

Into the Dark.

Unveiling Internal Site Search Abused for Black Hat SEO

Yunyi Zhang, Mingxuan Liu, Baojun Liu, Yiming Zhang, Haixin Duan,
Min Zhang, Hui Jiang, Yanzhe Li, Fan Shi



国防科技大学
NATIONAL UNIVERSITY
OF DEFENSE TECHNOLOGY



清华大学
Tsinghua University



Brief Summary

- Highlighted a new active yet overlooked Black Hat SEO technique, **ISAP (Internal site Search Abuse Promotion)**
- Proposed and implemented a lightweight ISAP detection scheme and found over **10,209 abused popular websites**
- Conducted a systematic **understanding of the ISAP ecosystem** to uncover its strategies and evaluate security risks

Promotion is Everywhere, Everywhen

Black Hat Search Engine Optimization

Getting exposure from search engines through unusual means

Promotion content

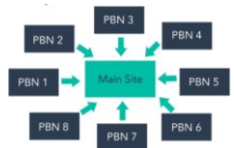


Illicit Drug Adult Service Gambling

Promotion Techniques



Keyword Stuffing Blog Comment Spam



Private Blog Networks Link Farms

Internal site Search Abuse Promotion (ISAP)



- ◆ Low-cost
No domain registration required
- ◆ Remarkably effective
Abuse the reputation of well-known domains

Internal Site Search

- ❖ A technique help users quickly locate the resources/subpages inside a website.
- ❖ 47.32% of Alex Top 1M websites implemented this function [1].

The screenshot shows a web browser at the URL <https://mastergardener.wsu.edu/?s=scholarship>. The page header includes the Washington State University logo and the text "Washington State University Extension Master Gardener Program". A navigation menu contains "Home", "Who We Are", "What We Do", "Get Involved", "Resources", and "News". The "Resources" menu item is highlighted with a red underline. Below the navigation is a large "Search" heading. A search input field contains the text "scholarship". A red search button with a magnifying glass icon is to the right of the input field. A blue box with a red border highlights the text "Search Keyword: scholarship" and points to the search input field. Another blue box with a red border highlights the text "Resources found within the wsu.edu website" and points to the search results area. The search results area shows a link for "Scholarship Information | Academics | Washington State University" with the URL <https://cahnrs.wsu.edu/academics/scholarships/>. Below the link is a small image of a person and the text "Four-year assistance. CAHNRS is committed to ensuring that its students have financial stability throughout their entire four years of college. Our goal is to ...".

[1] Kats D, Silva D L, Roturier J. Who Knows I Like Jelly Beans? An Investigation Into Search Privacy.

However, Internal Site Search has been abused to promote illegal content

Root Cause of Abuse

- ❖ A new URL under the same domain would be generated as the ISS result
 - ❖ Containing the search keywords, even for non-existent ones
- ❖ Abusers can inject promotional content (as keywords) to obtain subpages under popular websites

Generate a new URL under this domain

Reflection URL

Resources not found within the wsu.edu website

Inject promotional content into the content of this URL

Shantou Chaonan District Porn service (91pv.com)

