Understanding and Detecting Abused Image Hosting Modules As Malicious Services

ACM CCS (November , 2023)

Geng Hong, Mengying Wu, Pei Chen, Xiaojing Liao*, Guoyi Ye, Min Yang

Fudan University, Indiana University Bloomington*

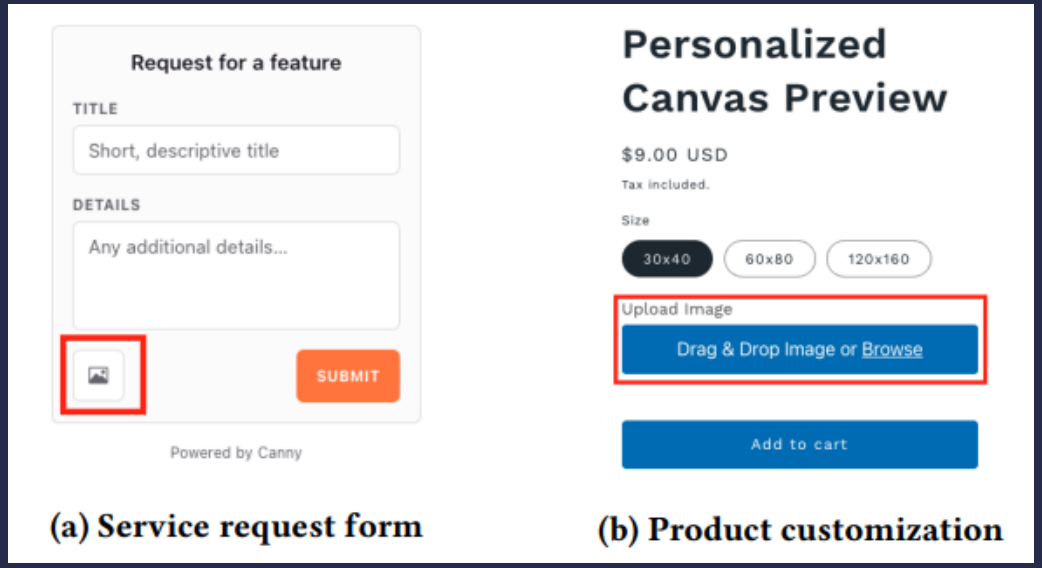
01

Introduction

Introduction

- Images are critical in facilitating communication and information sharing across diverse web services.

- Many web services rely on image hosting modules (IHMs) to manage user image uploading, hosting, and sharing.



(a) Service request form (b) Product customization

Abused IHMs as malicious services (AIMIE)

- The adversary exploits vulnerable IHM upload APIs of image hosting modules (IHMs) to create a malicious service.
- Miscreants can exploit vulnerable IHMs to store a considerable volume of images, wasting a large amount of storage space in the victim sites and a subsequent increase in storage expenses.
- The risk of miscreants leveraging abused IHMs to store and reference illicit content.

Detecting vulnerable IHMs in the wild

- Interested in understanding the prevalence of vulnerable IHM upload APIs.
- Developing Viola, a tool to assess the security of IHMs.
- First recognizes web services that implement IHMs in a given domain.
- Analyzing each stage of the image upload lifecycle
  – presubmit, preview, submit, and callback –
  to determine if there are any interfaces that can be accessed and exploited by third parties to upload and host malicious images

