
BLOX FEST
Infoblox

The Front Doors Are Locked, But DNS is Wide Open - Preventing Data Exfiltration via DNS

Jozef Krakora, Senior Product Manager, Security
Bin Yu, *Ph.D.*, Analytics Architect
T. J. Short, CISO and VP of Infrastructure, Everi

BLOX FEST

Topics of Discussion

- What kind of data is stolen and why
- Prevalence of data exfiltration and DNS tunneling
- How data is stolen via DNS tunneling
- Detection strategies
  - Behavior analysis (Patent pending)
  - Machine learning
  - Artificial intelligence
  - Architecture
- Detection results (Conference paper)
- Real-time detection and mitigation solution architecture
- Customer case study (T. J. Short, CISO, Everi)

Stealing Data – Why and What Kind?

PII (Personally Identifiable Information)	Information like social security numbers of employees or customers that cybercriminals can use to steal identity, or sell in the underground market for profit
Regulated Data	Data related to PCI DSS and HIPAA compliance that can be misused
Intellectual Property	Data that can give an organization a competitive advantage
Other Sensitive Information	Credit card numbers, company financials, payroll and emails

Hacktivism

Espionage

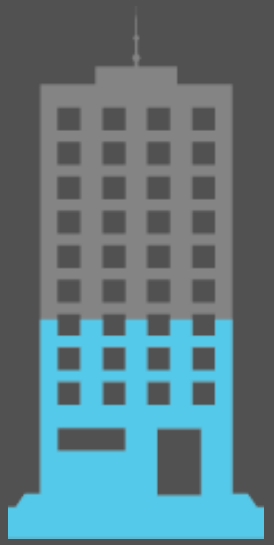
Financial Profit

BLOX FEST

DNS and Data Exfiltration

DNS tunneling attacks let infected endpoints or malicious insiders exfiltrate data.

Attackers have recently used DNS tunneling in cases involving the theft of **millions of accounts.**[5]



46%

of large businesses have experienced DNS exfiltration.[6]

$3.8 M

Average consolidated cost of a data breach[7]

Goal of Malicious Actors

- Hacktivism
- Espionage
- Financial

Data Targets

- Regulated Data
- PII (personally identifiable information)
- Intellectual property
- Company financials, payroll data

5. *SANS Institute paper referencing Ed Skoudis as speaker at RSA Conference,* June 2012
6. *DNS attacks putting organizations at risk, survey finds, SC Magazine,* December 23, 2014
7. Ponemon Institute, 2015 Cost of Data Breach Study

Customer Examples

A large developer of video games had malware inside the network that tried to exfiltrate data via DNS queries using spoofed addresses

A large automaker's main concern is loss of intellectual property that could erode their competitive advantage, and the company is very keen on preventing it from happening via DNS

A large bank failed an audit because of lack of protection for data over DNS

A large insurance company is concerned about liability because they are aware that DNS is not protected

BLOX FEST

DNS Tunneling

Recursive DNS

Request

www.google.com
www.apple.com
gmail.com

74.125.25.104
23.210.209.236
216.58.192.5

Response

Client

Recursive DNS

Request

7r3ncahnt3s.dnst.com
dulhvl8sfdq6rj.dnst.com
qth9zu6574uk3j.dnst.com
siv9dmlmunfb.dnst.com

acvacv19w1gt79t49w1ctd
3kbel9tsznfjbhiwi3kvauh
NXDomain
ServFail

Response

Compromised Client

Recursive DNS

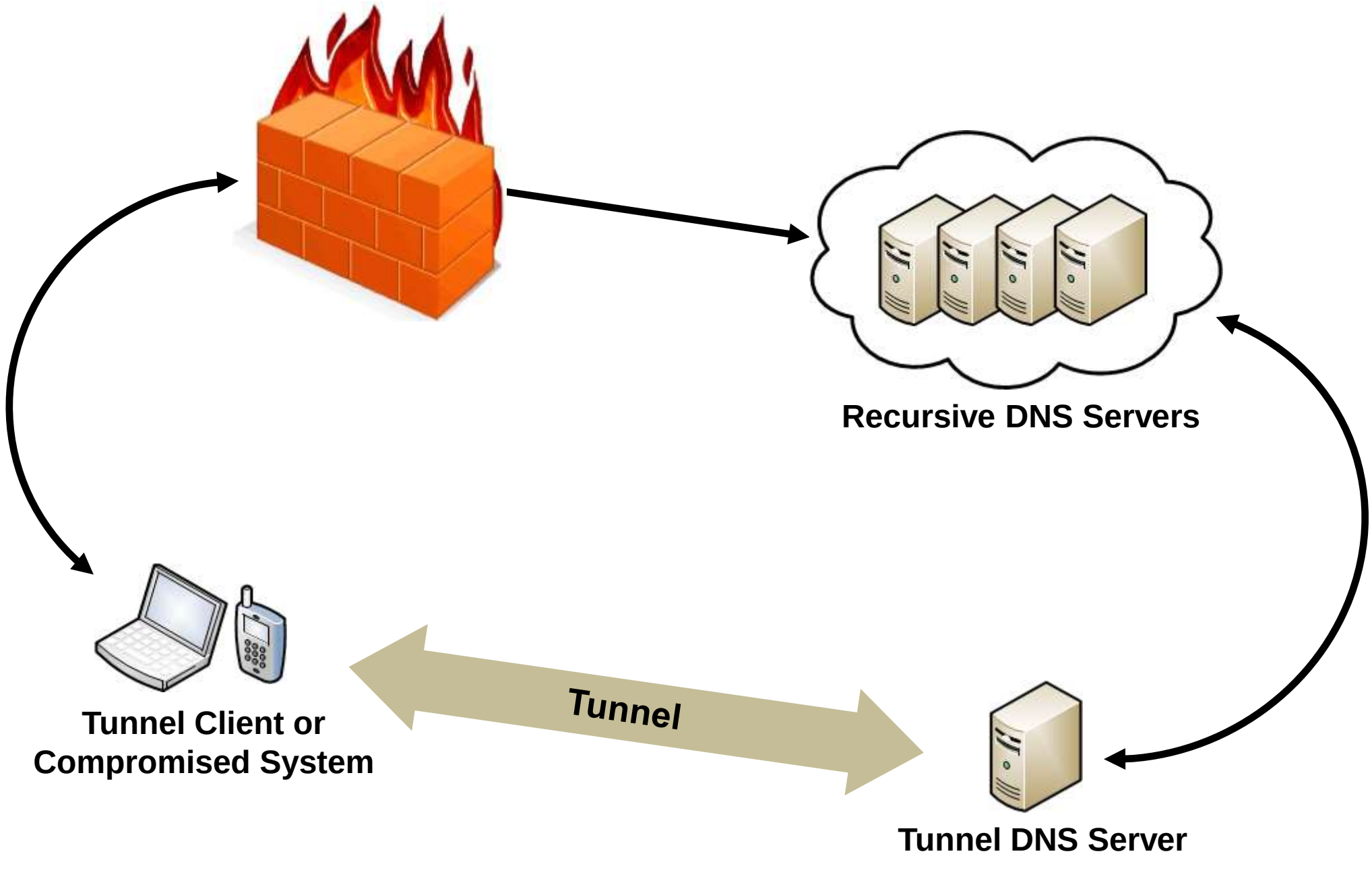
Request

7r3ncahnt3s.dnst.com
dulhvl8sfdq6rj.dnst.com
qth9zu6574uk3j.dnst.com
siv9dmlmunfb.dnst.com

