(Rayburn, 2021)

Automating Reconnaissance for More Efficient Penetration Testing

Table of Contents

- Why APIs?
- Vulnerability Detection
- Research Question
- Literature Review

Table of Contents

- Why APIs?
- Vulnerability Detection
- Research Question
- Literature Review

- Ethical Considerations
- Research Proposal
- Timeline
- Conclusion



















(Meta, 2023)





























Challenges to API Security



Table 1: Common API Vulnerabilities (Badhwar, 2021; Ball, 2022; Díaz-Rojas et al., 2021)

Common API Vulnerabilities	
Information Disclosure	Lack of Resources/Rate Limiting
Broken Object Level Authorisation (BOLA)	Broken Function Level Authorisation (BFLA)
Broken User Authentication	Mass Assignment
Excessive Data Exposure	Security Misconfigurations
Injections	Improper Assets Management
Business Logic Vulnerabilities	Insufficient Logging and Monitoring

• API vs Web App security (Ball, 2022; Siriwardena, 2020)

- API vs Web App security (Ball, 2022; Siriwardena, 2020)
 - Depth and scale
 - Endpoint exposure (Begum et al., 2018)
 - Micro-services (Irfan et al., 2023)
 - Data Sharing (Gu & Mendoza, 2018)
 - Cloud computing (Ariffin et al., 2020)

- API vs Web App security (Ball, 2022; Siriwardena, 2020)
 - Depth and scale
 - Endpoint exposure (Begum et al., 2018)
 - Micro-services (Irfan et al., 2023)
 - Data Sharing (Gu & Mendoza, 2018)
 - Cloud computing (Ariffin et al., 2020)

- Uncommon vulnerabilities
 - Traditional IDS ineffective (Ball, 2022)
 - Design-specific attack surface (Munsch & Munsch, 2021)
 - Customized architecture and zero-day logic flaws (Ball, 2022)

- Reconnaissance
 - Essential for vulnerability test reliability (Ball, 2022; Perumal et al., 2021)

- Reconnaissance
 - Essential for vulnerability test reliability (Ball, 2022; Perumal et al., 2021)
 - Time intensive (Akshay et al., 2022)
 - Manual search
 - Manual compilation

- Reconnaissance
 - Essential for vulnerability test reliability (Ball, 2022; Perumal et al., 2021)
 - Time intensive (Akshay et al., 2022)
 - Manual search
 - Manual compilation
 - Large breadth and depth (Akshay et al., 2022; Feng et al., 2020)
 - Multiple sources
 - Multiple data types

• API-Specific

- Fuzzing (Bogle et al., 2022)
- IoT (Dong et al., 2021)
- Error scenarios (Ceccato et al, 2021)

• API-Specific

- Fuzzing (Bogle et al., 2022)
- IoT (Dong et al., 2021)
- Error scenarios (Ceccato et al, 2021)

- Web App-Specific
 - Reconnaissance (Akshay et al., 2022)
 - Detection (Huang et al., 2021)
 - Exploitation (Ntantogian et al., 2019)

- API-Specific
 - Fuzzing (Bogle et al., 2022)
 - IoT (Dong et al., 2021)
 - Error scenarios (Ceccato et al, 2021)

- Web App-Specific
 - Reconnaissance (Akshay et al., 2022)
 - Detection (Huang et al., 2021)
 - Exploitation (Ntantogian et al., 2019)

- API-Specific
 - Fuzzing (Bogle et al., 2022)
 - IoT (Dong et al., 2021)
 - Error scenarios (Ceccato et al, 2021)

- Web App-Specific
 - Reconnaissance (Akshay et al., 2022)
 - Detection (Huang et al., 2021)
 - Exploitation (Ntantogian et al., 2019)

- Akshay et al.'s study (2022) in automated reconnaissance for web applications
 - Comprehensive recon

- Akshay et al.'s study (2022) in automated reconnaissance for web applications
 - Comprehensive recon
 - Web crawling \rightarrow subdomains, credentials
 - Multi-threading

- Akshay et al.'s study (2022) in automated reconnaissance for web applications
 - Comprehensive recon
 - Web crawling \rightarrow subdomains, credentials
 - Multi-threading
 - Malleable data output

