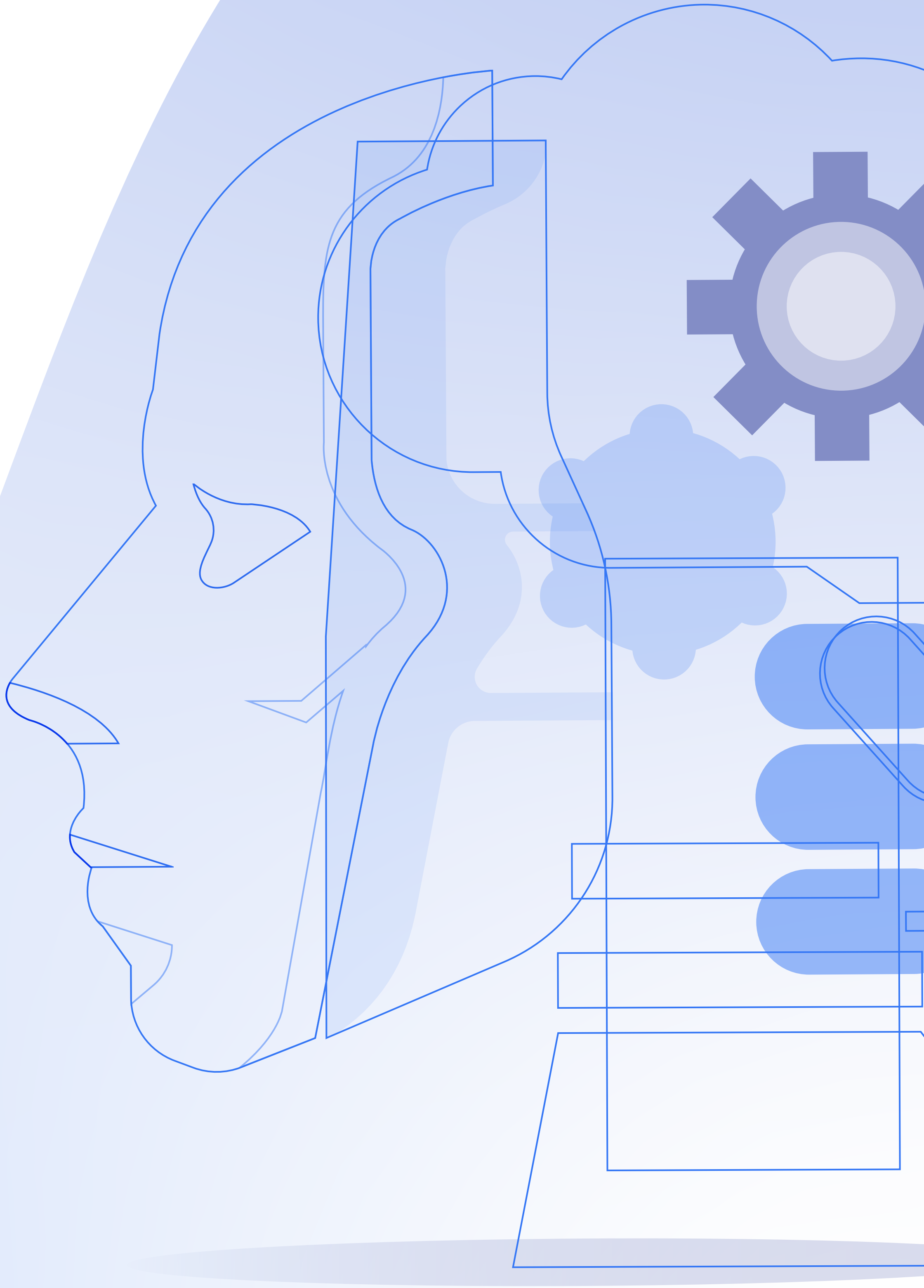
CUJO AI

Applying Artificial Intelligence for Operator Analytics and Digital Life Protection Services

2020

Table of Contents



Innovation is at the core of our digital lives, with Network Service Providers deploying Artificial Intelligence (AI) to power security tools by analyzing data from millions of cyber incidents and using it to identify potential threats.

AI is a widely applicable tool for new generation cyber defense systems. This white paper covers the increasingly important role of practical applications of AI in cybersecurity as well as explores five key Machine Learning (ML) algorithms used in CUJO AI security solutions for telecom operators.

