Sharing Information and Intelligence without Disclosing It

Private Search Set (PSS)

Alexandre Dulaunoy & Jean-Louis Huynen

CIRCL
https://circl.lu/

$ whoarewe

Alexandre Dulaunoy
adulau [1]

Jean-Louis Huynen
gallypette [2]

[1] https://github.com/adulau/
[2] https://github.com/gallypette

Introduction

- Google Safe Browsing Example:
  - Google Safe Browsing is some kind of database of malicious URL[3].
  - Initially implemented with Bloom Filters.
  - Bloom Filters provided probabilistic data checks, offering a "maybe" rather than a definitive "yes."
  - Shifted from Bloom Filters to more complex privacy-preserving solutions to match its requirements.
  - We believe that **sharing these kind of data structures has value in security/CTI**, mainly for when privacy is important.

[3] `https://safebrowsing.google.com/`

Early MISP Research and Evaluation

- MISP Privacy Aware Exchange (2017):
  - A project focused on enhancing indicator sharing with confidentiality and privacy[4].
  - Evaluated techniques like SACTI[5], improving performance over standard MPC[6].
  - Noted the high cost and complexity of MPC, which is typically online.

[4] *van de Kamp, T., Peter, A., Everts, M. H., & Jonker, W. (2016, October). Private Sharing of IOCs and Sightings.*

[5] `https://www.misp-project.org/2022/10/27/SACTI_Secure_aggregation_of_cyber_threat_intelligence.html/`

[6] Multi-Party Computation

Filters in Threat Intelligence

- Role of Filters in CTI:
  - Filters can be **attribute lists**[7] in tools like TIP, 'maltrail', or they can be rules in Snort, Suricata, and Yara.
- Data Structures in Security Tools:
  - Tools use **optimized data structures** such as datasets in Suricata, Bloom Filters in routers, and LRU caches.
- We decided to combine our experience with Bloom filters and provide the following capabilities: **confidentiality, privacy, watermarking, offline functionality, performance, and shareability**.

[7] IOCs, detection rules, skip-lists

Sharing these Data Structures

- Focus on:
  - **standardizing descriptions**: what data structure, what it contains, and how to use it,
  - **normalizing data**: how the data is stored within the data structure,
  - developing **storage solutions**: how the whole bundle is serialized.
- DCSO developed a **serialized format** for classical Bloom Filters[8], which served as our starting point.

[8] `https://github.com/dcso/bloom`

Starting Point - Hashlookup

- Development of Hashlookup[9]:
  - Started as a repository for known goodware with a web service for **offline forensic queries**.
  - Inspired by DCSO's implementations, it adopted serialized Bloom Filters for offline querying.
- Expansion into Yara:
  - The library was reimplemented in C and integrated into a Yara module,
  - it allows to filter out known-good files.
- AIL Project:
  - Utilized Bloom Filters **to share sensitive[10] data**, such as lists of onion services.

[9] `https://hashlookup.io`

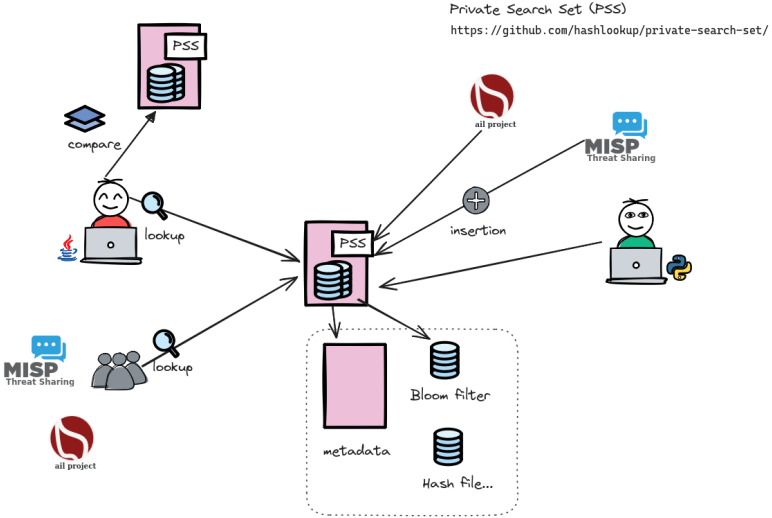
[10] or even commercially valuable information

