DNS TAPIR

Making Recursive DNS More Robust Through Cooperation

Johan Stenstam

Internetstiftelsen

October 14, 2024

Problem Statement

- DNS is increasingly being used for purposes that are not in line with the end user's interests and wishes.
- Various types of illegal activities use DNS both as a side channel and as a command and control mechanism.
- Various types of "tracking" activities now use DNS to a greater extent (in addition to traditional methods such as web cookies).
- The trend is negative, in the sense that such, umm, "bad", use of DNS is growing faster than traditional use of DNS. This problem is therefore getting worse.

What is "Threat Intelligence"?

"Threat Intelligence" is a rapidly growing market for information about "threats". It is often commercial, but there are also freely available sources.

A typical example is a service that basically says:

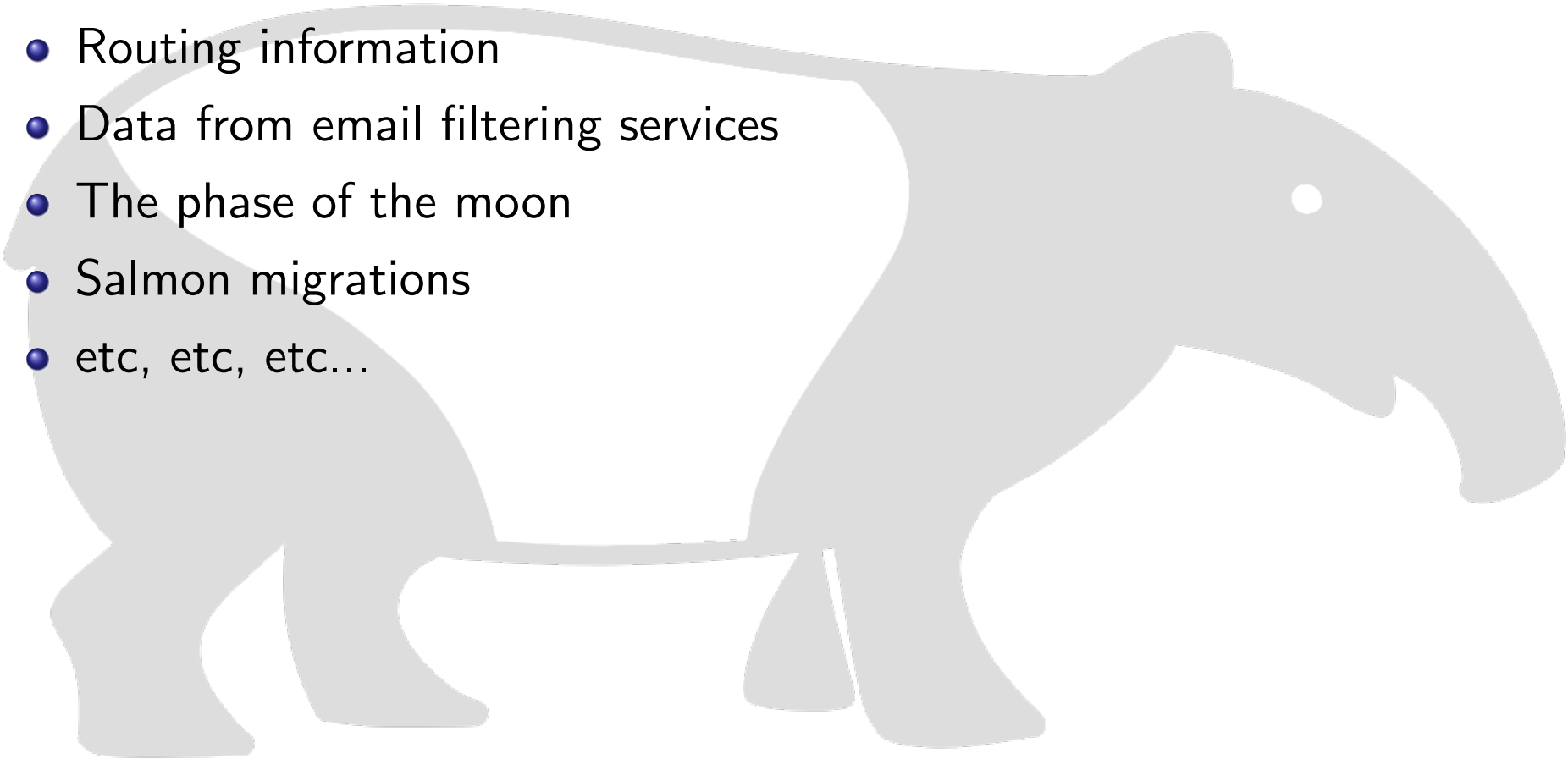
Block the bad domain name xyzzy.evil.empire.tld!

"Based on what?" is a good question to ask.

Data Sources for "Threat Intelligence"

Basically, all data is interesting when looking for threats.

- Web traffic
- Routing information
- Data from email filtering services
- The phase of the moon
- Salmon migrations
- etc, etc, etc...



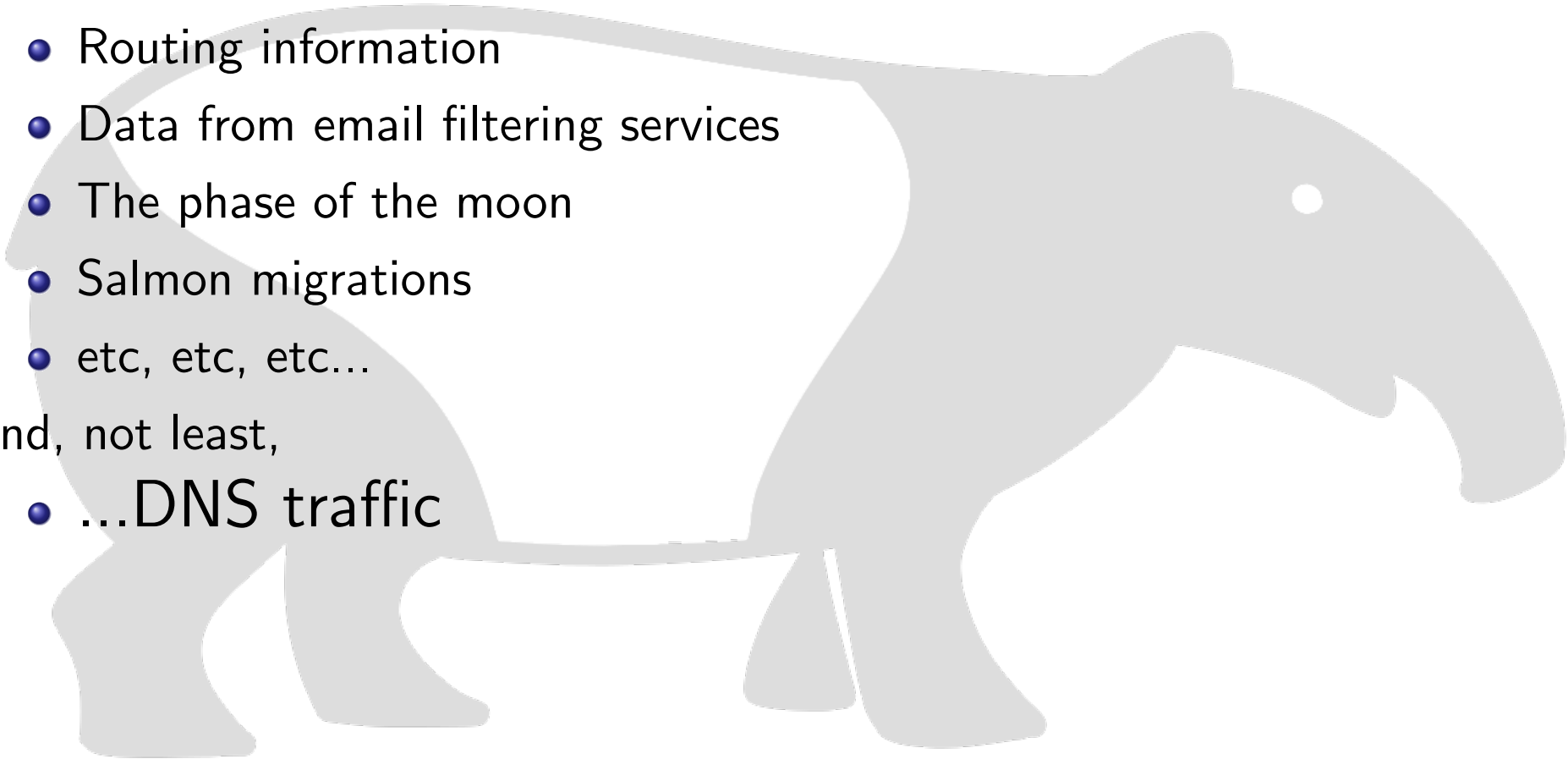
Data Sources for "Threat Intelligence"

