Collaborative Botnet Detection through Large-scale Network Traffic

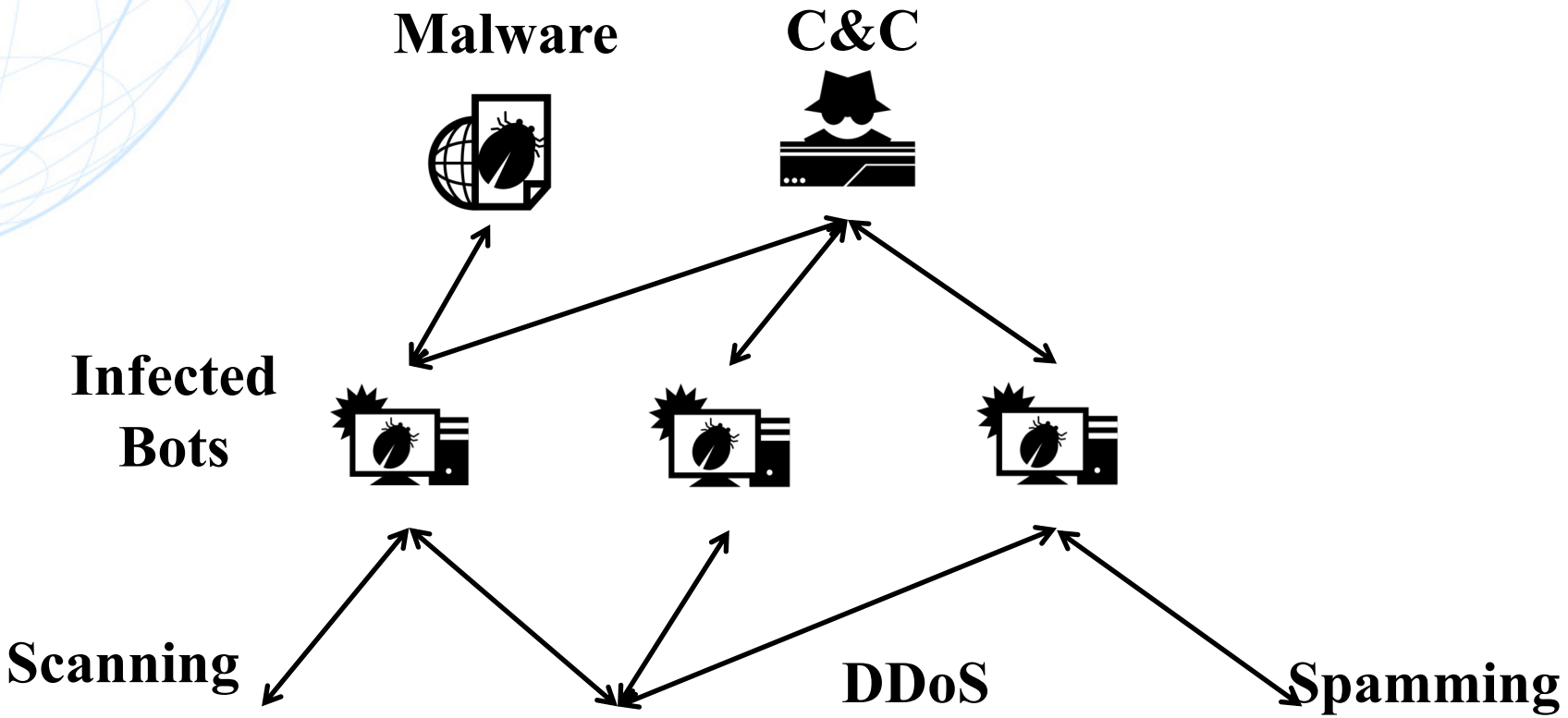
Bo Hu, Kazunori Kamiya (NTT)
Kenji Takahashi (NTT Ltd.)
Karel Mittig, Fabien Bignon(Orange)

Agenda

- Background
- Problem Statement
- Operator's TI approaches
  - Piper by NTT
  - Voodoo by Orange
- Collaborative Approach
- Collaboration Assessment
- Conclusion

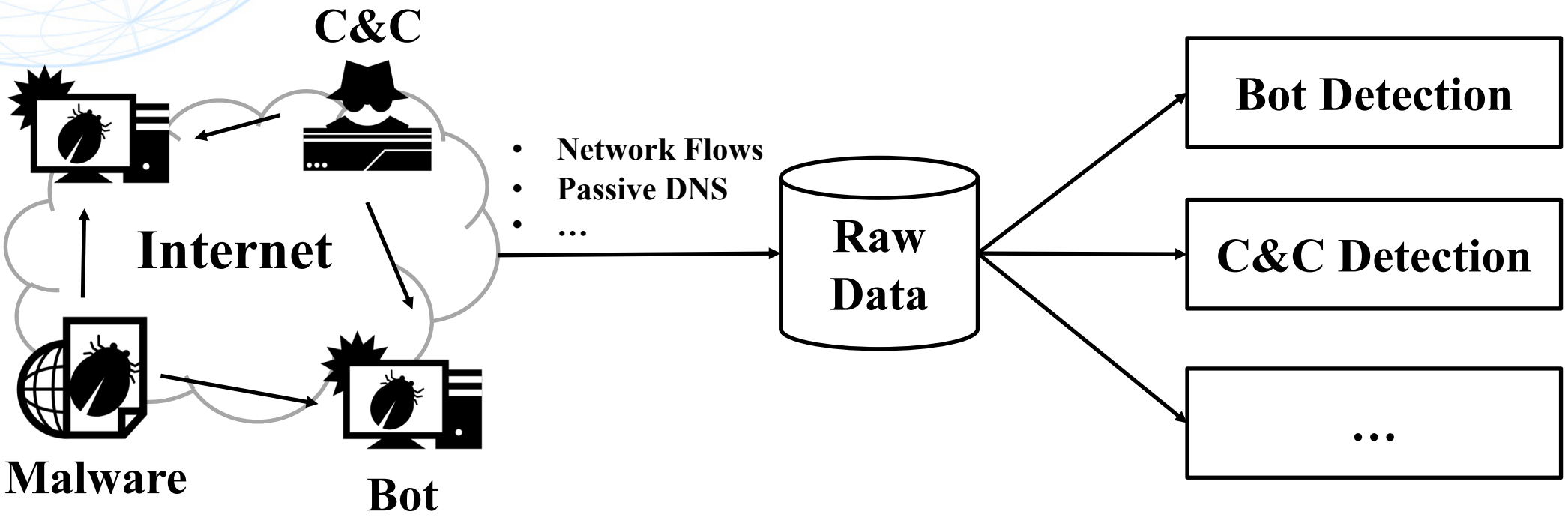
Background

- A botnet is a group of malware-infected hosts that collaborate together to launch various cyberattacks.



