AMRITA VISHWA VIDYAPEETHAM
DEEMED TO BE UNIVERSITY UNDER SECTION 3 OF UGC ACT, 1956

SCHOOL OF COMPUTING

A Proxy-Based Security Architecture for Mitigating Unauthorized LLM Access to Web Applications

***Domain:** Cybersecurity with a focus on Web Application Security and Content Protection*

Presented By:
Gundu Bhavana (CH.EN.U4CYS22017)
Dhanyasree Thallapalli (CH.EN.U4CYS22059)

Mentored By:
Dr. Saravanan [CSE - CYS]

Problem Identification

- In the era of LLMs and generative AI, news and media websites face rising threats from **advanced LLM based web scrapers** that semantically extract headlines, articles, and interviews for model training or content generation. Unlike traditional bots, these agents use **browser automation and mimic human behavior** to evade detection.
- This leads to **unauthorized use of proprietary content, undermines licensing models, and exposes platforms to reputational and legal risks** from hallucinated or misattributed information.

Problem Statement

- This project helps develop a **proxy-based middleware** to detect and control **LLM-powered web scrapers** targeting news and media sites. Deployed between users and the web server, it **intercepts HTTP requests** and uses **behavioral analysis, header fingerprinting, and honeypots traps** to identify suspicious activity.
- A **machine learning** model then classifies each session as **human, traditional bot, or LLM agent**. Based on the classification, the system dynamically blocks, deceives, or allows access — without disrupting legitimate users.

Supporting References

Case	Year	Description
Indian news agency ANI sues OpenAI for unsanctioned content use in AI training	2024	• Indian news agency ANI filed a lawsuit in the Delhi High Court in November 2024, accusing OpenAI of using its copyrighted content without permission to train ChatGPT and fabricating news stories attributed to the agency. • OpenAI responded by blocking ANI's website from future data scraping, but ANI contended that its content remains permanently stored in ChatGPT's memory with "no programmed deletion" .
NYT v. OpenAI: The Times's About-Face	2023	• The New York Times sued OpenAI and Microsoft in December 2023, alleging unauthorized use of millions of its articles to train AI models that reproduce content nearly verbatim, threatening its subscription-based business. • The case marks a major shift in the Times's legal stance, now fighting to protect its journalistic content and challenging the notion that copyrighted material can be freely used for AI training, with the court allowing most claims to proceed.

Supporting References

Author	Title	Journal/ Conference	Volume No	Year	Obejective
Jonne Y. Guyt, Hannes Datta, Johannes Boegershausen	Unlocking the Potential of Web Data for Retailing Research	Journal of Retailing	100	2024	This paper reviews web data applications in retailing, discusses its value, and provides a guide for researchers to incorporate web data collection into their routines. It also highlights the role of generative AI/LLMs in kickstarting web data collections
William Brach, Matej Petrik, Kristián Košt'ál, Michal Ries	Ghosts in the Markup: Techniques to Fight Large Language Model-Powered Web Scrapers	37th Conference of Open Innovations Association (FRUCT)	--	2025	This paper introduces a lightweight, two-step client-side defense strategy that combines content obfuscation and defensive prompt injection to effectively combat LLM-powered web scrapers by manipulating their content extraction behavior and protecting valuable web content

Supporting References

Author	Title	Journal/ Conference	Volume No	Year	Obejective
Zhengmian Hu, Gang Wu, Saayan Mitra, Ruiyi Zhang, Tong Sun, Heng Huang, and Viswanathan Swaminathan	Token-level adversarial prompt detection based on perplexity measures and contextuaL information	**International Conference on Learning Representations (ICLR)**	--	2025	This work proposes a novel approach to detecting adversarial prompts at a token level by leveraging an LLM's capability to predict the next token's probability. It aims to protect LLMs from being used in harmful ways and enhance their robustness against such attacks.
Julien Piet, Maha Alrashed, Chawin Sitawarin, Sizhe Chen, Zeming Wei, Elizabeth Sun, Basel Alomair, and David Wagner	Jatmo: Prompt Injection Defense by Task-Specific Finetuning	**European Symposium on Research in Computer Security (ESORICS)**	--	2024	This paper introduces Jatmo, a method for generating task-specific models resilient to prompt-injection attacks by fine-tuning non-instruction-tuned LLMs. It aims to significantly reduce prompt injection success rates while maintaining output quality.

Objectives

1. To **detect and differentiate LLM-powered web scrapers from human users and traditional bots** using ML, behavioral analysis and request fingerprinting.
2. To train and integrate a machine learning classifier that **identifies AI-powered scraping patterns** in real time.
3. To ensure **minimal latency** and no disruption to genuine user experience while maintaining high detection accuracy.

