INTERNET SCALE MALWARE ANALYSIS

ZACHARY HANIF, TAMAS K LENGYEL, AND GEORGE WEBSTER

OVERVIEW

- Who We Are
- Motivation
- Problem
- Framework
- Implementation
- Future Work

Zachary Hanif

- Director of Applied Data Science - Novetta
- Creator of Binary Pig - Feature extractor for large scale static analysis

NOVETTA

Tamas K Lengyel

- PhD student - University of Connecticut
- Senior Security Researcher - Novetta
- Creator of DRAKVUF - Dynamic analysis through VMI

NOVETTA

George Webster

- PhD student - Technical University of Munich
- Researcher - Cognitive bias in Cyber Defense

MOTIVATION

Create a system that can manage the full cyber analytic lifecycle for teams

- Fast: process large and historic sets of information
- Scalable: easily handle millions of samples
- Resilient: smartly handle errors
- Flexible: easily incorporate new methods

OVERVIEW: WHAT WE DID

Skald:

- A blueprint for creating file/malware analytic systems

Totem:

- Scalable static analysis

Drakvuf:

- Invisible, hypervisor level dynamic analysis

Current State

CURRENT STATE: TOOLS

Analysis tools focus on binaries

- Multi-stage droppers, metamorphic malware
- It costs nothing to create a new one

Defensive tools focus on signatures

- Snort, YARA, Antivirus..
- We are playing catch-up

CURRENT STATE: ACTOR

Actors are growing in sophistication

- Organized Crime
- Nation State

Malware no longer exists in a vacuum

- Teams with sophisticated tools, infrastructure, analysts, financial networks

CURRENT STATE: VOLUME

We can’t keep up

- In 2012, 200k samples a day
- In 2015, 1 million samples a day
- Cannot analyze historic data

Our tools are disjointed

- One shot wonders
- Heavy reliance on a single person to make the connections between datasets

The lone reverse engineer is not enough

ANALYSIS EVOLUTION: 5WS

We need to empower a research team to figure out what is happening

- Who is behind the action
- What are their goals
- Where is the infrastructure
- When do they operate
- Why are they conducting the operation
- How do we thwart their activities

ANALYSIS EVOLUTION: TOOLS

We are in the age of “Big Data”

- We need defensive tools that leverage Big Data

We need to use multiple techniques to make sense of what we gather

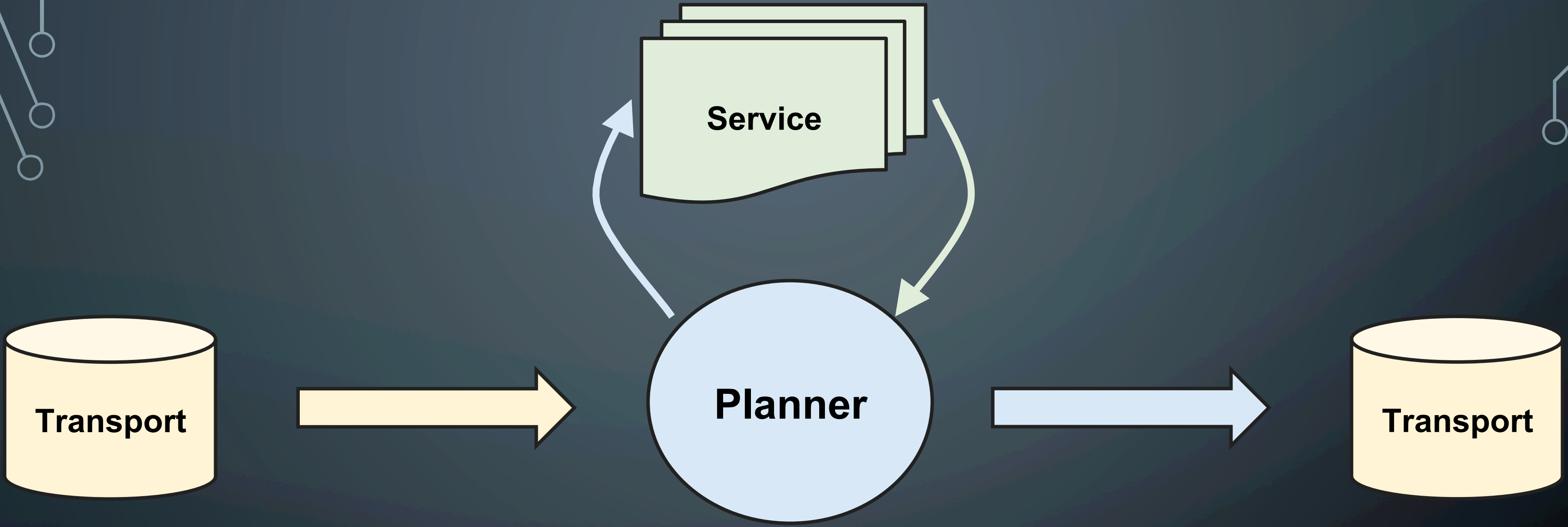
- Machine learning
- Graphical representation
- Reverse engineering