Reflection Page

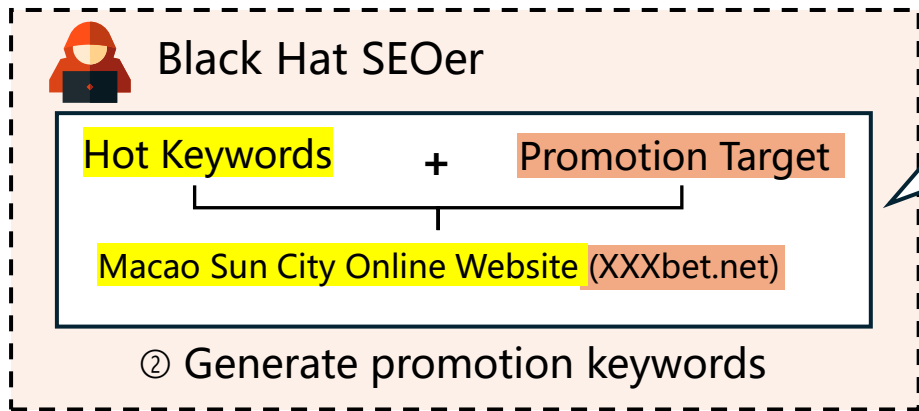
ISAP Process



ISAP Process



↓ ① Promotion target: XXXbet.net



Generate Promotion keywords

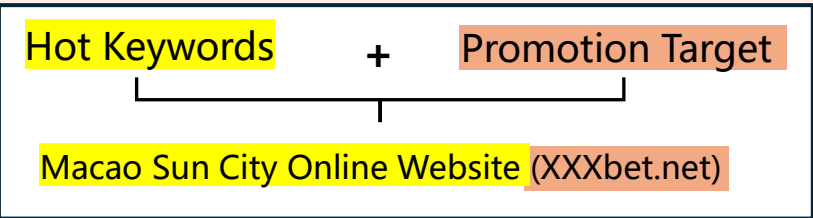
ISAP Process



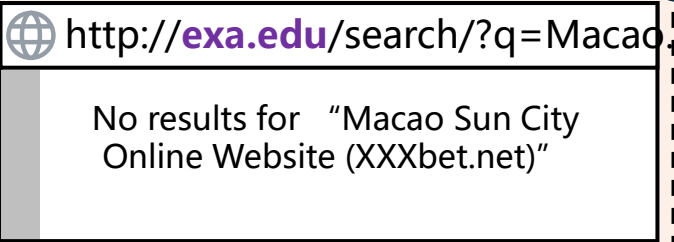
① Promotion target: XXXbet.net



Black Hat SEOer



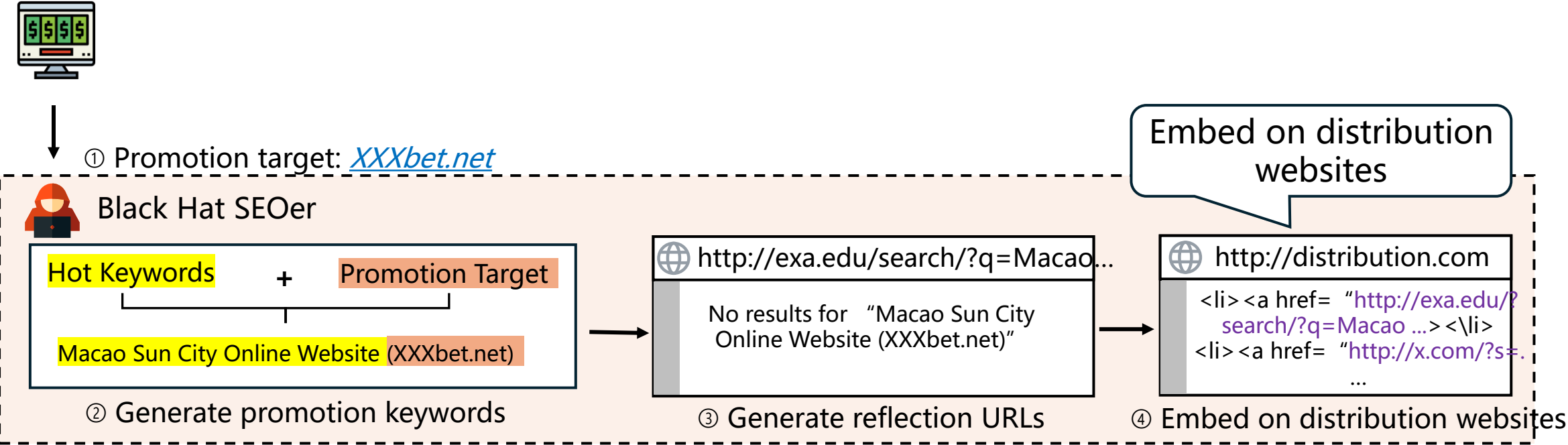
② Generate promotion keywords



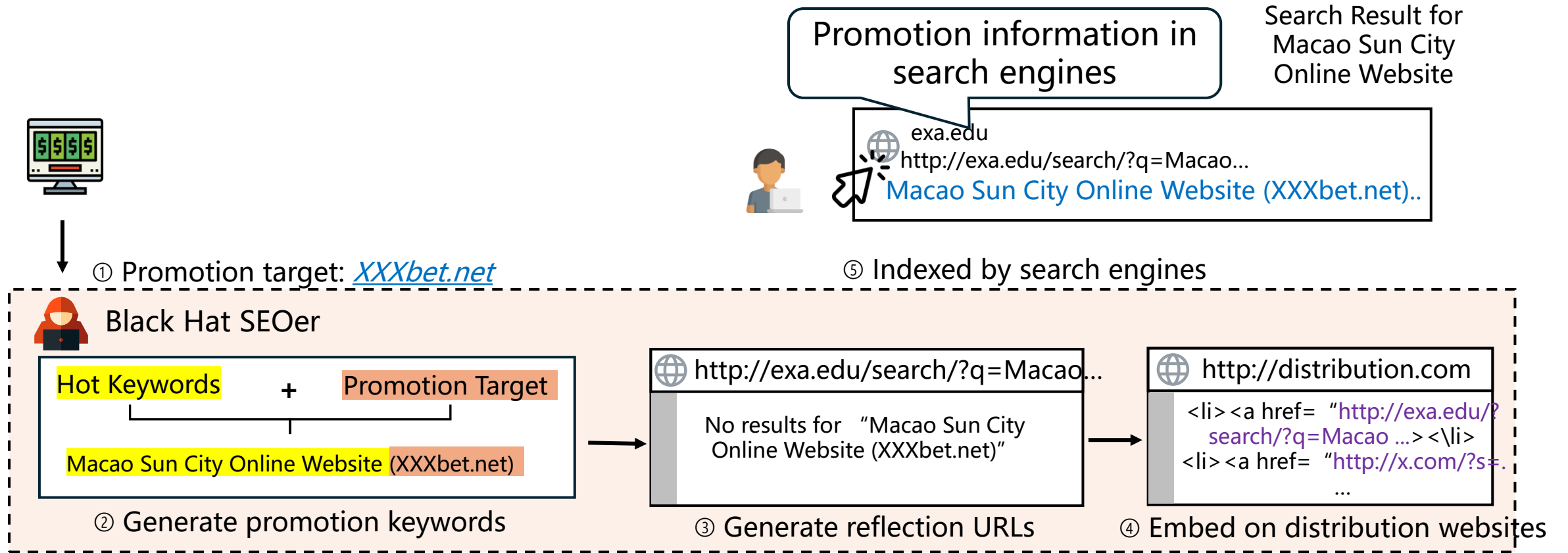
③ Generate reflection URLs

Find exploitable sites and generate reflection URLs

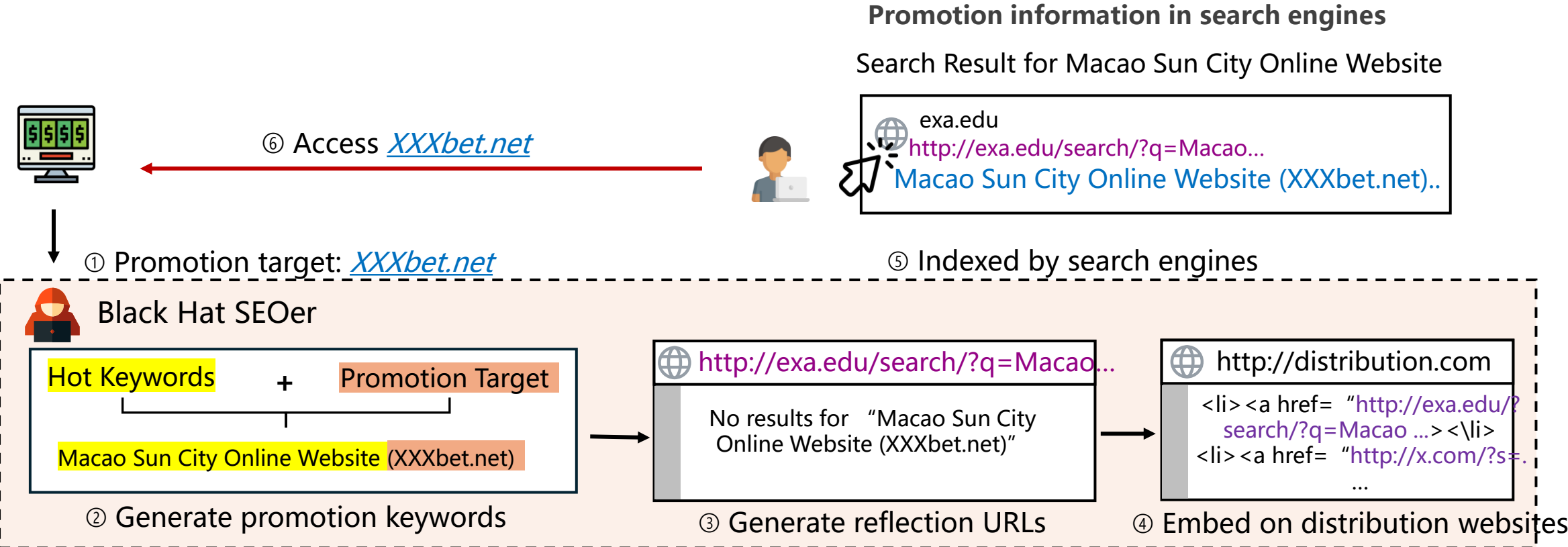
ISAP Process



ISAP Process



ISAP Process



Widespread Impact of ISAP

Looking student sex worker at Hohhot

All day Beijing automobile race

Tencent Cloud Agent; Accept USDT

Google
https://www.google.com/maps/viewer

呼和浩特玉泉区找小姐服务
呼和浩特玉泉区找小姐服务 谷歌地图担保人到付-河源有上门按摩服务电话-河源找明星上门按摩服务njJg.

Google
https://www.google.com/maps/viewer

呼和浩特怎么找美女预约小姐服务明星伴游(微信) ...
... 预约小姐服务明星伴游 呼和浩特按摩小姐服务 呼和浩特找小姐上门服务 呼和浩特哪里找小姐 呼和浩特找小姐学生妹过夜上门按摩服务444tnh.

Black Nova
https://blacknova.co/s=呼和浩特怎么找小姐学生妹过...

呼和浩特怎么找小姐学生妹过夜上门按摩服务 ...
Manage consent. BLACK JACK · ALBA · ANY · ARIA · AXES · Get Black Nova · Create your keypad · Login · ANY · ALBA · ARIA · AXES · Resources · News · CONTACT ...

American Driving Academy