- Akshay et al.'s study (2022) in automated reconnaissance for web applications
 - Comprehensive recon
 - Web crawling \rightarrow subdomains, credentials
 - Multi-threading
 - Malleable data output

API-specific

Research Question

Hypothesis:

Web crawling can automate key aspects of API reconnaissance to improve labour and time requirements for API vulnerability testing

• Aims

Objectives

32

- Aims
 - Robust automated reconnaissance

• Objectives

- Aims
 - Robust automated reconnaissance

- Objectives
 - Identify aspects most adept for automation

- Aims
 - Robust automated reconnaissance
 - Essential information
 - Target API

- Objectives
 - Identify aspects most adept for automation

- Aims
 - Robust automated reconnaissance
 - Essential information
 - Target API

Objectives

- Identify aspects most adept for automation
- Program algorithms
 - Identify, crawl, categorise
Research Question – Aims and Objectives

• Aims

- Robust automated reconnaissance
- Essential information
 - Target API
- Data output
 - Further investigation and manipulation

Objectives

- Identify aspects most adept for automation
- Program algorithms
 - Identify, crawl, categorise
- Program algorithms
 - Data sorting and presentation
 - .pdf & .csv

- API Reconnaissance
 - Endpoints
 - Gateway to API attack surface (Paxton-Fear, 2022)

- API Reconnaissance
 - Endpoints
 - Gateway to API attack surface (Paxton-Fear, 2022)
 - Two stages (Ball, 2022)

- API Reconnaissance
 - Endpoints
 - Gateway to API attack surface (Paxton-Fear, 2022)
 - Two stages (Ball, 2022)
 - Passive Recon
 - API endpoints
 - Credentials
 - Documentation

- API Reconnaissance
 - Endpoints
 - Gateway to API attack surface (Paxton-Fear, 2022)
 - Two stages (Ball, 2022)
 - Passive Recon
 - API endpoints
 - Credentials
 - Documentation

- Active Recon
 - Scanning and pinging
 - HTTP requests
 - API calls

- API Reconnaissance
 - Endpoints
 - Gateway to API attack surface (Paxton-Fear, 2022)
 - Two stages (Ball, 2022)
 - Passive Recon
 - API endpoints
 - Credentials
 - Documentation

- Active Recon
 - Scanning and pinging
 - HTTP requests
 - API calls

Passive API Reconnaissance

(Ball, 2022; Bhavsar & Chudasama, 2021; Paxton-Fear, 2022)

Phase One

- Breadth and depth
 - OWASP Amass
 - Google 'dorking'
 - Shodan
 - ProgrammableWeb

Phase Two

- Info consolidation
- Sensitive information
 - Github
 - Exploit DB (Offsec, 2023)
 - HackerOne (2023)
 - PasteHunter

Phase Three

- Documentation
 - Text document
 - Screenshots
- Task list
 - Active scanning
 - Exploitation

Passive API Reconnaissance

(Ball, 2022; Bhavsar & Chudasama, 2021; Paxton-Fear, 2022)

Phase One

- Breadth and depth
 - OWASP Amass
 - Google 'dorking'
 - Shodan
 - ProgrammableWeb

Phase Two

- Info consolidation
- Sensitive information
 - Github
 - Exploit DB (Offsec, 2023)
 - HackerOne (2023)
 - PasteHunter

Phase Three

- Documentation
 - Text document
 - Screenshots
- Task list
 - Active scanning
 - Exploitation







Web scraping



Web scraping



Web crawling



• Web scraping

- Web scraping
 - Uses HTTP protocol for data extraction (Khder, 2021)
 - Particularly relevant for API data (Ball, 2022)

Web scraping

- Uses HTTP protocol for data extraction (Khder, 2021)
 - Particularly relevant for API data (Ball, 2022)
- Can be keyword-specific (Hossain et al., 2020)
 - Targets specific topics throughout the internet

Web scraping

- Uses HTTP protocol for data extraction (Khder, 2021)
 - Particularly relevant for API data (Ball, 2022)
- Can be keyword-specific (Hossain et al., 2020)
 - Targets specific topics throughout the internet
- Varied sampling methods (Gupta et al., 2018)
 - Vertical or horizontal sampling

• Web crawling

- Web crawling
 - An extension of web scraping (Biswas & Nigam, 2021)
 - Link hopping or robot.txt