We need a new approach to create analytic systems

SKALD INTRODUCTION

Microservices based framework to cover the full analytic lifecycle

- Support 100’s of millions of objects
- Allow expanding to public/private infrastructure
- Provide the infrastructure needed to gather information and perform analysis on the data
- Support 3rd parties to provides parts of the system
- Able to incorporate existing tools



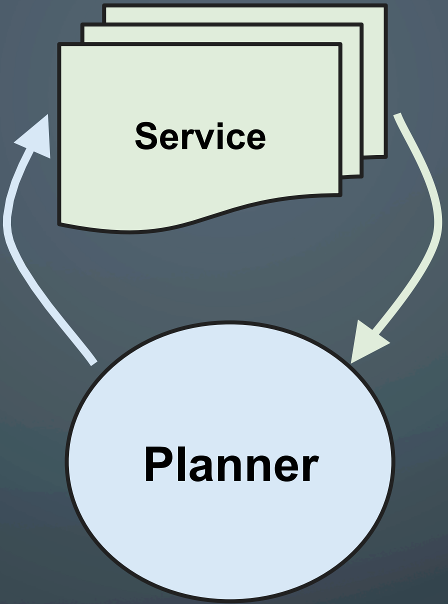
Core Components

Perform individual tasks

Independent and only interact with the Planner

Core Components

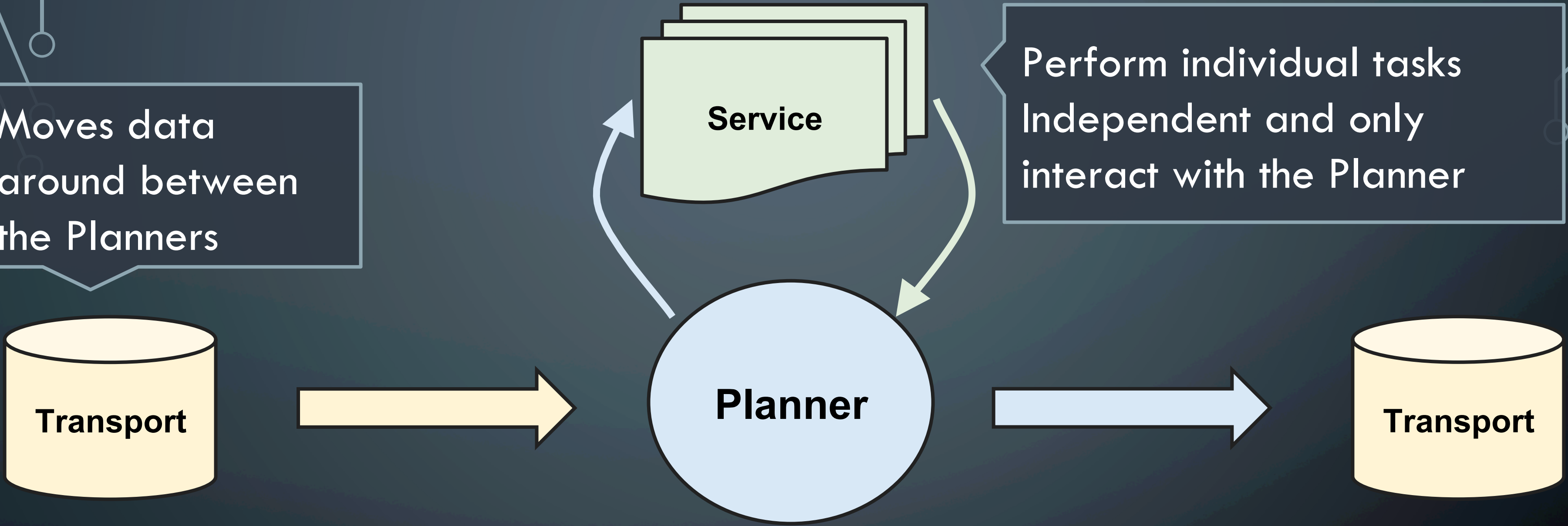
Core Components



Perform individual tasks
Independent and only interact with the Planner

Schedule Service execution
Perform QoS
Packages Services together for optimization