https://www.americandrivingacademy.com/s=呼和浩...

哪有妹子上门特殊按摩spa服务
CALL TODAY! 1-800-604-6741 · Login · Cart \$0.00 ...

Allcare Carnes Hill Medical Centre
https://allcarecarneshillmedicalcentre.com/s=呼和浩特...

呼和浩特找明星按摩上门服务 (找妹网) 101 可以人到再 ...

Google

网易新闻
https://news.163.com/news/search?keyword=%E5%85%A8%E5%A4...
全天北京赛车三码计划-全天北京赛车三码计划-【访问...
网页 全天北京赛车三码计划-全天北京赛车三码计划- () 全天北京赛车三码计划-全天北京赛车三码计划- () 全天北京赛车三码计划-全天北京...

辛加龙游戏
https://www.singalongktv.com/gonglue/186309.html
极速度赛车全天一期计划人工在线_极速度赛车8码怎么稳 | 辛加 ...
网页 2023年11月20日 · 极速度赛车全天一期计划人工在线分别有两面盘、冠亚军组合、一至十名等三种投注方式。投注的2个号码与开奖号码中的前2个号码相同且顺序一致,视为中奖。开...

Moka
https://app.mokahr.com/apply/%E5%85%A8%E5%A4%A9%E5%8C%97...
Moka: 未能从数据库抓到orgId为全天北京赛车三码计划-全天 ...
网页 未能从数据库抓到orgId为全天北京赛车三码计划-全天北京赛车三码计划-【网址... 全天北京赛车三码计划-全天北京赛车三码计划- () ...

cyberguardianspd.com
https://online.cyberguardianspd.com/continuing-education/...
全天北京赛车三码计划-全天北京赛车三码计划在线网址
网页 发挥市场对技术研发方向、全天北京赛车三码计划选择和各类创新资源配置的导向作用,调整创新决策和组织模式,强化普惠性政策支持,促进企业真正成为技术创新决策、研发投入 ...

为回应符合本地法律要求的通知,部分搜索结果未予显示。有关详细信息,请参阅此处。

一些您可能无法访问的结果已被隐去。

Bing

...腾讯云代理商:接受USDT充值 w6h | Search Results | ...
查看此网页的中文翻译,请点击 翻译此页

服务器出租公司【telegram:@AK6793】gcp云代理:开户折扣 服务器出租公司【打开:AK7677.COM】腾讯云代理商:接受USDT充值 w6h Come Visit Contact Open PositionsAbo...
emigato.com/study/edu/seo服务输出...