Basically, all data is interesting when looking for threats.

- Web traffic
- Routing information
- Data from email filtering services
- The phase of the moon
- Salmon migrations
- etc, etc, etc...

and, not least,

- ...DNS traffic

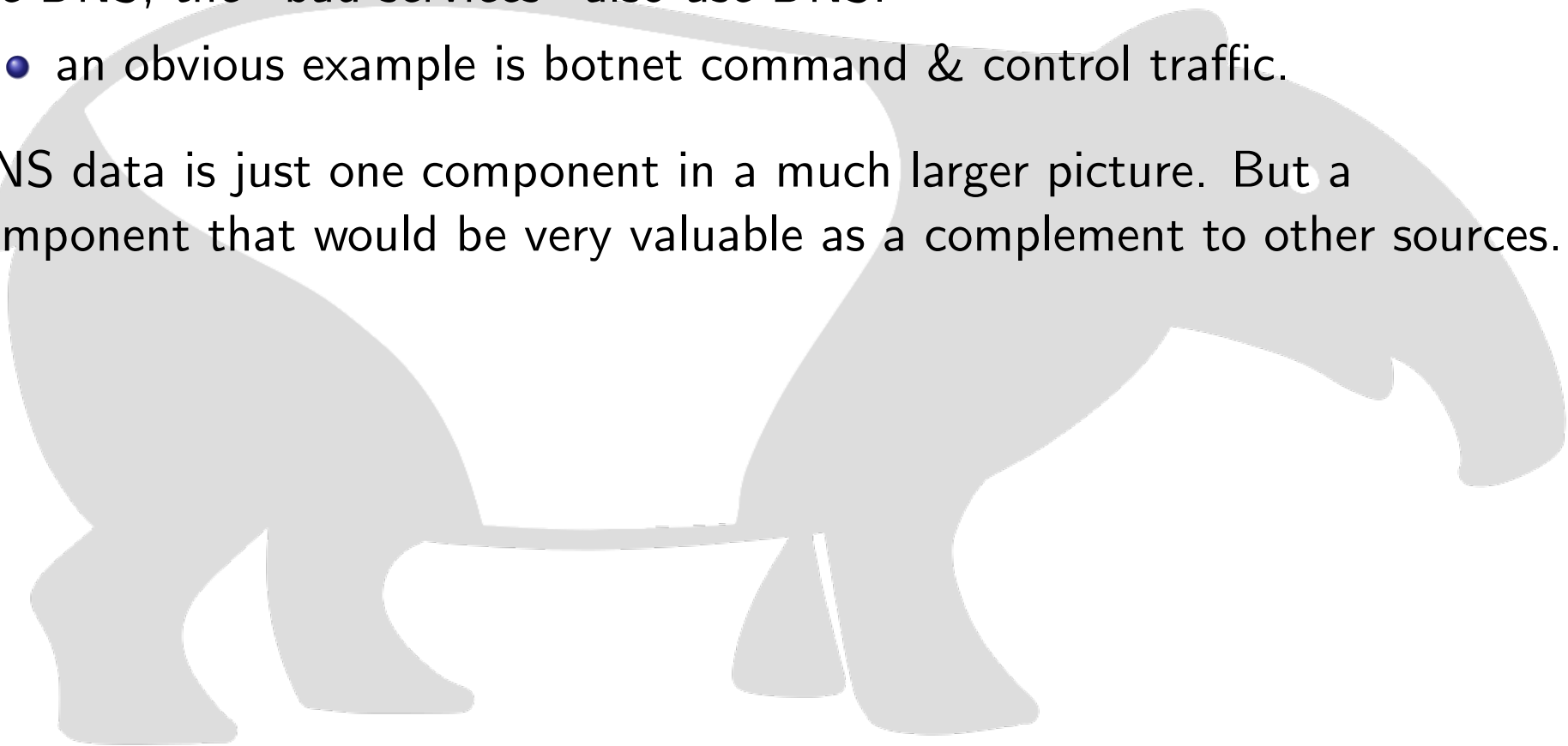


DNS Traffic Data as a Source for "Threat Intelligence"

DNS traffic data is very interesting, because it is so specific and detailed. Moreover, it is extremely common that in the same way normal services use DNS, the "bad services" also use DNS.

- an obvious example is botnet command & control traffic.

DNS data is just one component in a much larger picture. But a component that would be very valuable as a complement to other sources.



DNS Traffic Data as a Source for "Threat Intelligence"

DNS traffic data is very interesting, because it is so specific and detailed. Moreover, it is extremely common that in the same way normal services use DNS, the "bad services" also use DNS.

- an obvious example is botnet command & control traffic.