acvacv19w1gt79t49w1ctd
3kbel9tsznfjbhiwi3kvauh
NXDomain
ServFail

Response

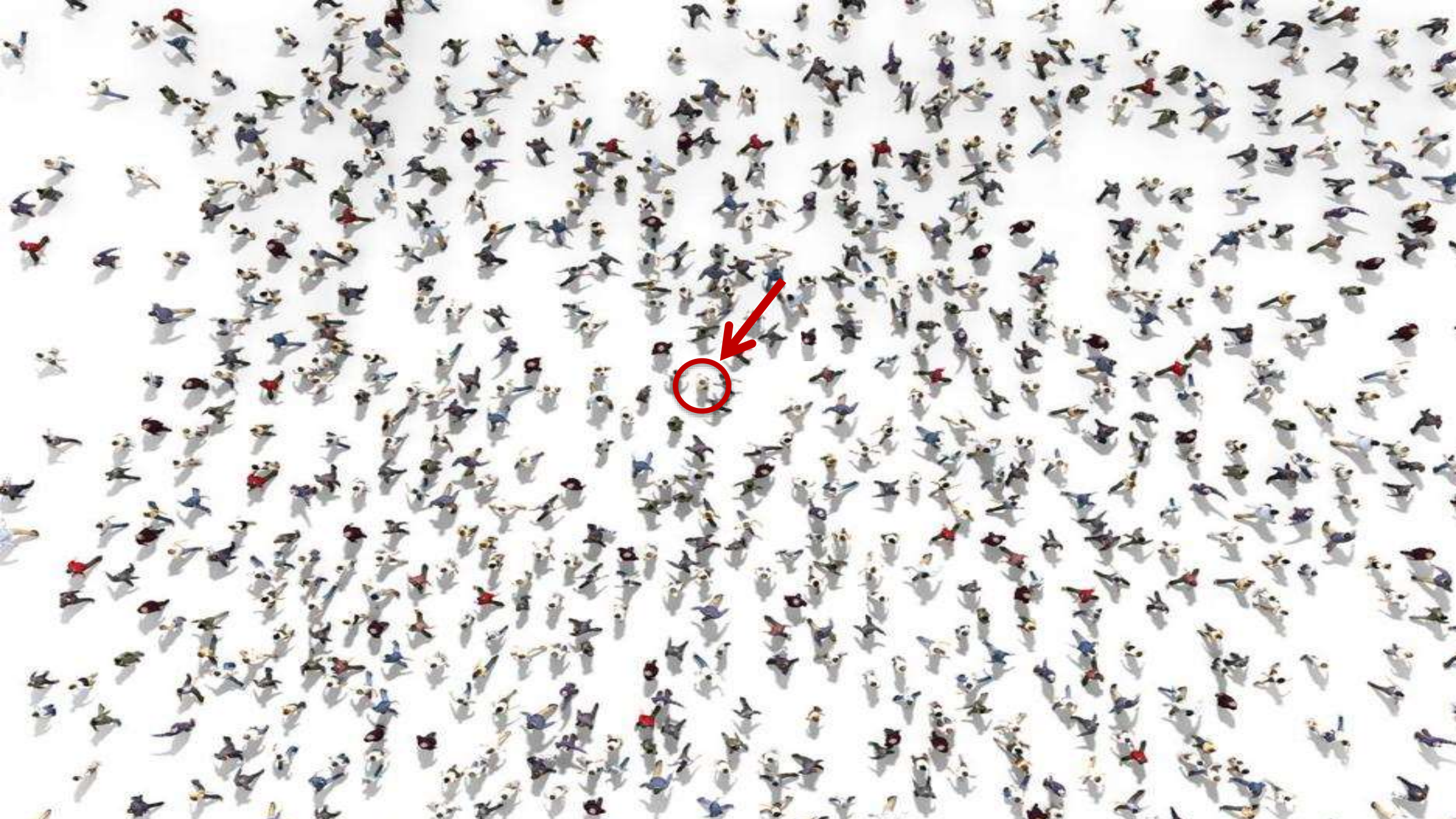
Compromised Client


Recursive DNS Servers
Tunnel
Tunnel Client or
Compromised System
Tunnel DNS Server

- Not firewalled
- Data exfiltration
- Command and control (C&C)
- Free hotspot

- DNS tunneling is a technique
- Legitimate vs malicious uses
- No consistent signatures