The following sections will look at:

- ✓ How conventional rule-based and machine learning-based security technologies differ
- ✓ The application of artificial intelligence in CUJO AI solutions
- ✓ How five CUJO AI machine learning algorithms empower four specific products

Rule-based vs. Machine Learning-based Methods

- ✓ Conventional methods in the cybersecurity field are rule based. A subcase of this is the blacklist & whitelist-based approach.
- ✓ Rule-based methods are rigid. They are not capable of identifying novel cases like zero-day threats, new browser fingerprinting techniques, or new IoT devices.
- ✓ ML-based methods are more flexible. They can learn patterns, model the relationship between the patterns and identify new cases based on their similarity to known ones.
- ✓ ML-based methods require less maintenance than traditional methods. Keeping a rule base up to date is more difficult than constantly feeding the ML model with new data.

Artificial Intelligence in the CUJO AI Portfolio

CUJO AI delivers state-of-the-art solutions for network security and privacy as well as for digital parenting, device identification and more.

CUJO AI Sentry	**CUJO AI Incognito**	**CUJO AI Compass**	**CUJO AI Explorer**
Network Security and Device Protection	**Privacy and Tracking Protection**	**Content Control and Digital Parenting**	**Advanced Device Identification & Classification**
Premium digital security service for consumers and businesses that uses AI to detect and block threats directed to any device connected to the network, while respecting the privacy of end users.	Operators utilize this AI-powered solution to automatically block third-party tracking software and ensure users have a private, safe and uncompromised browsing experience across all connected devices with no endpoint software needed.	Empowers families and businesses to define and manage how their members' online activity affects their everyday lives. Using AI to categorize network traffic, it provides detailed visibility of online habits and can enforce limits on time online, app usage and access to content.	Complete, programmatic access via APIs to all of the information collected and processed by the CUJO AI Platform. Explorer allows network operators to monitor and query all CUJO AI-generated data about connected devices, applications, content, threat and privacy.

CUJO AI's algorithms actively learn from network traffic, device behavior and other sources. The goal is to protect users, precisely identify devices and offer advanced content controls.

CUJO AI Sentry contains an ML classifier called [Web Sentry] to differentiate malicious web addresses from legitimate ones. Also, it includes a custom anomaly detection method called [Device Sentry] that recognizes the unusual behavior of IoT devices.

CUJO AI Incognito contains an ML classifier for detecting browser fingerprinting.

CUJO AI Compass includes an algorithm to detect when the various mobile applications or sites are actively used.

CUJO AI Explorer contains an ML-assisted system for recommending device identification rules for analysts.

CUJO AI Algorithms

Device Sentry and Web Sentry Algorithms in CUJO AI Sentry

Device Sentry

PATENTED

The Device Sentry algorithm is a custom anomaly detection method that we have developed to detect the unusual behavior of IoT devices. The algorithm is based on analyzing packet headers in the network traffic generated by the devices. No deep packet inspection is applied.

The algorithm inspects the communication endpoints of devices belonging to a given model.

> The algorithm is based on analyzing packet headers in the network traffic.