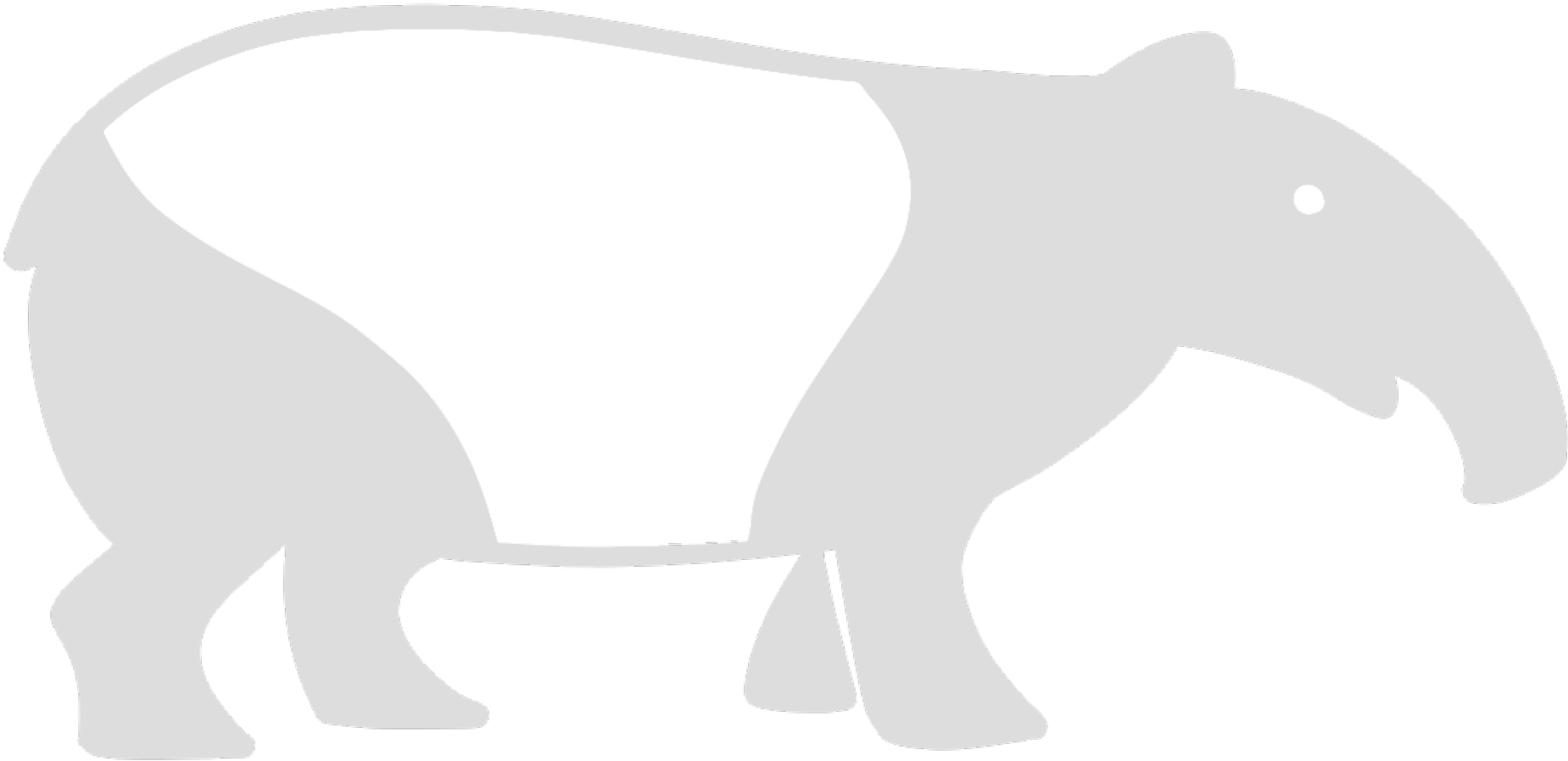
DNS data is just one component in a much larger picture. But a component that would be very valuable as a complement to other sources.

Due to its extremely sensitive nature, however, many (most?) operators of recursive name servers decline requests for access to DNS traffic data.

- this is of course to protect their users (which honors them)

Is There Any Way to Solve This Problem?

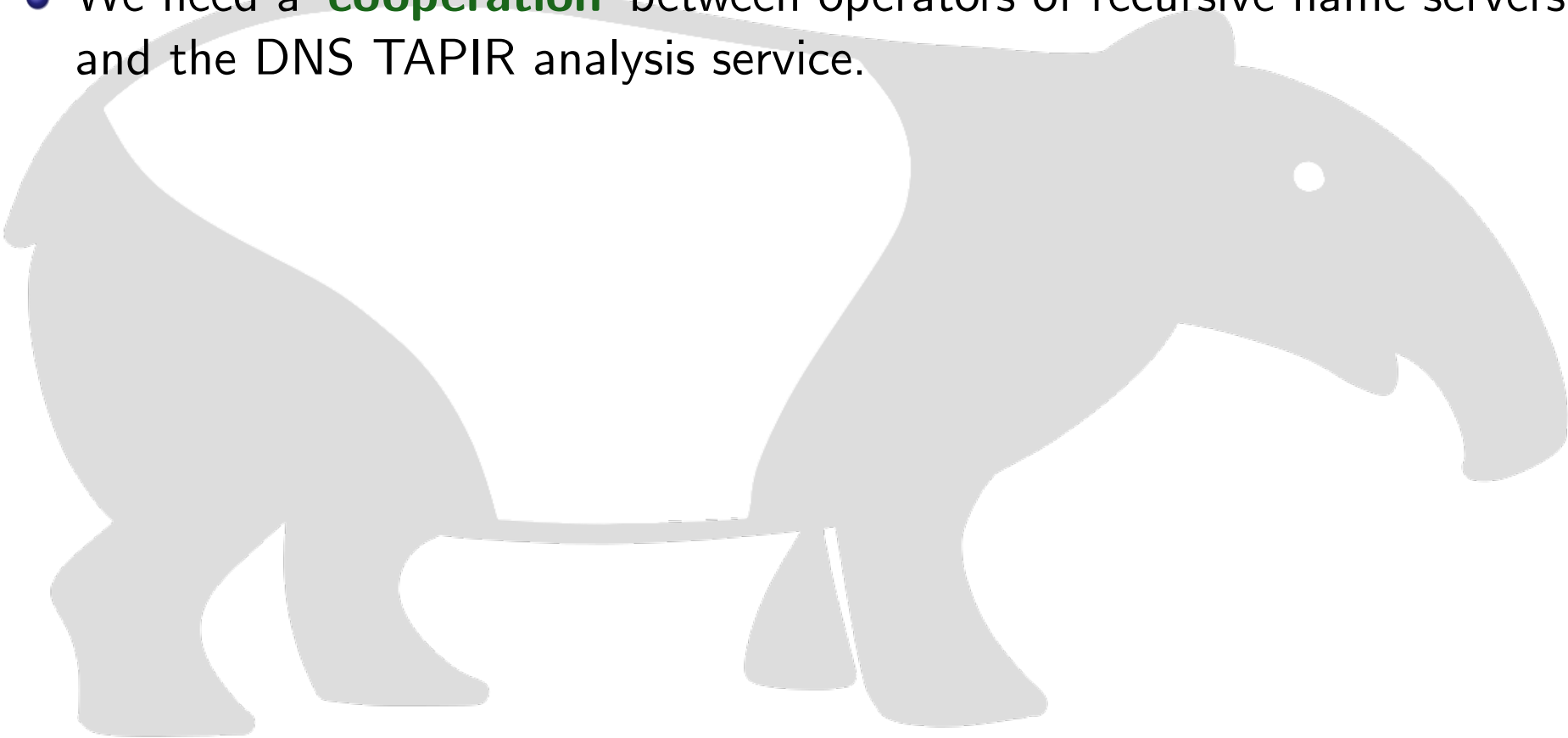
DNS TAPIR is a system being developed in a project funded by primarily PTS (the Swedish telecom authority). A main component of DNS TAPIR is an analysis service for DNS traffic data.



Is There Any Way to Solve This Problem? Yes!

DNS TAPIR is a system being developed in a project funded by primarily PTS (the Swedish telecom authority). A main component of DNS TAPIR is an analysis service for DNS traffic data.