* Patent pending




Machine Learning

Data Collection

Truth Marking

Feature Analysis

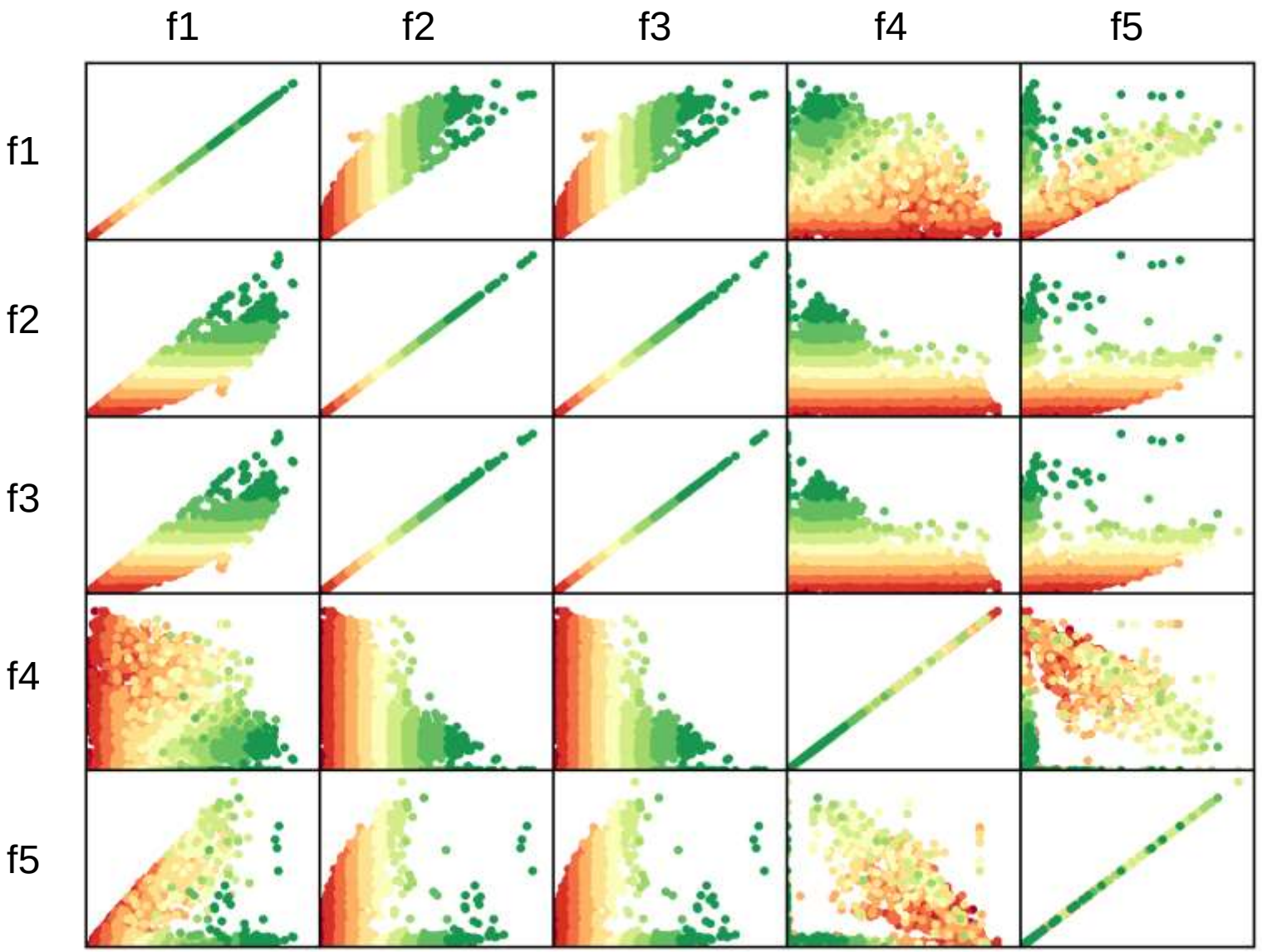
Training

Validation

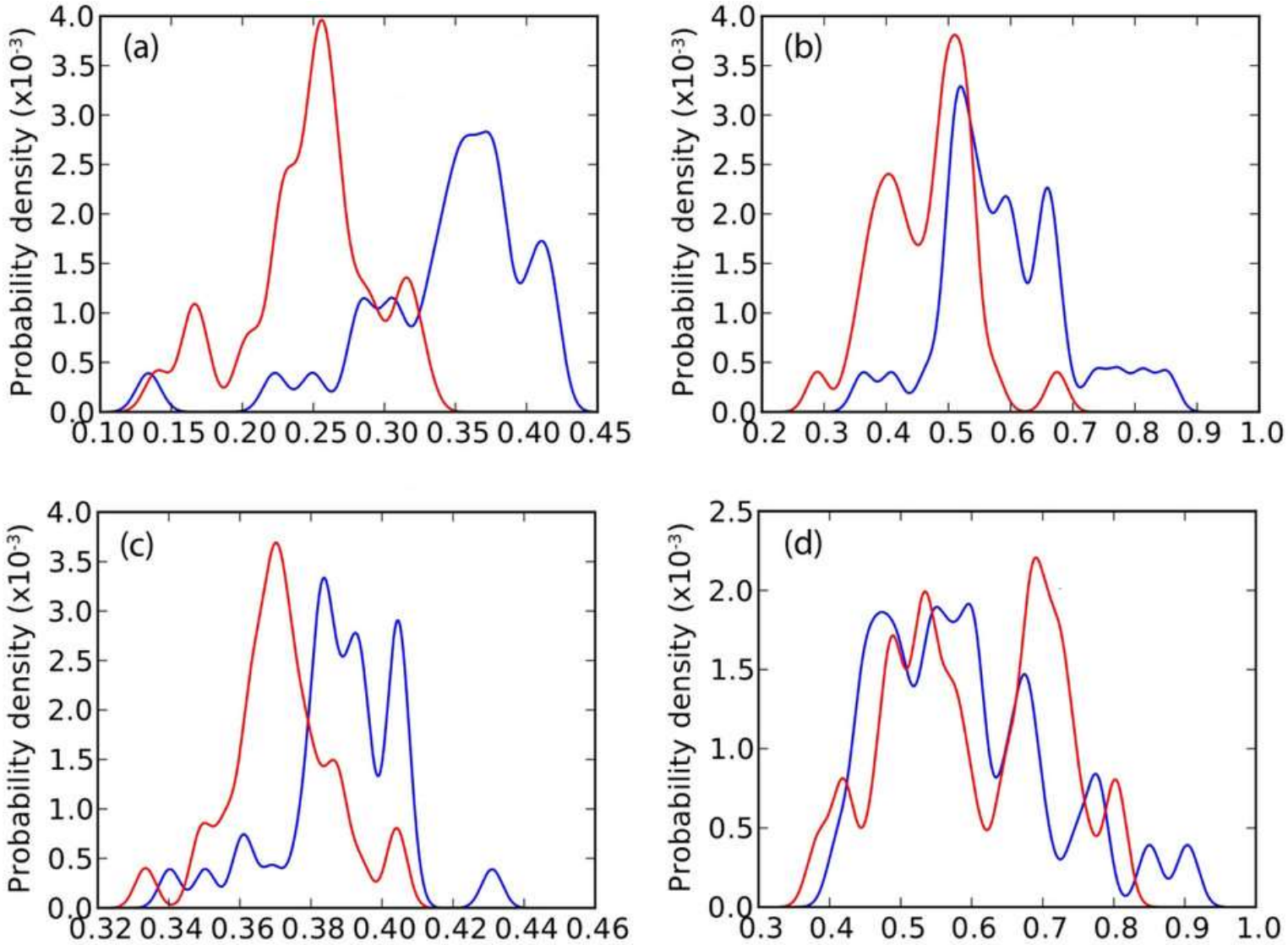
Offline

Model

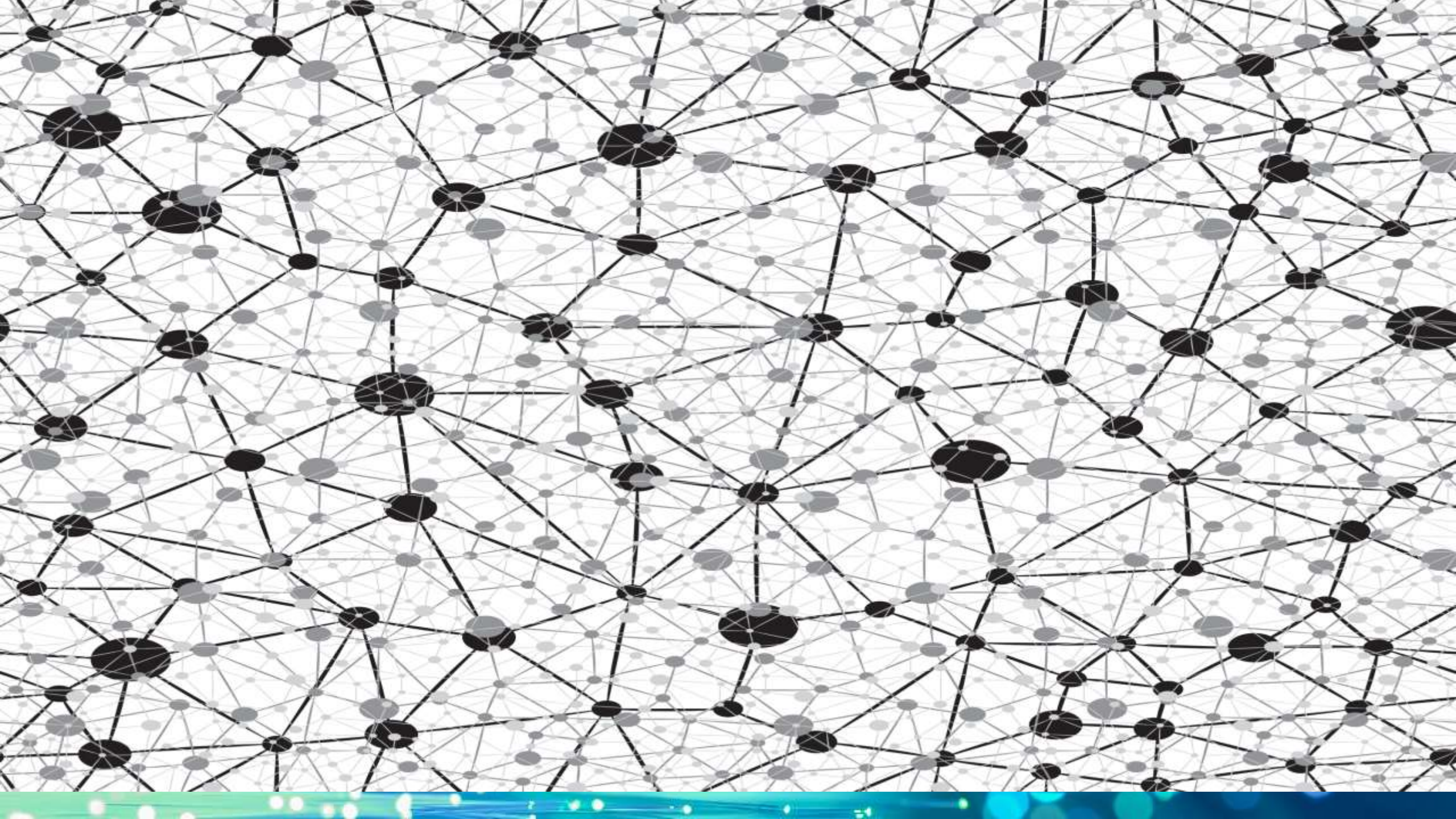
Online Deployment

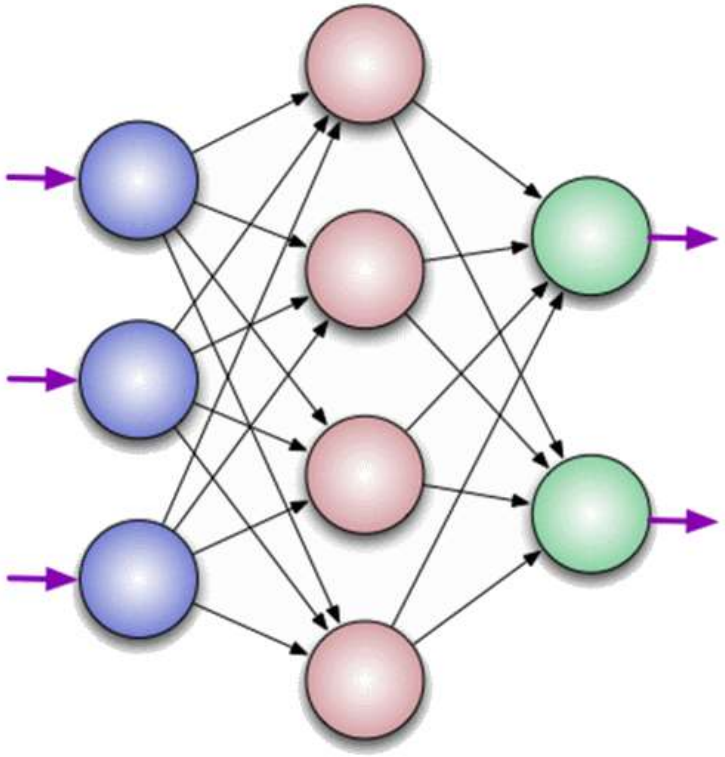


BLOX FEST


(a)
(b)
(c)
(d)
Probability density (x10⁻³)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45
0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
0.32 0.34 0.36 0.38 0.40 0.42 0.44 0.46
0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0