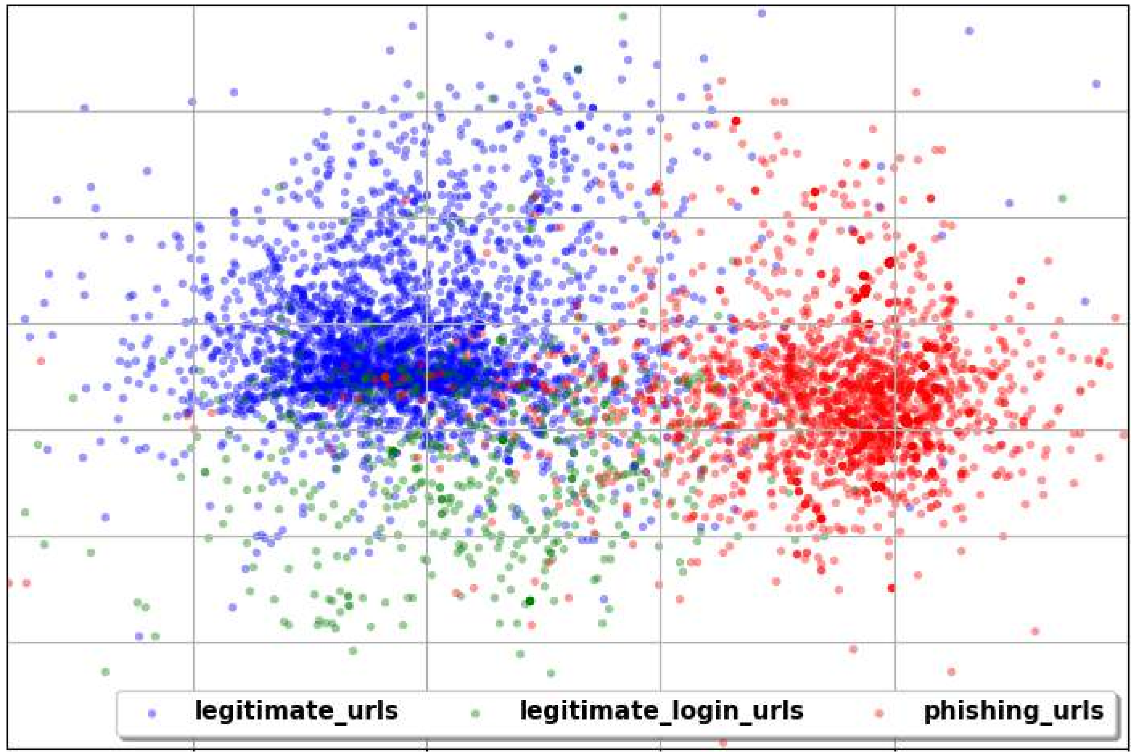
Web Sentry

PATENTED

The goal of Web Sentry is to differentiate malicious web addresses from legitimate ones. An example malicious web address is a fake PayPal login page that tries to steal passwords. The site can potentially be a newly appeared one, meaning that it is not yet listed in any threat intelligence feed. Therefore, it cannot be caught using the traditional DNS blacklisting method. Web Sentry applies machine learning to detect malicious URLs, including zero-day threats. The basis of the algorithm is a labeled data set of web addresses.

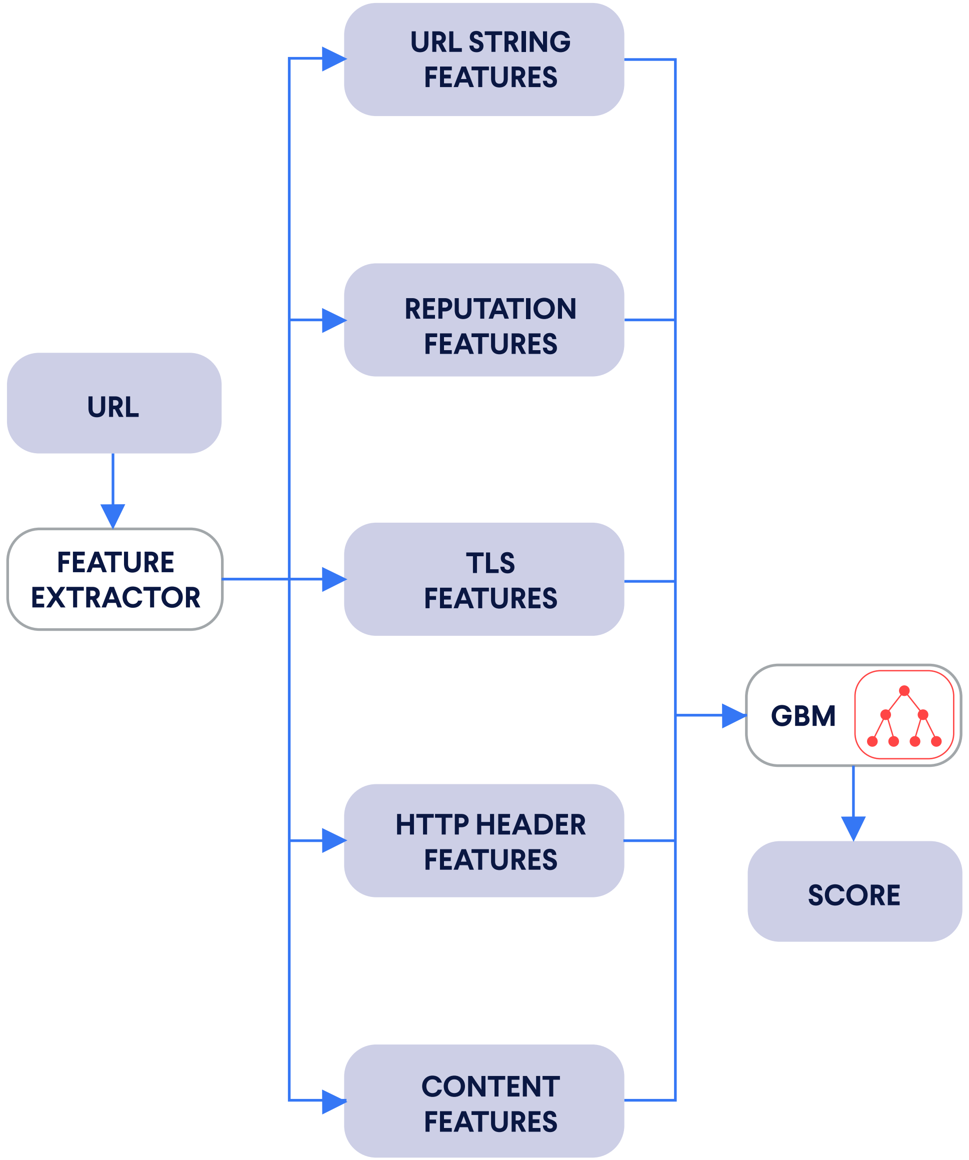
The classification is based on various attributes of the web address. These are called features in machine learning terminology.