- Web crawling
 - An extension of web scraping (Biswas & Nigam, 2021)
 - Link hopping or robot.txt
 - Surface and dark web (Amale et al., 2021)
 - The TOR channel and crawling depth

- Web crawling
 - An extension of web scraping (Biswas & Nigam, 2021)
 - Link hopping or robot.txt
 - Surface and dark web (Amale et al., 2021)
 - The TOR channel and crawling depth
 - General and focused crawlers (Arun et al., 2022)
 - Depends on needs of program

- Web crawling
 - Adapted from Hossain et al., 2020



- Web crawling
 - Adapted from Hossain et al., 2020
 - Focused crawler
 - Decision tree model



- Web crawling
 - Adapted from Hossain et al., 2020
 - Focused crawler
 - Decision tree model
 - Vertical sampling (Arauza et al., 2021)



- Web crawling
 - Adapted from Hossain et al., 2020
 - Focused crawler
 - Decision tree model
 - Vertical sampling (Arauza et al., 2021)
 - Fitness value (Navaneethan & Rajiv, 2021)
 - Naive Bayes classification



- Web crawling
 - Adapted from Hossain et al., 2020
 - Focused crawler
 - Decision tree model
 - Vertical sampling (Arauza et al., 2021)
 - Fitness value (Navaneethan & Rajiv, 2021)
 - Naive Bayes classification
 - Validity (Bouchard et al., 2017)
 - p-Value Statistics



be puthon the terms of terms o

• Python programming language

- Suitable for web scraping (Khder, 2021)
 - Popular and simple syntax

Python programming language

- Suitable for web scraping (Khder, 2021)
 - Popular and simple syntax
- Suitable for ethical hacking (Arnold & Seitz, 2021)
 - Object oriented programming
 - Extensive libraries

Python programming language

- Suitable for web scraping (Khder, 2021)
 - Popular and simple syntax
- Suitable for ethical hacking (Arnold & Seitz, 2021)
 - Object oriented programming
 - Extensive libraries
- Suitable for Automation (Shamunesh et al., 2023; (Chandukiran et al., 2023)

- Ease of malicious reconnaissance
 - Attack surface (Ball, 2022)
 - Unwilling target (Li, 2022)

- Ease of malicious reconnaissance
 - Attack surface (Ball, 2022)
 - Unwilling target (Li, 2022)
- Web Crawling
 - robot.txt (Biswas & Nigam, 2021)
 - Server operation (Arauza et al., 2021)

- Ease of malicious reconnaissance
 - Attack surface (Ball, 2022)
 - Unwilling target (Li, 2022)
- Web Crawling
 - robot.txt (Biswas & Nigam, 2021)
 - Server operation (Arauza et al., 2021)
- GDPR (2018)
 - Privacy and data



- Ease of malicious reconnaissance
 - Attack surface (Ball, 2022)
 - Unwilling target (Li, 2022)
- Web Crawling
 - robot.txt (Biswas & Nigam, 2021)
 - Server operation (Arauza et al., 2021)
- GDPR (2018)
 - Privacy and data



Research Proposal

- Automate API reconnaissance for vulnerability testing
 - Passive reconnaissance
- Automate API reconnaissance for vulnerability testing
 - Passive reconnaissance
- Python framework:

- Automate API reconnaissance for vulnerability testing
 - Passive reconnaissance
- Python framework:
 - Third party architecture
 - OWASP Amass

- Automate API reconnaissance for vulnerability testing
 - Passive reconnaissance
- Python framework:
 - Third party architecture
 - OWASP Amass
 - Original modules
 - Web crawling \rightarrow endpoints, credentials, documentation

- Automate API reconnaissance for vulnerability testing
 - Passive reconnaissance
- Python framework:
 - Third party architecture
 - OWASP Amass
 - Original modules
 - Web crawling \rightarrow endpoints, credentials, documentation
- Saved output
 - .pdf and .csv











- Ease of malicious reconnaissance
 - Only willing web applications

- Ease of malicious reconnaissance
 - Only willing web applications
- Web Crawling
 - Limited scrape rate
 - Limited crawl rate