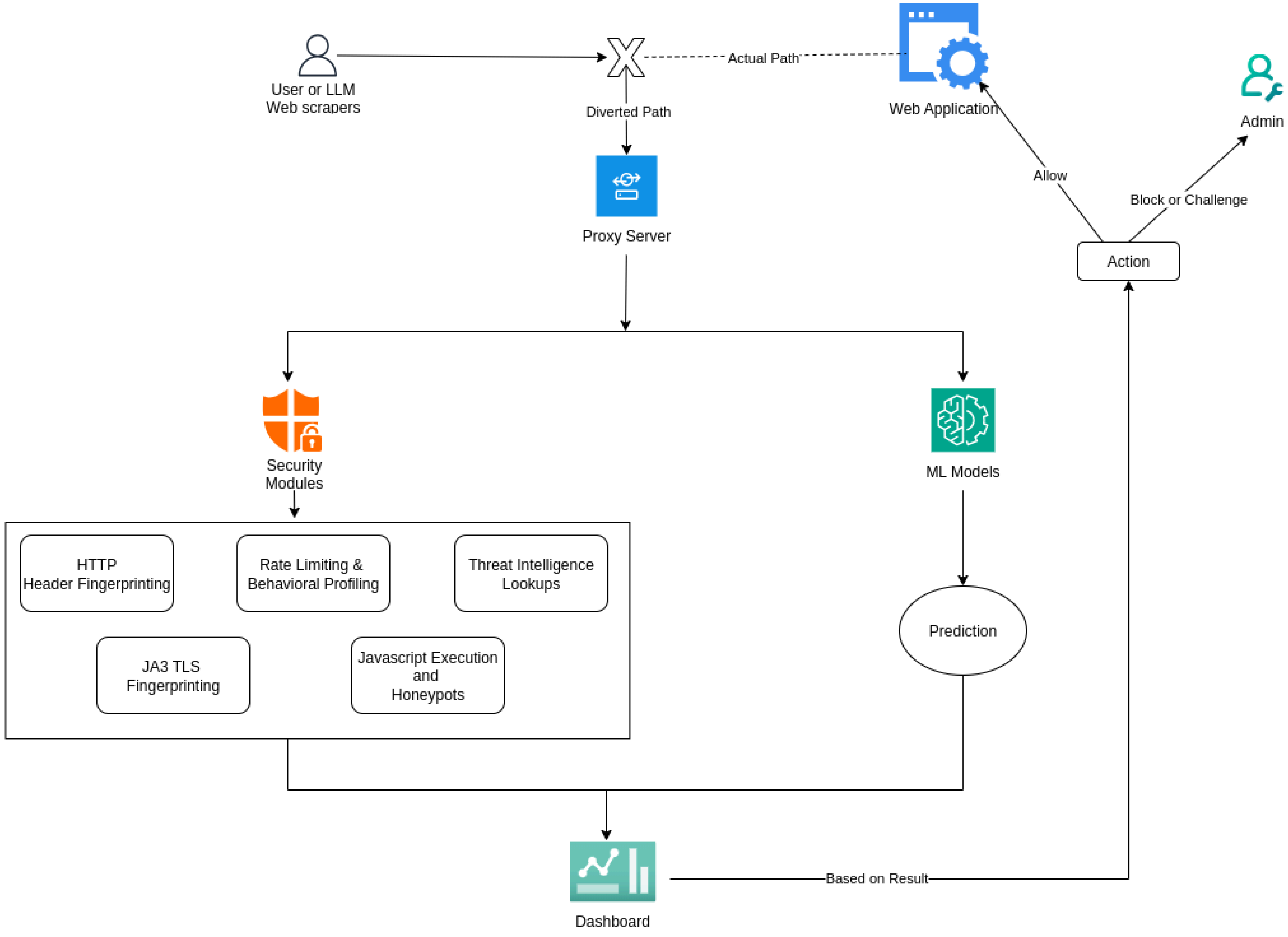
Scope

1. Applicable as a **drop-in proxy security layer** for any web application or enterprise API, requiring no modification to the core app.
2. Capable of **expanding into real-time traffic analysis, threat intelligence integration, and adaptive ML-based defense**, making it scalable for large platforms and data-sensitive domains (e.g., News Agencies, Legal teams, Media and content, compliance regulators).

Significance

1. While existing tools **focus only on blocking,** this system is uniquely designed to **monitor and analyze** scraper activity in real time, **offering the missing visibility and intelligence layer**—with built-in support for whitelisting trusted crawlers and domains, which no current CAPTCHA or bot defense provides.
2. While traditional defenses like **CAPTCHA** introduce noticeable **latency (often 300–600ms** or more due to **human interaction** and server-side validation), this system performs automated traffic analysis and **decision-making within ~100 - 150ms**, enabling real-time protection **without disrupting legitimate user experience.**

Architecture Diagram



AMRITA VISHWA VIDYAPEETHAM
DEEMED TO BE UNIVERSITY UNDER SECTION 3 OF UGC ACT, 1956
SCHOOL OF COMPUTING

Thank You

11 July, 2025