Neural Network






Artificial Intelligence

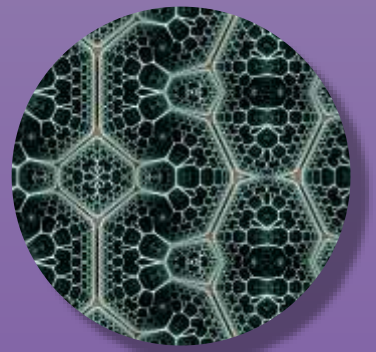

Google


JEOPARDY!
The IBM Challenge


Google DeepMind
AlphaGo


Siri

BLOX FEST

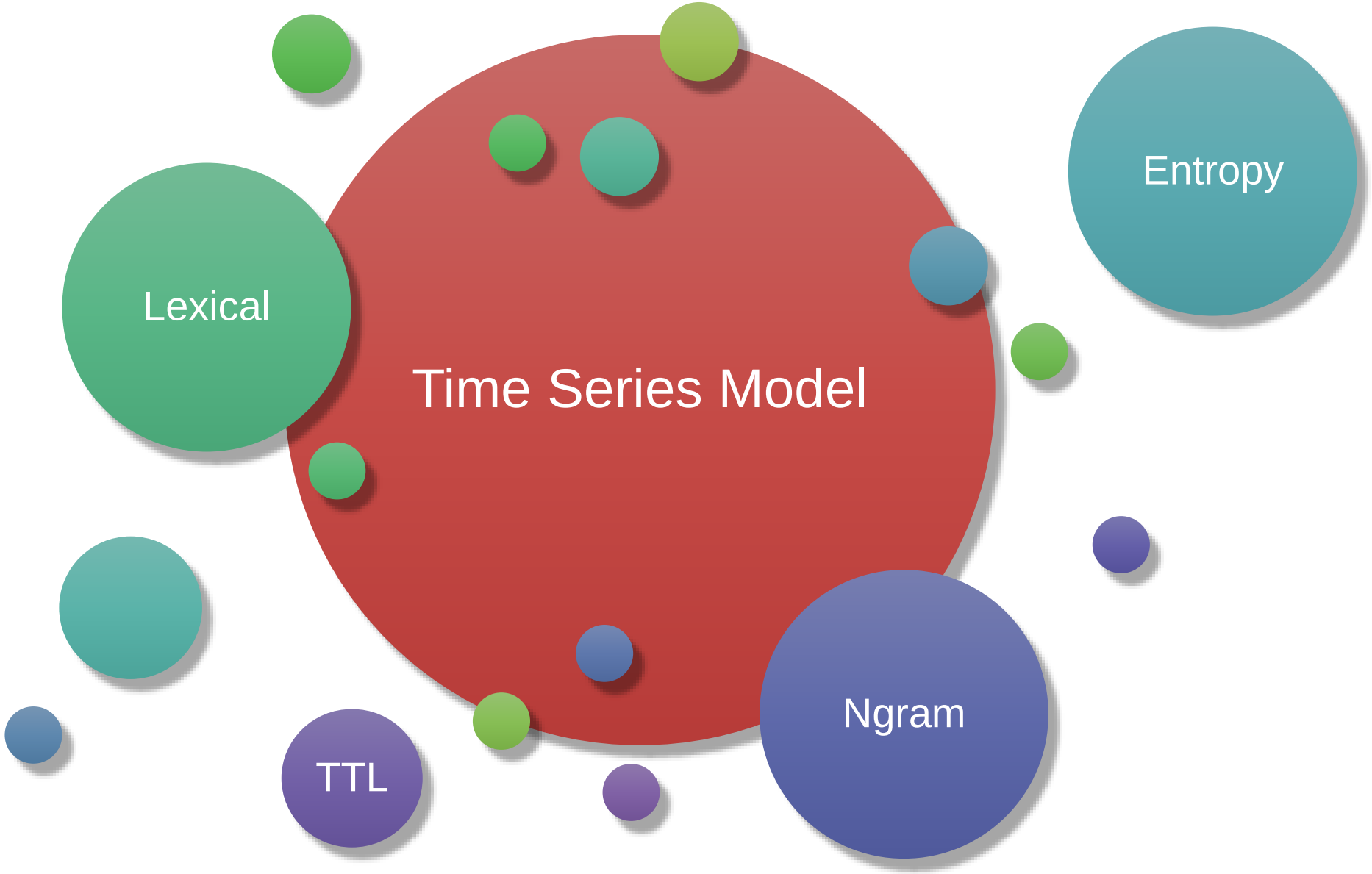


Data Modeling

Classification

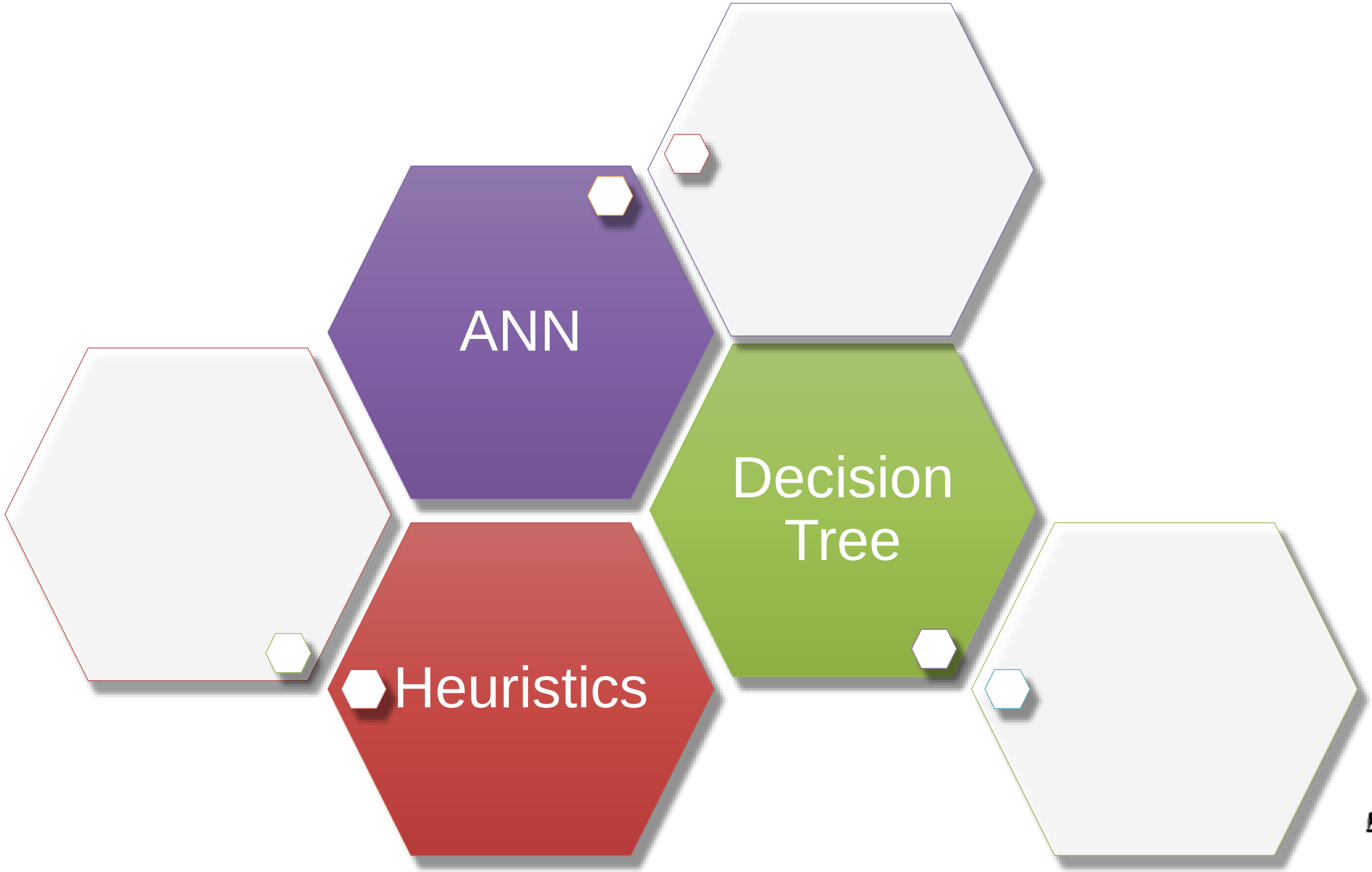
Artificial Intelligence