- Ease of malicious reconnaissance
 - Only willing web applications
- Web Crawling
 - Limited scrape rate
 - Limited crawl rate
- GDPR (2018)
 - Privacy and data parameters

- Ease of malicious reconnaissance
 - Only willing web applications
- Web Crawling
 - Limited scrape rate
 - Limited crawl rate
- GDPR (2018)
 - Privacy and data parameters



Dissertation Timeline (University of Essex, 2023)

Computing Department - MSc Project Roadmap

Modules	Research Methods and Professional Practice Module				Project Module						
Deadlines		Week 1	Week2	Week 3	Week 4	Weeks 5 to 7	Week 8	Weeks 9 to 28	Weeks 29	Week 30	
MILESTONES	Submission of Re			earch Outline		ubmission of Research Proposal and Ethics		Proposal and Ethics Disser	Online Conference Dissertation submission		
Preparation	Project Idea/potential supervisor search Research Project of	utline developm	ent								
Research and Development			1	F	Research Propo	osal Development		l de la companya de l			
								Literature Survey Project Development and Dissertation Write Up			
Defence								Artefact presentati	on - development & de	livery	



• Importance of APIs



- Importance of APIs
- Importance of reconnaissance



- Importance of APIs
- Importance of reconnaissance
- Web crawling utilisation



- Importance of APIs
- Importance of reconnaissance
- Web crawling utilisation
- Ethical concerns



- Importance of APIs
- Importance of reconnaissance
- Web crawling utilisation
- Ethical concerns
- Research question and proposal



- Importance of APIs
- Importance of reconnaissance
- Web crawling utilisation
- Ethical concerns
- Research question and proposal
- Timeline





Ahmed, R. et al. (2022) Machine Learning and Deep Learning Approaches for CyberSecurity: A Review. *IEEE Access*, 10: 19572 – 19585

Akshay S et al. (2022) Automation of Recon Rocess for Ethical Hackers. In: 2022 International Conference for Advancement in Technology, Goa, India, 21 – 22 January, 2022. IEEE: 1 – 6

Amale et al. (2021) SpyDark: Surface and Dark Web Crawler. 2021 Second International conference on Secure Cyber Computer and Communication. IEEE: 45 - 49

Arauza, O., Brewer, R., Hart, T., & Westlake, B. (2021) The Ethics of Web Crawling and Web Scraping in Cybercrime Research: Navigating Issues of Consent, Privacy, and Other Potential Harms Associated with Automated Data Collection. In: Holt, T. J. & Lavorgna, A. (eds.) *Researching Cybercrimes*. Switzerland. Springer: 435 - 456

References

- Ariffin, M. A. M., Ibrahim, M. F., & Kasiran, Z. (2020) API Vulnerabilities in Cloud Computing Platform: Attack and Detection. *International Journal of Engineering Trends and Technology*: 8 – 14
- Arnold, T & Seitz, J. (2021) Blackhat Python: Python Programming for Hackers and Pentesters. 2nd Ed. San Francisco, USA: No Starch Press.
- Arun, A. et al. (2022) An Automated Word Embedding with Parameter Tunde Model for Web Crawling. *Intelligent Automation & Soft Computing*, 32 (3): 1617 - 1632
- Badhwar, R. (2021) Intro to API Security Issues and Some Solutions! In: *The CISO's Next Frontier*. Cham, CH. Springer: 239 244



- Ball, C. J. (2022) Hacking APIs: Breaking Web Application Programming Interfaces. San Francisco, CA, USA. No Starch Press.
- Begum, A., Bhuiyan, T., Hadid, I., & Rahman, S. (2018) API Vulnerabilities: Current Status and Dependencies. *International Journal of Engineering & Technology*, 7: 9 -13
- Biswas, P. & Nigam, H. (2021) From Web Scraping to Web Crawling. In: A. Choudhary et al. (eds.) *Applications of Artificial Intelligence and Machine Learning*. Singapore. Springer: 97 - 112
- Bhavsar, R. & Chudasama, D. (2021) Technical Methods of Information Gathering. *Journal of Web Engineering & Technology*, 8 (3): 1 - 5

References

Bogle, A., Mahmood, R, Pennington, J., Tran, T., & Tsang, D. (2022) A Framework for Automated API Fuzzing at Enterprise Scale. In: *IEEE Conference on Software Testing, Verification and Validation*. IEEE: 377 – 388