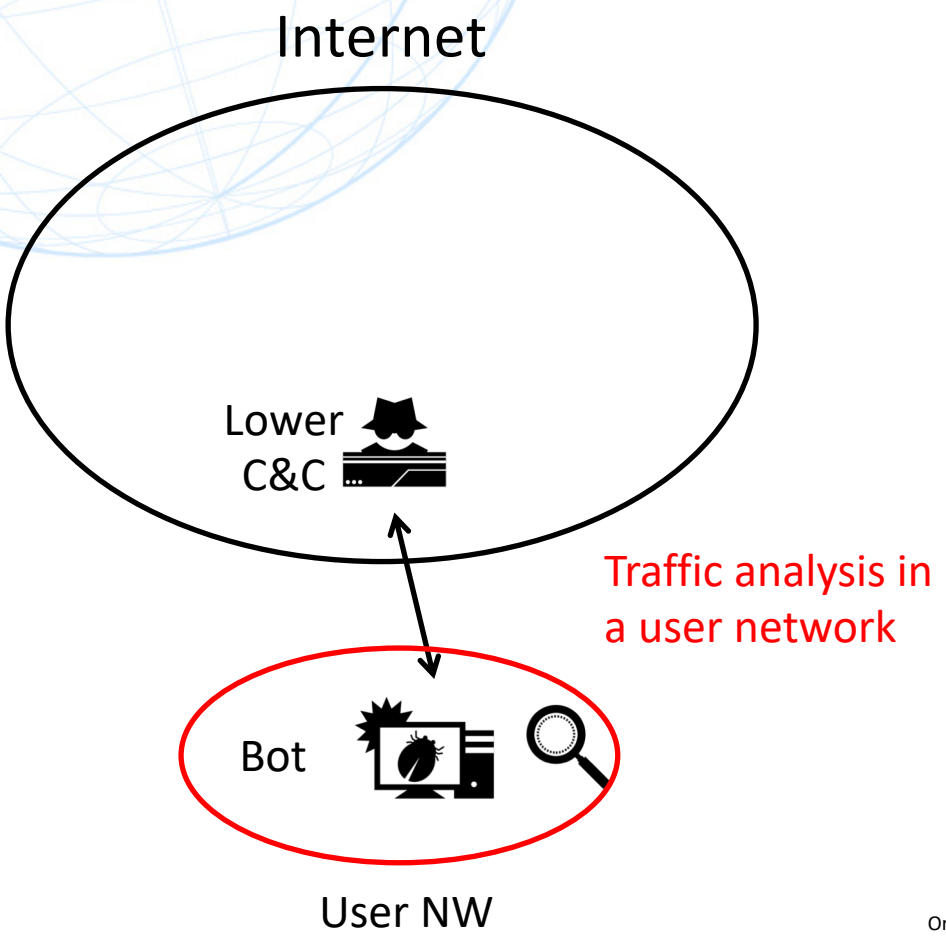
Background

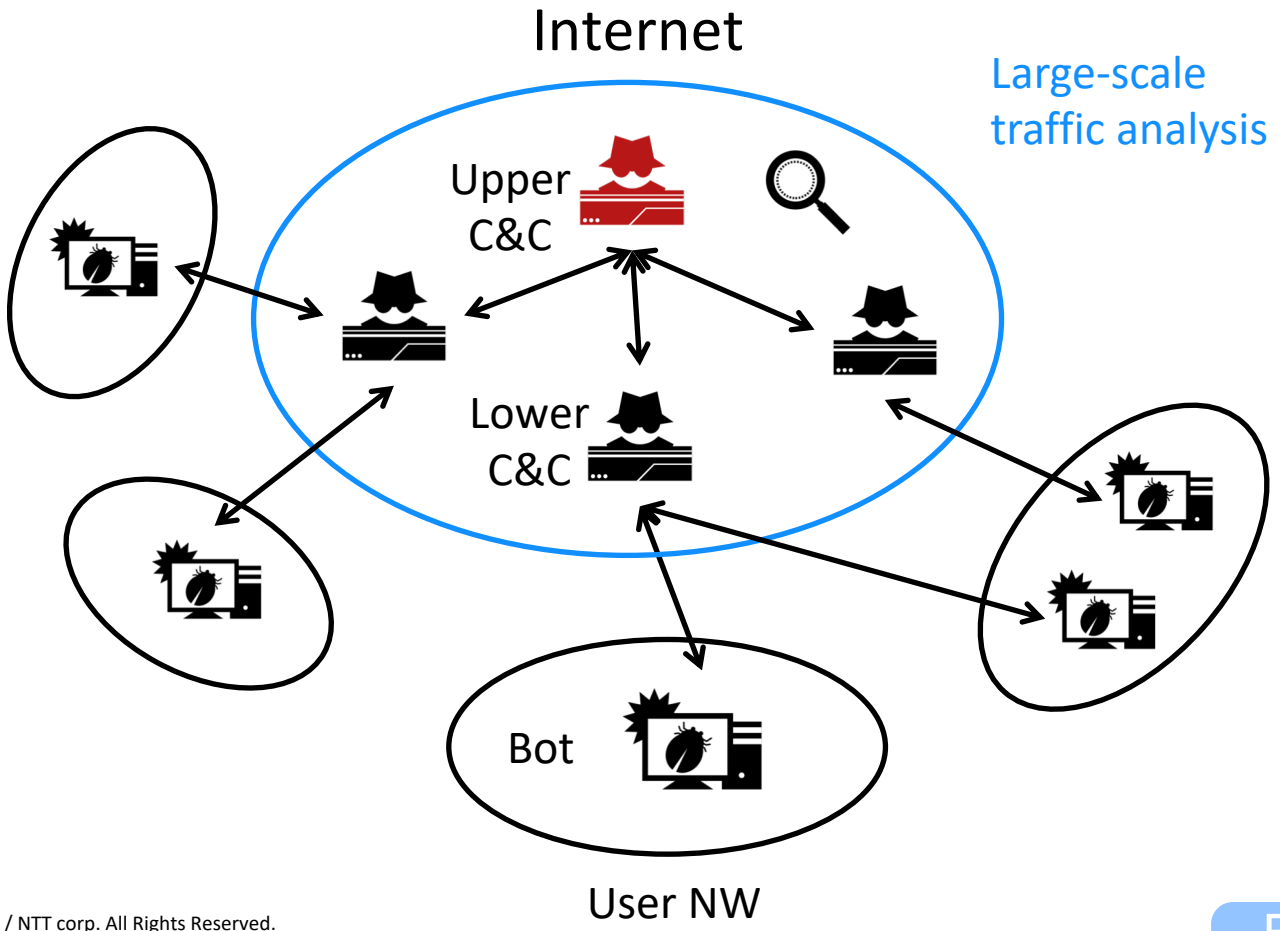
- **Machine learning** can automatically extract intelligence from large-scale network traffic
- **Botnet detection** is its one important application field.
  - e.g., bot detection, Command & Control (C&C) server detection



Problem Statement

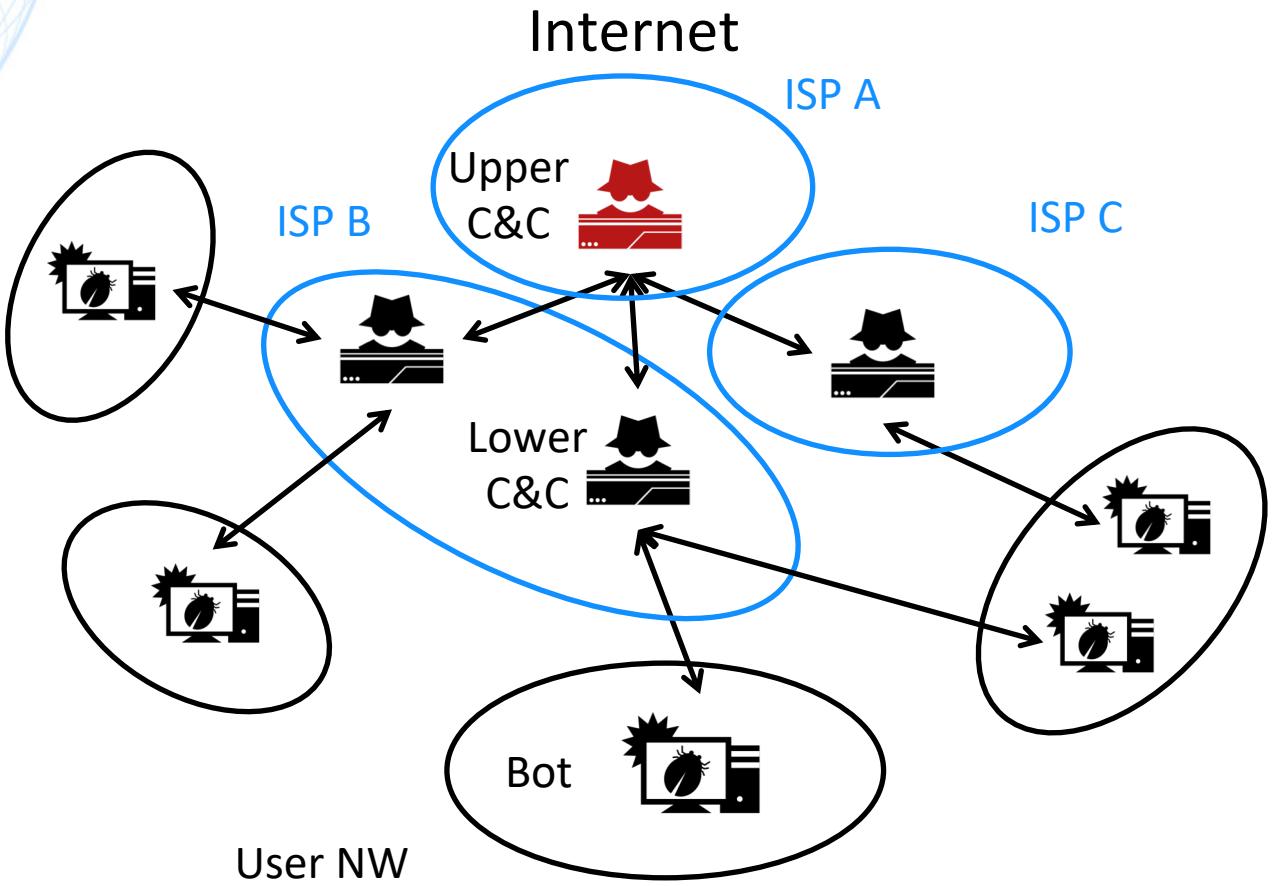
- Many existing methods focus on traffic analysis in a user network.
- However, they lack the visibility of layered and distributed botnet infrastructure.
- Large-scale traffic analysis at the Internet backbone is necessary.