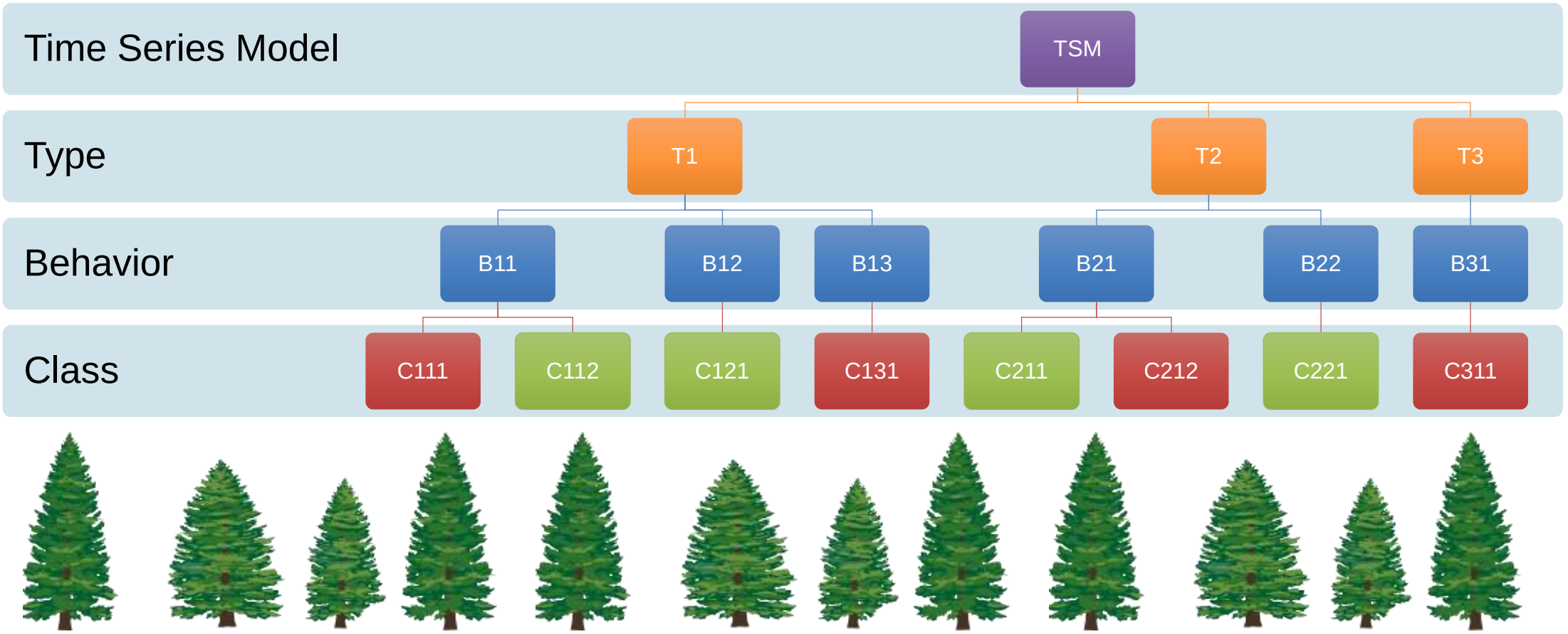
BLOX FEST


Lexical
Time Series Model
Entropy
Ngram
TTL

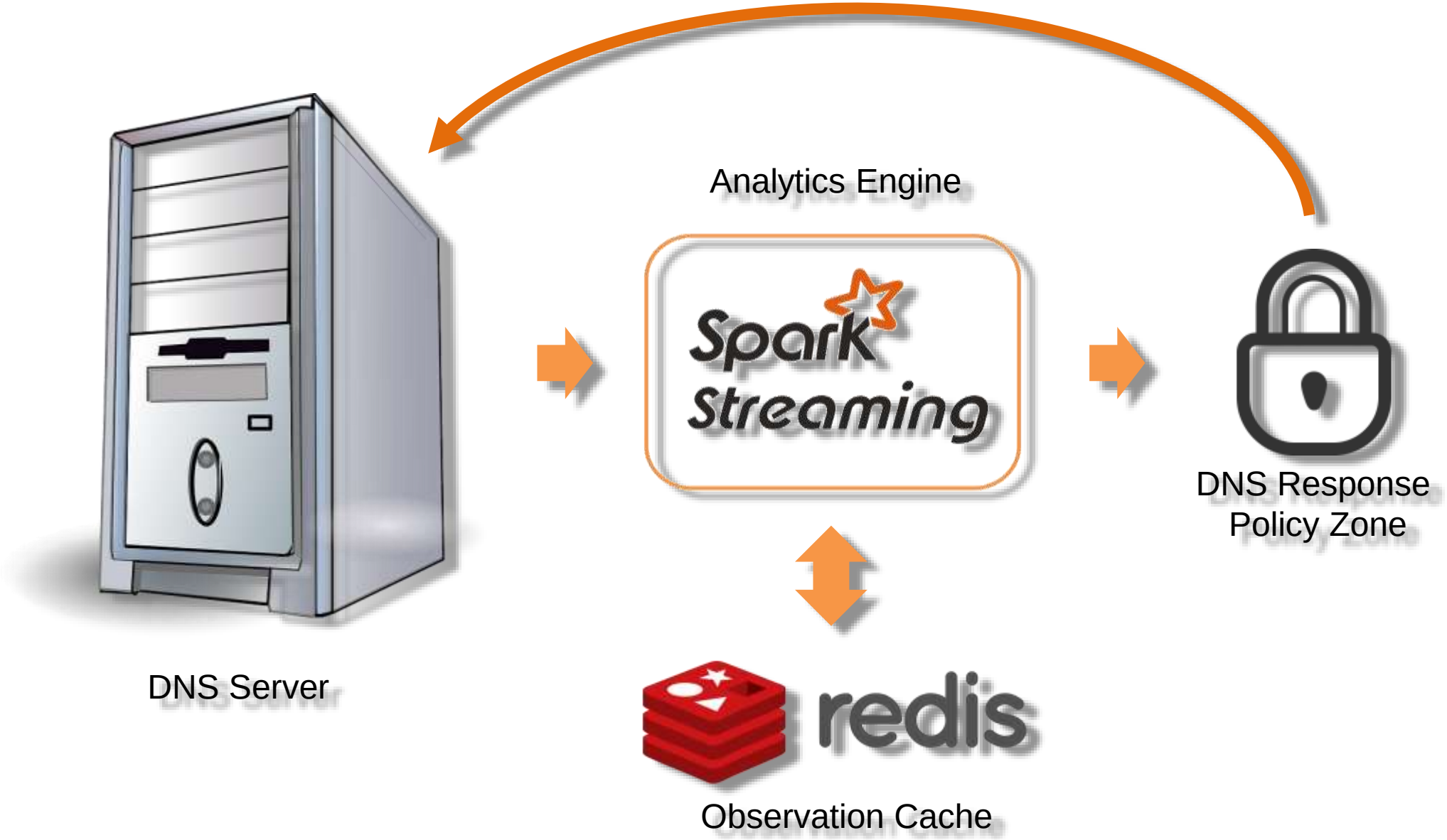

pos.err
neg.err
pos.rt2
neg.rt2
pos.rt3
neg.rt3
pos.ratz
neg.ratz
pos.ten
neg.ten
pos.gin
neg.gin
pos.ster
neg.ster
pos.ntb
neg.ntb
pos.esc
neg.esc