Bouchard, M., Frank, R., & Westlake, B. (2017) Assessing the Validity of Automated Webcrawlers as Data Collection Tools to Investigate Online Child Sexual Exploitation. *Sexual Abuse*, 29 (7): 685 - 708

Ceccato, M. et al. (2022) Automated Black-box Testing of Nominal and Error Scenarios in RESTFUL APIs. *Software Testing, Verification and Reliability*, 32 (5): 433 – 442

Díaz-Rojas, J. A., Limón, X., Ocharán-Hernández, J. O., & Pérez-Arriaga, J. C. (2021) Web API Security vulnerabilities and Mitigation Mechanisms: A Systematic Mapping Study. In: *9th International Conference in Software Engineering Research and Innovation*. IEEE: 207 – 218



- Dong, L et al. (2021) IoT-APIScanner: Detecting API Unauthorized Access Vulnerabilities of IoT Platform. *IEEE Internet of Thing Journal*, 8 (13): 10327 - 10335
- Feng, H., Fu, X., Sun, H., Wang, H., Zhang, Y. (2020) Efficient Vulnerability Detection Based on Abstract Syntax Tree and Deep Learning. In: *IEEE Conference on Computer Communications Workshops, Toronto, ON, CA*. IEEE: 722 - 727
- GDPR (2018) General Data Protection Regulation (GDPR). General Data Protection Regulation (GDPR). [Available Online]: https://gdpr-info.eu/
- Gu, G. & Mendoza, A. (2018) Mobile Application Web API Reconnaissance: Web-to-Mobile Inconsistencies & Vulnerabilities. In: *2018 IEEE Symposium on Security and Privacy*. IEEE: 756 – 769



- Gupta, A., Singh, K. B., & Singh, R. K. (2018) Web Crawling Techniques and It's Implications. *Globus: An International Journal of Management & IT*, 9 (2): 1 7
- HackerOne (2023) *Hacktivity* | *HackerOne.* hackerone.com [Available Online] https://hackerone.com/hacktivity/overview
- Hossain, S. A., Nobel, N. I., Rahman, A. K. M. S., Shamrat, F. M. J. M., Tasnim, Z. (2020) An Effective Implementation of Web Crawling Technology to Retrieve Data from the World Wide Web (WWW). *International Journal of Scientific & Technological Research*, 9 (1): 1252 1256
- Huang, G., Li, J., Ren, J., Zhang, B. (2021) Efficiency and Effectiveness of Web Application Vulnerability Detection Approaches: A Review. *ACM Computer Survey*, 54 (9): 1 – 35



Irfan, MD et al. (2023) API Traffic Anomaly Detection in Microservice Architecture. In: *IEEE/ACM 23rd Symposium on Cluster, Cloud, and Internet Computing Workshops*. IEEE: 206 – 213

Khder, M. A. (2021) Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application. *International Journal of Advances in Soft Computing & Application*, 13 (3): 144 - 168

Li, V. (2021) Bug Bounty Bootcamp: The Guide to Finding and Reporting Web Vulnerabilities. San Francisco, USA: No Starch Press.

Munsch, A. & Munsch, P. (2021) The Future of API (Application Programming Interface) Security: The Adoption of APIs for Digital Communications and the Implications of Cyber Security Vulnerabilities. *Journal of International Technology and Information Management*, 29 (3): 25 – 45

References

- Navaneethan, C. & Rajiv, S. (2021) Keyword Weight Optimization Using Gradient Strategies in Event Focused Web Crawling. *Patter Recognition Letters*, 142: 3 - 10
- Ntantogian, C., Stasinopoulos, A., & Xenakis, C. (2018) Commix: Automating Evaluation and Exploitation of Command Injection Vulnerabilities in Web Applications. *International Journal of Information Security*, 18: 49 – 72
- OFFSEC (2023) *OFFSEC's Exploit Database Archive* | *Exploit Database*. exploitdb.com. [Available online] https://www.exploit-db/com/
- Paxton-Fear, K. (2022) *API Hacking Toolbox w/ Dr. Katie Paxton-Fear* | *Traceable AI*. youtube.com. [Available online] https://www.youtube.com/watch? v=qC8NQFwVOR0



Permual, A., et al. (2021) Cybercrime Issues in Smart Cities, Networks and Prevention Using Ethical Hacking. In: C. Chakraborty et al. (eds.) *Data-Driven Mining, Learning and Analytics for Secured Smart Cities*. Switzerland. Springer: 333 - 358

Siriwardena, P. (2020) Advanced API Security: OAuth 2.0 and Beyond. 2nd Ed. New York, NY, USA. Apress.