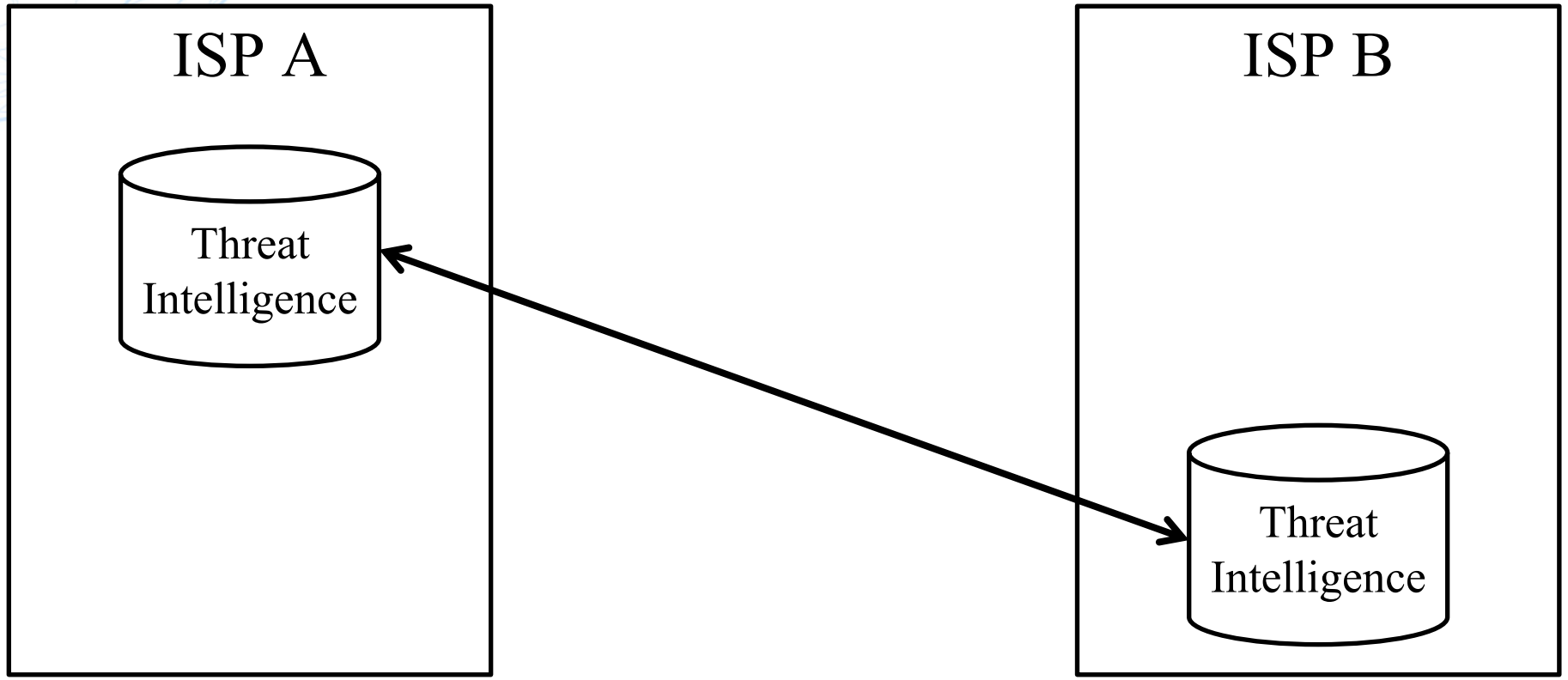
Problem Statement

- To comprehensively detect botnets through the Internet, collaborations among multiple ISPs are expected.



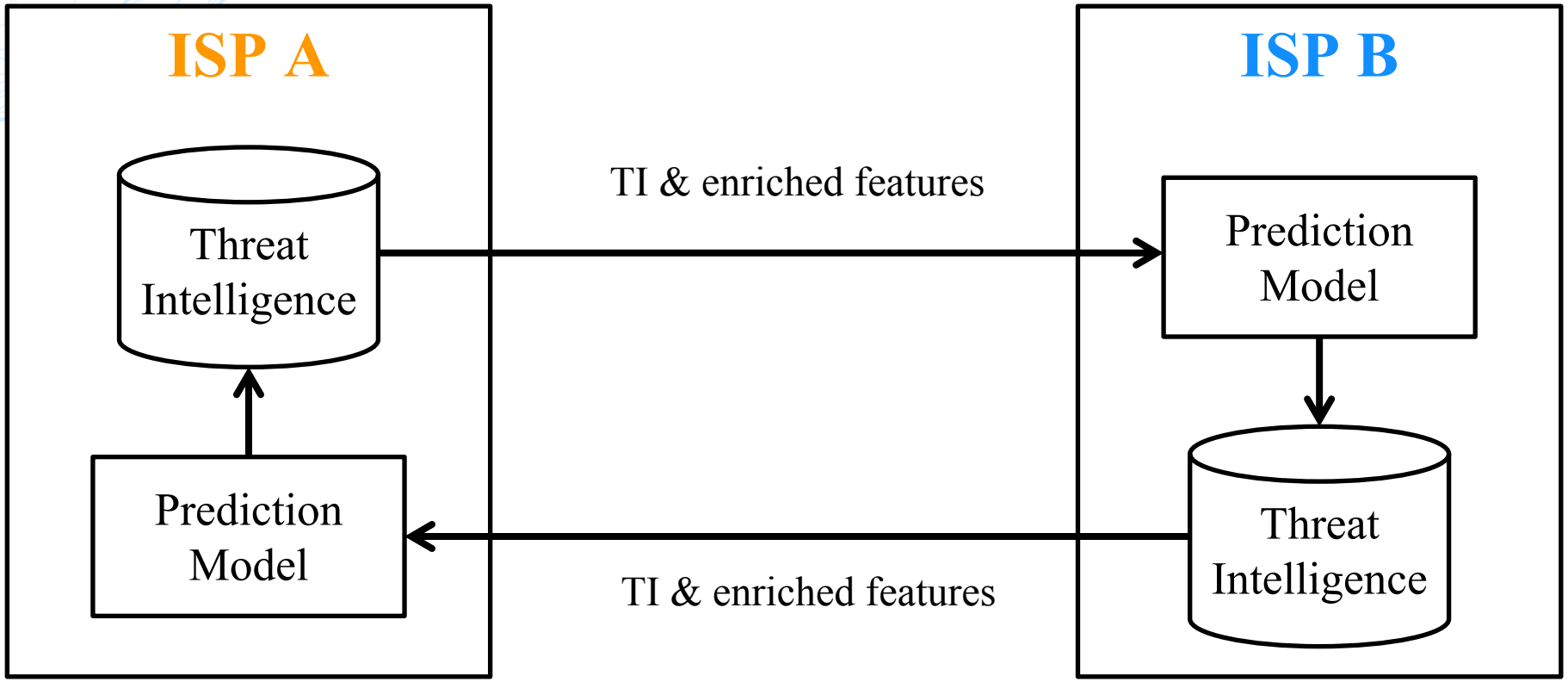
Problem Statement

- However, this is challenging, since each ISP has different techniques , traffic data and threat intelligence (TI).
- Considering the user privacy and difference of techniques, existing collaborations are limited to information exchanges.



