Current State of API Security and Machine Learning

Fatima Hussain, Brett Noye, and **Salah Sharieh,** *Royal Bank of Canada, Toronto*

May 2019

I. INTRODUCTION

The adaptation of application program interface (API)s in every enterprise is the emerging business trend, and at the same time it diversifies the threat domain for businesses. APIs are becoming the new and most important infrastructure layer on the Internet and are the most vulnerable points of attack in modern systems. Each API adds new dimensions to security threats and attack vectors to corporate data and applications, therefore critically forfeiting the business systems. Traditional security features for API protection are provided through API gateways, and it had been nothing more than API keys and username/password combinations (HTTP authentication). On the other hand, intruders and hackers are getting smarter. Combining the proliferation of social engineering platforms with recent technological advancements, the ability to gain access to confidential data has become both easier and common [1], [2]. APIs funnel data among applications, a multitude of various API users, and cloud infrastructure, therefore sensitive or confidential information might get exposed to unauthorized users, if API security is not carefully crafted. Using a holistic approach to securing APIs not only addresses the vulnerability issues, but offers protection for all of the infrastructure, networks and information.

With the recent advancement in technology and processing and computational power, Machine Learning (ML), a subcategory of Artificial Intelligence (AI), is being used in cyber security for application, network security, and performing threat and risk analysis from structured and unstructured data sources [3]. It can identify relationships among various types of threats, suspicious IP addresses and abnormal behaviors by careful data analysis. As real time applications are time critical, the ability of ML to quickly perform security analysis and to make critical decisions and launch remedying responses makes it an excellent choice for API developers and managers. In this article, we will briefly address concerns related to API security and recent ML-enabled API security trends.

II. API SECURITY

With the advent of digital transformation and recent emerging trends of cyber physical systems, new business trends are evolving from traditional enterprise. New business styles and models are being introduced and promoted. Development of unique and modern API is the enabler for most of the latest styles of business models. According to a report [4] more than 20,000 APIs are reported by January 2018. This tremendous increase in use of APIs is emerging from a trend that over 60% of companies acknowledge the critical importance of API integration to their business models.

The Internet of Things is another driving force behind rapid growth in development of new and smart APIs as these APIs act as interface between smart devices and the Internet. The latest business trends incorporate the use of APIs, which was not seen before in some of businesses earlier, such as:

- Banking institutions are incorporating the use of APIs for their clients for better service experience. Banks are adopting API agile model for efficient and adoptable financial and secure architectures.
- Healthcare practitioners use various APIs for patients and clients to provide integrated healthcare services and allow interoperability across the organization.

- Retailers are using APIs for smarter e-commerce platforms for their customers, such as mobile payments, etc.

Broadly speaking, three main types of APIs are available in practice:

- Public APIs are publicly available and can be access from outside the organization.
- Partner APIs are not publicly available and are exposed only to strategic business partners.
- Internal APIs are exposed to internal systems and not visible to users or public. These APIs are meant to be used among various internal development teams.

A. Traditional API Security

A traditional security model mainly relies on authentication, throttling, and communication security as shown in Table 1, while AI supported API security keeps track of traffic insights, historical trend along security with existing foundational security features.

AI Driven API Security	Deep Insight of API Traffic: Trend Analysis, Historical Attack, Anomaly Detection
	Application and Data Attacks
	API Login Attacks
Traditional, Foundational API Security	Access Control (token. Key authentication)
	Load Balancing and Rate Limiting (client throttling)
	Communication & Network Privacy (SSL/TLS)

Table 1: AI-driven API Security Model

All major social networks and services have their own APIs; Facebook, Twitter and Google are a few prominent ones to name. Are these really secure and well protected? Recent security breaches in Facebook and Cambridge and also API attacks in Tesla and Snapchat are thought provoking. It raises various questions such as whether these breaches are caused by API vulnerabilities or because of lack of security and privacy.

A few considerations are important to make before designing security for any type of an API: basic attributes like “level of exposure” of an API (who will access it, is it an internal application or anyone on Internet can access it), “sensitivity of data” being exchanged (is it non-user related data or sensitive data like credit card or health information) and “data integrity”. With the use of APIs, we are exposing our resources and end systems to outside users. In a typical web page request, very little information is shared, i.e, in HTTP address and forms, however in the case of API calls, not only HTTP URLs but also HTTP header, queries and parameters are exposed to the outside world. Thus, granularity boundaries are moved from secure internal boundaries to external devices.