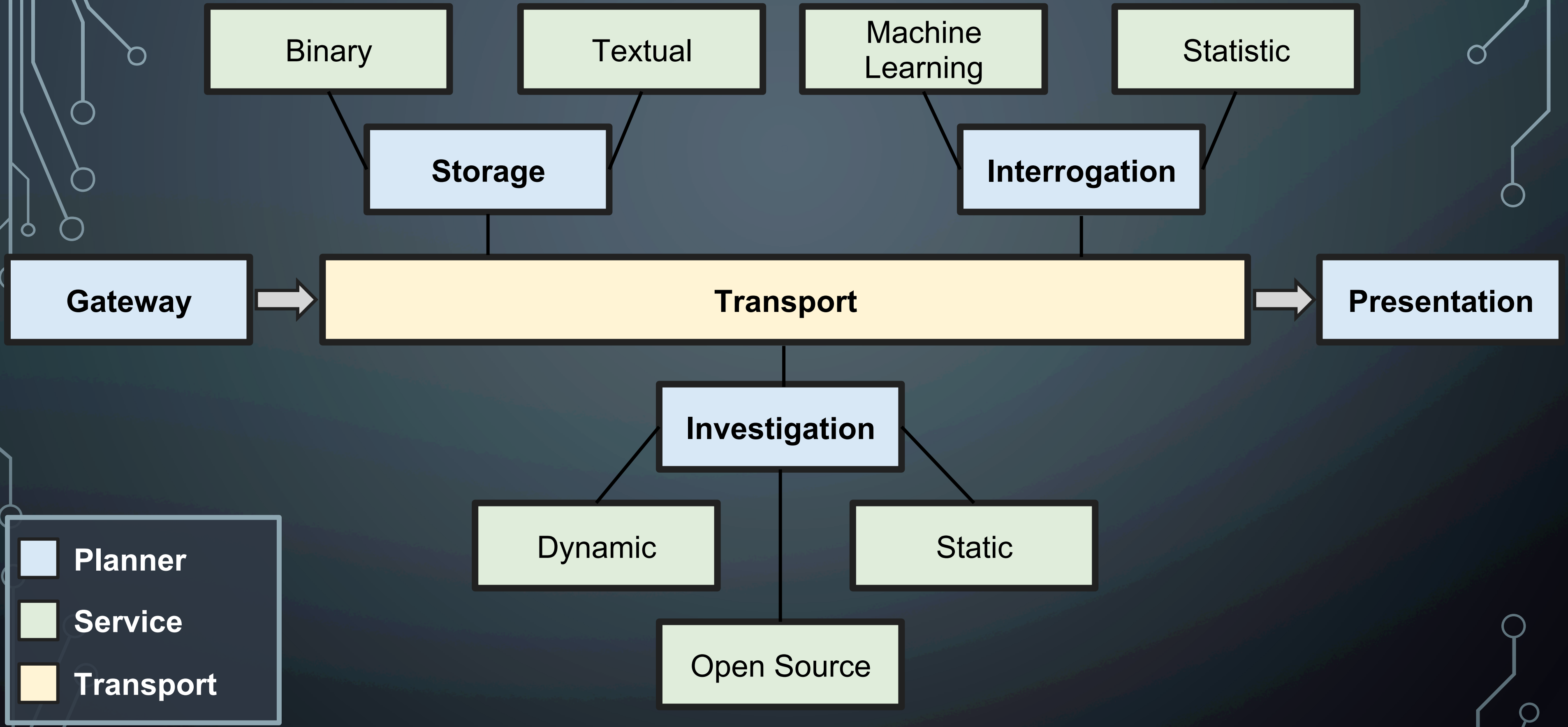
Core Components



Moves data around between the Planners

Perform individual tasks
Independent and only interact with the Planner

Schedule Service execution
Perform QoS
Packages Services together for optimization



The Big Picture

Planner: Interrogate

PLANNER: INTERROGATE

Focuses on analyzing one object

- Static analysis
- Dynamic analysis
- Gather 3rd party information

For example: Cuckoo, Drakvuf, PEInfo, VirusTotal, Yara, …

Benefit: Allows you to easily add new feature extraction techniques

Planner: Investigate

PLANNER: INVESTIGATE PART 1

Focuses on how to analyzing already gathered information

- Machine Learning
- Statistical analysis

For example: Clustering, pattern matching, behavioral analysis, …

Benefit: Allows you to plug in analytic methods to run over already existing data

PLANNER: INVESTIGATE PART 2

Focuses on how to display the information in the system

- Service dedicated to displaying data
- API interface, Web portal, etc

For example: a Service to display information to IDAPro, Maltego, custom web frontend, etc