Overview of Proposed Collaboration

- We introduce a collaborative framework which
  - Leverages TI/enriched features of malicious servers from other operators to enhance the existing prediction models, and
  - Shares newly generated TI/enriched features to others,
  - While preserving the privacy and confidentiality of communications

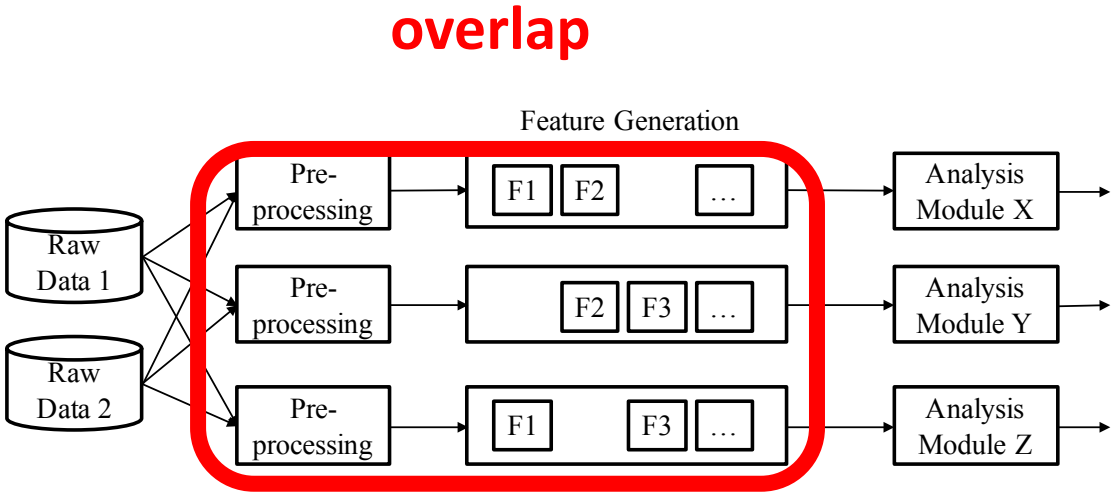


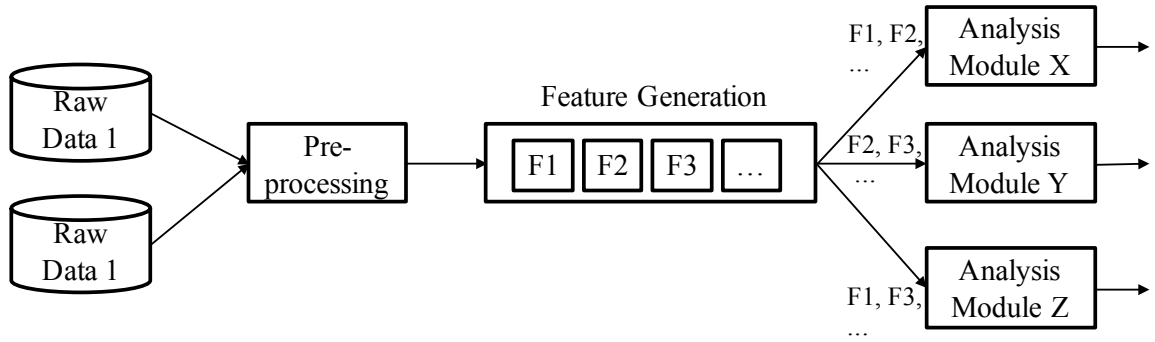
Piper

by NTT R&D

Overlap in Analysis

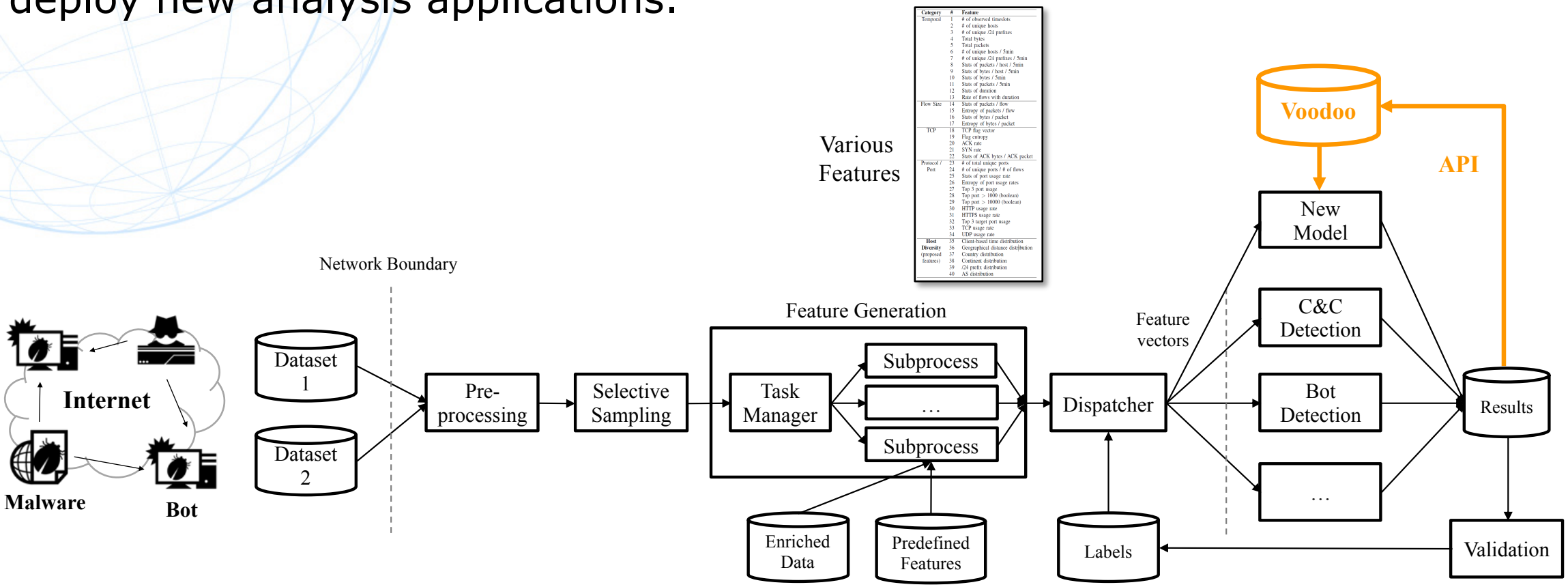
- Applying multiple applications **separately** to the same data may result in a huge **overlap**
- An **efficient** and **practical** platform for Internet-scale traffic is required





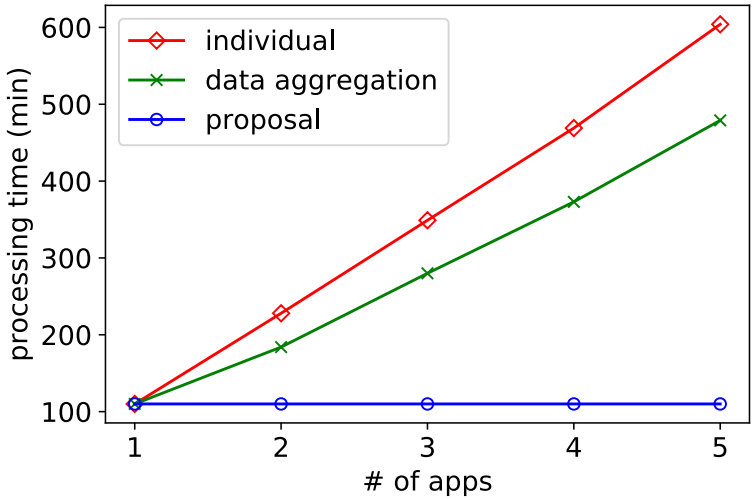
Piper: A Machine Learning Pipeline

- Piper: Unified machine learning pipeline for Internet-scale traffic analysis and threat detection.
- With its fast pipeline and predefined capabilities, ISPs/enterprises can easily deploy new analysis applications.