02

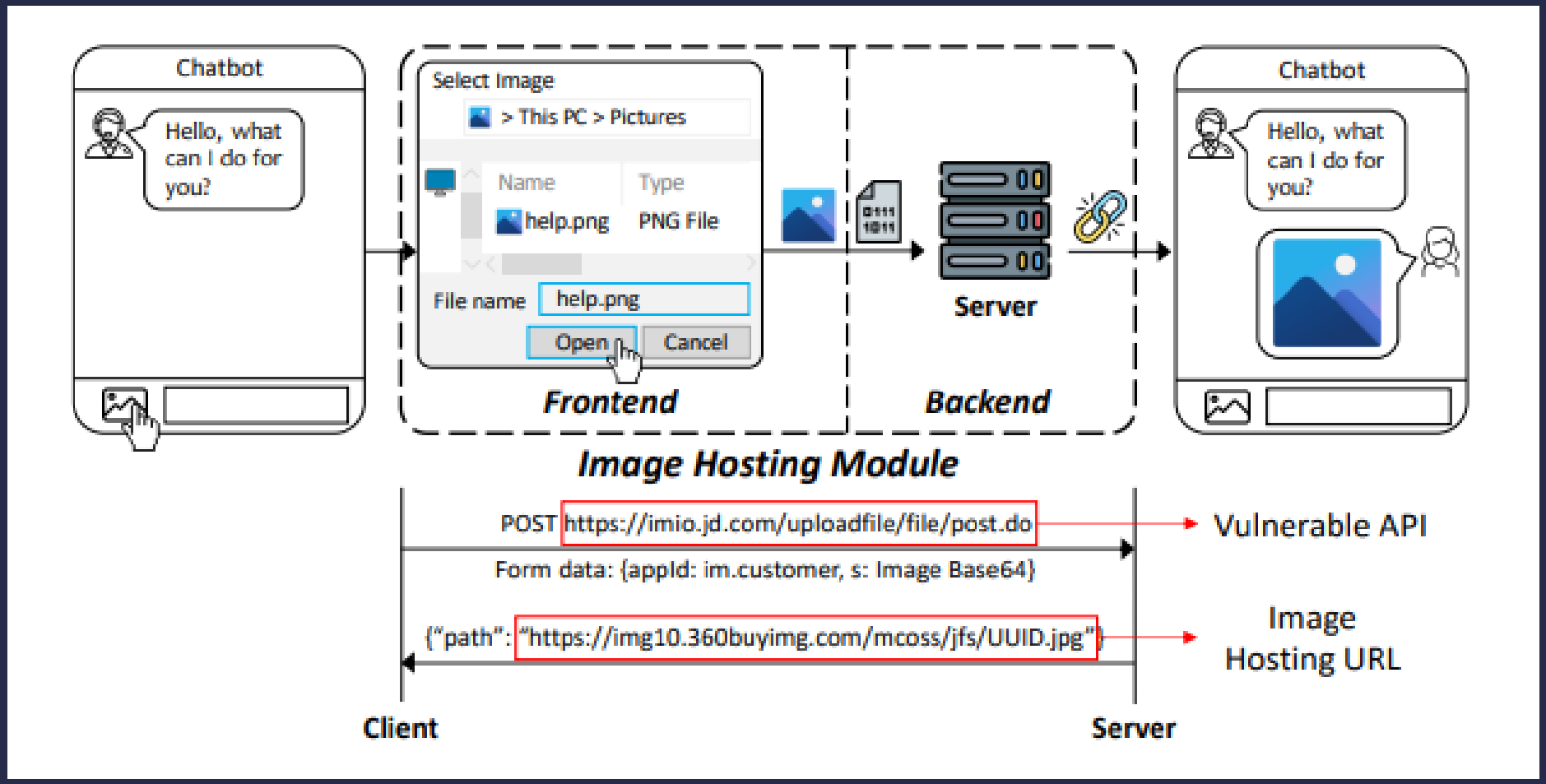
Background

Image Hosting Module (IHM)

- Web image hosting service can be categorized as types of infrastructure-as-a-service or code module-as-a-service.

- Once the image is hosted on the server, it can be accessed or shared.
- An IHM is not intended to function as a standalone image hosting platform.
- The hosting web service of an IHM sometimes restricts the type of images that can be uploaded or shared via IHMs in alignment with the rules and policies of the hosting web service.