Shamunesh P, Srinivas, L. N. B., Vinoth S (2023) Cybercheck – OSINT & Web Vulnerability Scanner. In: *Second International Conference on Edge Computing and Applications*. IEEE: 275 – 279

University of Essex (2023) Computing Department – MSc Project Roadmap. [Available Online]: https://www.my-course.co.uk/course/view.php?id=10163



- Castro, A. (2020) *Amazon Logo* | *The Verge.* theverge.com [Available Online] https://www.theverge.com/2020/7/30/21348368/amazon-q2-2020-earnings-covid-19coronavirus-jeff-bezos
- CMM (n.d.) *Under Lock and Key*. Care Management Matters [Available Online] https://www.caremanagementmatters.co.uk/feature/under-lock-and-key-making-thenhs-and-social-care-cyber-safe/
- Green Imaging (n.d.) *Healthcare* | *Green Imaging.* greenimaging.net [Available Online] https://greenimaging.net/what-is-going-on-with-healthcare/
- IBM (2023) *IBM Logo* | *Forbes*. forbes.com [Available Online] https://www.forbes.com/sites/moorinsights/2023/07/07/ibm-watsonx-empowersbusinesses-to-build-tune-and-deploy-reliable-generative-ai-models/?sh=35383a022fda



- Kazmierski (2019) *UN Flags* | *Shutterstock.* Ranking Digital Rights [Available Online] https://rankingdigitalrights.org/2019/09/26/what-should-governments-do/
- LogicRays (2020) Python Logo | LogicRays Academy Blog. logicraysacademy.com [Available Online] https://www.logicraysacademy.com/blog/what-is-pythonprogramming-language/
- Meta (2023) *facebook Logo. facebook.com* [Available Online] https://th-th.facebook.com/
- Netflix (2023) *Netflix Logo*. netflix.com [Available Online] https://about.netflix.com/en/news/announcing-basic-with-ads-us





- Newsom, N. (2010) A Red Graph on the Rise Over Stacks of Gold Coins. Alamy [Available Online] https://www.alamy.com/stock-photo-a-red-graph-on-the-riseover-stacks-of-gold-coins-35005260.html?imageid=2669D11F-627A-4CE6-A542-2191DE571843&p=136117&pn=1&searchId=b63614e611cfdf67bf8919cedd2b1f5 4&searchtype=0
- Planview (2023) Salesforce Logo. planview.com [Available Online] https://www.planview.com/products-solutions/products/hub/integrations/ salesforce/
- Rayburn, D (2021) API | Streaming Media Blog. Streamingmediablog.com. [Available Online] https://www.streamingmediablog.com/2021/09/apistreaming.html



- ScrapingBot (2023) *How to Build a Web Crawler*. scraping-bot.io [Available Online] https://www.scraping-bot.io/how-to-build-a-web-crawler/
- Sofi (2022) *Commercial Banking* | *SoFi Learn.* sofi.com [Available Online] https://www.sofi.com/learn/content/what-is-commercial-banking/
- StickPNG (n.d.) *Download Malicious Hacker Transparent PNG*. stickpng.com [Available Online] https://www.stickpng.com/img/icons-logos-emojis/emojis/malicious-hacker
- ThreatPost (2018) *Navigating an Uncharted Future, Bug Bounty Hunters Seek Safe Harbors*. threatpost.com [Available Online] https://threatpost.com/navigating-an-uncharted-future-bug-bounty-hunters-seek-safe-harbors/133202/



- Uber Technologies, Inc. (2023) Uber Logo | Uber. Google Play Store [Available Online] https://play.google.com/store/apps/details? id=com.ubercab&hl=th
- Walgreens (2023) *Walgreens Logo* | *Newsroom.* Walgreens Newsroom [Available Online] https://news.walgreens.com/