In total, we extract around 100 features from each web address.



Visualization of the training data set for [Web Sentry]. To prepare the images, we applied deep neural network with a bottleneck of two neurons. The axes show the values of the bottleneck neurons when the network is fed with all features. According to the machine, the phishing sites (red dots) can be separated from legitimate sites quite well.

The decision logic in Web Sentry is a Gradient Boosting Machine (GBM). A GBM can be viewed as an ensemble of decision trees, where the trees are arranged in a sequence and each tree tries to learn the error of the previous tree. The input of the training process is a large data set of web addresses, where each web address is labeled as legitimate or malicious and all features computed for the web addresses. The result of training is a model that learns the relationship between the features and the label. This model can then be used to predict the maliciousness of new web addresses based on their features.



Predicting the maliciousness of a web address (URL) with [Web Sentry]

> Web Sentry processes around 10 million classification requests per day and raises approximately 1000 alerts per day, protecting users from visiting malicious websites.

ML Classifier in CUJO AI Incognito

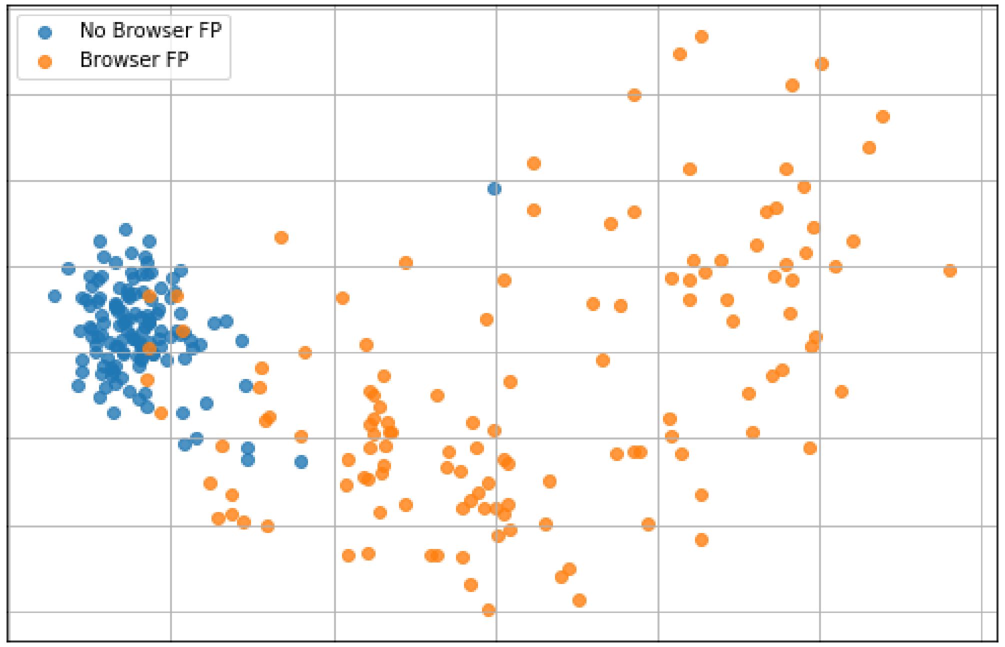
As the traditional, cookie-based tracking of users becomes more difficult, the tracking business is moving toward different techniques such as browser fingerprinting. The idea behind browser fingerprinting is to collect information about the browser and its environment for the purpose of identification. CUJO AI Incognito contains an ML-based browser fingerprinting detector.

Browser fingerprinting is typically implemented in JavaScript (JS). Those JS files can contain functions that are responsible for the fingerprinting. However, the presence of a function definition does not prove the act of fingerprinting. The fingerprinting function is not necessarily called when the website is loaded.

Incognito contains two separate ML classifiers for detecting browser fingerprinting.

The **static** classifier treats JS files as simple text objects and determines if they contain indicators of fingerprinting or not.