IHM Workflow

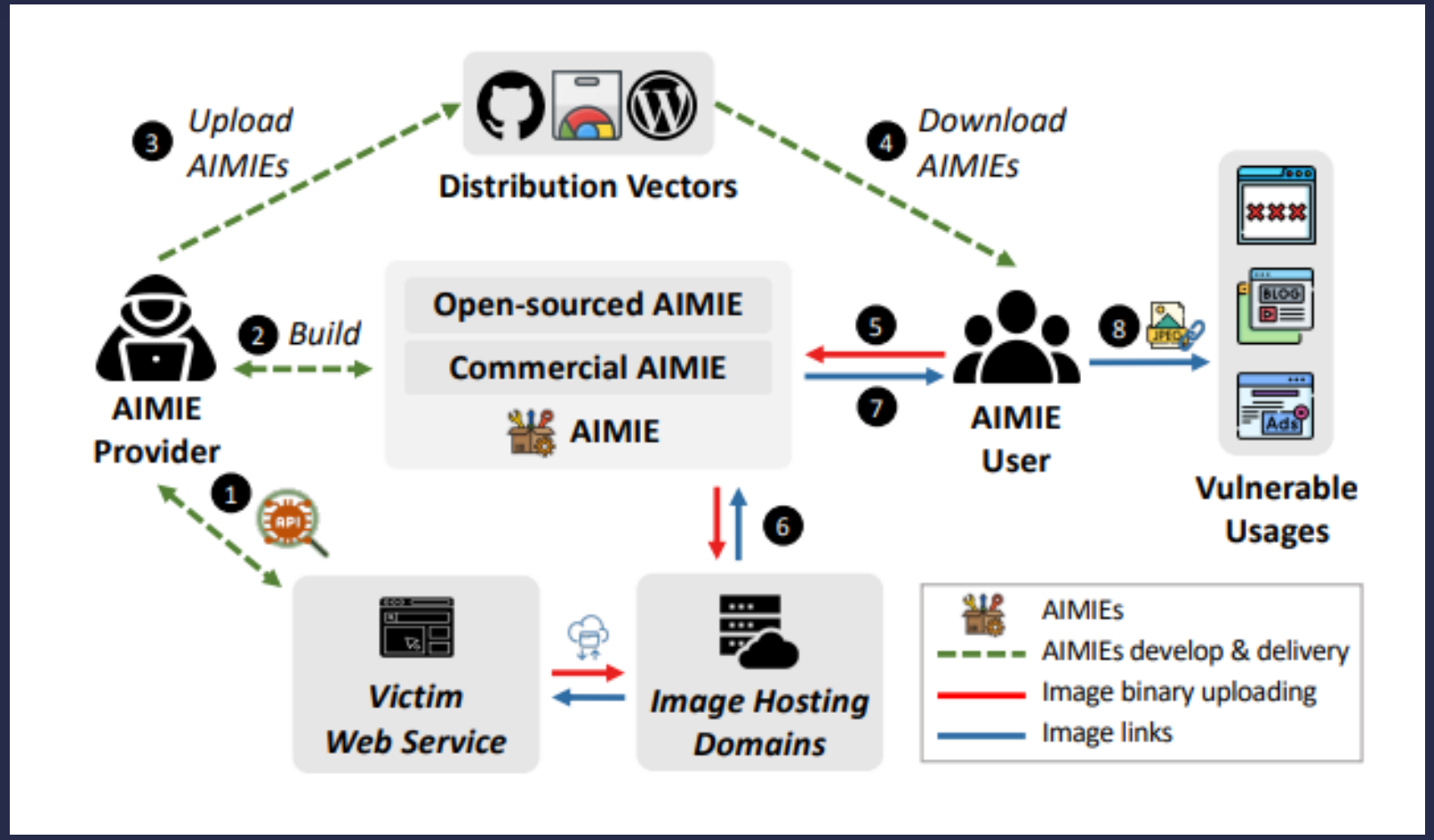


03

AIMIEs

Abused IHM as Malicious Service

Threat Model - AIMIE Workflow



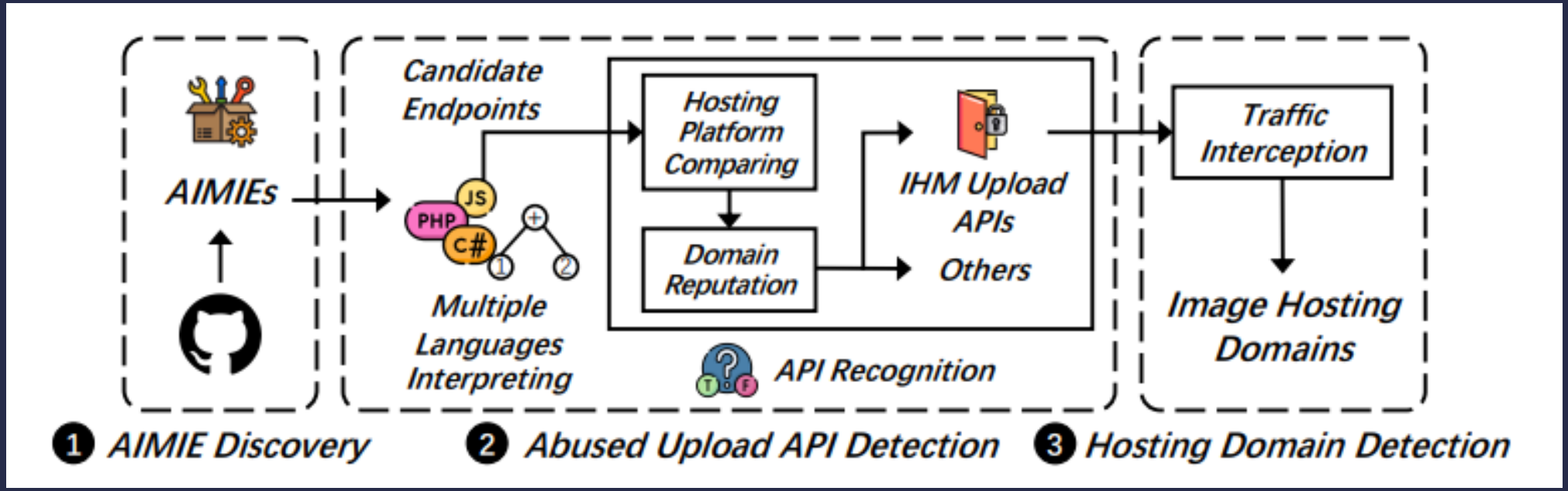
AIMIE upload API

- An AIMIE upload API is an interface offered by an AIMIE service, facilitating users in uploading images.
- These upload APIs are associated with vulnerable IHMs, image hosting platforms, or other open-sourced/commercial AIMIEs.
- An upload API is usually in the form of domain and path,
  e.g., victimhost.com/uploadfile/file/post.do,
- Specifying in the HOST and PATH fields in the POST request, respectively.

Image hosting Domain

- The domain of a victim image hosting server, which hosts and stores uploaded images.
- The AIMIE ecosystem is flourishing with various entities, including open-sourced AIMIEs, such as auxpi, CDNDrive and commercial AIMIEs, e.g., yum6.
- Commercial AIMIEs usually offer a variety of hosting plans and flexible service models.

Methodology overview



Methodology - AIMIE discovery

- Collecting a set of open-sourced AIMIEs for analysis.
- Founding 508 open-sourced AIMIE candidates.
- We studied documentation, git commit logs, and comments to confirm both the abusive behavior and the services provided by the AIMIE.
- Results:
  - Confirmed 89 open-sourced AIMIEs by examining 508 candidates.

Methodology - Abused upload API recognition

- Developing a multi-language interpreter to extract abused APIs scattering over the open-sourced AIMIE.
- The interpreter will parse the source code of each open-sourced AIMIE.
- Build our interpreter on top of open-source parsers to the five most popular programming languages (JavaScript, PHP, Python, Golang, and Java) in open-sourced AIMIEs.
- To determine whether those URLs are candidate APIs , we investigate semantic information (e.g., FQDN, path) of each URL.

Methodology - IHM upload APIs triage

- With a set of API candidates output by the above multi-language interpreter, our approach further classify abused IHM upload APIs.

- we compile a list of image hosting platforms by utilizing search engines, resulting in a list of 98 domain names associated with such platforms.