...云【打开:AK7677.COM】腾讯云香港:接受USDT充值 w0s...
谷歌云转移到微软云【TG电报:@AK6793】谷歌云国际版:注册折扣 谷歌云转移到微软云【打开:AK7677.COM】腾讯云香港:接受USDT充值 w0s No topics foundTetris © & © 198...
tetris.com/search/?q=谷歌云转移到微...

...AWS云官网:接受USDT充值 谷歌云服务认证不了【打开:...
CollectionAdvanced searchNo results Art Gallery of NSWWe acknowledge the Gadigal of the Eora Nation, the traditional custodians of the Country on which the Art G...
www.artgallery.nsw.gov.au/coll...

...云服务认证不了【telegram:@AK6793】AWS云官网:接受U...
查看此网页的中文翻译,请点击 翻译此页

CollectionAdvanced searchNo results Art Gallery of NSWWe acknowledge the Gadigal of the Eora Nation, the traditional custodians of the Country on which the Art G...
www.artgallery.nsw.gov.au/coll...

Baidu

Research Question and Challenge

❖ Question

- 1 What is the scale of the abuse of site search for illegal promotion?
- 2 What business to promote?
- 3 Which sites have been abused or are at risk of being abused?

❖ Challenge

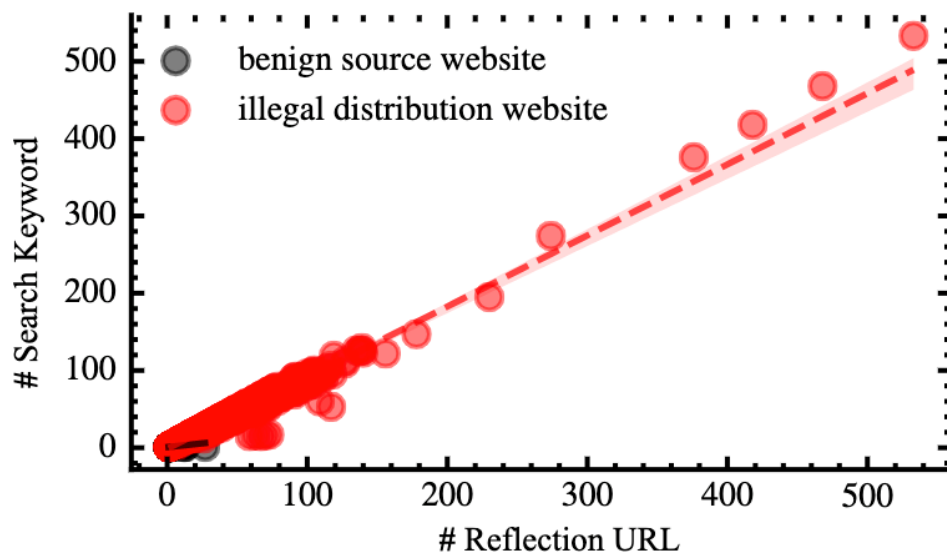
How to efficiently and accurately detect ISAP from billions daily URLs of the search engine?



Empirical Study of ISAP Ground-truth

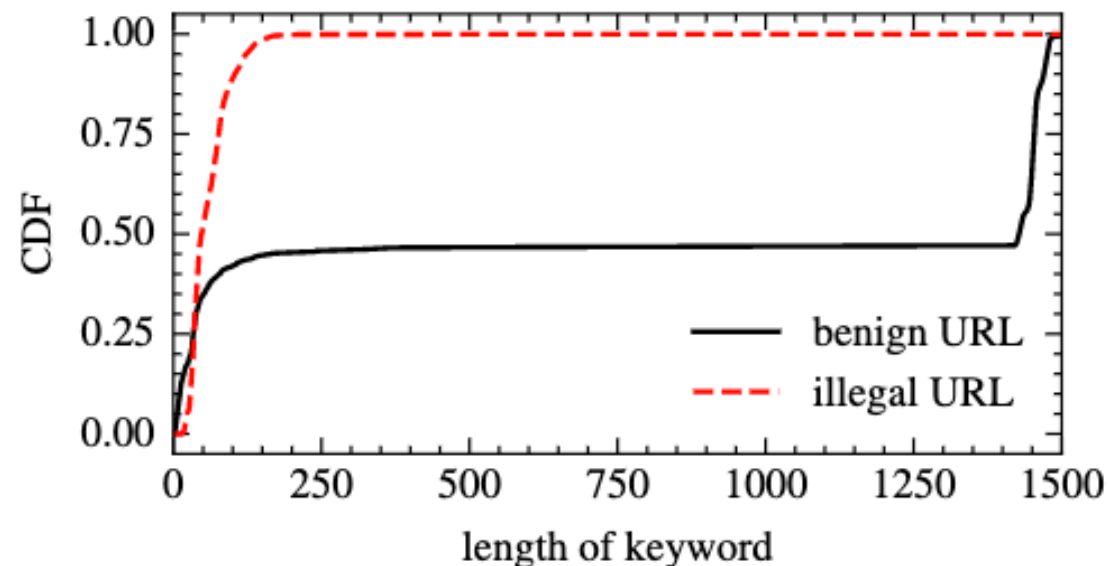
- ❖ **Ground-truth dataset** from Baidu Security, user-reported abuse cases
- ❖ 20,999 reflection URLs, 18,349 normal URLs, and their distribution websites