Benefit: Allows you easily extend or change your system display

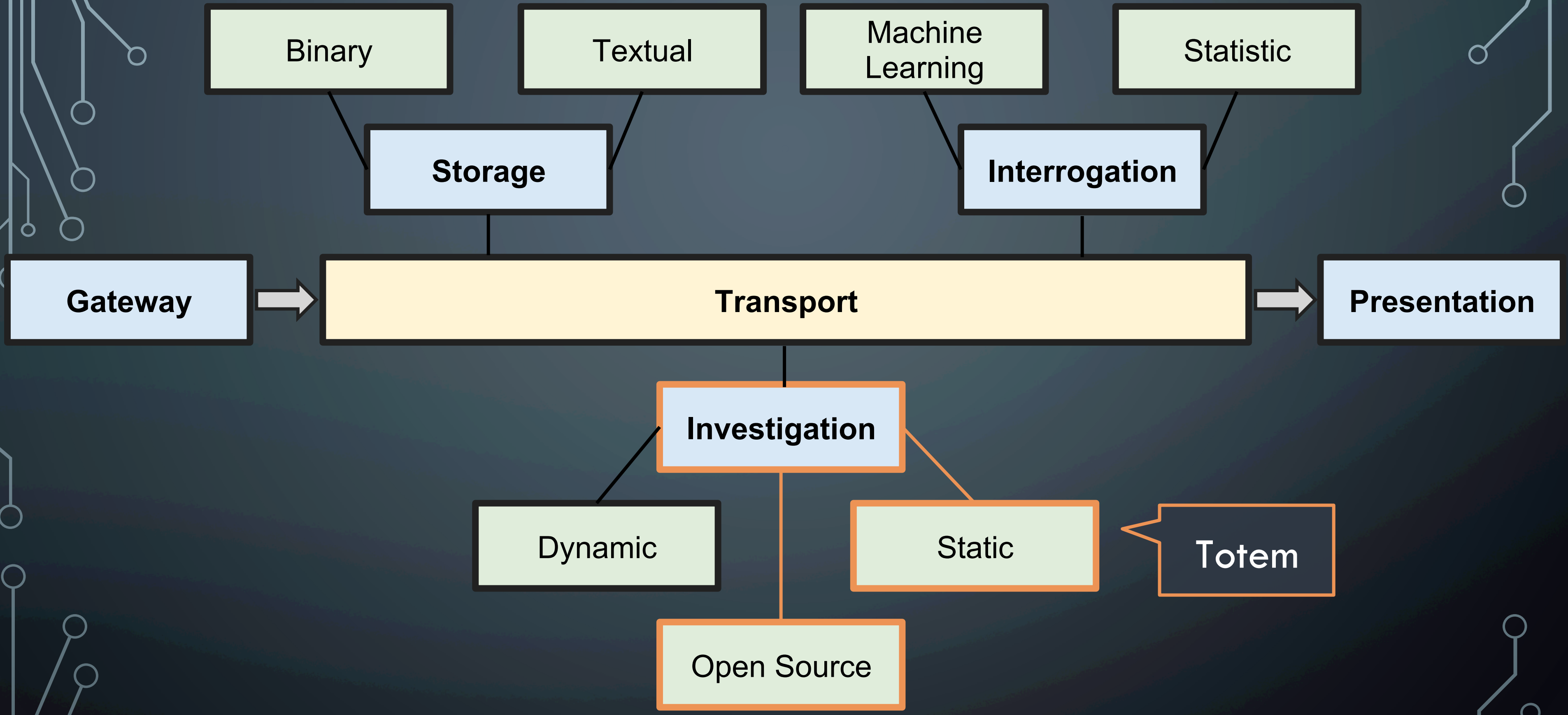
We created the framework … now what

WE CREATED THE FRAMEWORK … NOW WHAT

We do not have a complete solution. We created the design principles with Skald and are working to put the pieces together.

Will discuss:

1. Implementation for static analysis, TOTEM
2. Implementation for dynamic analysis, DRAKVUF



The Big Picture - Totem

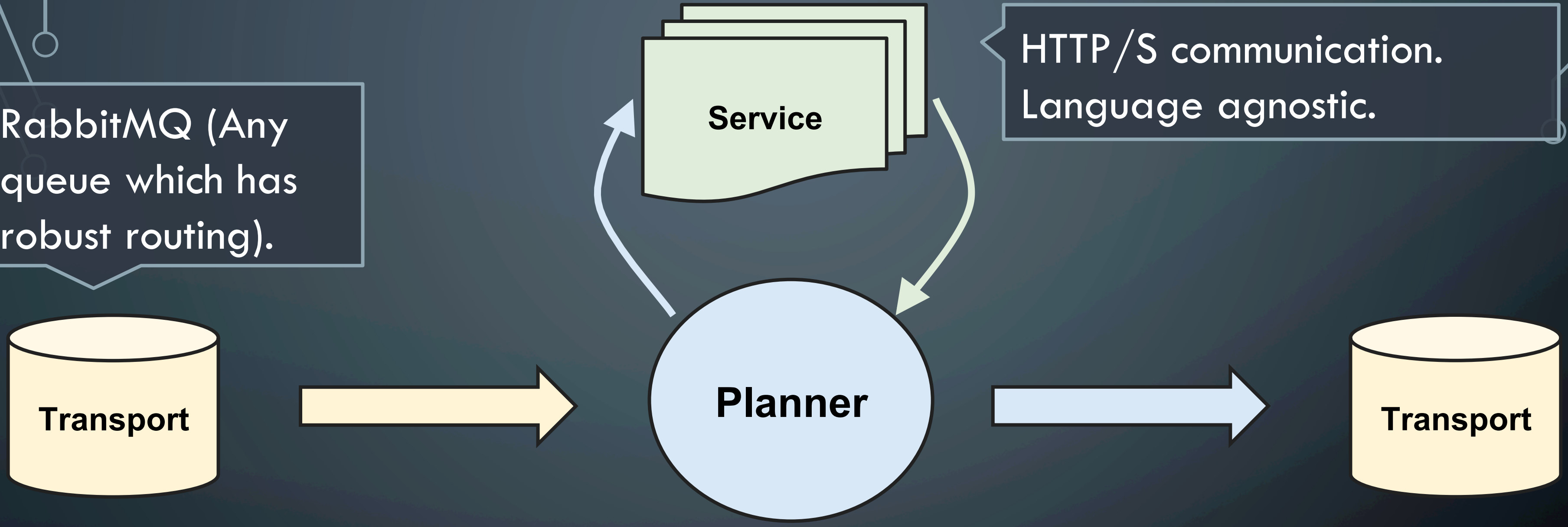
TOTEM

Static analysis on large file datasets

- Historical and streaming analytical workloads

Implemented with the Skald framework

- Asynchronous, easily extendable
- Scalable, fast, flexible, resilient to failures



RabbitMQ (Any queue which has robust routing).