- Results :
  - Among 89 open-sourced AIMIEs, we found most of them implemented in JavaScript (46), followed by PHP (16), Python (7) and Golang (7), and Java (6).

Methodology - Abused hosting domain detection

- Deploying 14 open-sourced AIMIEs, which covered all 109 IHM upload APIs found in our study, in virtual machines with Ubuntu 20.04.
- Triggering each IHM upload API in open-sourced AIMIEs and log response traffic using burpsuite and Proxychains.
- Building the mapping between abused IHM upload APIs and the associated image hosting URLs
- Analyzing the payload of each abused IHM upload API's POST response to extract the URL within it.
- Results:
  - We confirm 76 valid abused IHM upload APIs and map them to 122 image hosting domains

04

Measurement

Open-source AIMIEs

- 89 unique open-sourced AIMIEs, which exploit 109 IHM upload APIs and abused 127 image hosting domains.

- Their lines of code (LoC) range from 717 to 206,891. We observe open-sourced AIMIEs as a thriving ecosystem: the earliest open-sourced AIMIE found in our study was in 2017.

Developer of open-sourced AIMIEs

- Referring to authors in the commit logs.
- Using the email address of the commit author to pinpoint an individual code contributor.
- In total, we found 294 developers among the 89 open-sourced AIMIEs
- Founding about 56.2% of open-sourced AIMIEs were updated more than three times per month, and 13.5% were updated more than 10 times per month.

Impacts of open-sourced AIMIEs

- Studying the number of stars and forks of open-sourced AIMIEs' GitHub repositories.
- The auxpi library has been forked over 380 times. Piles of abusive repositories are created based on it for various purposes, such as online shopping or content management system (CMS).

open-sourced AIMIE	# of Stars	# of Forks	Language	# of Abused APIs	# of Host Domains
0xDkd/auxpi	2,636	377	Go	12	22
apachecn/CDNDrive	668	90	Python	11	34
Mikubill/transfer	630	91	Go	9	12
ShareX/CustomUploaders	372	228	sxcu	2	0
szvone/imgApi	183	47	PHP	2	13
iAJue/Alibaba_pic	173	88	PHP	1	4
BlueSkyXN/KIENG-FigureBed	119	153	PHP	9	0

IHM Upload API Analysis

Domain	Path	# of *open-sourced AIMIE*	% of *open-sourced AIMIE*
kfupload.alibaba.com	/mupload	23	25.84%
you.163.com	/xhr/file/upload.json	13	14.61%
search.jd.com	/image	11	12.36%
mp.toutiao.com	/upload_photo	11	12.36%
pic.sogou.com	/pic/upload_pic.jsp	10	11.24%
cdn-ms.juejin.im	/v1/upload	9	10.11%
review.suning.com	/imageload/uploadImg.do	9	10.11%
picupload.service.weibo.com	/interface/pic_upload.php	8	8.99%
prntscr.com	/upload.php	7	7.87%
changyan.sohu.com	/api/2/comment/attachment	7	7.87%
shopapi.io.mi.com	/homemanage/shop/uploadpic	7	7.87%

Images hosted via AIMIEs

- Images uploaded through specific IHM upload APIs are typically stored in the same resource directory path, which helps developers maintain an organized file structure.

- Checking whether URLs of explicit images matched the patterns of URLs of uploaded images that we detected

- For example, on Tencent clouds customer service, images uploaded through the abused API yzf.qq.com/fsnb/kf-file/upload_wx_media will be hosted on the victim server with the following pattern:
  yzf.qq.com/fsnb/kf-file/kf_pic/UUID.jpg.

Top 10 abused companies

Company	Sample FQDN	# Related FQDNs	# Image Links	# Blocklist Domains
Tencent	p.qlogo.cn	7	428	56,421
Alibaba	ae05.alicdn.com	16	443	51,201
Ctrip	dimg04.c-ctrip.com	1	314	29,502
Toutiao	p1.toutiaoimg.com	7	35	13,340
Baidu	wkphoto.cdn.bcebos.com	34	352	11,834
JD	dd-static.jd.com	13	399	3,170
Sina	tvax1.sinaimg.cn	19	297	2,034
Sohu	i2.itc.cn	19	130	687
Meituan	p0.meituan.net	2	51	519
360	p1.qhimg.com	17	65	34

