Threats in Crowdsourcing Threat Intelligence for Practical Threat Triaging

Afsah Anwar, Yi Hui Chen, Roy Hodgman, Tom Sellers,
Engin Kirda, Alina Oprea

Evolving Threat Landscape

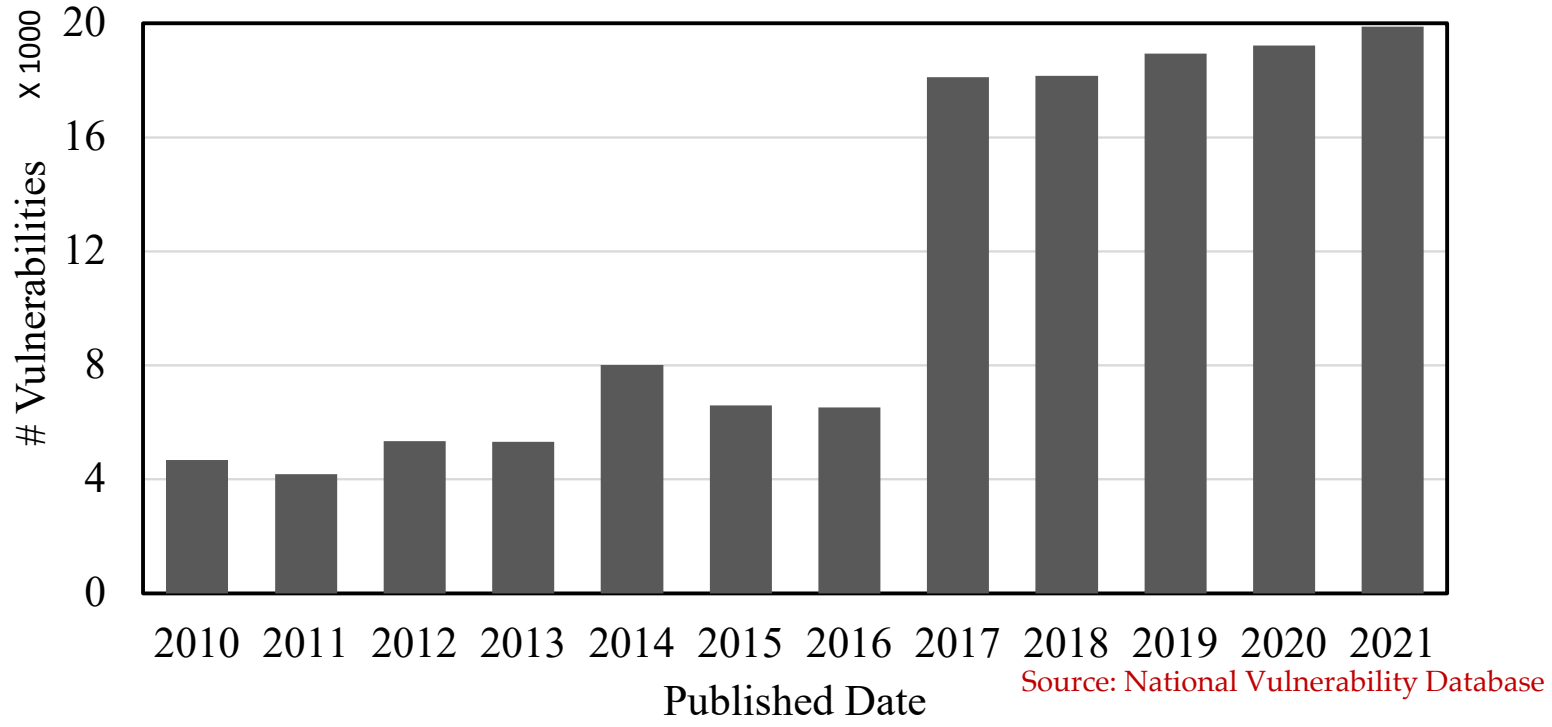
Northeastern University
Khoury College of
Computer Sciences

By Fortinet, Jonas Walker, and Derek Manky | July 27, 2022

Nowadays, threat actors are leaning on new tools and techniques to improve the efficiency of their attacks. With attacks increasing in speed, agility, and sophistication, it is critical to maximize artificial intelligence and machine learning approaches to defend against evolving attack techniques.