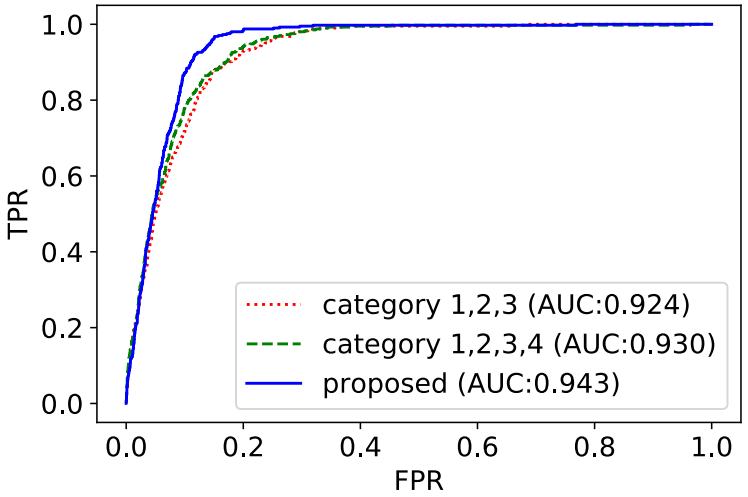


Piper: A Machine Learning Pipeline

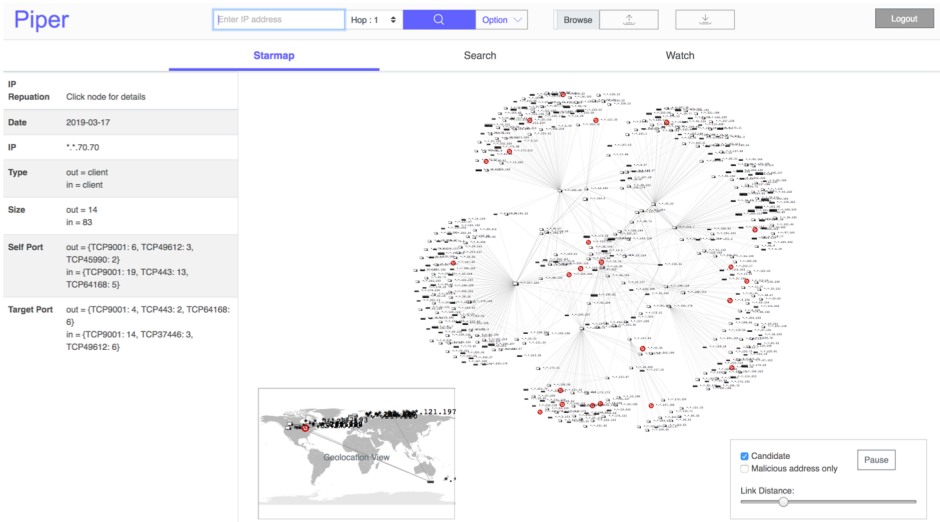
Faster processing



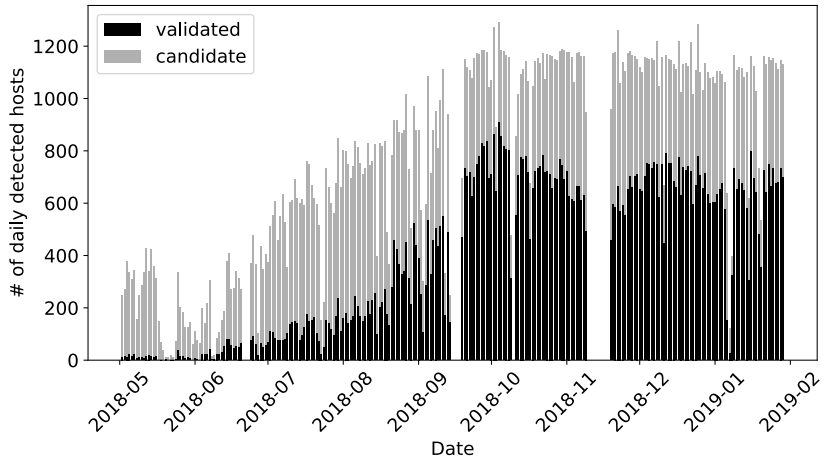
Higher Accuracy



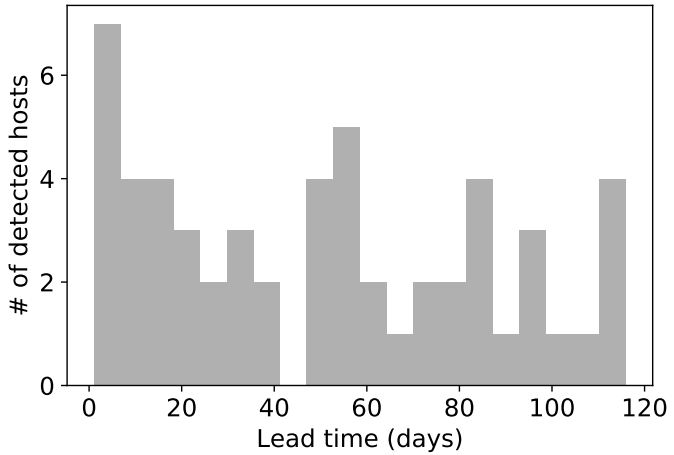
Graph-based GUI



Continuous Detection



Early Detection

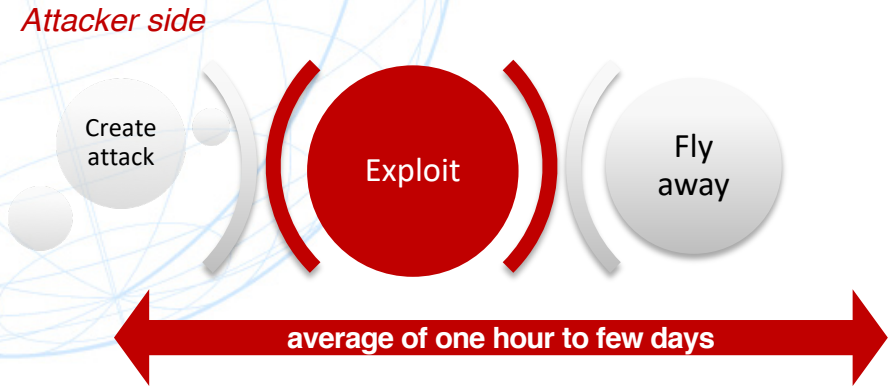


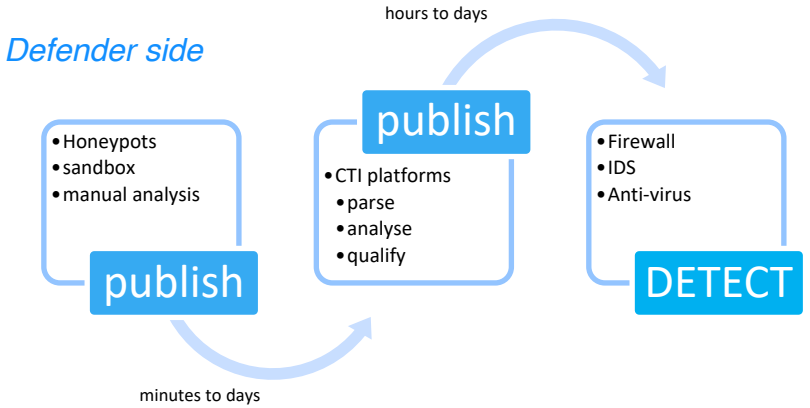
Voodoo

by Orange Labs

Voodoo
Origin

Going beyond of the Reactive CTI limits
Towards proactive CTI





77%* of indicators are already obsolete when reaching Threat Intelligence platforms:

- Information must be discovered, processed and published as fast as possible
- New information must come with an associated threat and context

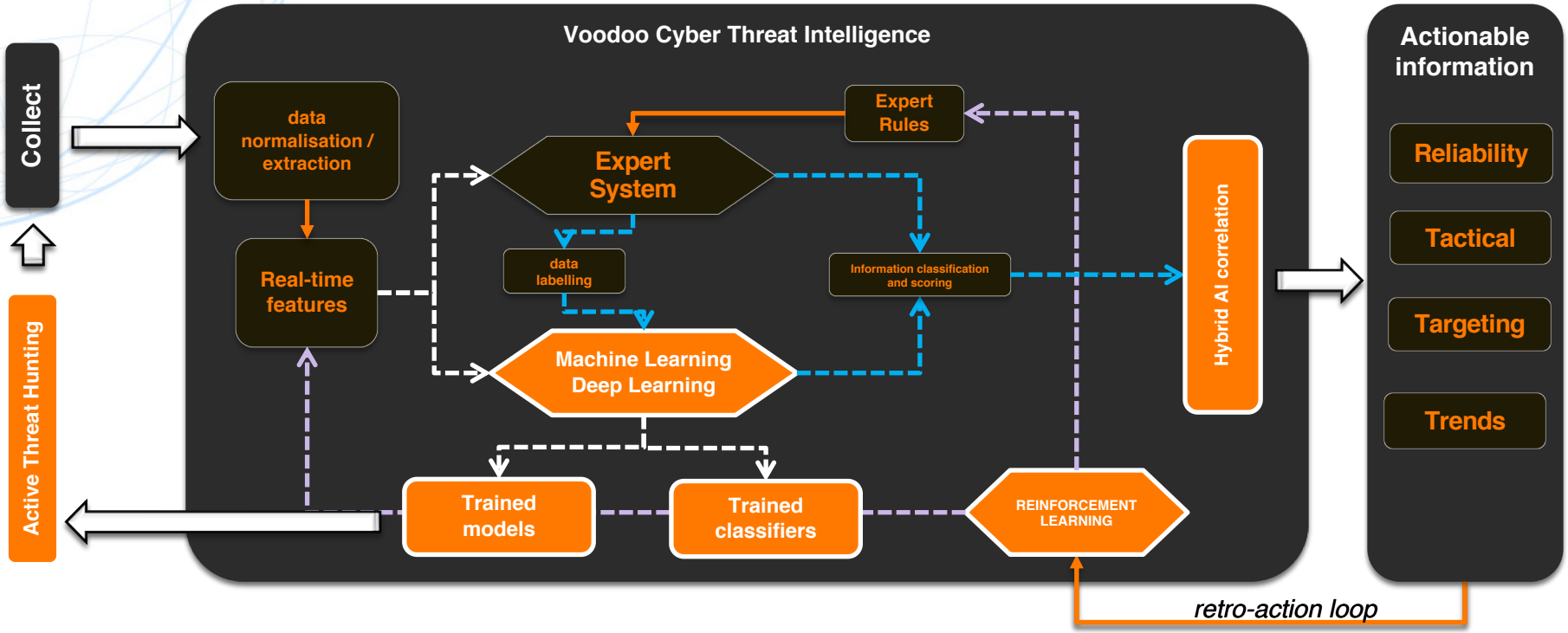
→ Requires new Threat Hunting algorithms based on threats and hackers profiling