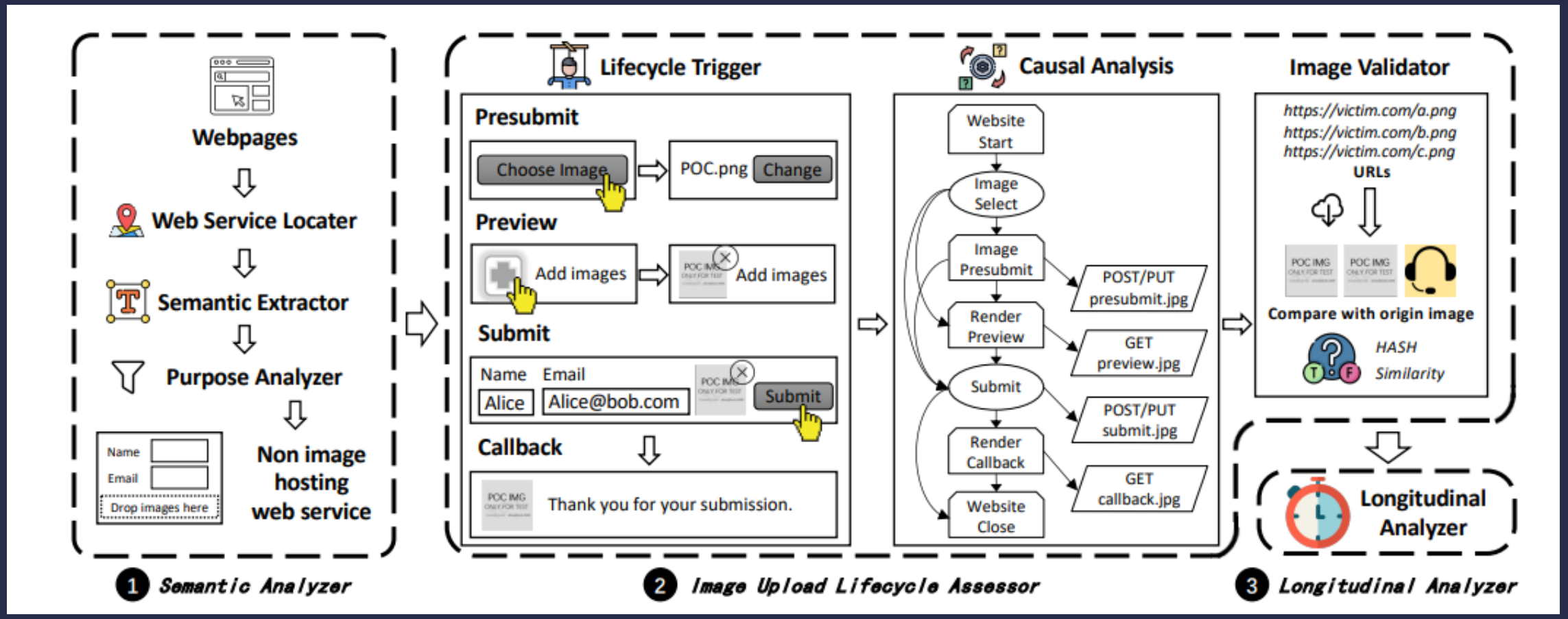
05

Viola

Viola

- Wondering about the prevalence of vulnerable IHM upload APIs in the wild.
- Consisting of a semantic analyzer, an upload lifecycle assessor, and a longitudinal analyzer
- Recognizes all IHM-enabled web services within the domain using semantic analyzer.
- Through upload lifecycle assessor, it assesses each stage of the image upload lifecycle (presubmit, preview, submit, and callback) to identify vulnerable IHM upload APIs using assessor to track the lifecycle.

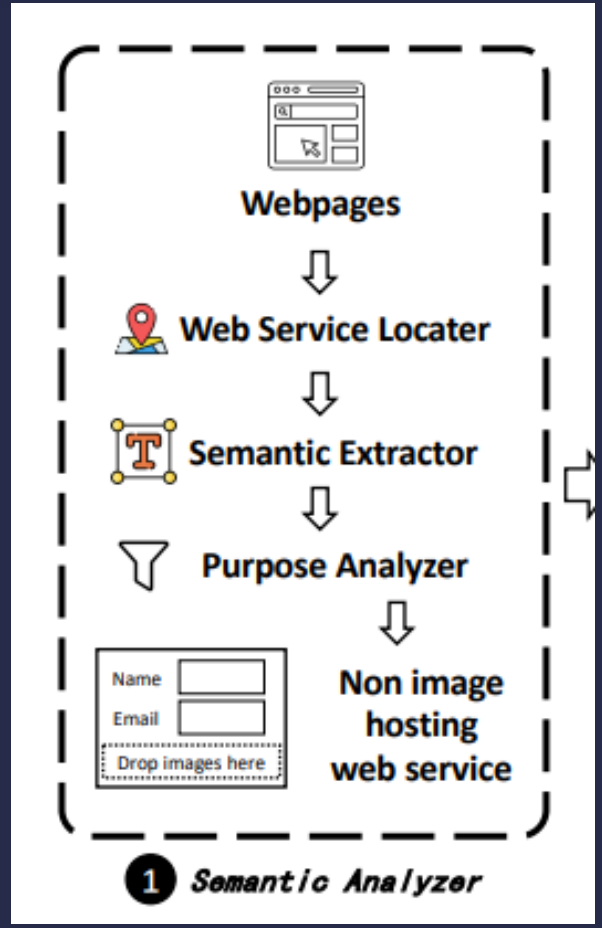
Overview of Viola



Semantic Analyzer

- Locating host web services:
  - Locating the client-side user interface of IHM.
  - Traversing the DOM tree to determine the host.
  - Using <input type="file"> locate user interface.

- Checking semantic inconsistency:
  - Constructing the semantic vector.
  - Comparing the Euclidean distance between the candidate web service and the IHMs semantics.
  - If the minimum distance is larger than a threshold, we take it semantics of web is inconsistent with IHMs.