Vulnerability Disclosure Rate

- More and more vulnerabilities are disclosed now



Source: National Vulnerability Database

Malware Investigations - Traditional

Northeastern University
Khoury College of
Computer Sciences

Dissecting Android Malware: Characterization and Evolution

Yajin Zhou
Department of Computer Science
North Carolina State University
yajin_zhou@ncsu.edu

Xuxian Jiang
Department of Computer Science
North Carolina State University
jiang@cs.ncsu.edu

Oakland 2012

***Abstract*—The popularity and adoption of smartphones has greatly stimulated the spread of mobile malware, especially on the popular platforms such as Android. In light of their rapid growth, there is a pressing need to develop effective solutions. However, our defense capability is largely constrained by the limited understanding of these emerging mobile malware and the lack of timely access to related samples.**

In this paper, we focus on the Android platform and aim to systematize or characterize existing Android malware. Particularly, with more than one year effort, we have managed to collect more than 1,200 malware samples that cover the majority of existing Android malware families, ranging from their debut in August 2010 to recent ones in October 2011. In addition, we systematically characterize them from various aspects, including their installation methods, activation mechanisms as well as the nature of carried malicious payloads. The characterization and a subsequent evolution-based study of representative families reveal that they are evolving rapidly to circumvent the detection from existing mobile anti-virus software. Based on the evaluation with four representative mobile security software, our experiments show that the best case detects 79.6% of them while the worst case detects *only* 20.2% in our dataset. These results clearly call for the need to better develop next-generation anti-mobile-malware solutions.

The goals and contributions of this paper are threefold. First, we fulfil the need by presenting the first large collection of 1260 Android malware samples[1] in 49 different malware families, which covers the majority of existing Android malware, ranging from their debut in August 2010 to recent ones in October 2011. The dataset is accumulated from more than one year effort in collecting related malware samples, including manual or automated crawling from a variety of Android Markets. To better mitigate mobile malware threats, we will release the entire dataset to the research community at http://malgenomeproject.org/.[2]

Second, based on the collected malware samples, we perform a timeline analysis of their discovery and thoroughly characterize them based on their detailed behavior breakdown, including the installation, activation, and payloads. The timeline analysis is instrumental to revealing major outbreaks of certain Android malware in the wild while the detailed breakdown and characterization of existing Android malware is helpful to better understand them and shed light on possible defenses.

Adversarial Learning Attacks on Graph-based IoT Malware Detection Systems

Ahmed Abusnaina Aminollah Khormali Hisham Alasmary
Jeman Park Afsah Anwar Aziz Mohaisen
University of Central Florida

***Abstract*—IoT malware detection using control flow graph (CFG)-based features and deep learning networks are widely explored. The main goal of this study is to investigate the robustness of such models against adversarial learning. We designed two approaches to craft adversarial IoT software: off-the-shelf methods and Graph Embedding and Augmentation (GEA) method. In the off-the-shelf adversarial learning attack methods, we examine eight different adversarial learning methods to force the model to misclassification. The GEA approach aims to preserve the functionality and practicality of the generated adversarial sample through a careful embedding of a benign sample to a malicious one. Intensive experiments are conducted to evaluate the performance of the proposed method, showing that off-the-shelf adversarial attack methods are able to achieve a misclassification rate of 100%. In addition, we observed that the GEA approach is able to misclassify all IoT malware samples as benign. The findings of this work highlight the essential need for more robust detection tools against adversarial learning, including features that are not easy to manipulate, unlike CFG-based features. The implications of the study are quite broad, since the approach challenged in this work is widely used for other applications using graphs.**

system to identify whether a given software is malicious or benign [9]. Moreover, the type of the malicious software can be identified through malware family-level classification and label extrapolation, a concept widely applied [10].