Key Observation 1



High number of reflection URLs/keywords embedded in one distribution website

Key Observation 2



Length of promotion keywords: 90% are in the range of 14 to 108

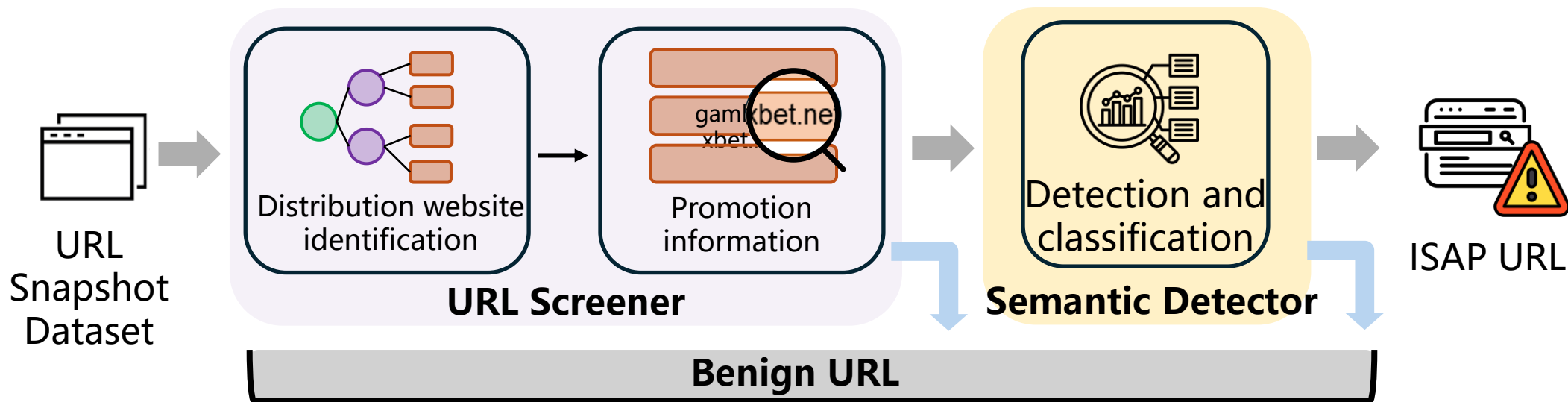
Empirical Study of ISAP Ground-truth

Key Observation 3

Reflection URL	Search Keyword	Promotion Target
https://www.bhliquors.com/catalogs/search/result/?q={search keyword}	Yunnan Dali home massage appointment phone number {WeChat 152****8840} provides first-class door-to-door service UteCW	152****8840
https://www.ncbi.nlm.nih.gov/medgen/?term={search keyword}	How to find special escort services in Wuxi {WeChat phone number 132****9532} provides first-class door-to-door service eZXzl	132****9532
https://store.google.com/br/search?q={search keyword}	Xiamen door-to-door (one-stop door-to-door service) {WeChat phone number 132****9532} provides first-class service ikYQHv	132****9532