HTTP/S communication.
Language agnostic.

TOTEM worker.
Manager RMQ communications, Service communications, and result resolution.

Core Components

TOTEM

Scale is a pain - dynamic deployment is a hard problem in practice

- async, loosely coupled architecture lets us scale to large datasets and workloads

Totem's speed comes from the ability to join workers dynamically

TOTEM - PERFORMANCE

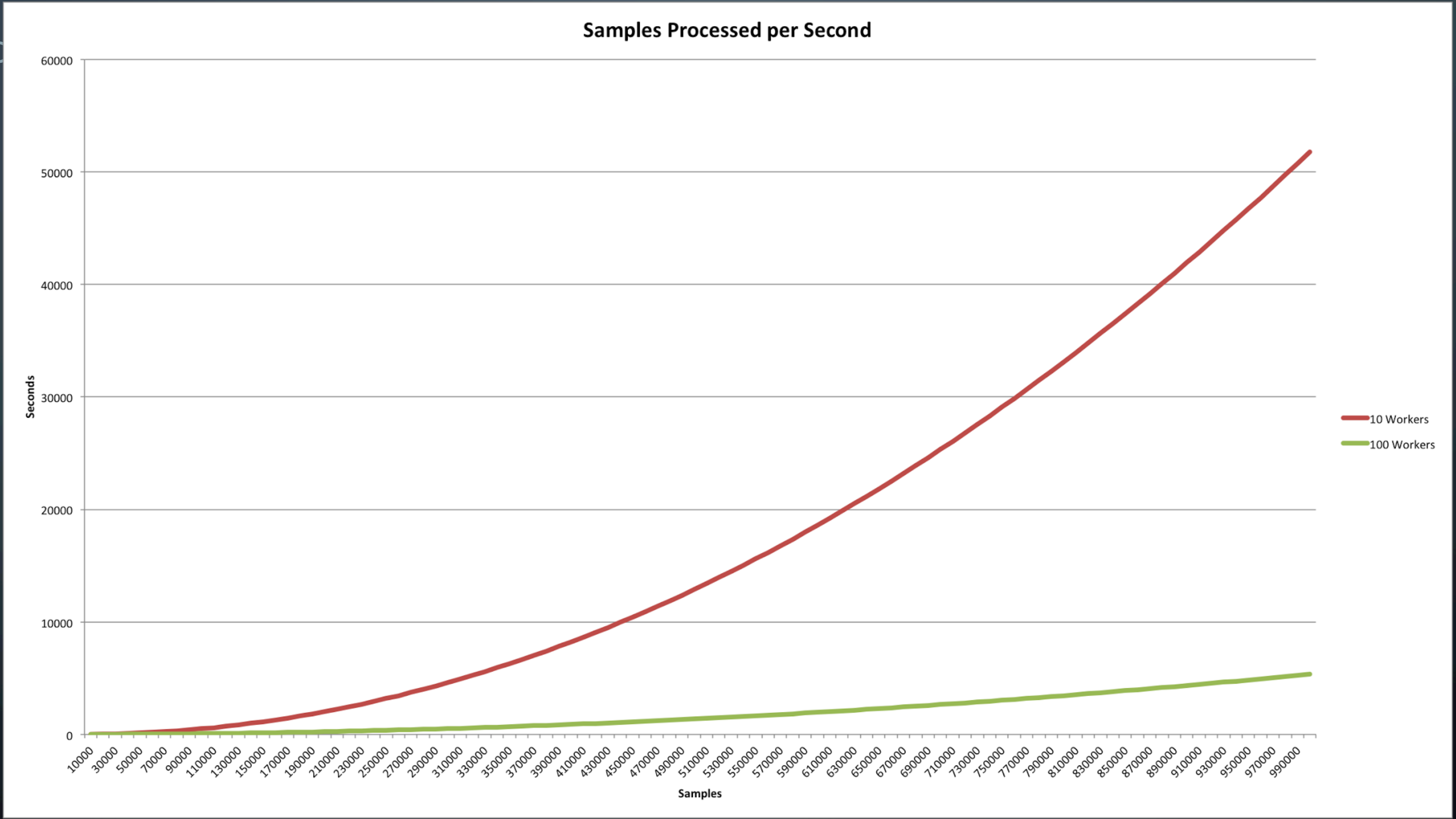
	1K Samples		5K Samples		10K Samples		50K Samples	
Framework	**Time**	**Error**	**Time**	**Error**	**Time**	**Error**	**Time**	**Error**
CRITS	2.8000	0	3.1774	0	3.3781	151	1.1929	17012
TOTEM 3 Workers	0.0502	0	0.0558	0	0.0616	0	0.1303	0
TOTEM 100 Workers	0.0032	0	0.0032	0	0.0032	0	0.0025	0