Machine learning algorithms, specifically deep learning networks, are actively used in the process of detecting/classifying malicious software from benign ones [10], [11]. Generally, machine/deep learning networks, thanks to their high performance, are widely used in a wide range of applications, such as health-care [12], finance [13], industry [14], [15], computer-vision [16], and cyber-security [17], [18]. For instance, machine learning theory is leveraged into the process of software graph analysis to build more powerful analysis tools [19]. One such application is exploring IoT malware using both graph analysis and machine learning [9]. These models not only can learn the representative characteristics of the graph, but also can be utilized to build automatic detection system to predict the label of the unseen software. However, the rise in the utilization of deep learning models in security-related

Malware Investigations - Traditional

Northeastern University
Khoury College of
Computer Sciences

Dissecting Android Malware: Characterization and Evolution

Yajin Zhou
Department of Computer Science
North Carolina State University
yajin_zhou@ncsu.edu

Xuxian Jiang
Department of Computer Science
North Carolina State University
jiang@cs.ncsu.edu

Oakland 2012

***Abstract*—The popularity and adoption of smartphones has greatly stimulated the spread of mobile malware, especially on the popular platforms such as Android. In light of their rapid growth, there is a pressing need to develop effective solutions. However, our defense capability is largely constrained by the limited understanding of these emerging mobile malware and the lack of timely access to related samples.**

In this paper, we focus on the Android platform and aim to systematize or characterize existing Android malware. Particularly, with more than one year effort, we have managed to collect more than 1,200 malware samples that cover the majority of existing Android malware families, ranging from their debut in August 2010 to recent ones in October 2011. In addition, we systematically characterize them from various aspects, including their installation methods, activation mechanisms as well as the nature of carried malicious payloads. The characterization and a subsequent evolution-based study of representative families reveal that they are evolving rapidly to circumvent the detection from existing mobile anti-virus software. Based on the evaluation with four representative mobile security software, our experiments show that the best case detects 79.6% of them while the worst case detects *only* 20.2% in our dataset. These results clearly call for the need to better develop next-generation anti-mobile-malware solutions.

The goals and contributions of this paper are threefold. First, we fulfil the need by presenting the first large collection of 1260 Android malware samples[1] in 49 different malware families, which covers the majority of existing Android malware, ranging from their debut in August 2010 to recent ones in October 2011. The dataset is accumulated from more than one year effort in collecting related malware samples, including manual or automated crawling from a variety of Android Markets. To better mitigate mobile malware threats, we will release the entire dataset to the research community at http://malgenomeproject.org/.[2]

Second, based on the collected malware samples, we perform a timeline analysis of their discovery and thoroughly characterize them based on their detailed behavior breakdown, including the installation, activation, and payloads. The timeline analysis is instrumental to revealing major outbreaks of certain Android malware in the wild while the detailed breakdown and characterization of existing Android malware is helpful to better understand them and shed light on possible defenses.

Adversarial Learning Attacks on Graph-based IoT Malware Detection Systems

Ahmed Abusnaina Aminollah Khormali Hisham Alasmary
Jeman Park Afsah Anwar Aziz Mohaisen
University of Central Florida