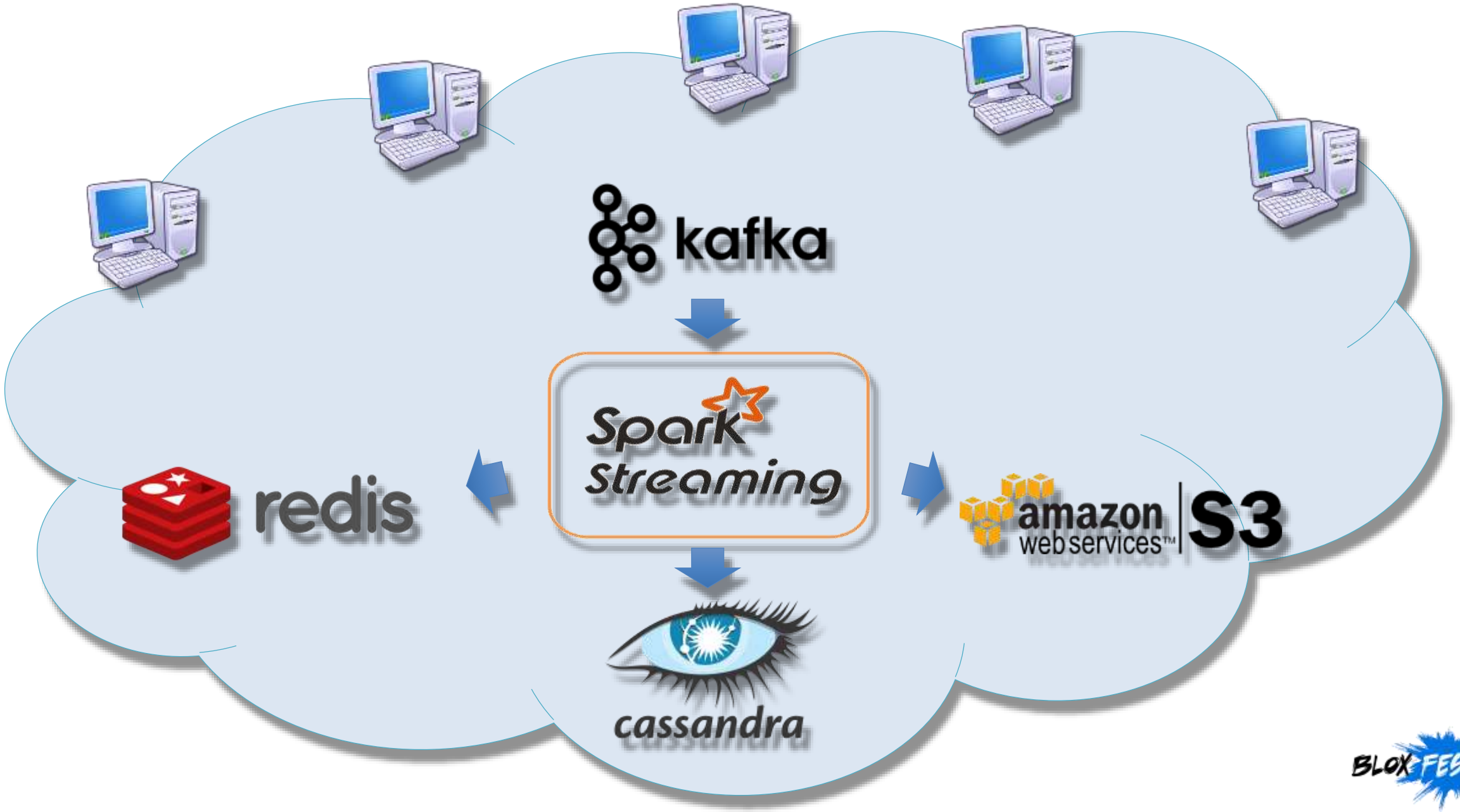
BLOX FEST


ANN
Decision Tree
Heuristics


Time Series Model
TSM
Type
T1
T2
T3
Behavior
B11
B12
B13
B21
B22
B31
Class
C111
C112
C121
C131
C211
C212
C221
C311

Big Data Architecture


Analytics Engine
Spark Streaming
DNS Response Policy Zone
DNS Server
redis
Observation Cache


kafka
Spark Streaming
redis
amazon web services™ S3
cassandra

Detection Results

Behavior Analysis based DNS Tunneling Detection and Classification with Big Data Technologies

Bin Yu, Les Smith, Mark Threefoot and Femi Olumofin
CTO Office, Infoblox Inc., 3111 Coronado Dr, Santa Clara, California 95054, U.S.A.
{binyu, lwsmith, mthreefoot, folumofin}@infoblox.com

Keywords: Behaviour Analysis, Time Series, Big Data Analytics, DNS Security, Data Exfiltration, Anomaly Detection, Classification.

Abstract: Domain Name System (DNS) is ubiquitous in any network. DNS tunnelling is a technique to transfer data, convey messages or conduct TCP activities over DNS protocol that is typically not blocked or watched by security enforcement such as firewalls. As a technique, it can be utilized in many malicious ways which can compromise the security of a network by the activities of data exfiltration, cyber-espionage, and command and control. On the other side, it can also be used by legitimate users. The traditional methods may not be able to distinguish between legitimate and malicious uses even if they can detect the DNS tunnelling activities. We propose a behaviour analysis based method that can not only detect the DNS tunnelling, but also classify the activities in order to catch and block the malicious tunnelling traffic. The proposed method can achieve the scale of real-time detection on fast and large DNS data with the use of big data technologies in offline training and online detection systems.

1 INTRODUCTION

Domain Name System (DNS) that mainly services a domain name resolution to IP addresses on UDP is a service ubiquitous in every network. Because DNS is not intended for data transfer, people can overlook it as a threat for malicious communications or for data exfiltration. Most networks, public or private, do not firewall DNS traffic which creates security vulnerability. Tunnelling data over DNS or TCP over DNS is a technique that can be used as a way to circumvent access and security policies in firewalled networks. A typical example is to illegally browse the web through public hotspot while free service is not provided. There are many free software tools available for people of interest to setup a DNS tunnelling system quickly. One of the most popular tools is Iodine (Iodine). The fact that information bypasses a network first line security mechanism makes DNS tunnelling very attractive also in contexts other than free web browsing. Such examples include command and control and data exfiltration in cyber-espionage attacks in which it is fundamental for an attacker to have an available but inconspicuous communication channel.

DNS tunnelling works by encapsulating data into DNS packets. Typically, the tunnel client encapsulates the data to be sent in a query for a specific domain name. The DNS resolver treats the tunnel traffic as a regular request by starting the lookup process for the requested domain name, possibly recursively consulting other DNS resolvers, as shown in Figure 1. At the end of this operation, the request is processed by the tunnel server. The server retrieves the encapsulated data and responds to DNS queries by enclosing tunnel data in the answer section of the DNS response message.

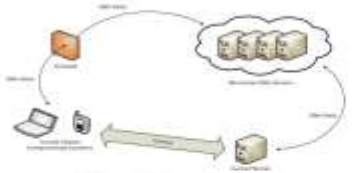


Figure 1: DNS tunnelling setup.

Although most DNS tunnelling techniques use TXT type queries in DNS that can maximize the payload in response packets, there are implementations that make use of DNS query types other than TXT such as A, AAAA, CNAME, NS,

Yu, B., Smith, L., Threefoot, M. and Olumofin, F.
Behavior Analysis based DNS Tunneling Detection and Classification with Big Data Technologies
In *Proceedings of the International Conference on Internet of Things and Big Data (IoTBD 2016)*, pages 276-282
ISBN: 978-989-758-183-0

...ure, Gini index is another way
...f the data that is defined as
$1 - \sum D^2(x)$.
...e entropy feature, Gini index
...lue is bounded within a range

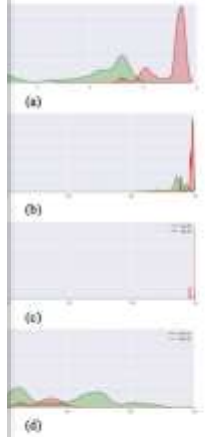
... Error

...sure the diversity of a data set
...n error. Like the Gini index
...this feature is also bounded
...The definition is as follows.
$1 - \max\{D(x)\}$.

...abels

...st feature is the number of
...QDN payload named as *nlb* to
... and malicious payloads.

...enc is the output of a neural
...of the above features as input.
...tribed in the next section.



...in this paper is collected from Farsight
... receives passive DNS from a large
...tributors worldwide, mainly in the US.
...mple filtering logics such as DNS type,
...series length, and whitelisting, a set of
... extracted and reviewed by security
...th labelling. About 2000 samples are
...ining and testing a tree classifier that is
...d to minimize the false positive rate.

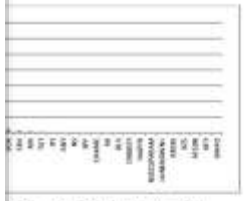
...-TIME DETECTION ...EM

...s that were trained in offline system
...yed in an online real-time detection
...2014) that is designated to deal with
...e streaming data. In an enterprise
...he throughput can be up to 1-3 million
...per second. The throughput can reach
...econd in a cloud based deployment.
...e horizontal scalability is one of the
...t factors in design.
... in Figure 6, the incoming stream is
... real-time with Storm or Spark
...d inserted into the observation cache
... extracted features that are indexed by
... address and SLD. The observation
...in-memory layer and an on-disk layer
...use is dependent on the data size. The
...be triggered by event or scheduled by
...cost effective.

...eal-time detection system architecture.

...LTS AND CONCLUSION

...onth DNS data collected from Farsight
... 2013 at a rate of 1.8B/day is used in
... process. In total, 126K tunnels are
...re a tunnel is defined as from one
... IP address to one unique destination
... Table 1 shows the summery of the
...ted. About 70% detected tunnels are



...NS query distribution by query type.

...RE EXTRACTION AND ...TION

... DNS traffic typically has very small
... the reason many approaches detect
... on payload size (Farnham, 2013,
...However, when space and bandwidth
...ore and more legitimate users are
... domain names. Since the main
...e tunnelling technique is to convey
... the tunnel in a way as efficient as
...ntropy metrics become good features.
...e, human readability of domain names
...tor in tunnel detection (Born, 2010).