*Based on Orange Innovation Study, 2020

Voodoo Insight

Hybrid AI based Threat Intelligence
mixing machine learning, deep-learning and expert system altogether

Voodoo: leverage CTI through Hybrid AI to proactively discover threats and produce valuable information



Voodoo Assets

Research Project Evaluation

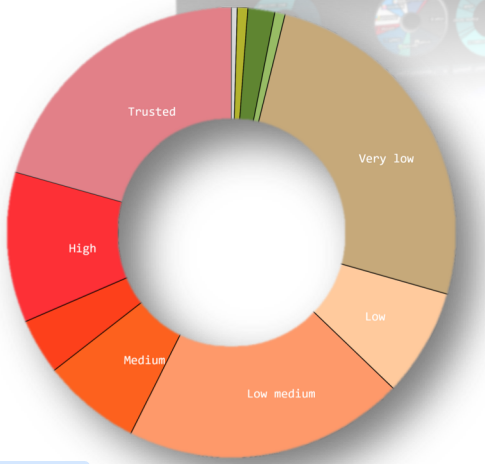
- 10M active IoCs (with an average of 1M renewed daily)
- 13M active IoAs (with an average of 1,5M renewed daily)

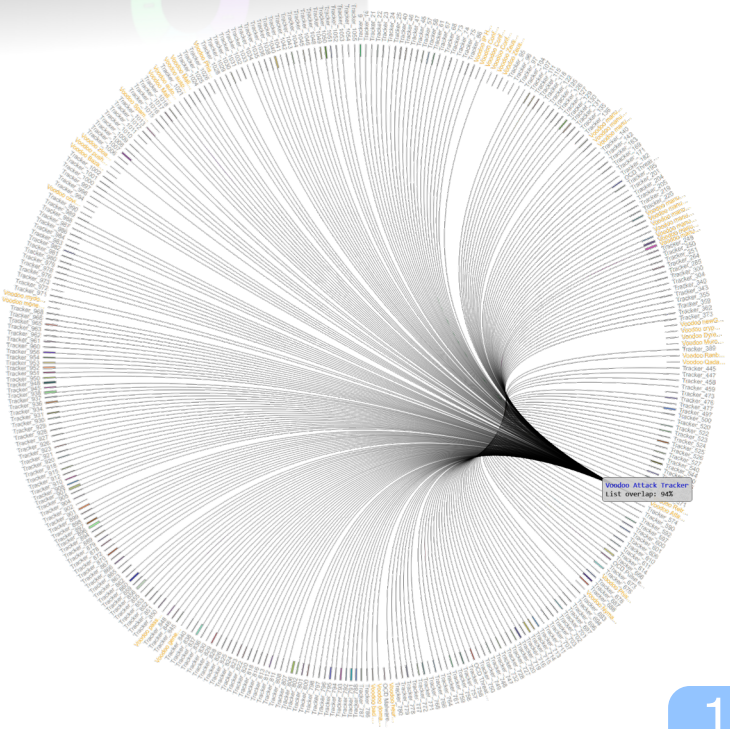
43 Voodoo sources (hunting, trackers, manual, ...)

correlated with 900+ external information sources

25% of Voodoo information is unique

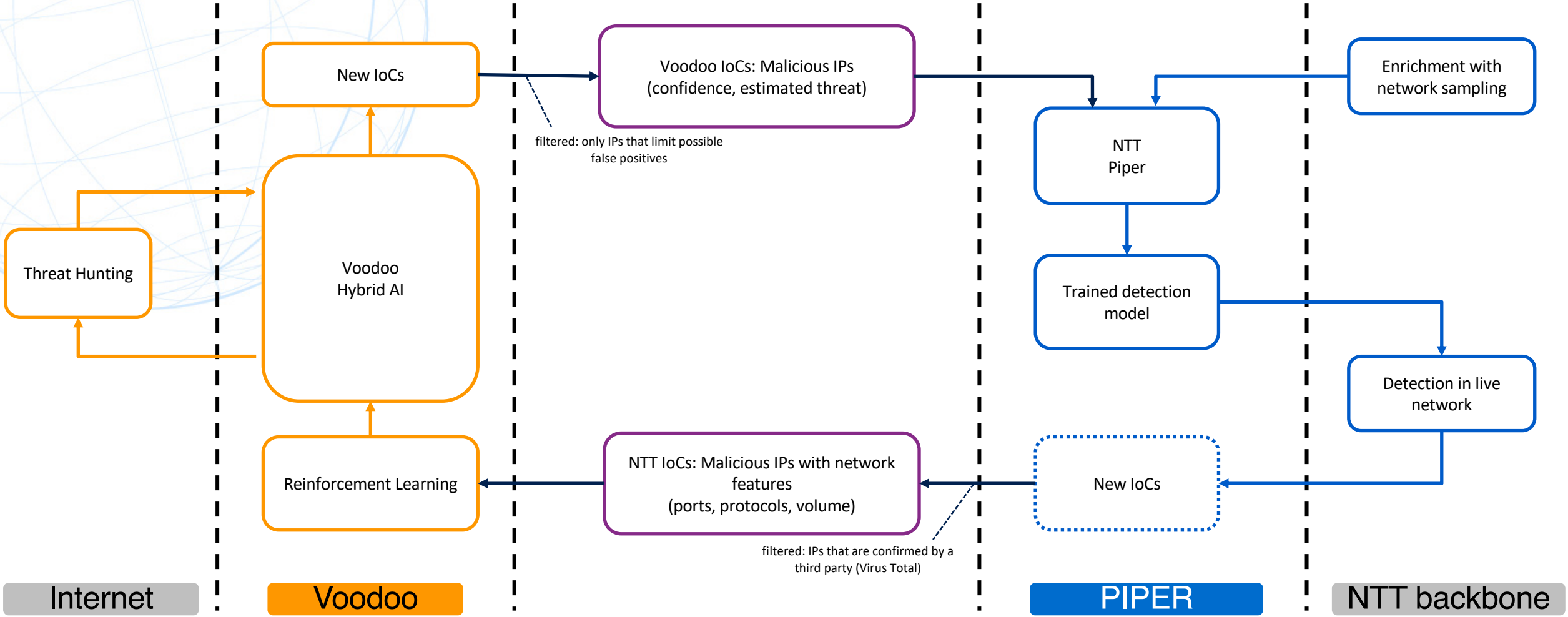
Information classified by threats, confidence, context, ...