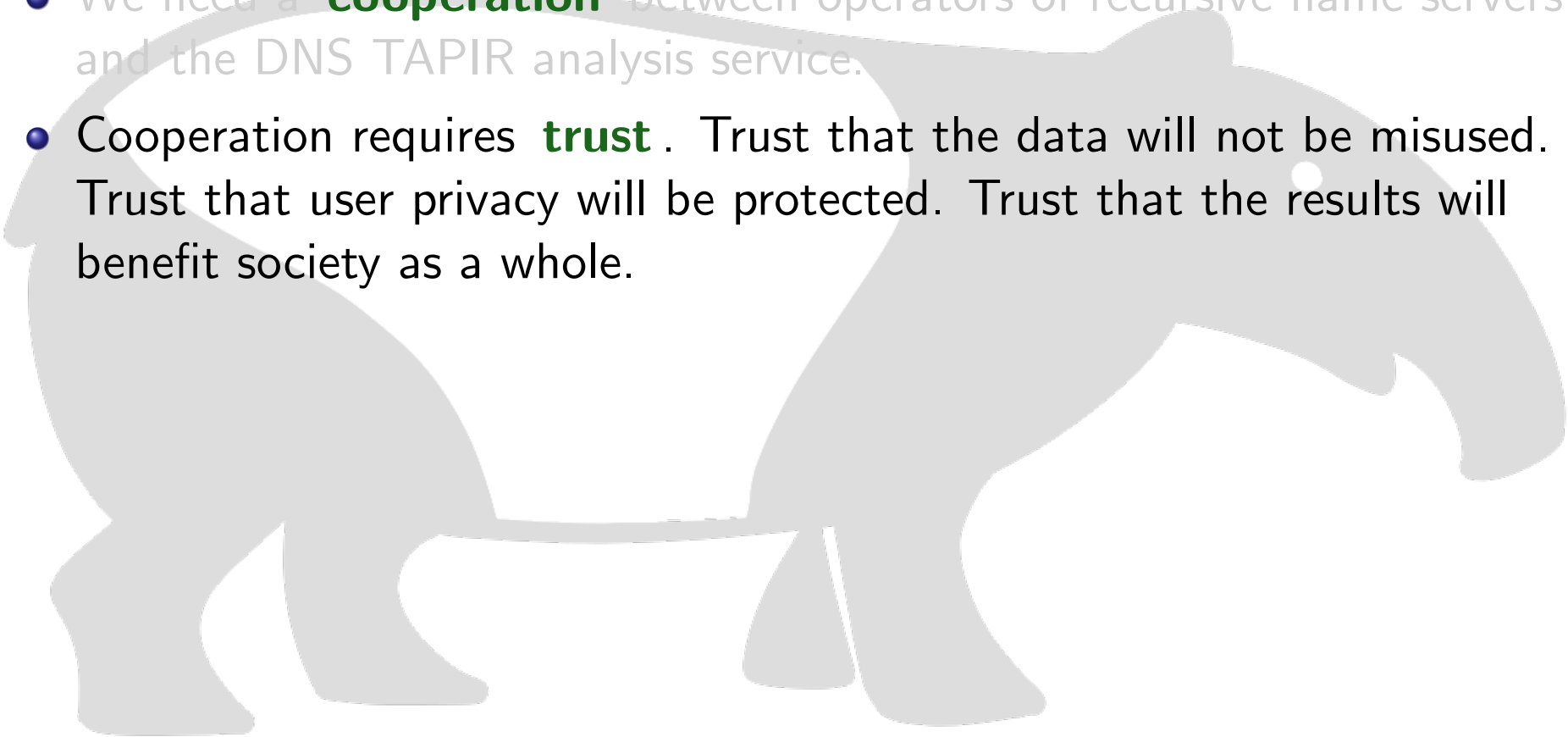
- We need a **cooperation** between operators of recursive name servers and the DNS TAPIR analysis service.



Is There Any Way to Solve This Problem? Yes!

DNS TAPIR is a system being developed in a project funded by primarily PTS (the Swedish telecom authority). A main component of DNS TAPIR is an analysis service for DNS traffic data.

- We need a **cooperation** between operators of recursive name servers and the DNS TAPIR analysis service.
- Cooperation requires **trust**. Trust that the data will not be misused. Trust that user privacy will be protected. Trust that the results will benefit society as a whole.



Is There Any Way to Solve This Problem? Yes!

DNS TAPIR is a system being developed in a project funded by primarily PTS (the Swedish telecom authority). A main component of DNS TAPIR is an analysis service for DNS traffic data.

- We need a **cooperation** between operators of recursive name servers and the DNS TAPIR analysis service.
- Cooperation requires **trust**. Trust that the data will not be misused. Trust that user privacy will be protected. Trust that the results will benefit society as a whole.
- Trust is established through **transparency**. DNS TAPIR must be completely transparent about exactly how privacy is protected, etc.

Is There Any Way to Solve This Problem? Yes!

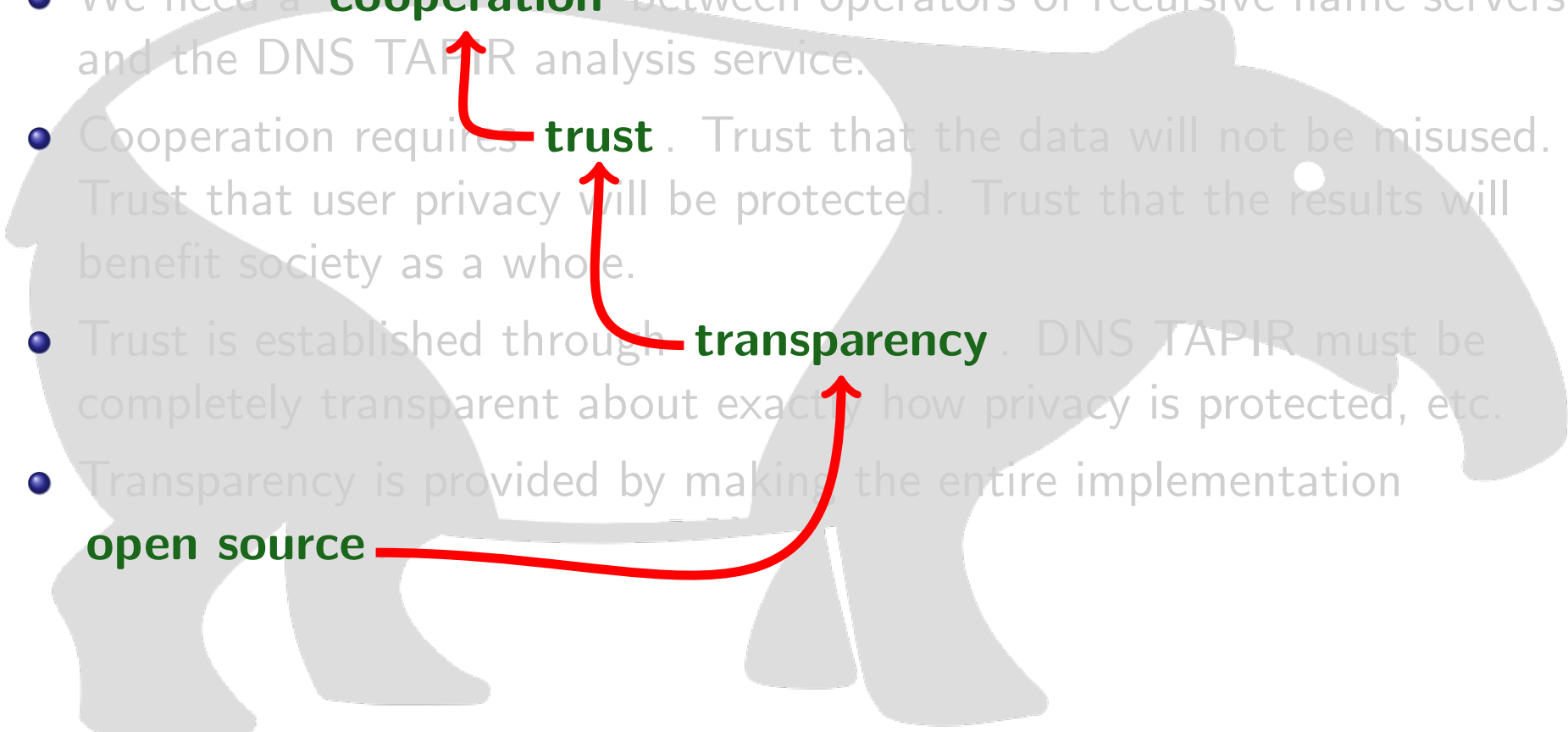
DNS TAPIR is a system being developed in a project funded by primarily PTS (the Swedish telecom authority). A main component of DNS TAPIR is an analysis service for DNS traffic data.