...ve Payload

... types of DNS queries. A tunnel will
...e to carry outbound payloads. The
...ds are carried in many different ways
... the DNS resource record type. For
...T type, the payload is encoded in the
... other types, such as A, AAAA, or
...payload is carried in one or more
...e legitimate DNS queries that have
... query and response, a malicious
...change the payload from message to
...ffective payload is a string that is
... its original with common prefix,
...ed middle segments removed so that
...an stand out.

...on Features for Inbound and ...und

...d features common for both inbound
... traffics are extracted. Figure 3
...eature analysis results for inbound
...ature analysis results for outbound

...a resolver or DNS server
...rnal client IP address
... information availability.
...ted into an observation
...a TTL pre-set to remove
...ries. It also has a capacity
... remove old data points
...a hit the capacity though
...TTL criterion. This is to
...s and reserve the storage
... recycled. Applying the
...h of the messages within
...et that is denoted as a 2-

$f_{(x}$.

...s the k^{th} feature on the i^{th}
... inbound payloads. The
...r features are the basic
...ures on the series that can

$...(f_{(x}$

... is the collection
...g] to represent the
... features across the time
...tropy values on effective
...ayloads are calculated,
...e fact that the payload of
... change as much as the
...e series.

...ON

...sification. In the first tier,
...d on identifying encoded
...er is for tunnel detection.

...fication

...sifiers are designed and
...indicating if a payload is
... inbound and outbound
...ach of the classifiers is
...les with truth labelled by
...d on independent sets of
... classifiers have a single
...eurons and each uses a
...defined as follows.

$...(f_k w_k + w_0)$

...more malicious
...s. This makes
... future research

...g DNS tunneling,
...m.
...twork (SPF) for
...mail, version 1,
...tml/rfc4408.
... Semi-supervised
...ne flux domain
...ic, in *the 10th ...ining and Machine* ...258-271.
...er DNS (EDNS0),
...s/rfc2 671.txt.
...: detecting covert
the 11th ACM Communications

...garelli, L., 2007.
...tical mechanisms,
Conference on

...garelli, L., 2008.
... across network
the 4.3rd IEEE ...unications, Beijing

...08. A preliminary
...in *Proceedings of* ...nce on Computer
...t. Thomas, U.S.

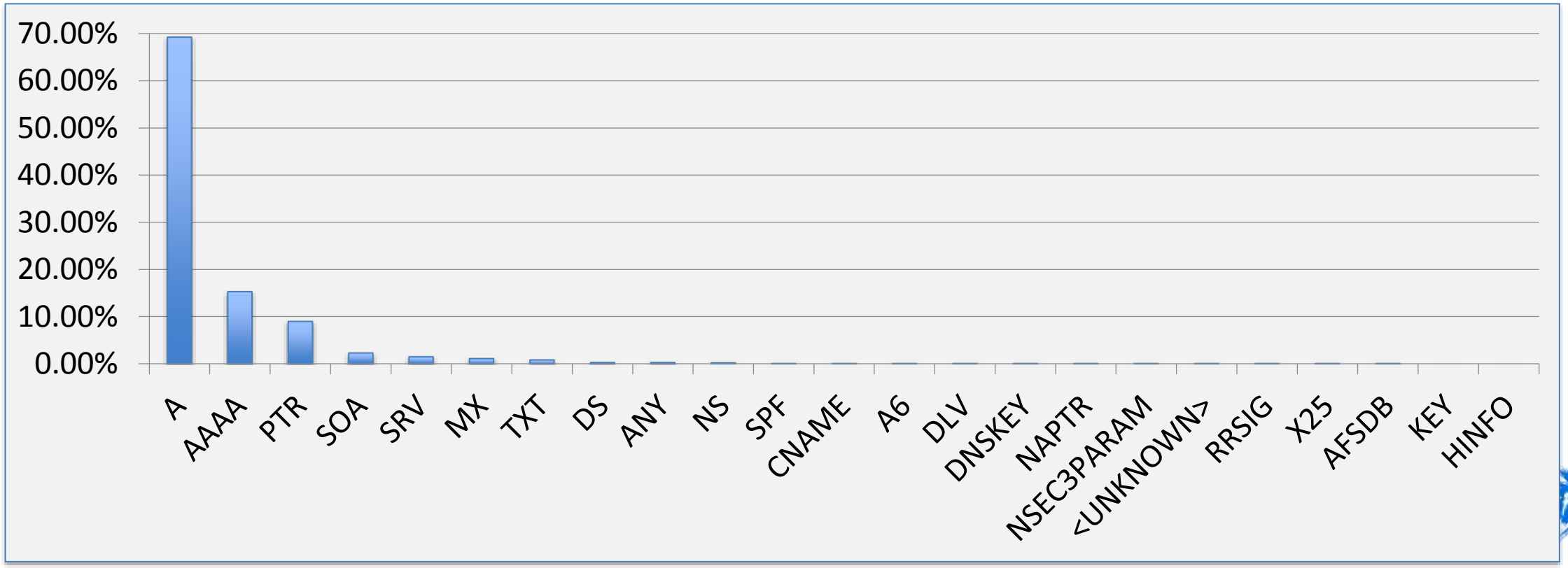
...ols with AI, in
...a, Nevada.
...ing DNS tunnels
...in *Proceedings of* ...Las Vegas, NV.
...., Schomaus, H.,
...013. Flow-based
...ging Management
Lecture Notes in ...124-135.
...e: http://www.isc.

...www.farsightsecu

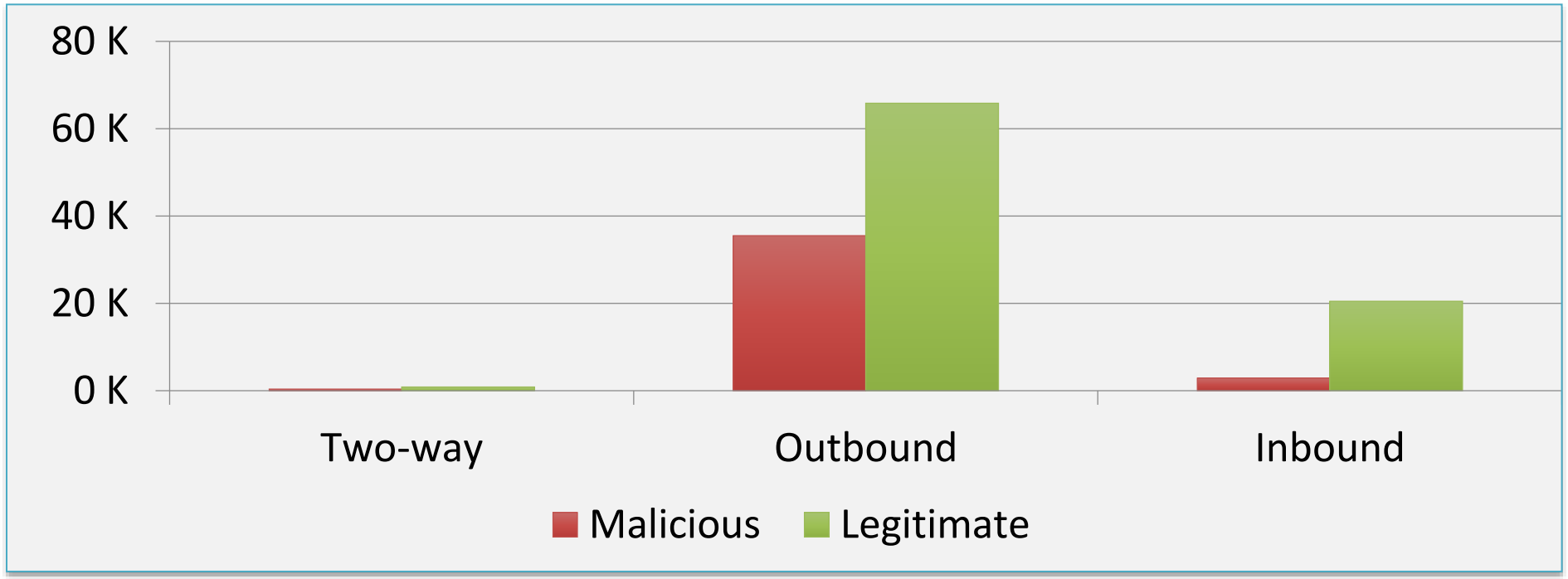
...ical Theory of
...ical Journal*, Vol.