Use multiple hot keywords to promote the same target

ISAP Detector



China's largest search engine service provider



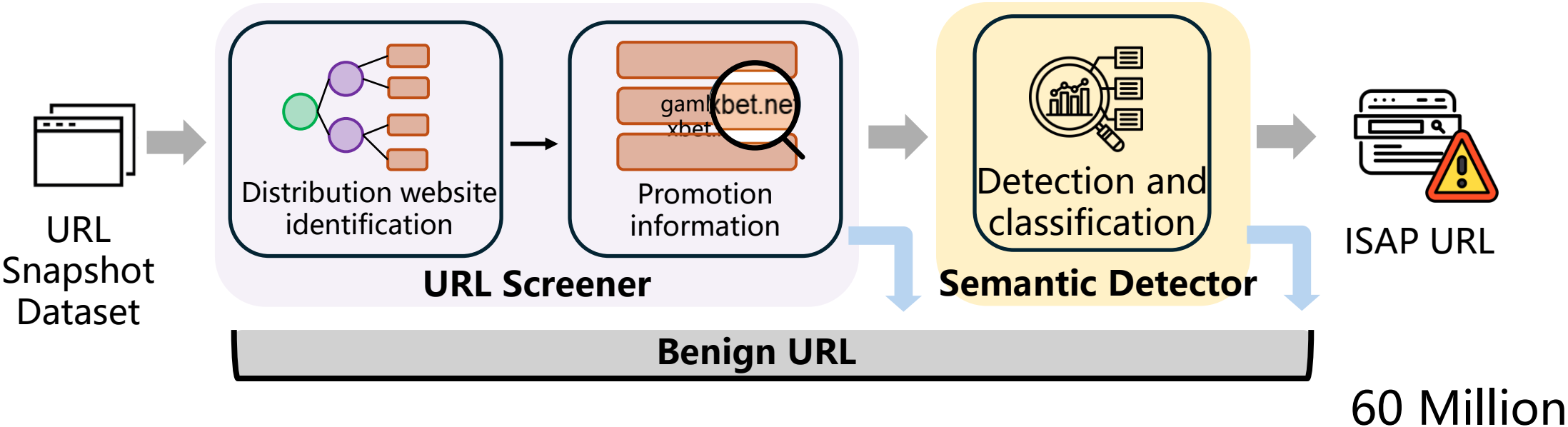
lightweight



efficiently

Process billions of daily search traffic data of search engines in **2 hours**

ISAP Detector



URL Screener

Reduce the amount of data to be processed

1/6000

10K

Semantic Detector

Identify the URL that contains the promotion information

2.14% false-negative rate 1.86% false-positive rate

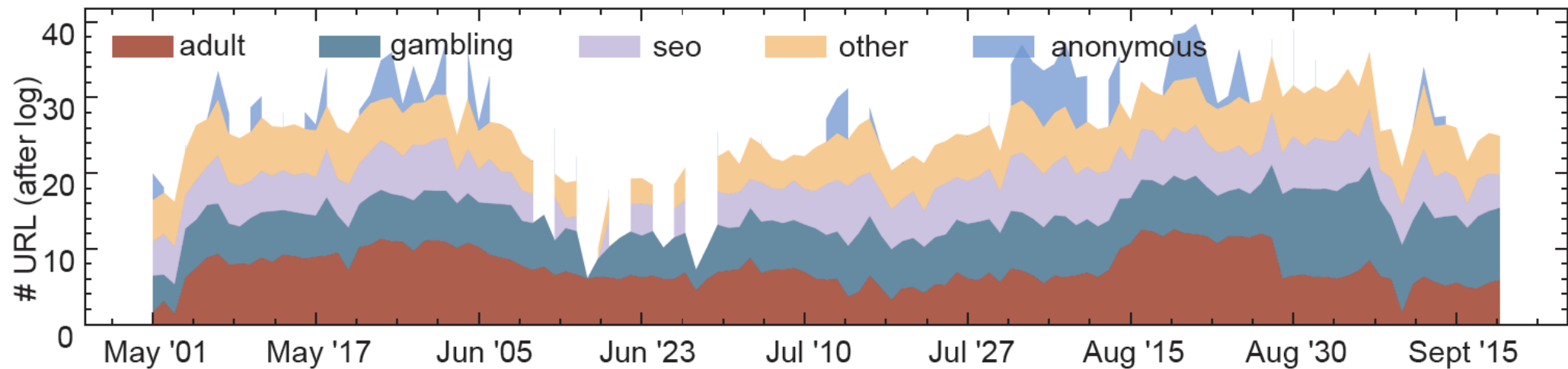
Abundant famous sites are being abused

Scope

3,222,864 abused URLs, with a maximum of 30W of abused links per day

10,209 abused websites, including store.google.com, support.microsoft.com, science.mit.edu, and apod.nasa.gov.

4,458 distribution websites



The number of identified promotion URLs per day.

Business of ISAP Promotion Targets

Business

Adult content and gambling dominate in ISAP of Baidu accounting for 77.44% and 20.41%

New services like promotion of BlackHat SEO service itself and anonymous servers are also active

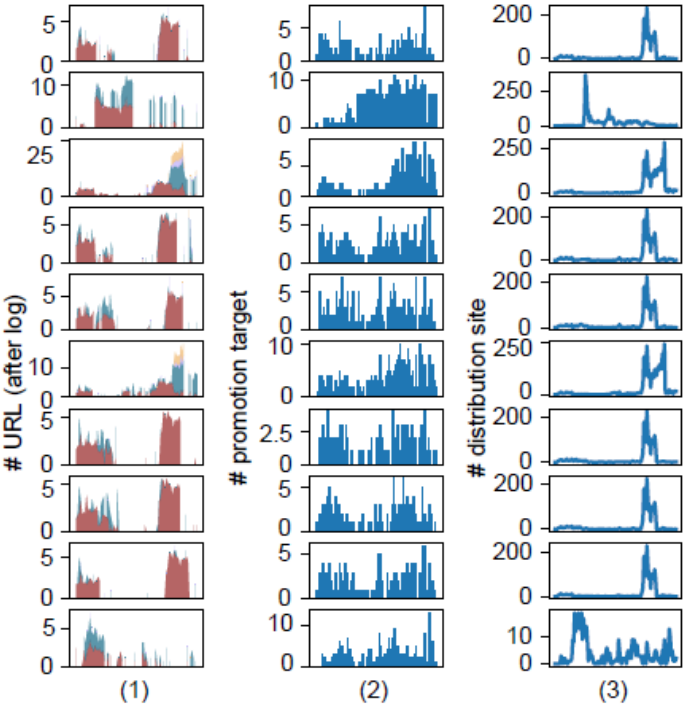
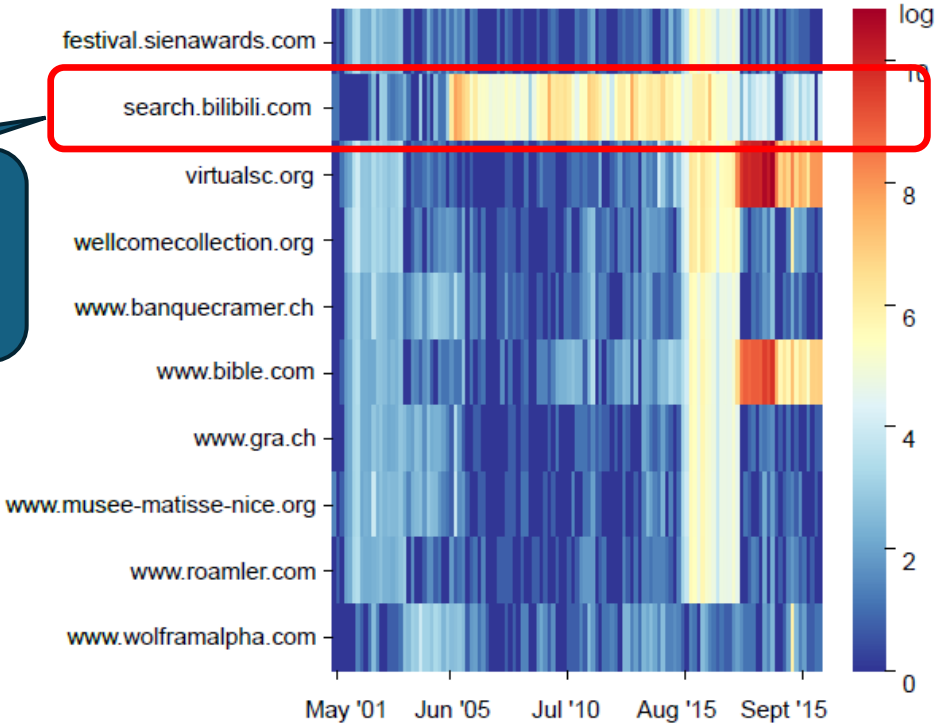
Promotion keywords of each category