TOTEM - PERFORMANCE

	1K Samples		5K Samples		10K Samples		50K Samples	
Framework	**Time**	**Error**	**Time**	**Error**	**Time**	**Error**	**Time**	**Error**
CRITS	2.8000	0	3.1774	0	3.378[illegible]	[illegible]151	1.1929	17012
TOTEM 3 Workers	0.0502	0	0.0558	0	0.0616	0	0.1303	0
TOTEM 100 Workers	0.0032	0	0.0032	0	0.0032	0	0.0025	0

Scale Issues

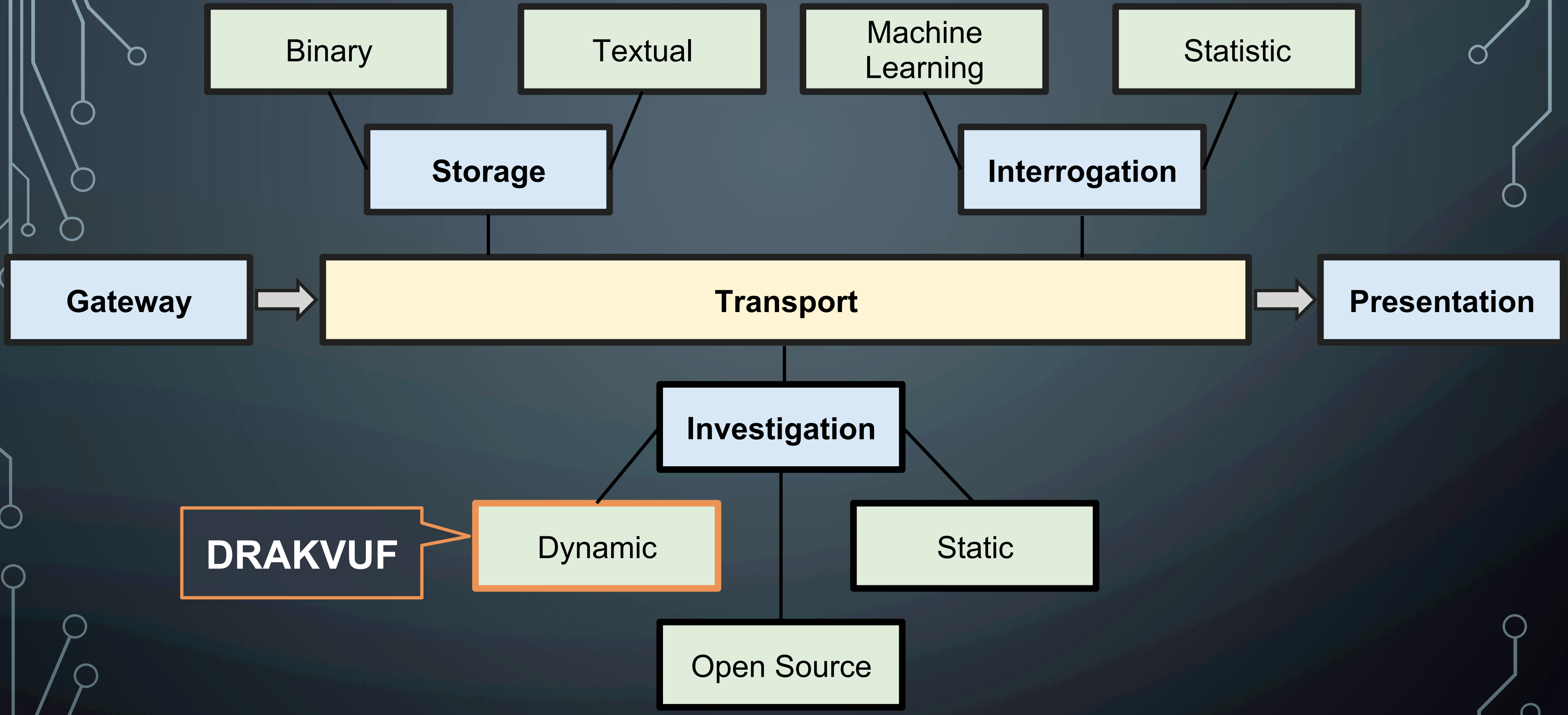
Cache Speed-up



TOTEM – Performance Million +

ANALYTIC SERVICES

- CRITs Services parity
- Support for 3rd party services: VT, etc.
- Resource extraction (PE32, PDF, etc)
- PDF, Office, HTML, JS parsing and analytics support



The Big Picture - DRAKVUF

DRAKVUF

- Agentless dynamic analysis
- Open source: http://drakvuf.com
- Monitoring via the Xen Hypervisor
- Natively supported - no custom patching!