- We need a **cooperation** between operators of recursive name servers and the DNS TAPIR analysis service.
- Cooperation requires **trust**. Trust that the data will not be misused. Trust that user privacy will be protected. Trust that the results will benefit society as a whole.
- Trust is established through **transparency**. DNS TAPIR must be completely transparent about exactly how privacy is protected, etc.
- Transparency is provided by making the entire implementation **open source**.

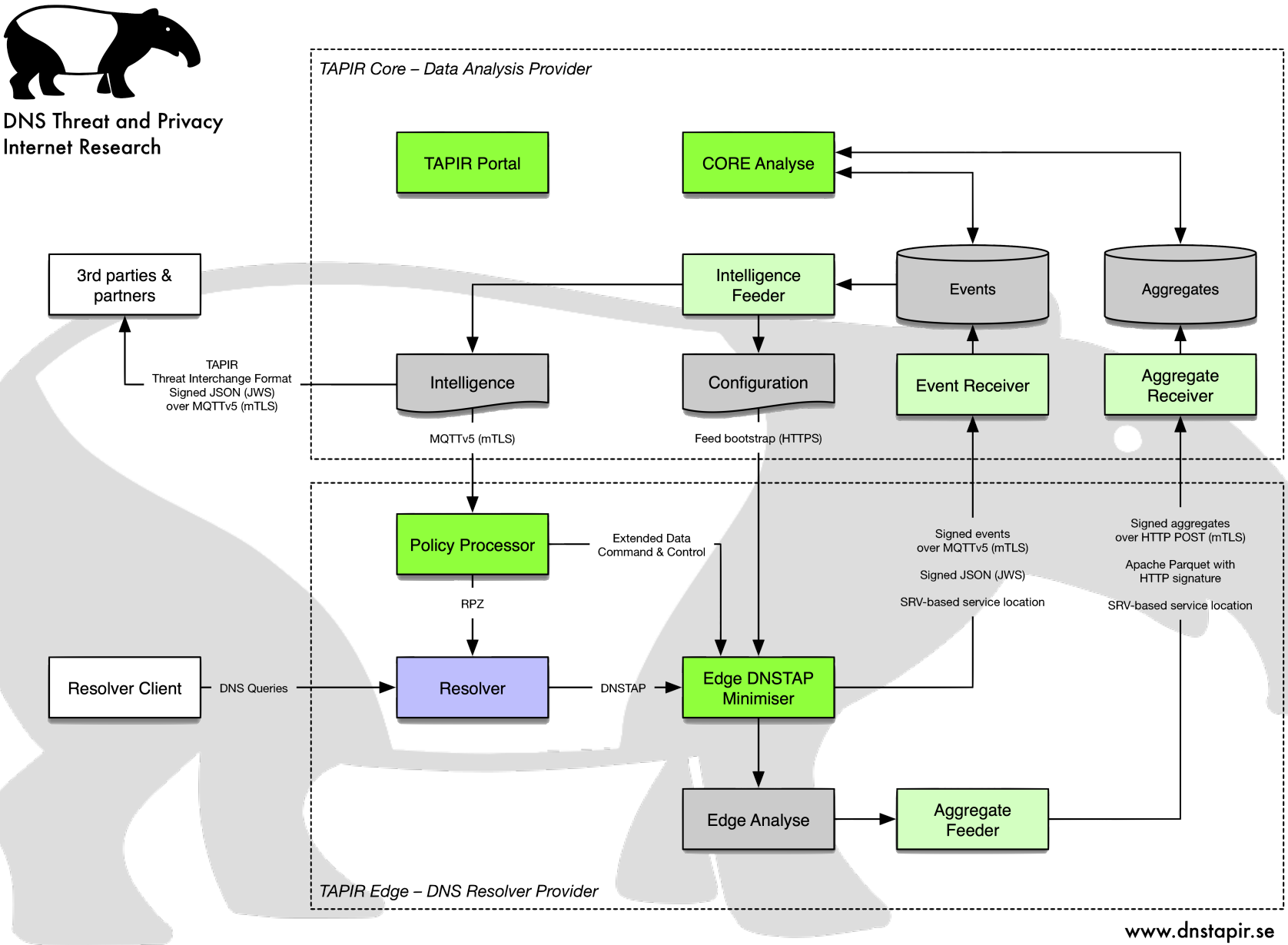
Is There Any Way to Solve This Problem? Yes!

DNS TAPIR is a system being developed in a project funded by primarily PTS (the Swedish telecom authority). A main component of DNS TAPIR is an analysis service for DNS traffic data.