The **dynamic** classifier analyzes the behavior of the JS files by inspecting the function call tree when the user visits a specific web site. It is more accurate than the static classifier, but it is computationally more expensive.

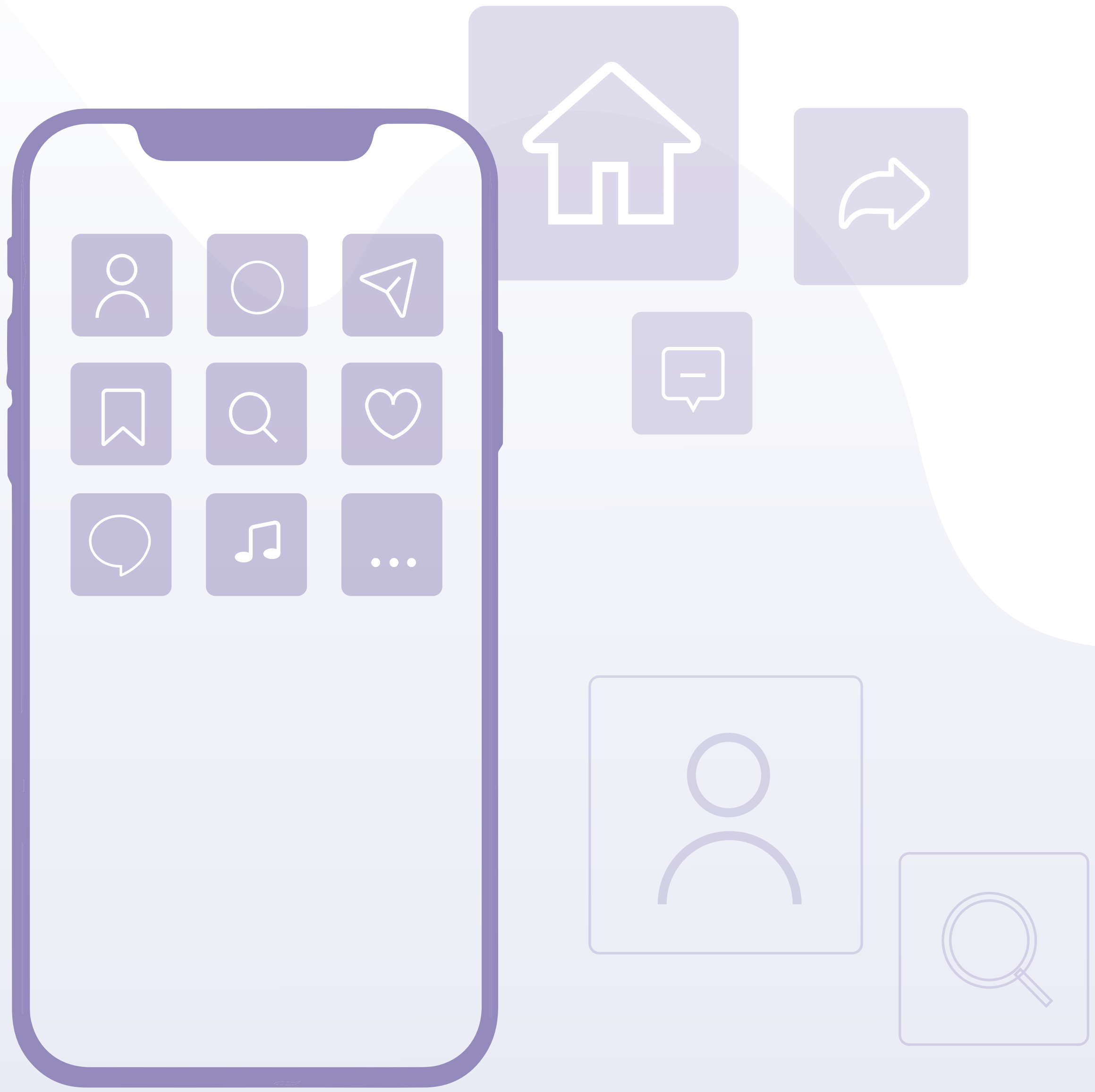


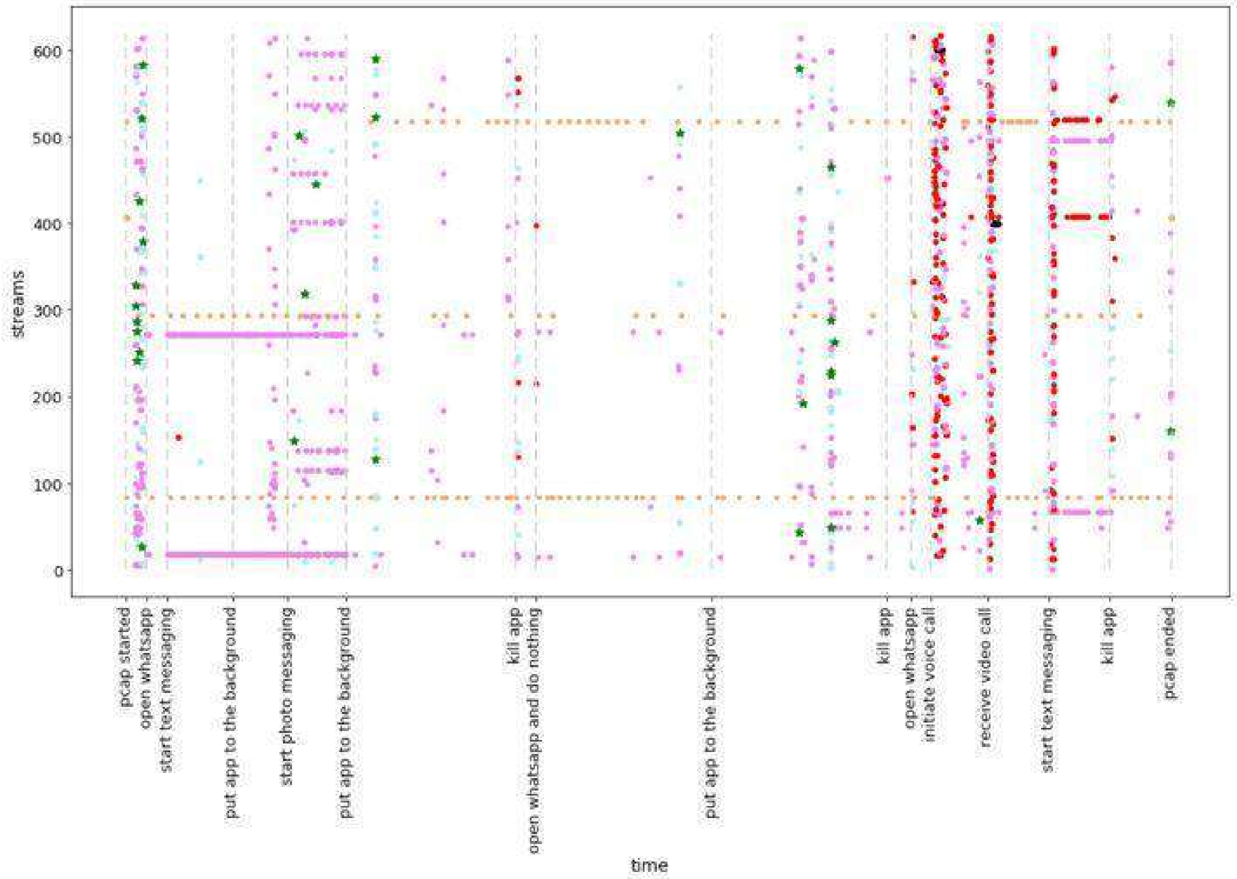
PCA-based visualization of the static browser fingerprinting data set. The two classes can be separated fairly well. The JavaScript files that apply browser fingerprinting (orange dots) show more diversity than the negative cases (blue dots).

Semiautomatic Rule-based System in CUJO AI Compass

PATENT PENDING

CUJO AI Compass can detect when the various mobile applications are actively used, without applying deep packet inspection. The approach is ruled based and semiautomatic. First, an algorithm recommends app detection rules for human analysts, and then the analysts finalize and maintain the rules. This section describes the automatic part of the approach.





Visualization of the network streams observed during a WhatsApp session. The horizontal axis is time; the vertical axis enumerates the streams by their index. Each dot represents network activity. Stars are TLS handshake events. In the first interval of activity (from "start text messaging" to "put an app to the background"), the bulk of communication can be associated with two streams. Afterward, the pattern is less regular. This figure illustrates that app usage detection is a challenging task because the signal is mixed with a significant amount of noise.