DRAKVUF

Stealthy

- Monitoring via Intel virtualization extensions

Scalable

- Copy-on-write disk and memory

Resilient

- Complete, unhindered view of the execution

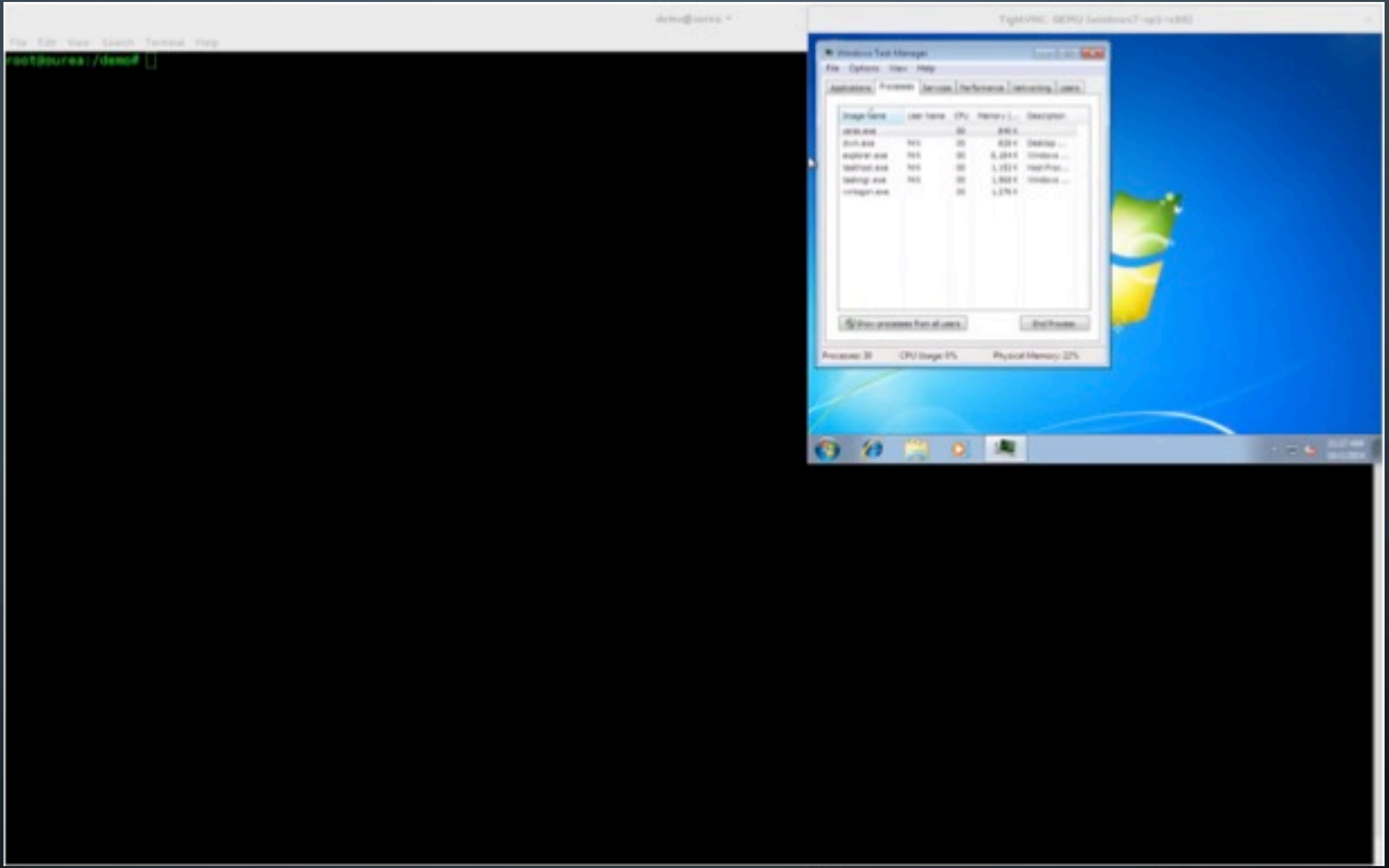
DRAKVUF

Monitor all types of malware, including kernel-mode rootkits

- Monitor system calls, heap allocations & scheduling

Start the execution of the malware by injecting a new process into the VM!

- Hijack any existing process to start the sample



```
root@ourea:/demo# 
```

File Edit View Search Terminal Help
TightVNC: QEMU (windows7-sp1-x86)
Windows Task Manager
File Options View Help
Applications Processes Services Performance Networking Users
Show processes from all users
End Process

Image Name	User Name	CPU	Memory (...	Description
csrss.exe		00	840 K	
dwm.exe	[illegible]	00	828 K	Desktop ...
explorer.exe	[illegible]	00	[illegible]	Windows ...
taskhost.exe	[illegible]	00	[illegible]	Host Proc...
taskmgr.exe	[illegible]	00	[illegible]	Windows ...
winlogon.exe		00	[illegible]	