Upload Lifecycle Assessor

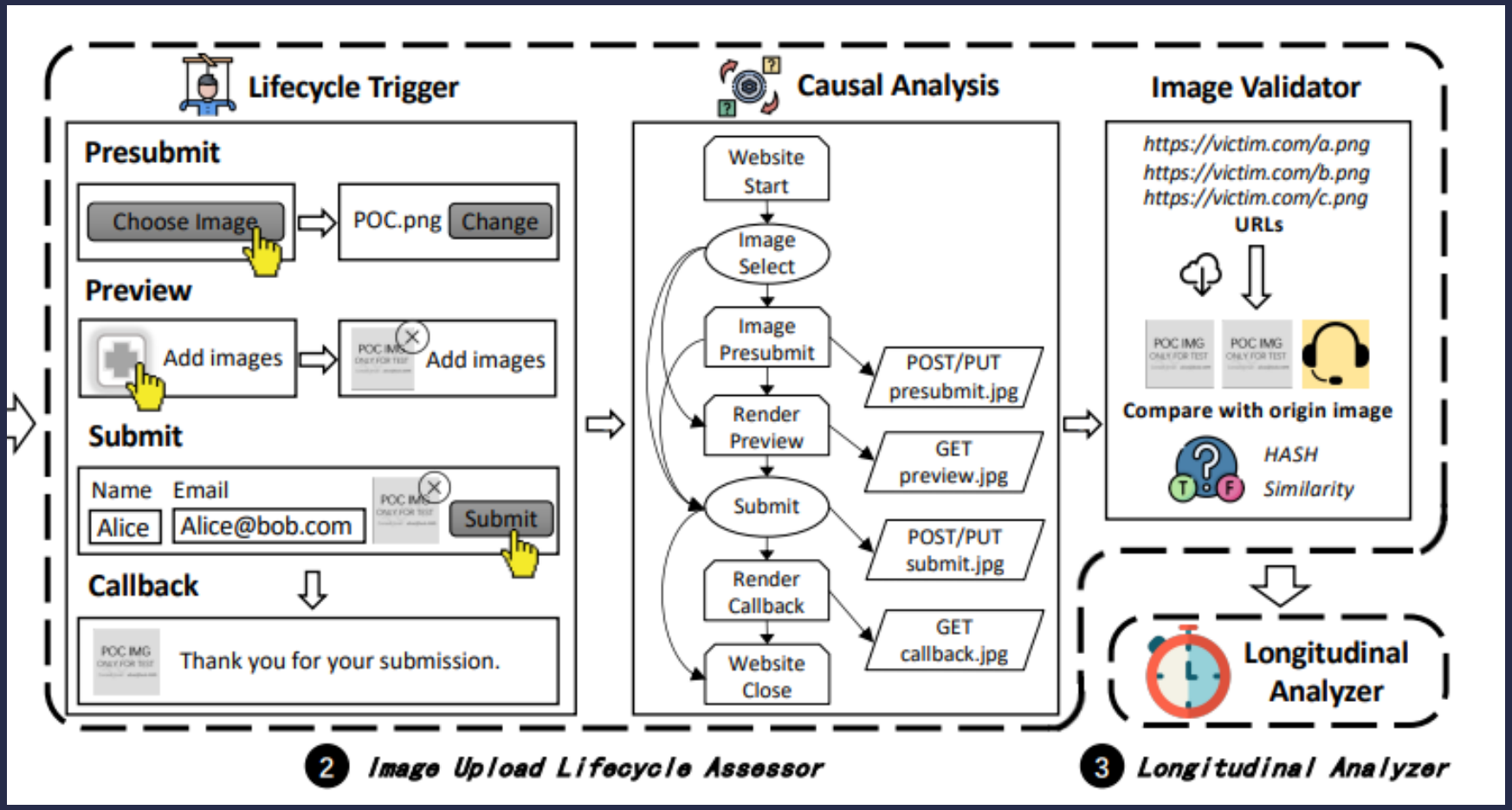
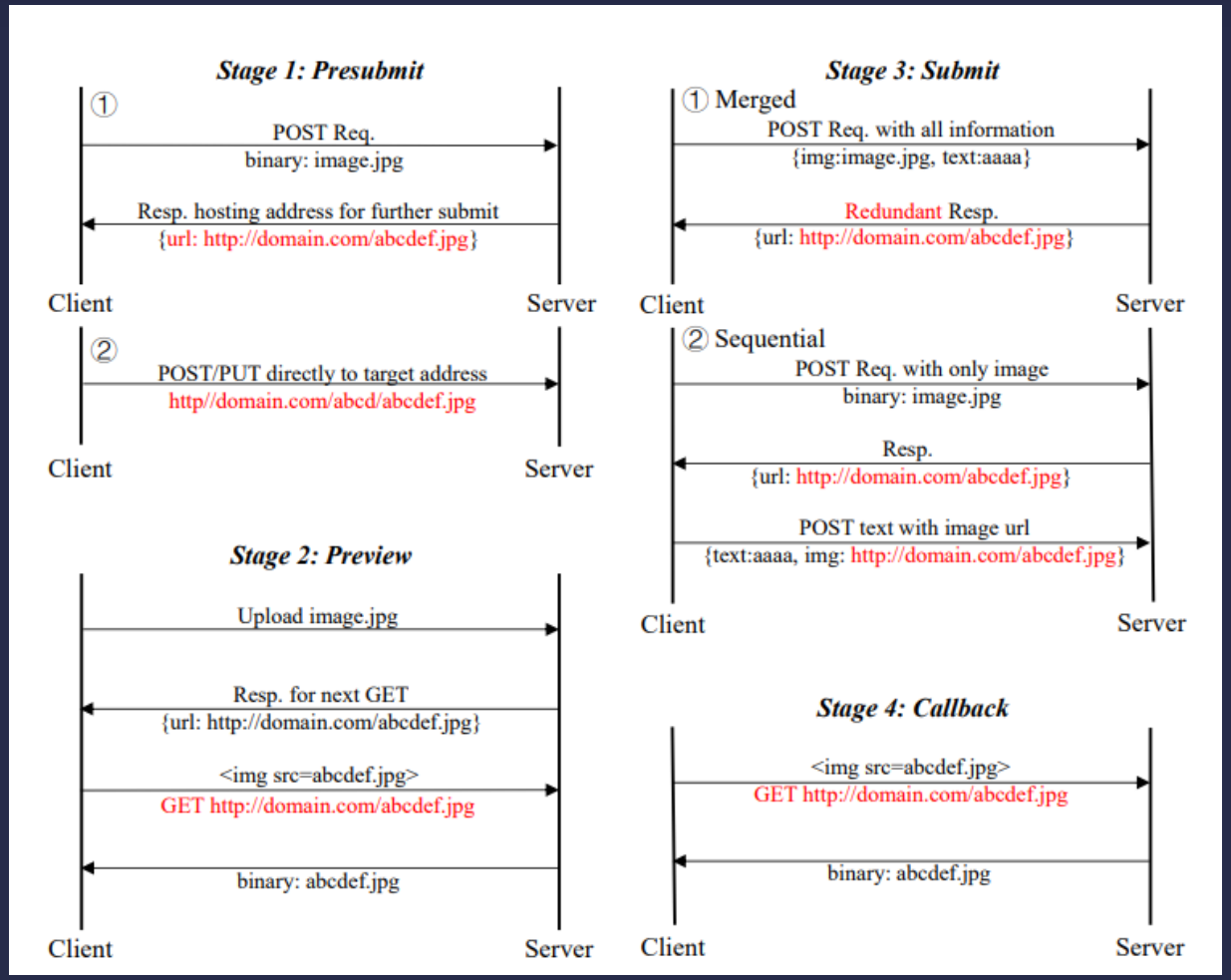
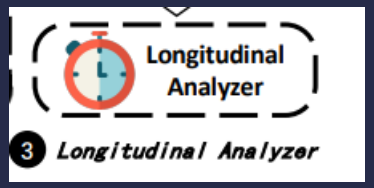