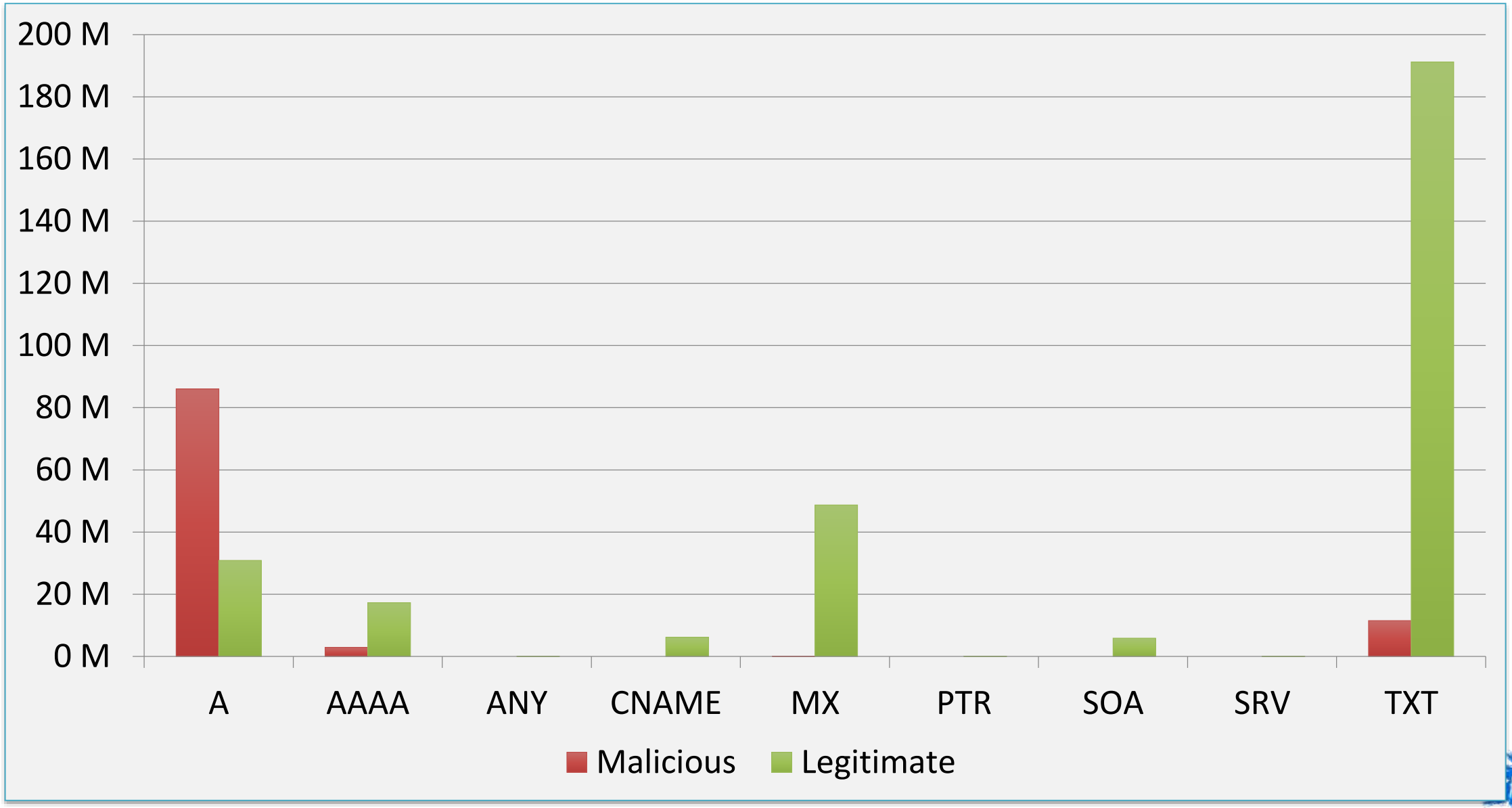
...ooks/rgrams/book

- Farsight (ISC)
- 2012.12 – 2013.08
- 1.8 billion per day
- 150 TB

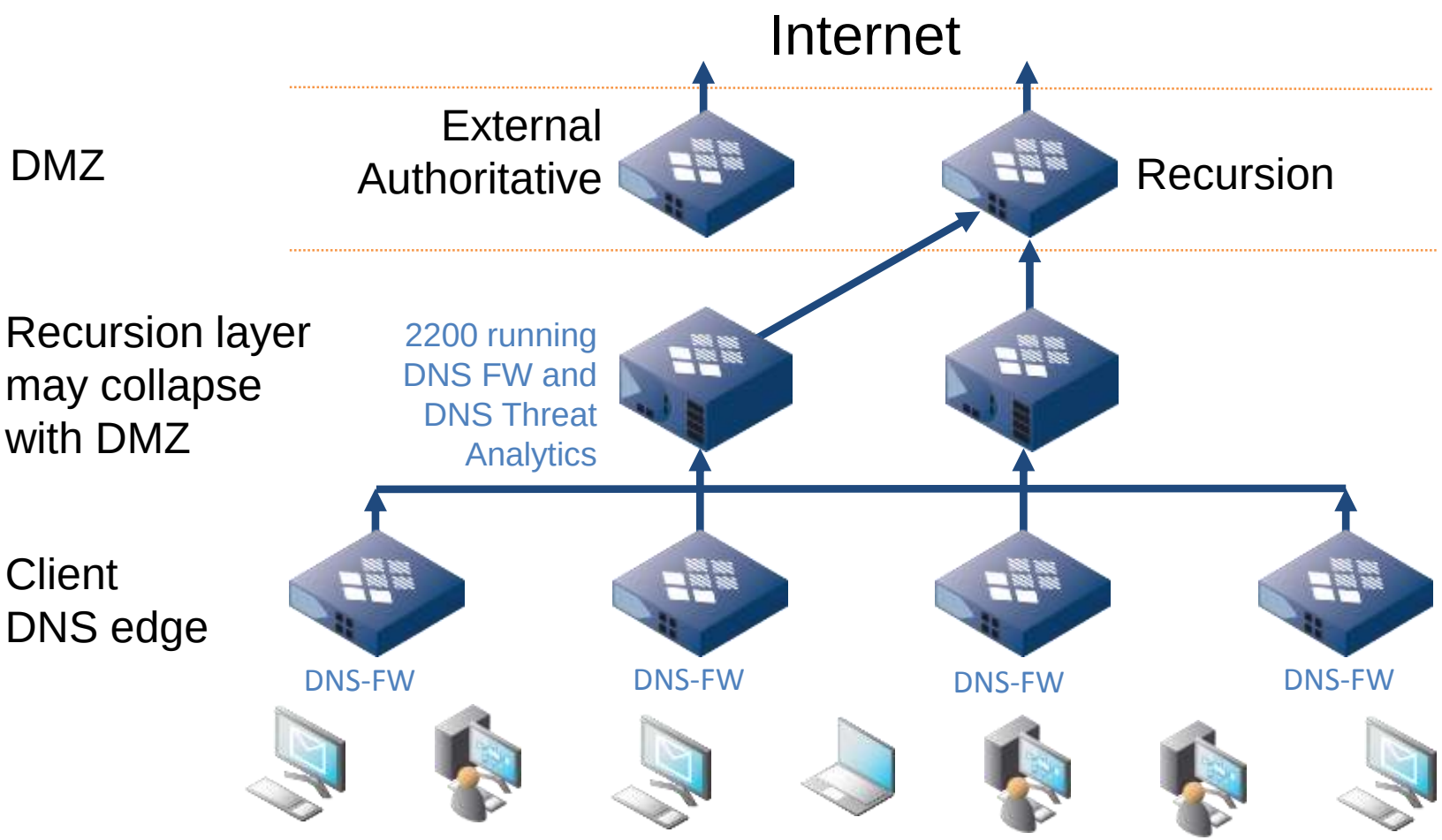


	Malicious	Legitimate	All
Two-way	356	869	1,225
Outbound	35,478	65,820	101,298
Inbound	2,845	20,504	23,349
Total	38,678	87,193	125,871




200 M
180 M
160 M
140 M
120 M
100 M
80 M
60 M
40 M
20 M
0 M
A
AAAA
ANY
CNAME
MX
PTR
SOA
SRV
TXT
Malicious
Legitimate

DNS Threat Analytics in Recursion Layer



- Central detection of tunnels and data exfiltration
- Scaling of enforcement to all edge Grid members once destinations are on RPZ list
- Pinpoint infected systems at edge
- Lower platforms can be deployed at edge

Customer Case Study:

T. J. Short, CISO and VP of Infrastructure,

Everi Holdings, Inc.

“The attackers are getting smarter every day. They’re getting new tools, new ideas, new concepts. So we have to have defenses that are leading edge, that can change, adapt, and update very quickly. Infoblox Internal DNS Security does that.”

Questions?