Processes: 30 CPU Usage: 0% Physical Memory: 22%

DRAKVUF - Demo

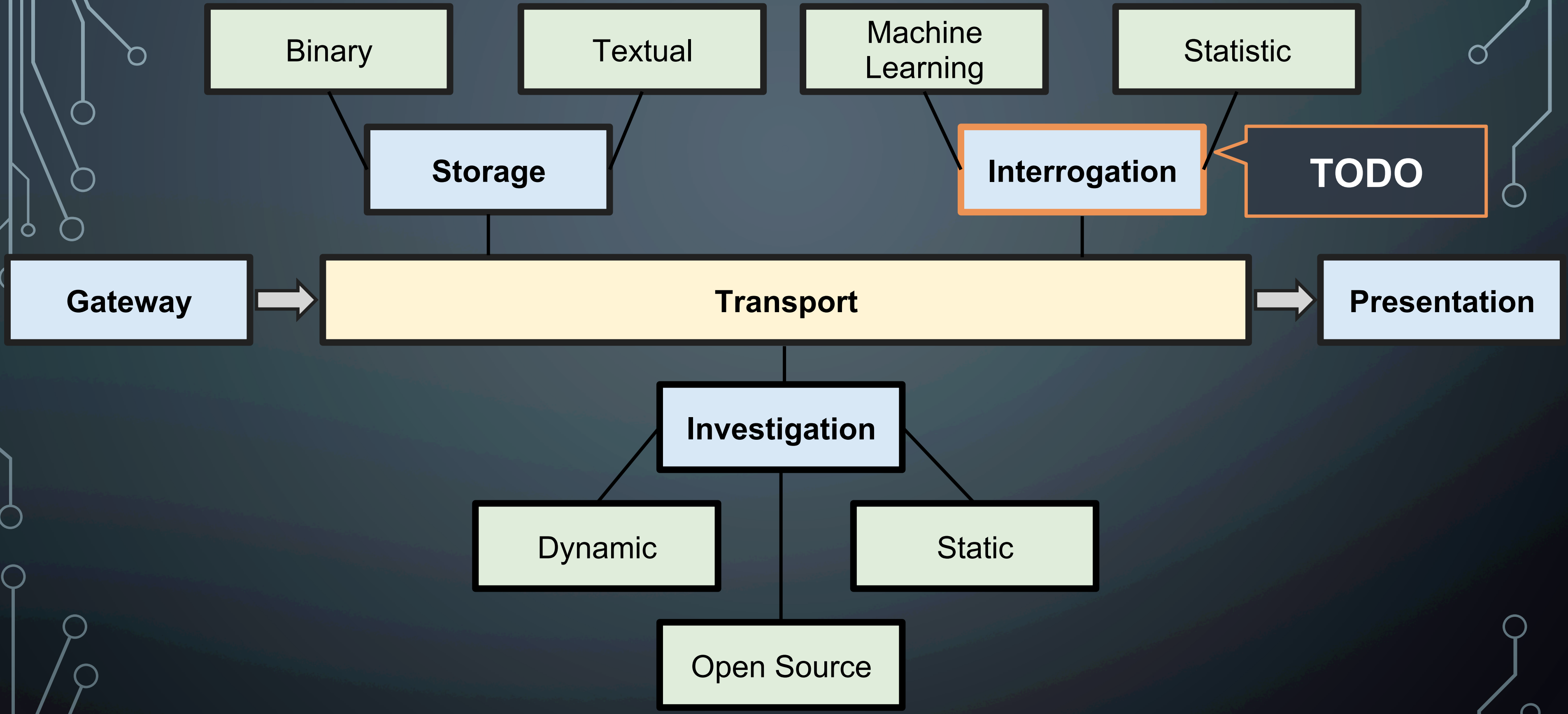
DRAKVUF

Cloud ready

- Base features are all available since Xen 4.3
- Major rework and extensions in Xen 4.6
- Ask your provider to deploy Xen Security Modules!

Analyzing mobile malware

- Basic tracing implemented for ARM in Xen 4.6
- Work in progress



The Big Picture - DRAKVUF

INTERROGATION GOALS

We got the data.. now what?

- Focusing on individual samples is a losing battle
- Modern malware has an infrastructure behind it

Fighting malware effectively requires context

The Age of Big Data

- Identify families, heredity, campaign actors and infrastructure

INTERROGATION SERVICES

Make data accessible in different forms

- GUI for human analyst
- Structured data for Machine Learning

Retain historical datasets

- A data point may mean nothing today but everything in 6 months
- Has to be searchable

OPEN SOURCE DUMP

TOTEM (new) NOVETTA

- Service analytics
- HTTP fileserver
- Static analysis transfer framework

DRAKVUF (already opensource)

- Log parser for TOTEM being released now

FIND US AT BLACK HAT

We will be located at the Novetta Booth. Would love the opportunity to discuss this further with you all.

BLACK HAT SOUND BYTES

We made the foundations for large scale malware analysis

We are releasing the first steps:

- Static analysis, TOTEM - http://totem.novetta.com?
  - 100k samples with 3 analytics in 9 minutes
- Dynamic analysis, DRAKVUF - http://drakvuf.com
  - Stealthy hypervisor based dynamic analysis

THANK YOU!

- Claudia Eckert
- Andre Ludwig
- Sebastian Vogl
- Joe LeGasse
- Yara Exchange
- Novetta
- DARPA
- Volatility crew
- Zentific
- VirusTotal
- And countless others