Image upload Lifecycle



Longitudinal Analyzer



- The AIMIE users abuse the IHM upload APIs to host their images for free and/or illegal purposes.
- Notice that a few images will be inaccessible after a fixed period due to the mitigating of websites.
- For each image URL, we continuously monitor the image URL at regular intervals within a period.
- If an uploaded image through the IHMs exists for more than 3 days, we consider it a vulnerability.

06

Evaluation

Effectiveness of Viola

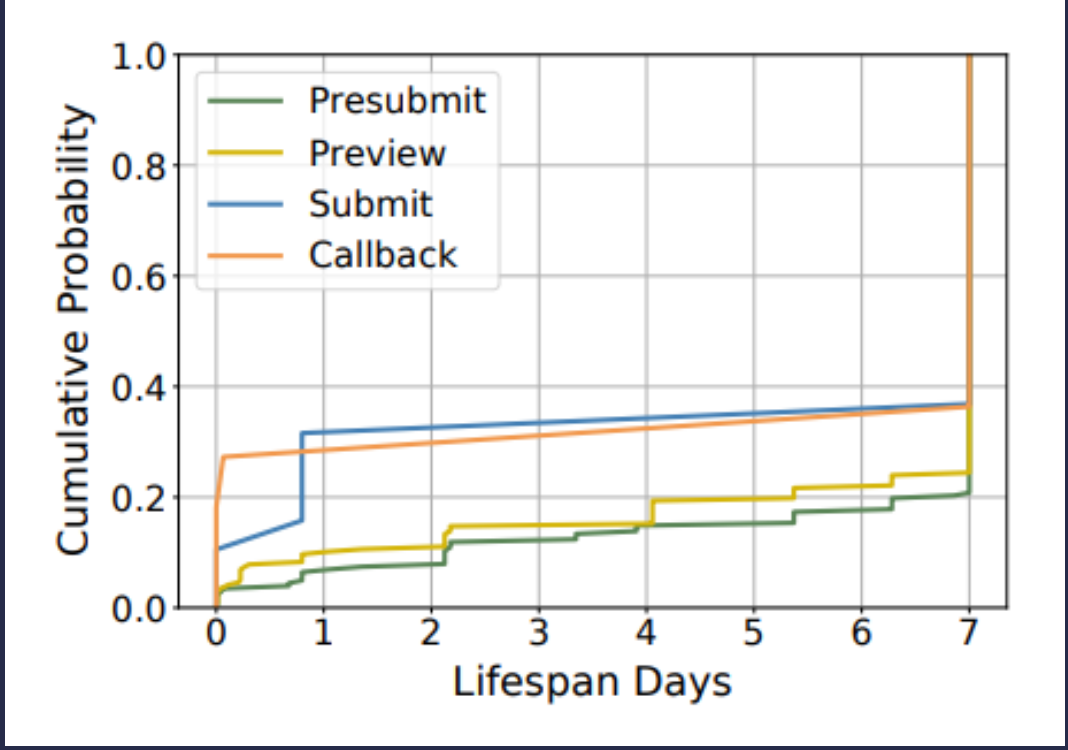
- Run all the experiments on a Ubuntu 18.04 server with Intel Xeon 2.8G, 32 cores CPU and 192 GB memory.
- Semantic Analyzer and Image Upload Lifecycle Assessor cost 17 hours and 25 hours, respectively.
- We assess precision by manually inspecting 100 randomly selected cases from the Viola results; 97 cases are true positive, and 3 are false positive.

