHub page: `https://spf-measurement.github.io`.

The stable URL of the GitHub Repo is: `link` [1].

A.2.3 Hardware dependencies

[Mandatory] We are providing our own environment with everything set up for the reviewers to run our scripts and reproduce our results.

A.2.4 Software dependencies

[Mandatory] We are using PySpark for big data processing, python3 for regular scripting, and gnuplot for plotting scripts into a PDF. You can find some other software dependencies in the given requirements.txt file.

A.2.5 Benchmarks

[Mandatory] N/A

A.3 Set-up

[Mandatory] Please run the following commands in your terminal.

- Login to our environment: `ssh -D 8080 usenix24-ae@pharah.cs.vt.edu -p 2222`. Password is: `usenix24-1393`
- Login to one of the internal servers with packages preinstalled: `ssh pharah01`
- `cd submission28-copy/`
- `source venv/bin/activate`

A.3.1 Installation

[Mandatory] N/A

A.3.2 Basic Test

[Mandatory] N/A

A.4 Evaluation workflow

[Mandatory for Artifacts Functional & Results Reproduced, optional for Artifact Available]

[1] https://github.com/spf-measurement/spf-measurement.github.io/commit/a65cc2fac4deaa320af477054339be07c62d4819

A.4.1 Major Claims

[Mandatory for Artifacts Functional & Results Reproduced, optional for Artifact Available] Following are the major claims made in Section 5:

(C1): Table 1: According to our latest snapshot, 63% .com domains with MX records have a SPF TXT record and 60% of them have an include mechanism.

(C2): Table 2: The percentage of include mechanism in SPF-enabled domains have increased by 4% on average from our first snapshot back in Nov, 2021.

(C3): Figure 3: More than 9.2% of SPF records with include requires more than 10 DNS look ups.

(C4): Table 3: Syntax errors in SPF records are minimal (0.2%).

(C5): Figure 4: Only six SPF records appear in the include mechanisms for 50% of all SPF records while just two SPF records—managed by two specific hosting providers—account for 83.7% of the SPF records that require more than 10 DNS lookups for evaluation.

(C6): Figure 7: 40% servers violate the total lookup limit recommendation in practice.

(C7): Section 5.2: 84K (6.9% of 1.2M) initiate SPF queries prior to issuing the DATA command.

(C8): Section 5.2: We find that 903 servers (1.1%) send more than 51 queries, but they do so by issuing duplicated SPF queries.

(C9): Section 5.2: We find that 199 SMTP servers queried for all 50 include in our SPF record; this strongly suggests that some SMTP servers use SPF validators that do not have any total lookup limit.

(C10): Section 5.2: Out of these 199 servers, 164 of them maintain a void lookup limit as they aborted SPF resolution right after querying the set number of domains that incur `NXDOMAIN` responses in Exp. #2; the rest (35) requested all 36 domains that result in void lookups, which indicates that these servers could potentially serve as reflectors for launching DNS queries against a targeted victim's authoritative server.

A.4.2 Experiments

[Mandatory for Artifacts Functional & Results Reproduced, optional for Artifact Available]

(E1): *[Reproducing Tab. 1, Tab. 2 and verifying claims C1, C2] [5 human-minutes + 1 hours compute-hour + 400GB disk space]*

How to: *Following are the steps to perform the experiment and to collect and organize the results as expected from the paper.*

Preparation: N/A

Execution: For reproducing Table 1 and 2, run `PYSPARK_DRIVER_PYTHON=‘which python‘ PYSPARK_PYTHON=‘which python‘ spark-submit generate-table1-table2-data.py`.

Results: Match the outputs of the script with table 1 and table 2 data. Numbers are followed with explanations, which are expected to be very similar with the above claims.

(E2): *[Reproducing Tab. 3 and Fig. 3; verifying claim C3] [5 human-minutes + 2 hours compute-hour + 400GB disk space]*

How to: *Following are the steps to perform the experiment and to collect and organize the results as expected from the paper.*

Preparation: N/A

Execution:

- For reproducing Tab. 3 and generating data to plot Fig. 3, run `PYSPARK_DRIVER_PYTHON=‘which python‘ PYSPARK_PYTHON=‘which python‘ spark-submit generate-table3-fig3-data.py`.
- For plotting Fig. 3, do the following in the terminal. **The corresponding plotting script can be found at ‘artifact-appendix/plots/cdf-vs-no-of-lookups-include.plot’ file.**
- `cd artifact-appendix/`
- `make clean`
- `make`
- `pdflatex usenix-24.tex`

Results:

- Match the outputs of the script with table 3 data. Numbers are followed with explanations, which are expected to be very similar with the claim C3.
- Check the Generated Figures subsection in the generated `usenix24.pdf` file and match the figure with the corresponding one in our final version.
- `cd ..`

(E3): *[Reproducing Fig. 4 and verifying claims C4, C5] [negligible]*

How to: *Following are the steps to perform the experiment and to collect and organize the results as expected from the paper.*

Preparation: N/A

Execution:

- For producing temporary data needed to plot Figure 4, run `python3 generate-fig4-data.py`
- For plotting Fig. 4, do the following in the terminal. **The corresponding plotting script can be found in ‘artifact-appendix/plots/cdf-vs-include-centralization.plot’ file.**
- `cd artifact-appendix/`
- `make clean`
- `make`
- `pdflatex usenix-24.tex`

Results:

- Check the Generated Figures subsection in the generated `usenix24.pdf` file and match the figure with the corresponding one in our final version.
- `cd ..`

(E4): *[Reproducing Fig. 7 and verifying major claims from C6 to C10] [10 human-minutes + 30 mins compute-hour + 2GB disk space]*

How to: *Following are the steps to perform the experiment and to collect and organize the results as expected from the paper.*

Preparation: N/A

Execution:

- For producing temporary data needed to plot Figure 7, run `python3 generate-fig7-data.py`
- For plotting Fig. 7, do the following in the terminal. **The corresponding plotting script can be found at 'artifact-appendix/plots/cdf-vs-vuln-scan-lookups.plot' file.**
- `cd artifact-appendix/`
- `make clean`
- `make`
- `pdflatex usenix-24.tex`

Results:

- For verifying claims from C7 to C10, match the outputs with the claims. Numbers are followed with explanations.
- Check the Generated Figures subsection in the generated `usenix24.pdf` file and match the figure with the corresponding one in our final version.
- `cd ..`

(E4): *[Reproducing Fig. 9] [negligible]*

How to: *Following are the steps to perform the experiment and to collect and organize the results as expected from the paper.*

Preparation: N/A

Execution:

- For Fig. 9 verification, you should find our survey responses as a CSV file in the `temp` directory after removing columns unrelated to these figures. You can verify the consistency of the survey figure by manually checking the CSV file.
- For plotting Fig. 9, do the following in the terminal. **The corresponding plotting script can be found at 'artifact-appendix/plots/survey.plot' file.**
- `cd artifact-appendix/`
- `make clean`
- `make`
- `pdflatex usenix-24.tex`

Results:

- Check the Generated Figures subsection in the generated `usenix24.pdf` file and match the figure with the corresponding one in our final version.
- `cd ..`

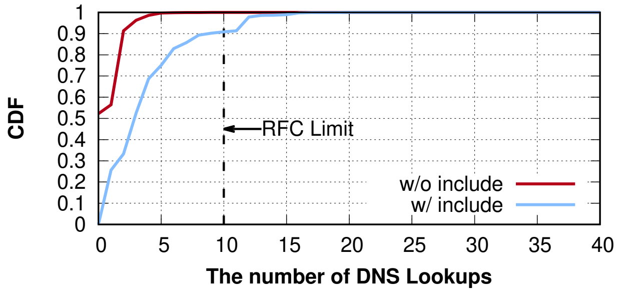

Figure 1: This figure corresponds to Figure 3 in our paper. More than 9.2% of SPF records with include requires more than 10 DNS look ups. Note that the x axis extends to 175!

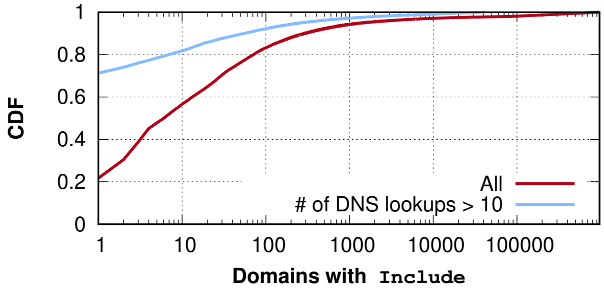

Figure 2: This figure corresponds to Figure 4 in our paper. As of March 27th, 2023, only six SPF records appear in the include mechanisms for 50% of all SPF records. Remarkably, just two SPF records—managed by two specific hosting providers—account for 83.7% of the SPF records that require more than 10 DNS lookups for evaluation.

A.5 Version

Based on the LaTeX template for Artifact Evaluation V20231005. Submission, reviewing and badging methodology followed for the evaluation of this artifact can be found at https://secartifacts.github.io/usenixsec2024/.

A.6 Generated Figures

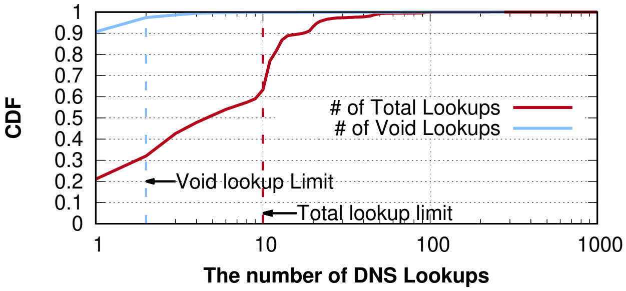

Figure 3: This figure corresponds to Figure 7 in our paper. CDF of the number of DNS requests that our DNS authoritative server receives; 40% servers violate the total lookup limit recommendation in practice. Note that *x* axis extends to 25,117![3]

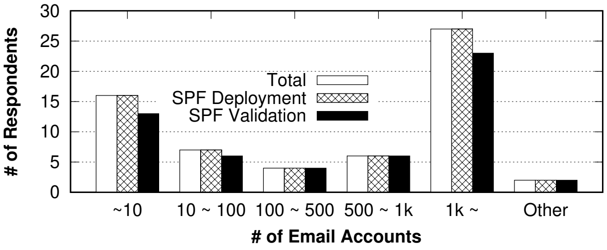

Figure 4: This figure corresponds to Figure 9 in our paper. The figure shows the distribution of the number of email accounts managed by each of the 62 respondents who answered both questions regarding SPF deployment and SPF validation support; note that all respondents serving as SMTP administrators confirmed the deployment of SPF records.