ML-assisted System in CUJO AI Explorer

PATENT PENDING

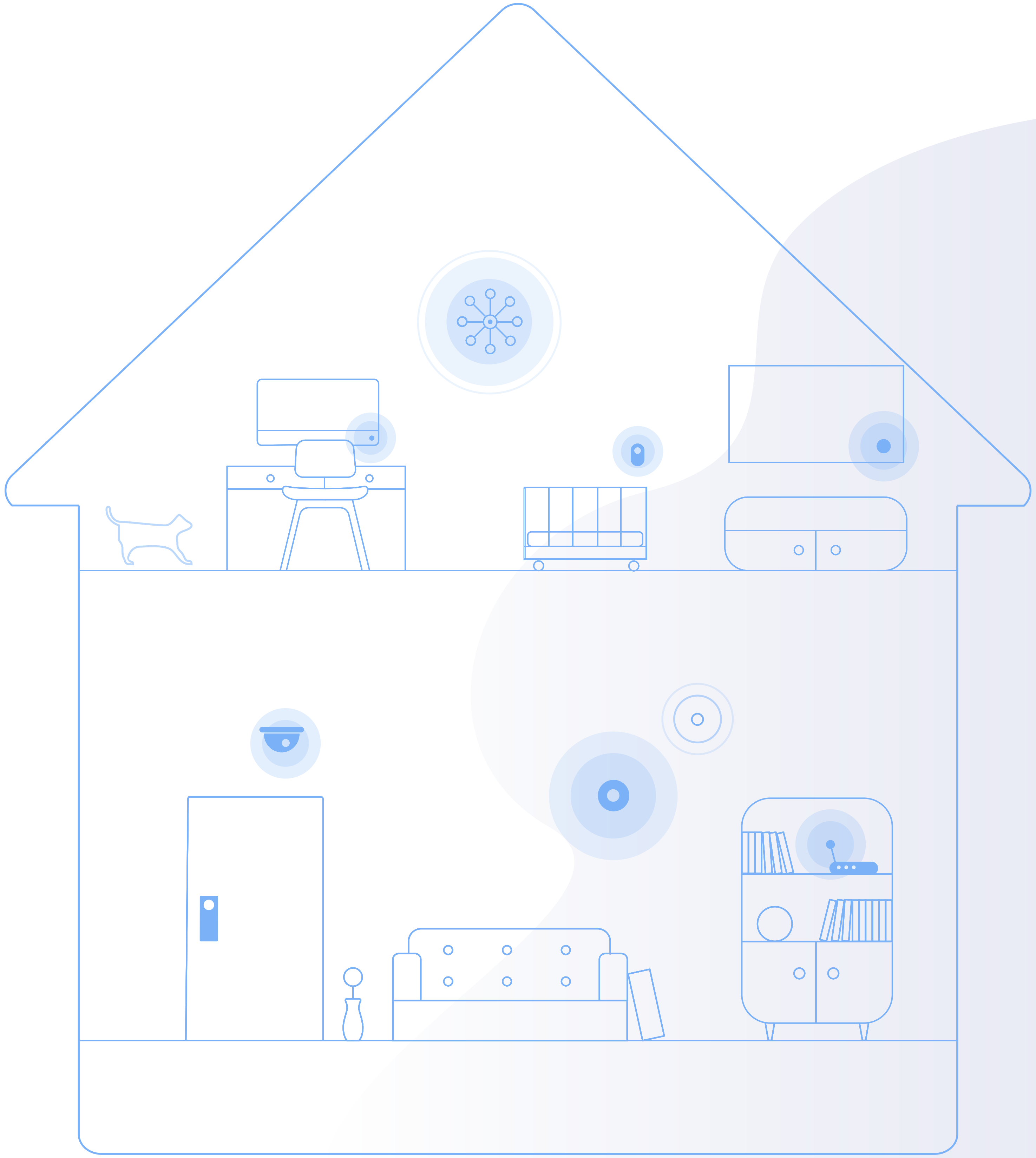
CUJO AI Explorer is based on an extensive set of device model detection rules. As there is a significant volume of new devices appearing on the market, it is a challenge to keep the rule base up to date. To tackle this problem, we have developed an ML-based system that helps Device Intelligence analysts by automatically recommending new rules and device models.

CUJO AI accesses data for over 1 billion devices.

Device identification is based on a device fingerprinting mechanism that is built in to CUJO AI-powered routers. Our approach treats the fingerprints as text documents and applies modern natural language processing techniques to analyze them.

CUJO AI accesses data for over 1 billion devices. This allows us to build custom language models for device fingerprints from anonymous data.

Using the language model, the rule recommender identifies relevant-looking fingerprint segments that are not yet used in the production system and recommends new device model descriptions with generated rules for the analysts.



Summary

Rule-based vs. Machine Learning

Conventional rule-based network security methods are **not capable of identifying novel malicious activity** like zero-day threats, browser fingerprinting techniques, or new IoT devices.

ML-based methods are considered **more flexible and efficient** since they can learn patterns, model the relationship between the patterns and identify new cases based on their similarity to known ones.