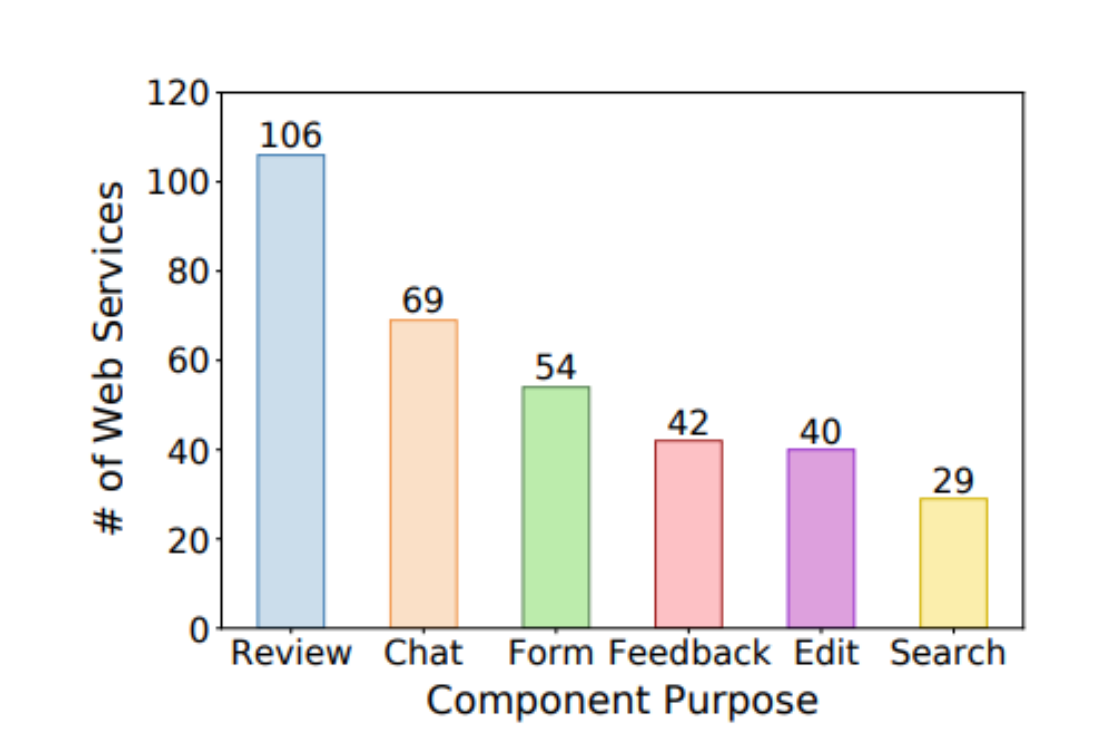
Effectiveness of Semantic Analyzer

		Semantic Distance Threshold				
		4.5	5.0	5.5	6.0	6.5
Segmentation Threshold	0.30	48.48%	67.53%	71.43%	77.55%	79.25%
	0.34	48.48%	70.00%	74.42%	80.00%	83.63%
	0.38	52.17%	72.29%	74.73%	**84.40%**	80.00%
	0.42	51.43%	71.43%	73.91%	83.63%	78.63%
	0.46	51.43%	71.43%	73.92%	83.63%	78.63%

Findings

- Image Lifespan & Vulnerable Web Service study





07

Mitigation

Root Cause

- Concerning the upload API aspect, developers often lack adequate control over user-uploaded images.
  - The absence of access control
  - IP-based rate limiting for request frequency,
  - the frequency of image uploads.

- The backend handling of images frequently overlooks the investigation of abnormal usage patterns.
  - In cases where images are uploaded during online chats with customer service on an e-commerce platform, such images should be accessible only from specific IP addresses and for a limited number of times.

Mitigations (1)

- Leverage the client-side cache:
  - In Preview or Callback stage, developers usually allow users to examine the result after selection or submission.
  - Adopting the Blob object and data URLs.

- Align the image hosting content and resource to purpose:
  - Developers should ensure that the image hosting content and associated resources align with their intended purpose.

Mitigations (2)

- Set up the access control for uploaded images:
  - The attacker inevitably received the server-side image links.
  - Recommend the developer set a proper expired time for the temporarily uploaded images.

- Hide image resource path:
  - Refers to the uploaded image, instead of the full genuine image resource paths.

08

Conclusions

Conclusions

- Miscreants are increasingly abusing image hosting modules as malicious services (AIMIEs) to host illicit images and disseminate harmful content
- the first measurement study of AIMIE services to provide an inside view of such vulnerability.
- We find that 1,151 explicit images are uploaded through 26 IHM upload APIs.
- we model the image upload lifecycle and construct a vulnerability Scanner, which can effectively and accurately discover 477 vulnerable IHM upload APIs in the wild.