III. MACHINE LEARNING AND API SECURITY

A. AI Enabled Security Solutions

As discussed above, traditional security measures for APIs focus on API access using authorization, authentication, rate limiting and network privacy. These are powerful tools but not comprehensive solutions for addressing specialized threats like API-specific denial of service (DoS), and application, data and log-in attacks. For instance, API specific DoS originates from poorly designed APIs in which rate limiting is not enforced. Sometimes a few API endpoints are computationally heavy to run, such as authentication logic requiring a hashing algorithm. Attackers purposefully exploit and spam such endpoints and take down the entire system.

A comprehensive API security solution does not require only security capabilities, but also anomaly detection ability. This is a first line of defense, and malicious behavior is detected immediately [1].

AI and ML are excellent tools for development of such comprehensive and intelligent APIs and can be used to manage challenging and new emerging threat models. These include identifying and flagging anomalous behaviors and malicious data trends and identifying and blocking API attacks and abnormal behavior patterns under multiple environments and circumstances. As a result, continuous learning capabilities are added to the APIs, and anomalous behavior is flagged without prior knowledge of attacks and written policies.

Various machine learning algorithms such as Naïve Bayes, K-Nearest Neighbors, Decision Tree, Random Forest and Support Vector Machine, Deep Learning and Neural Networks are recommended and are being used for API security [5], [6]. Various standardization bodies and government agencies are actively working in cyber security or more specifically in API security; such as, National Institute of Standards and Technology's (NIST) for security of mobile applications, National Information Assurance Partnership (NIAP) for profile protection of application software, and Mobile Technology Tiger Team (MTTT) for mobile application security criteria and standards [7].

B. Latest Trends and Platforms for API Security using Machine Learning

In the following lines, we will summarize a few of the well-known available platforms/tools used in application security incorporating machine learning:

- ***Amazon AI***: Amazon ML and AI provides visualization tools and wizards for developing ML models without diving deep into ML/AI technology. Amazon Sagemaker is a platform used to build, develop and train any kind of AI/ ML enabled application without worrying about custom code generation, system and infrastructure support. These tools can be used to integrate ML enabled application security and predictive models from data stored in Amazon S3, Amazon Redshift or Amazon RDS, which can help identify potentially fraudulent retail transactions, or detect fraudulent or inappropriate item reviews, etc. [8], [9]
  1. ***AWS Deep Learning AMIs***: provide infrastructure support and tools to develop and accelerate deep learning models for web applications in the cloud.
  2. ***Open GraphQL***: is another AI enabled application integration platform.
  3. ***Amazon EC2 P3 Instance***: is used to provide high performance computing services to ML/AI enabled applications. It supports various ML platforms such as, TensorFlow, Apache MXNet, Microsoft Cognitive Toolkit (CNTK), Theano, etc. [8]
- ***Wallarm***: Wallarm is an adaptive, dynamic and automated security platform for web applications. It performs behavioral analysis and syntax learning and provides run-time API monitoring and protection. Wallarm analyzes API traffic, API logic, and data boundaries, and detect anomalies in application payloads [10], such as:
  - Distinguish between legitimate and malicious bots/automation and payloads
  - Protect RESTful and mobile API
  - Decode XML, JSON, Base64 encoded and nested protocols
- ***Google AI Cloud***: It provides ML-enabled pre-trained models and is used to generate specialized applications and tailored models. It uses neural-net-based ML services for developing efficient and secure business applications. Various tools and platforms such as, Cloud Deep Learning VM Image Beta, Cloud Datalab, Cloud Natural Language API, Cloud Auto ML BETA are available for developing secure and smart business and private applications [11].
- ***COMBAT API***: COMBAT stands for continuous monitoring of behavior to protect from mobile application threats. It is used to check whether any API is malicious or not by applying Explainable Analytics (algorithm). When APIs are being developed, COMBAT is called like an API and will return the threat score of an API in the subject.
- ***CASTRA***: Castra is based on behavioral biometrics and is used for access control in

Internet of Things networks. It recognizes the walking patterns and interaction behaviors with the device and authenticates users to have access on apps.

- ***Ping Intelligence*** *(Formally ElasticBeam)*: It provides API behavioral security and monitors end to end flow in a multi-cloud environment at various times and across various clients. This software is comprised of an API behavioral security AI engine along with an API security enforcer (ASE) for traffic processing and enforcement. It can be installed on API servers, gateways (Node Js, Tomcat, Websphere) and can learn and manage API activity on all the API management platforms. It mainly detects and protects the following attacks [12], [13].
    - Attacks on applications, systems, and data such as data theft, deletion, or poisoning.
    - Log-in attacks, including pre-log-in activity.
    - API-specific DoS/DDoS attacks such as cookie management.