Private Search Sets

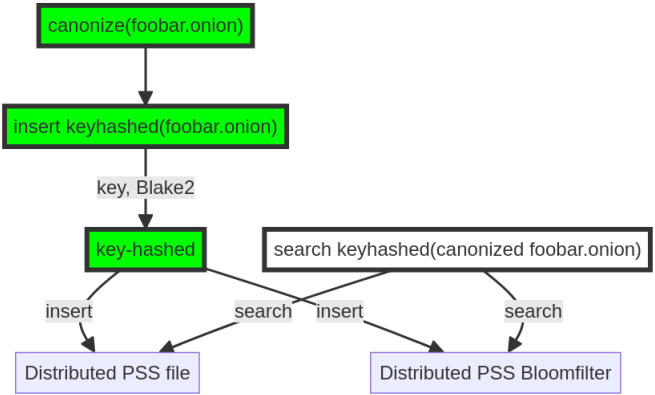
- Introduction to PSS:
  - Facilitates **sharing and interoperability** between tools and CTI processes.
  - Designed for fast, private lookups and easy distribution.
  - Features include watermarking, offline searching, and a **flexible meta-format**.
  - The PSS format is succinctly described in a document spanning 4 pages[11].

[11] `https://github.com/hashlookup/private-search-set`

PSS Overview



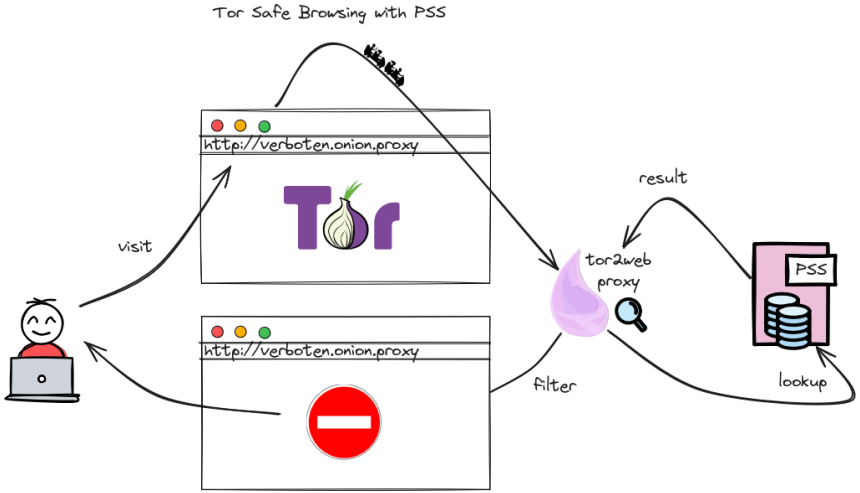
How Bloom filters in PSS are created?



Safe For Work Tor Browser Use Case

- Demonstrates PSS with a Tor proxy setup for safe browsing of the darknet.
- Prevents access to harmful hidden services by analysts, crucial for monitoring ransomware activities.
- Uses a combination of PSS of known malicious (in a bloomfilter) / known good (in a set) onion services.
- Similar lists can be used by forensic investigators or CTI analysts to **facilitate detection/sharing** without the need to validate the content directly.

PSS - Tor safe browsing



Integration

- The first version of the PSS standard has been published, along with a minimal Python library to handle the PSS meta-format.
- Initial **integrations with different open-source tools** developed by CIRCL include:
  - MISP integration for exporting selectors/indicators such as financial details and sensitive information in PSS format.
  - The AIL project handles the tracking of terms using PSS Bloom filters, **offering a confidential alternative to traditional keyword matching lists**.

Future Works

- Publishing the first **IETF Internet-Draft** for the PSS meta-format.
- Developing a MISP 3 **privacy-aware correlation** (Private Set Intersection) database that uses PSS for inter-instance sharing within a community.
- Gathering feedback and exploring use cases for PSS to **refine the format** before finalizing the first version.
- Plans to extend support to other data structures such as LRU caches.

Conclusion

- PSS **facilitates the sharing of information and intelligence** that was originally restricted due to confidentiality or legal reasons.
- Provides easy access to an open standard and open source implementation to integrate it into your software or CTI processes.
- Probabilistic data structures are still full of potential opportunities for CTI.

References

- PSS `https://github.com/hashlookup/private-search-set`
- Poppy - Rust implementation of the DCSO Bloom filter and updated Bloom filter format `https://github.com/hashlookup/poppy`
- Poppy - Poppy a new Bloom filter format and open source library `https://www.misp-project.org/2024/03/25/Poppy-a-new-bloom-filter-format-and-project.html/`

Joint Threat Analysis Network (JTAN)

- Connecting Europe Facility grant
  - Consortium: CERT.PL, CIRCL (LU), Corexalys (FR), CERT.LV, CERT.AT, SK-CERT, CERT-EE, DNSC (RO)
  - 2021.07 – 2024.06
- Goal:
  continuous proactive information sharing among European CSIRTs & beyond

Co-financed by the Connecting Europe Facility of the European Union