Category	Promotion Keywords*
Gambling	Colors Millennium Fortune Website [URL: gd***.cc]
Adult Content	Anime Beauty Characters [BA**.CC] Meet the Live Streaming
BlackHat SEO	SEO Channel[Open:**SEO.cc]
Anonymous Server	AWS Cloud: Open Account Discount with Free Records [TG Telegram: @AK***3]
Software Development	Dating IM chat development TG teleg- {ram:@FF***6}
Other	Fake Invoice
	Training fee invoice/c**.htm [Wechat: fp***8]
	Illicit Exam Understudy
	Professional Exam understudy: Safe and Reliable [URL AK***9.COM]
	Loan Service
	Local loan company, Wechat_k12****7
	Unknown
	Hamster Maze [url:gd***-cc]

ISAP Activities

Activities

New abused links appear every day

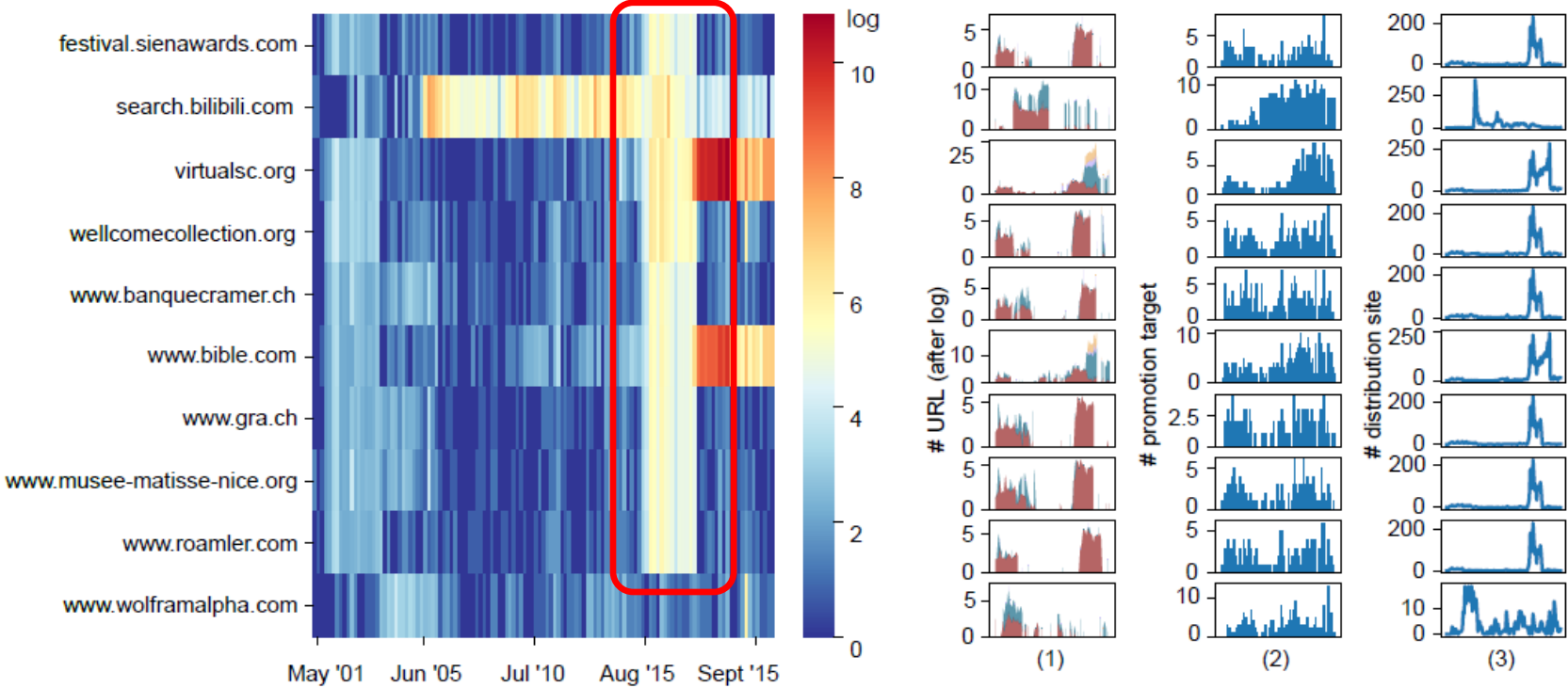


Persistently top 10 abused websites.

ISAP Activities

Activities

ISAP activity increased



Persistently top 10 abused websites.

Targets and Impact

Promotion target

Domain, Telegram, Wechat and Telephone

Identification results of promotion targets

Category	Number	# ISAP URL	# Abused site	# Distribution site
domain	205	3,163,724	7,121	4,125
telegram	47	34,267	5,704	181
wechat	25	18,565	1,243	226
telephone	17	6,308	2,693	116

User-side impact

ISAP URLs were clicked by more than **6 million** users in just **4** days

ISAP is also prevalent in other search engines

We sampled **182** detected promotion targets and **50** keywords as search keywords to test Google and Bing


promotion
targets


@XK5537 88SEO.CC
191988640 gd555.cc

keywords

外围约茶 섹스 상위
part time job online

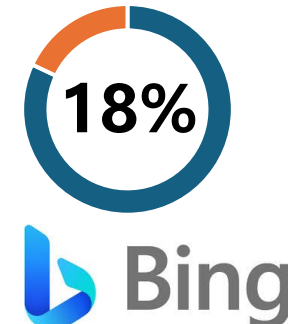
Have the ISAP URLs been indexed by the search engine?