Collaborative Framework

Collaborative Framework



Joint Experiment
Results

Orange & NTT joint experiment

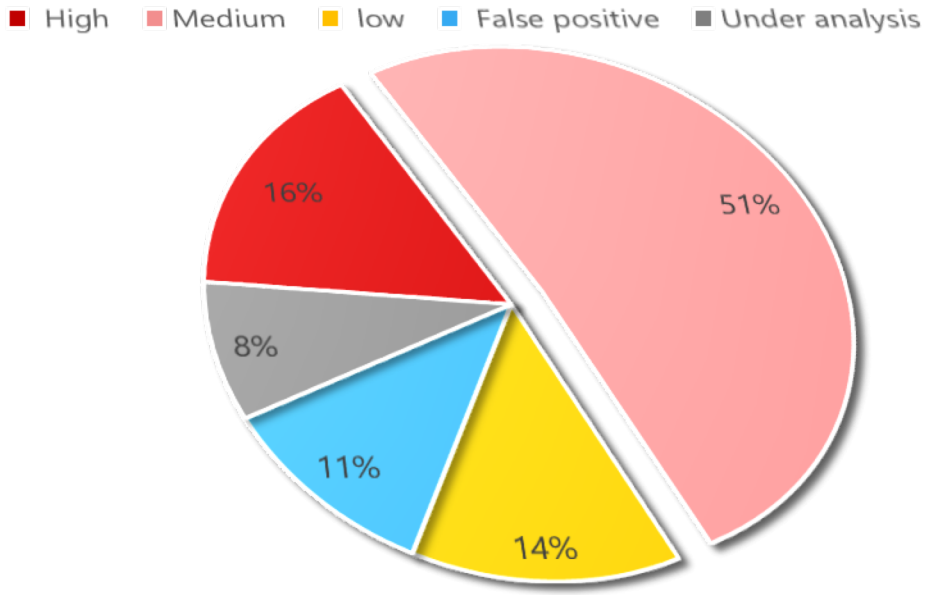
Voodoo
General statistics

652 distinct IP addresses reported by Piper in one month experiment

- **120 unique entries** - unknown by Voodoo
- **209 IPs referenced first** by Piper

Feed name	% Piper overlap
Virus total	27.9%
Warui ip malware	19.5%
Cert gov Georgia	14.8%
Zerocert	14.3%
Surbl	13,8%

Top 5 sources cross referencing



Computed reliability of Piper unique entries by Voodoo

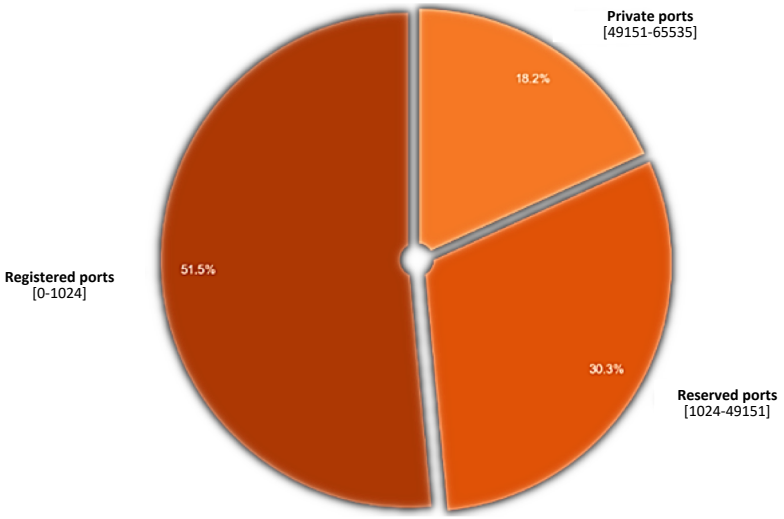
➔ 18.4% unique entries from Piper
➔ Ovelapped entries covered by an average of 11 feeds

Orange & NTT joint experiment

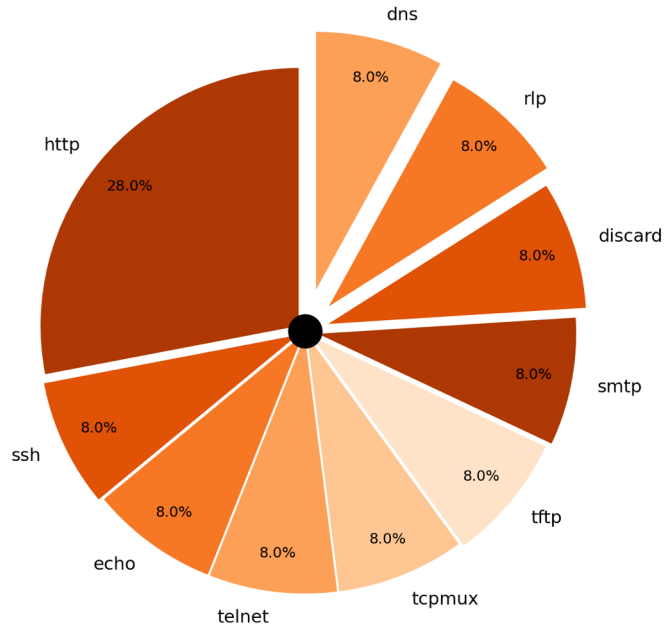
Voodoo
Enriched Threat Intelligence

- Voodoo has a limited access to live malicious traffic
- The collaboration with Piper helps to enrich Voodoo knowledge regarding malicious ports