IV. FUTURE DIRECTION AND CONCLUSION

Organizations are leveraging superior functionalities of their existing applications by utilizing smart AI enabled APIs. These APIs support self-learning cognitive capabilities and also transform their business models. AI is being used to provide smarter security and privacy solutions for these APIs. API behavioral security is need for today's globally connected cyber world. Numerous players in this domain, in addition to above mentioned platforms, are striving hard to present versatile and secure application development models and infrastructure support. We can conveniently conclude that API security is the utmost need of today cyber world and ML/AI are being used as an effective and smart tool for achieving API security at various layers of protocol stack. However, more research and development efforts for AI supported APIs is required, in terms of API business models, analytical and technical blueprints and above all compliance and standardization issues.

REFERENCES

1) Tang et al, "Multi-factor web API security for securing Mobile Cloud," in IEEE International Conference on Fuzzy Systems and Knowledge Discovery, 2016.
2) Pajouh et al. , "A two-layer dimension reduction and two-tier classification model for anomaly-based intrusion detection in iot backbone networks," in IEEE Transactions on Emerging Topics in Computing, 2018.
3) "Using machine learning and AI to develop API-based security solutions." https://sdtimes.com/security/using-machine-learning-and-ai-to-develop-api-based-security-solutions/, 2018.
4) http://apihound.com/apifinder
5) Shi et al, "Active Deep Learning Attacks under Strict Limitations for Online API Calls", in IEEE HST -2018
6) Ganesh et al, "Malware Detection using API Calls; Topic Models and Machine Learning" in IEEE CASE-2015
7) https://www.nist.gov/
8) https://aws.amazon.com/aml/
9) https://aws.amazon.com/ec2/instance-types/p3/
10) https://wallarm.com/
11) https://cloud.google.com/products/ai/building-blocks/
12) https://www.pingidentity.com/en/platform/apiintelligence.html
13) https://www.elasticbeam.com/

Dr. Fatima Hussain is working as Security Analyst in "API Security and Governance" squad, Royal Bank of Canada. She is leading the development and promotion of new API and API development learning curriculum along with API security and governance duties. Dr Hussain's background includes number of distinguished professorships at Ryerson University and University of Guelph, where she has received awards for her research teaching and course development accomplishments within Wireless Telecommunication, Internet of Things, and Machine Learning. She has a long list of research publications in top tier conferences, books and journals. Dr. Hussain holds Doctorate and Master of Science, degrees from Ryerson University in Electrical and Computer Engineering.

Brett Noye is currently the director, responsible for API Security & Governance within Royal Bank of Canada's API Platforms team. His 27 years of IT experience spans the Cable TV, Brokerage, Market Data, and Banking sectors across multiple roles, platforms and technologies. Leveraging multiple COEs within RBC has enabled a broader enterprise understanding of API security over the traditional models that normally just focus on access controls and code assessments.

Dr. Salah Sharieh is a senior Director at Royal Bank of Canada with extensive experience in business, technology and digital transformation. Salah holds the degree of Doctor of Philosophy from McMaster University. He has more than twenty-five years of industry experience. He has several peer-reviewed publications and has contributed to several books. Salah is a member in the Yeates School of Graduate Studies at Ryerson University where he supervised Ph.D. and Master students.

EDITOR:

Waleed Ejaz is a Senior Research Associate at the Department of Electrical and Computer Engineering, Ryerson University, Toronto, Canada. Prior to this, he was a Post-doctoral fellow at Queen's University, Kingston, Canada. He received his Ph.D. degree in Information and Communication Engineering from Sejong University, Republic of Korea in 2014. He earned his M.Sc. and B.Sc. degrees in Computer Engineering from National University of Sciences & Technology, Islamabad, Pakistan and University of Engineering & Technology, Taxila, Pakistan, respectively. He worked in top engineering universities in Pakistan and Saudi Arabia as a Faculty Member. His current research interests include the Internet of Things (IoT), energy harvesting, 5G cellular networks, and mobile cloud computing. He is currently serving as an Associate Editor of the Canadian Journal of Electrical and Computer Engineering and the IEEE ACCESS. In addition, he is handling the special issues in IET Communications, the IEEE ACCESS, and the Journal of Internet Technology. He also completed certificate courses on Teaching and Learning in Higher Education from the Chang School at Ryerson University.