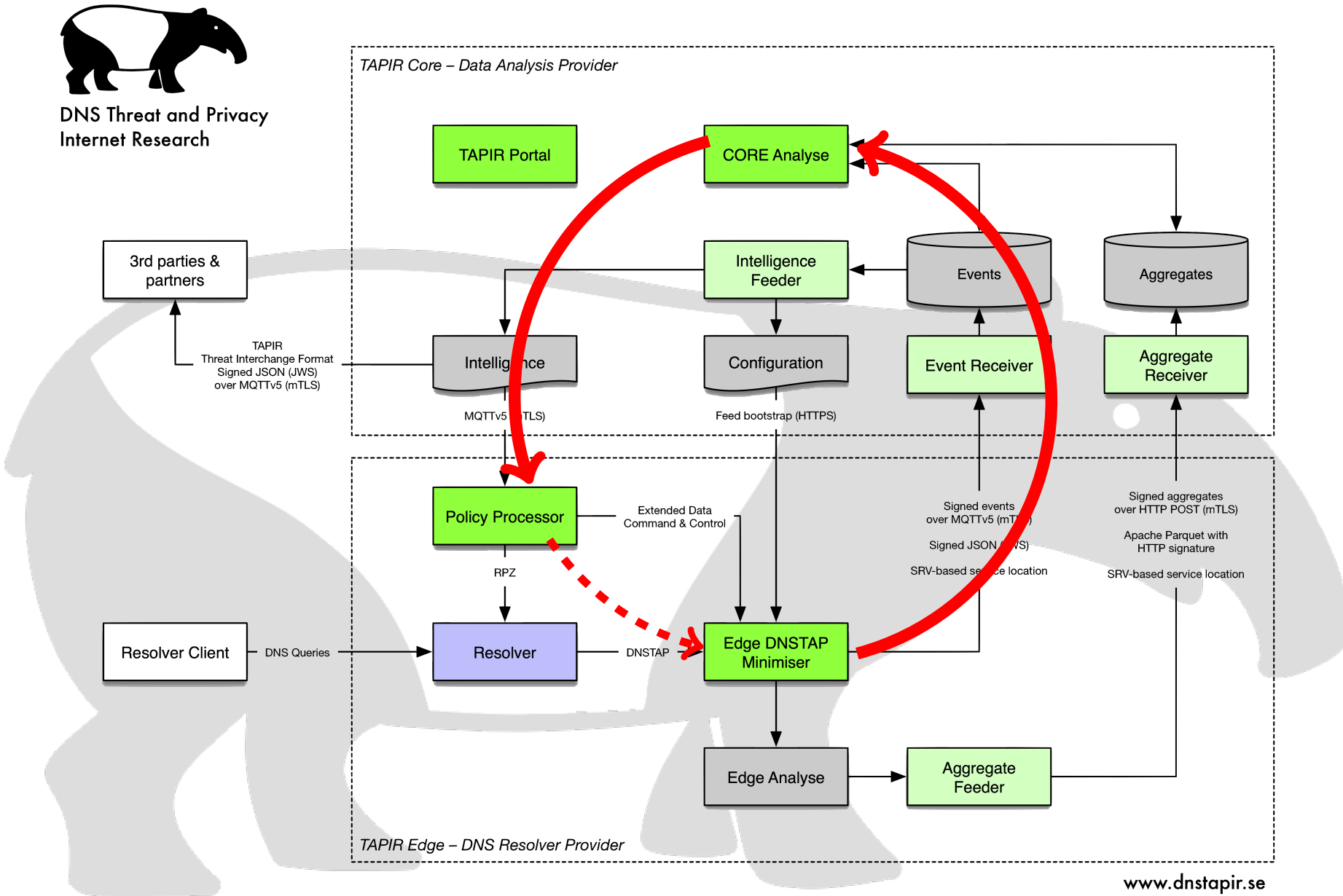
- We need a **cooperation** between operators of recursive name servers and the DNS TAPIR analysis service.
- Cooperation requires **trust**. Trust that the data will not be misused. Trust that user privacy will be protected. Trust that the results will benefit society as a whole.
- Trust is established through **transparency**. DNS TAPIR must be completely transparent about exactly how privacy is protected, etc.
- Transparency is provided by making the entire implementation **open source**



Current DNS TAPIR Design



Current DNS TAPIR Design



TAPIR-POP: TAPIR Policy Processor

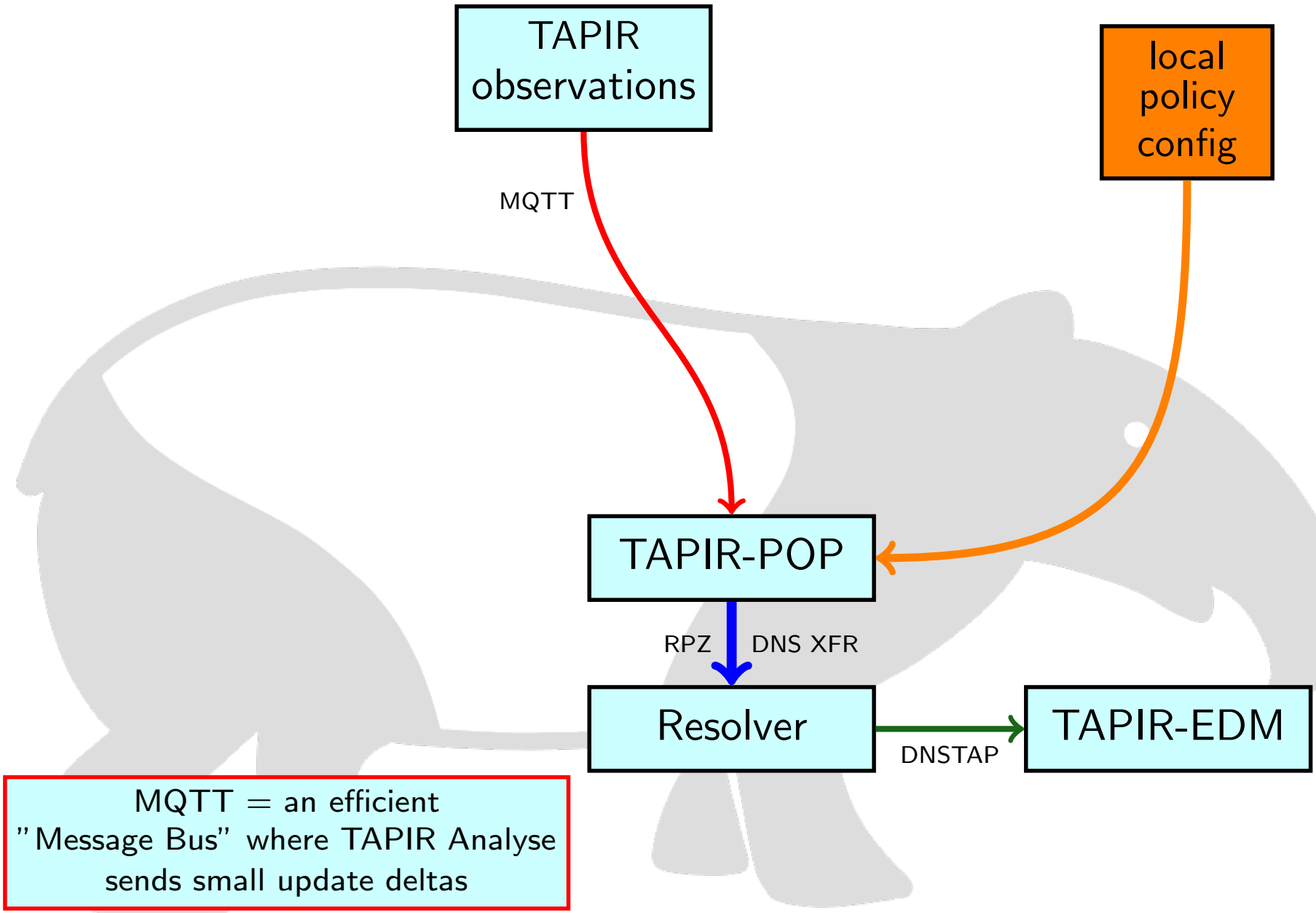
The main task of the TAPIR Policy Processor is to:

- consume the "observations" data that comes from TAPIR Analyse,
- combine this with data from other data sources,
- clean the result from all domain names that for various reasons should always be allowed,
- apply local policy configuration for filtering
- as all filtering decisions are temporary, do background garbage collection of old data
- transform the result into a single, minimal RPZ feed.