➔ Piper allowed to enrich shared IoCs with 543 new ports of interest



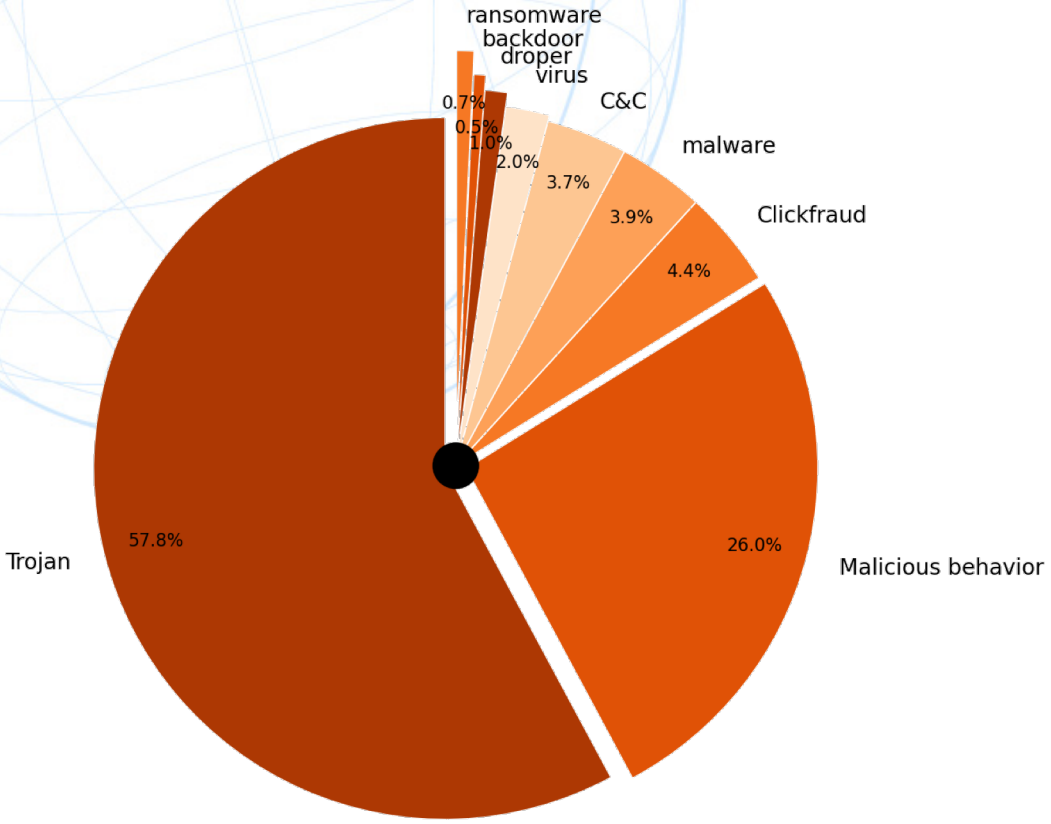
Ports reported by Piper



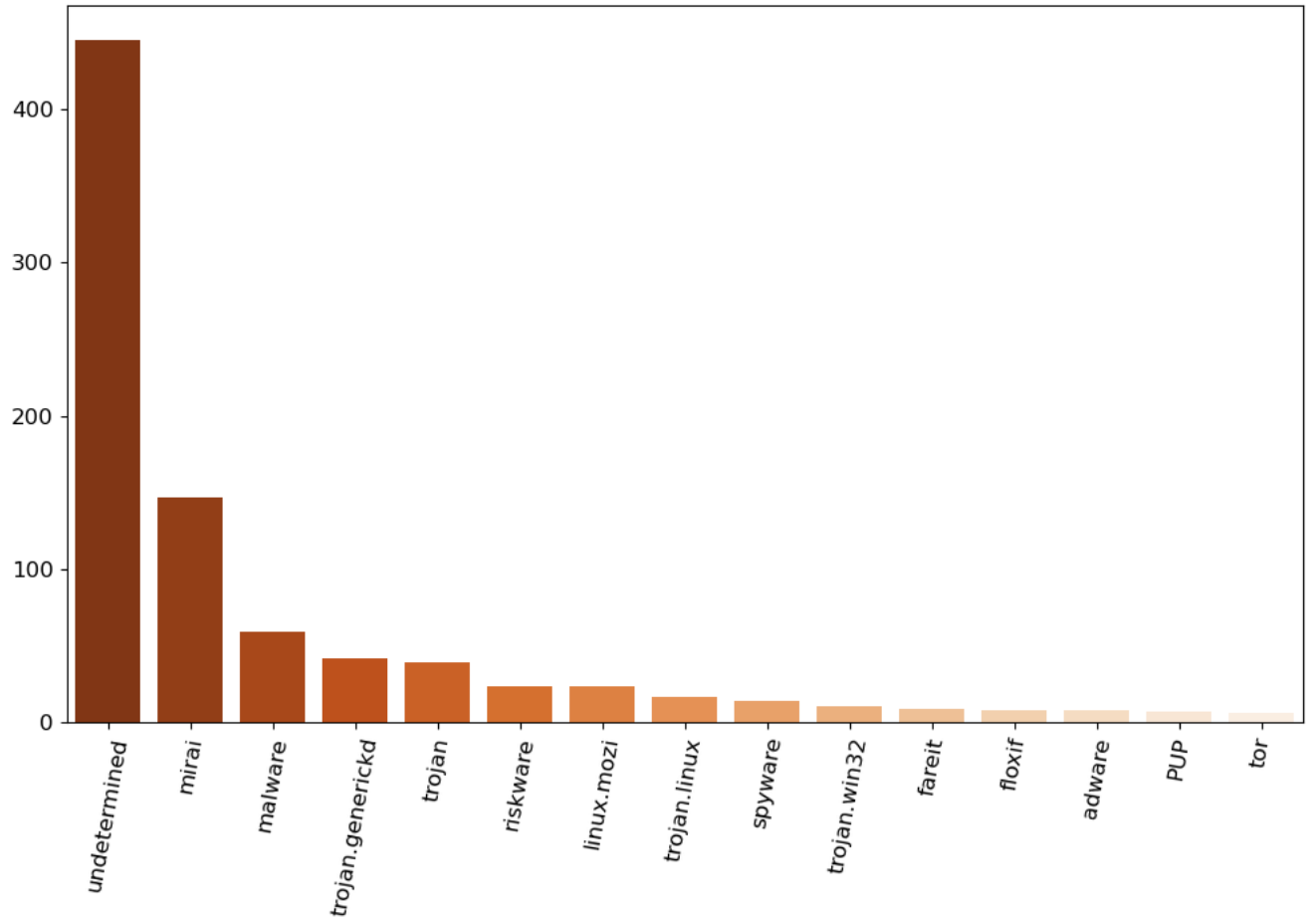
Top 10 registered ports

Orange & NTT joint experiment

Voodoo
Enriched Threat Intelligence



Threats family computed by Voodoo



Threats names computed by Voodoo

Orange & NTT joint experiment

Reinforcement learning
Reliability correction

Scope: IPs addresses reported by Piper, and already known (or with related URLs known) by Voodoo

False POSITIVE Correction

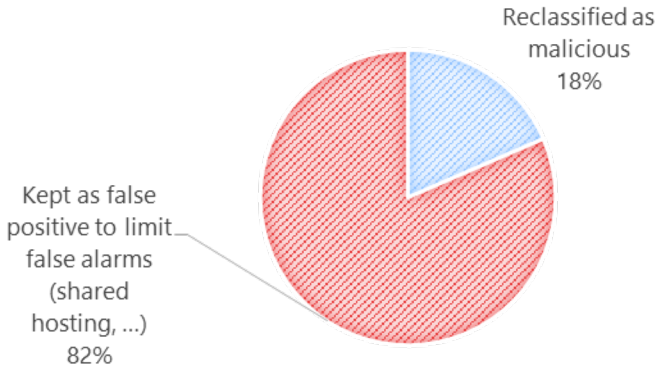
IoCs initially classified by Voodoo as false positive because there is :

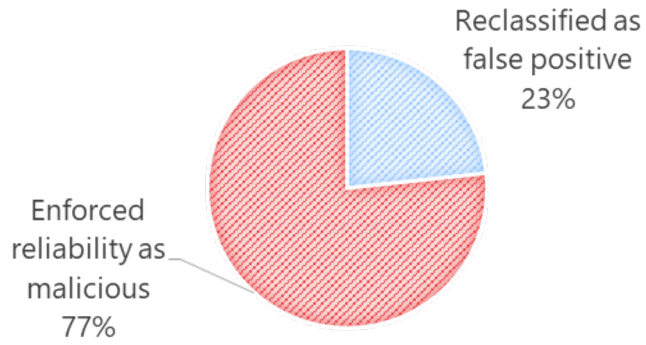
- no longer a feed referencing the IP addresses anymore
- not enough information context to classify IP addresses as malicious

False NEGATIVE Correction

IoCs classified with a low confidence on Voodoo side because there is no longer a feed referencing the IP addresses anymore.

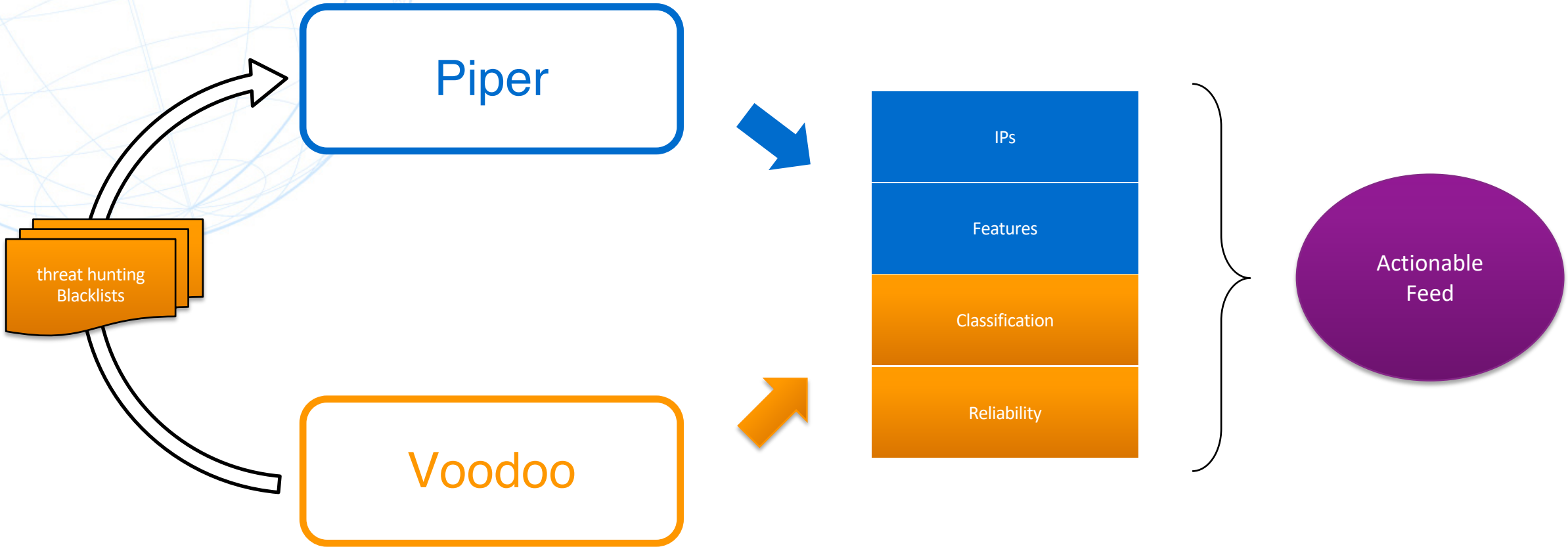
➔ Collaboration between Voodoo and Piper allows to reconsider these IoC's as active





Orange & NTT joint experiment

Collaboration processing
Win-win computation



Conclusion

Conclusion

- As joint experiments, NTT and Orange have introduced new approach to enrich and extend TI while preserving the privacy and confidentiality of communications.

- According to our preliminary results, the collaboration allows to detect new C&C servers and botnet related activities.

- This provides new opportunities to improve cybersecurity among ISPs, and to discover and extend detections to other malicious activities.

Thank you