ML-based methods also require **less maintenance** than traditional methods: Constantly Feeding the ML model with new data is less difficult than keeping a rule base up to date.

AI Is the Trending Solution

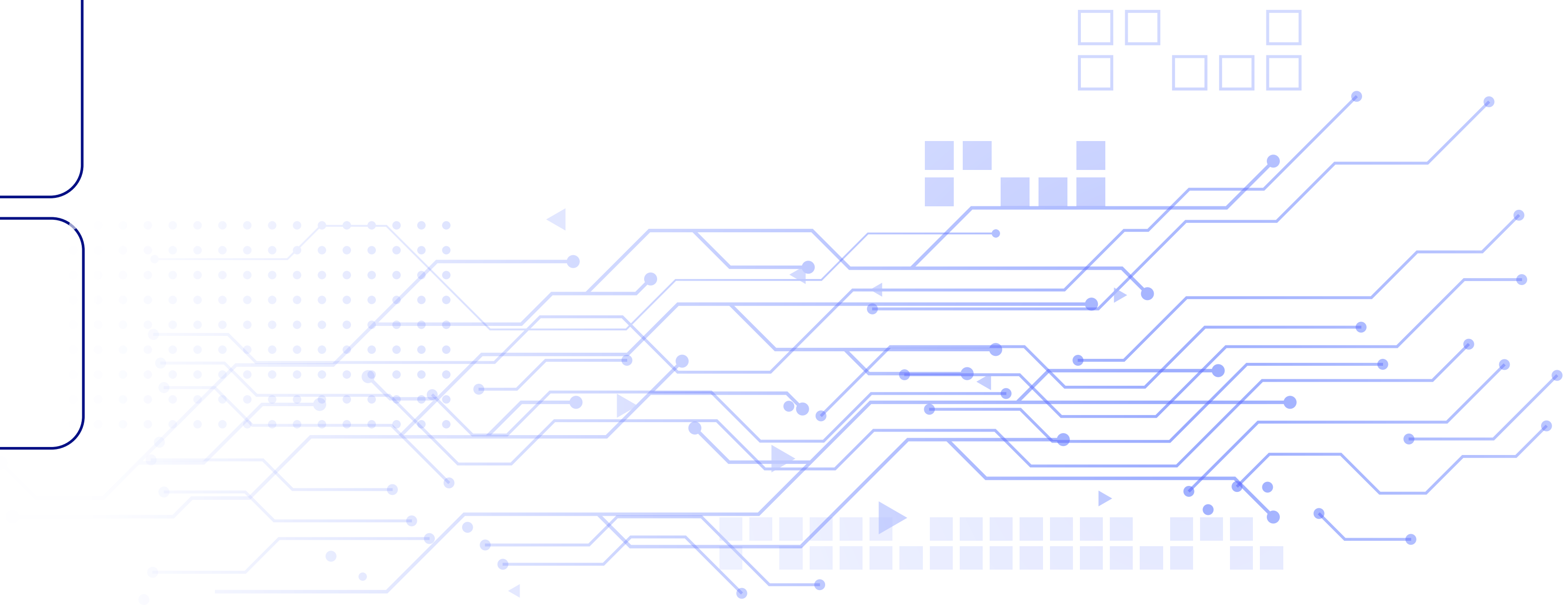
As a response to the evolution of cybersecurity threats, ML-based network security solutions are no longer simply nice to have. Instead, they are crucial to stop modern, well-resourced and sophisticated cyberattacks.

AI empowers telecom operators to **detect threats, react to threat actors and block malicious activities within minutes**, as opposed to the previously standard processes where a human analyst interaction was necessary.

Artificial Intelligence in CUJO AI Portfolio

CUJO AI's ML algorithms **actively learn from network traffic, device behavior** and other sources to protect the users, precisely identify the devices and offer advanced content controls.

Our solutions for telecom operators contain **ML classifiers** to differentiate malicious web addresses from legitimate ones as well as detect browser fingerprinting (which is impossible with conventional rule-based methods), including an **ML-based custom anomaly detection** method that recognizes the unusual behavior of IoT devices and a **ML-assisted system** for recommending device identification rules for analysts.



CUJO AI

CUJO AI is the global leader in the development and application of artificial intelligence to improve the security, control and privacy of connected devices in homes and businesses.

CUJO AI brings to fixed network, mobile and public Wi-Fi operators around the world a complete portfolio of products to provide users with a seamlessly integrated suite of Digital Life Protection services while improving their own network monitoring, intelligence and protection capabilities.

More information: **connect@cujo.com**

Media inquiries: **press@cujo.com**

cujo.com