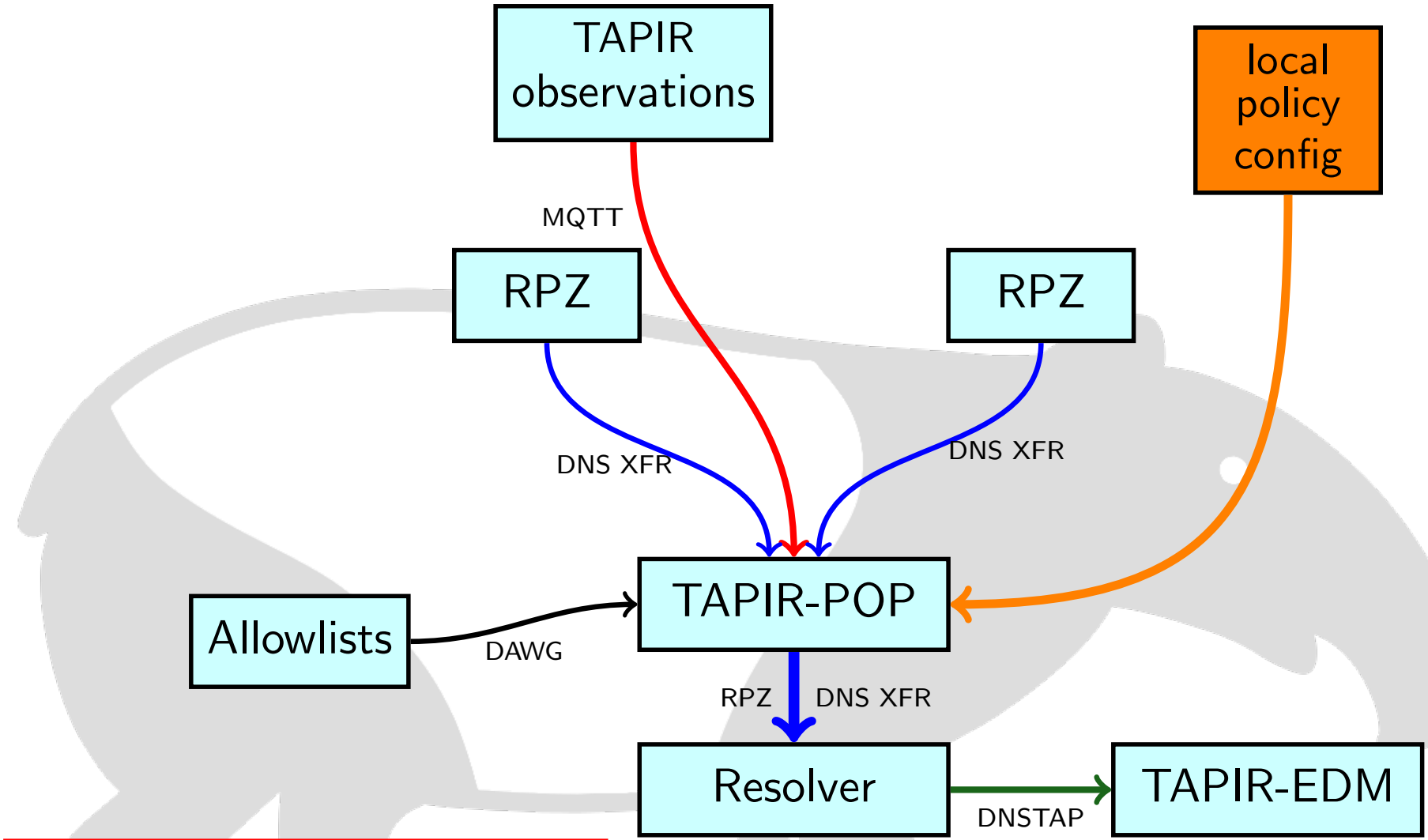
It sounds simple, but it gets quite complex.

"Response Policy Zone", a standard format for formulating resolver policies in the form of DNS zone data