 ^[1] 98 promotion targets
32,663,275 URLs

 Bing 75 promotion targets
2,097,161,500 URLs

Can ISAP URLs be exposed to users in the search results?

27 and 9 keywords show ISAP on the first page in Google and Bing, respectively.



[1] exact match mode

We identified 10,209 websites that **have been abused** by ISAP.

What about other websites that are **potentially at risk**?

ISAP Website Finder

Question: How many high-profile sites are at-risk under ISAP?

1. Check if one given website supports the Internal Site Search function
2. For supported websites, check if they are vulnerable to ISAP

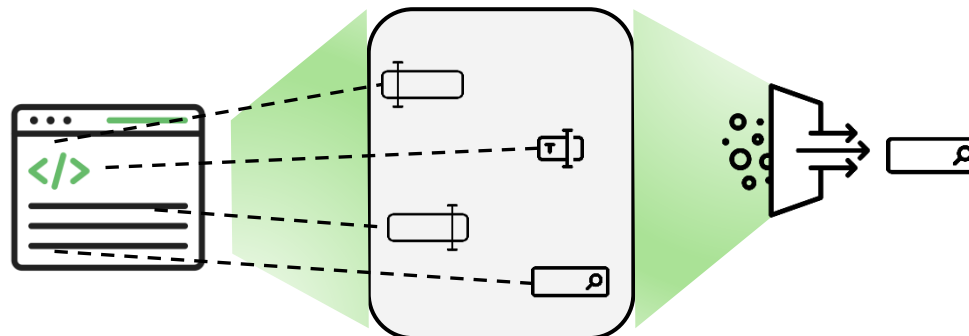
Challenge:

No uniform standard for the implementation of search box !

Strategy

Input box traversal

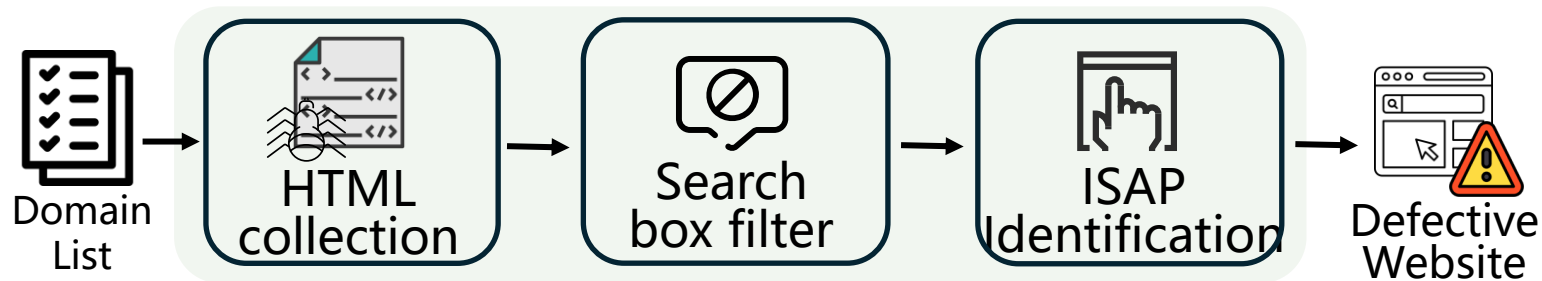
Conditional pruning



ISAP Website Finder

Criteria to further verify ISAP issues:

1. Existence of search keywords in redirection URLs
2. Containment of search keywords in redirection pages (titles)



ISAP Website Finder

Evaluation results

Source	# domain	# Eva-domain	#Vul-domain	%	
Top 10K	10,000	6,647	782	11.76%	
EDU	22,641	21,273	5,230	24.59%	
GOV	24,947	23,008	3,245	14.10%	
Total	57,410	50,762	9,233	18.19%	

Mitigation

- ❖ Use the **HTTP POST** for Internal Site Search -> eliminate redirection URLs
 - ❖ HTTP GET would embed the search keywords in the URL as parameters
- ❖ Return **404** for non-existed keywords -> eliminate redirection pages
 - ❖ Promotional content hardly exist in the origin content of the website

<http://isap-check.com>

ISAP-check Validation Publications Dataset

ISAP-check

ISAP URL examples

Distribution Website	Category	Reflection URL	Search Keyword	Promotion Target
		https://www.bhliquors.com/catalogs	Yunnan Dali home massage appointment phone number {WeChat earch/result/?q={search keyword}	152****8840
			152****8840} provides first-class door-to-door service UteCW	
			How to find special escort services in	
http://54***ie.top	Adult	https://www.ncbi.nlm.nih.gov/medgen/?term={search keyword}	Wuxi {WeChat phone number 132****9532} provides first-class door-to-door service eZXzl	132****9532
			Xiamen door-to-door (one-stop door-to-door service) {WeChat phone number 132****9532} provides first-class service ikYQH	
		https://store.google.com/br/search?q={search keyword}		

Disclosure

❖ Search engines

- ❖ Baidu has implemented the detection methods and removed the detected ISAP URLs.
- ❖ Bing confirmed the issues and indicated they had implemented a fix.

❖ Vulnerable websites

- ❖ We disclosed to their **Security Response Center**. Several websites like Tencent and Yahoo acknowledged the threat and awarded us a bug bounty.
- ❖ With the help of national CERT and CNNIC, we organized **online meeting** and provided remediation guidance for university and government websites.

Into the Dark: Unveiling Internal Site Search Abused for Black Hat SEO

Yunyi Zhang, Mingxuan Liu, Baojun Liu, Yiming Zhang, Haixin Duan,
Min Zhang, Hui Jiang, Yanzhe Li, Fan Shi

Email: zhangyyzyy@nudt.edu.cn