TAPIR-POP: Data Sources

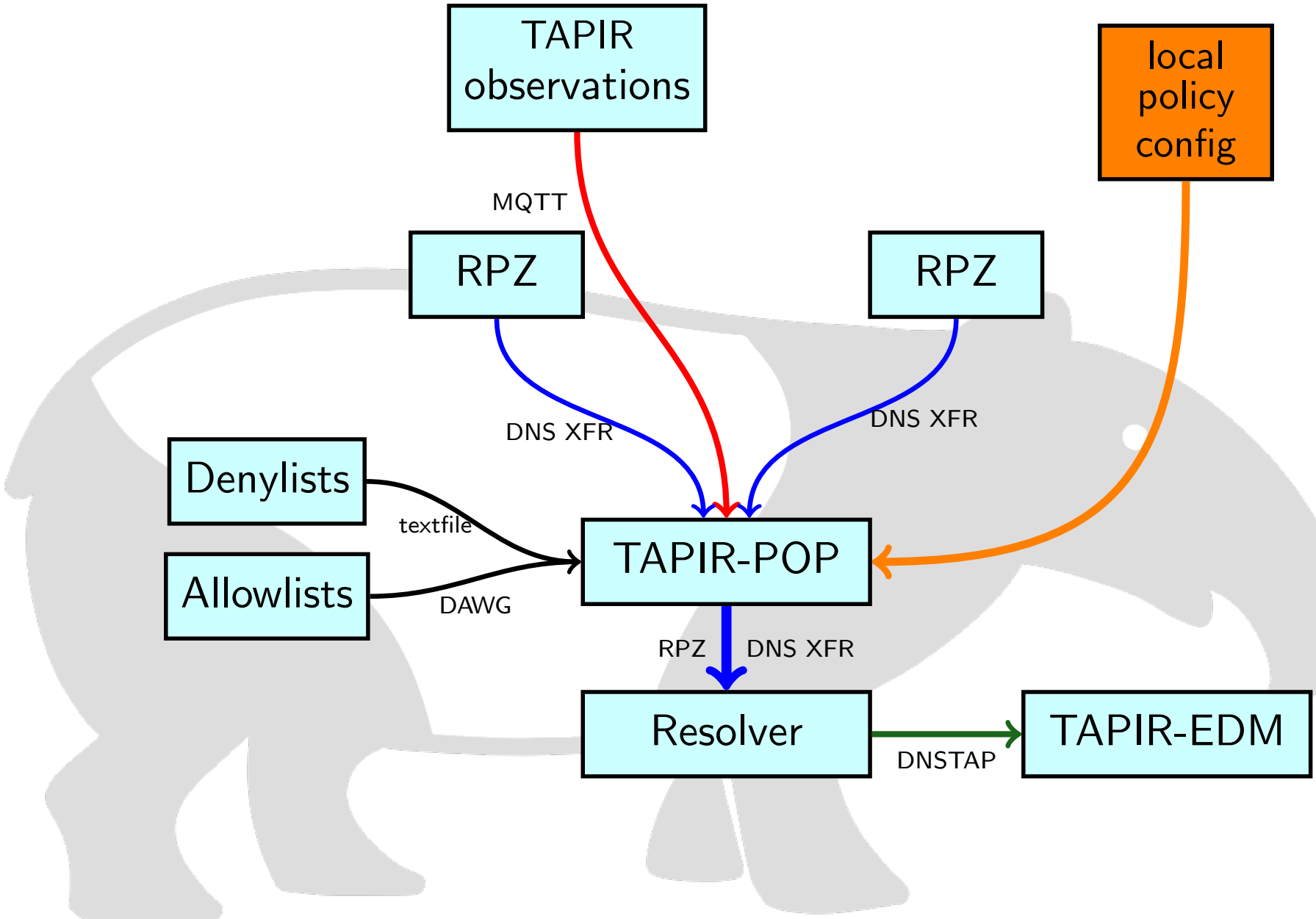


TAPIR-POP: Data Sources

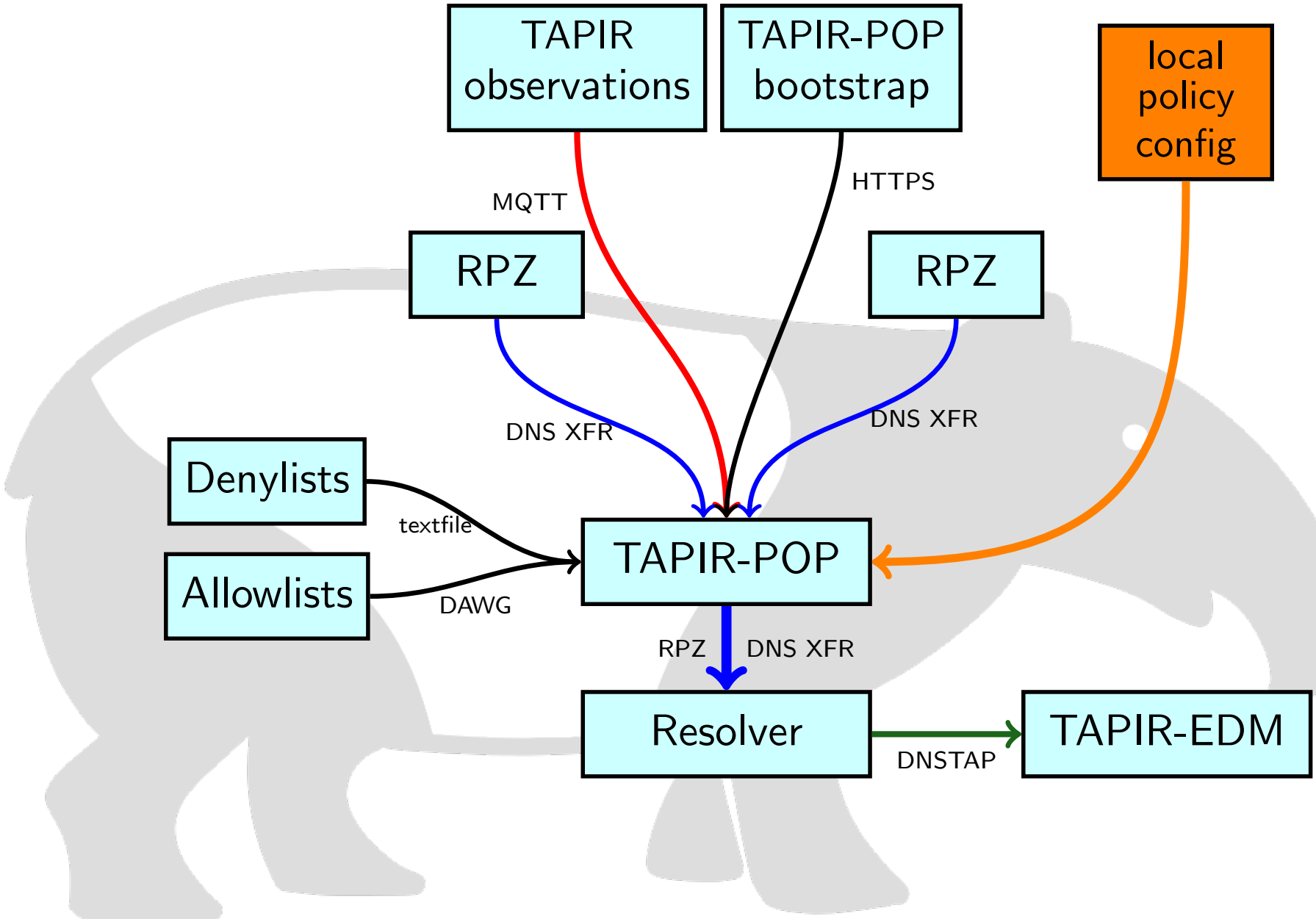


DAWG = Directed Acyclic Word Graph, an extremely compact format that is "mmap"-ed directly into POP

TAPIR-POP: Data Sources



TAPIR-POP: Data Sources



Who Decides What Should Be Filtered?

The stream of information from TAPIR Analyse only reports "observations" for interesting domain names (via flags).

- Other data sources may or may not contain "blocking instructions"

But in the end, it is TAPIR-POP that makes a decision, based on:

- All available information about a domain name
- The local filtering policy

The local policy is determined by the operator of the resolver. The goal is to really, really avoid so-called false positive results while capturing as much of the unwanted traffic as possible.

What is a Local Filtering Policy?

Examples of currently possible policy rules:

Filter domain names that:

- "are flagged by at least N different sources"
- "have more than M different TAPIR observations"
- "have at least the specific TAPIR observations X, Y, and Z"
- "regardless of everything else, never ever filter these domain names..."

Along with improving the analysis in TAPIR Core, the configuration "language" for expressing local policy is an area we expect to develop significantly in the future.

Current Status of DNS TAPIR

We have an almost complete prototype, including a public resolver:

`resolver.dnstapir.se`

- All "Edge" components are implemented:
  - TAPIR Edge Data Minimiser: mature and robust, quite "ready".
  - TAPIR Policy Manager: less mature, but architecturally ok.
- All "Core" components are implemented:
  - TAPIR Analyse: is a bit of a special case, which will never be "finished". But what we have works well.
  - Various storage and communication infrastructure: robust.
- We are testing the prototype in a "production environment" (see above).

Next Steps for DNS TAPIR

- The first phase of the project was to build a complete prototype. This is now done and we have submitted a final report.
- The next phase involves "productizing" DNS TAPIR, moving from prototype to production and primarily initiating cooperation with operators to get "real data" into TAPIR Analyse.
- This requires a new round of funding (this work is ongoing).

The biggest difference between "phase 1" and "phase 2" is that "phase 1" was a relatively isolated work while "phase 2" will involve many other actors.

We cannot do this without cooperation with operators and other stakeholders.

Summary

- The days of "neutral DNS resolvers" are most likely essentially over
- In the future, some form of "policy filter" between users and the Internet will be the norm, rather than an exception

The two difficult problems that must be addressed are therefore:

- **How to collect information from the sensitive DNS query data while protecting user privacy?**
- **How to ensure that the design of the policy filter remains a local decision?**

DNS TAPIR is our proposal for a solution. But it will only work in a **close cooperation with the resolver operators**.

Thanks & Contact

Johan Stenstam	johani@dnstapir.se
Johan Stenstam	johan.stenstam@internetstiftelsen.se
Web	https://dnstapir.se
Code	https://github.com/dnstapir