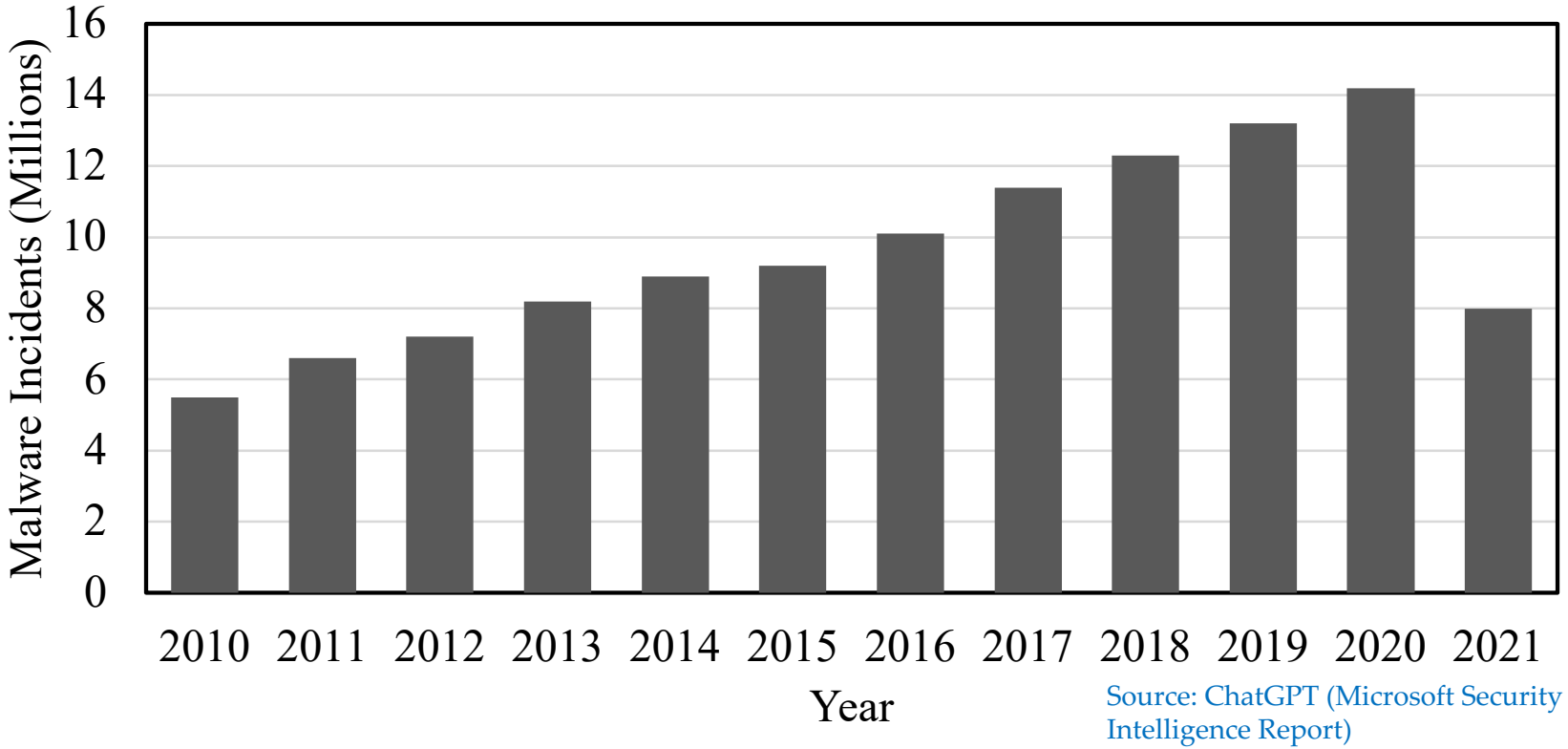
ICDCS 2019

***Abstract*—IoT malware detection using control flow graph (CFG)-based features and deep learning networks are widely explored. The main goal of this study is to investigate the robustness of such models against adversarial learning. We designed two approaches to craft adversarial IoT software: off-the-shelf methods and Graph Embedding and Augmentation (GEA) method. In the off-the-shelf adversarial learning attack methods, we examine eight different adversarial learning methods to force the model to misclassification. The GEA approach aims to preserve the functionality and practicality of the generated adversarial sample through a careful embedding of a benign sample to a malicious one. Intensive experiments are conducted to evaluate the performance of the proposed method, showing that off-the-shelf adversarial attack methods are able to achieve a misclassification rate of 100%. In addition, we observed that the GEA approach is able to misclassify all IoT malware samples as benign. The findings of this work highlight the essential need for more robust detection tools against adversarial learning, including features that are not easy to manipulate, unlike CFG-based features. The implications of the study are quite broad, since the approach challenged in this work is widely used for other applications using graphs.**

system to identify whether a given software is malicious or benign [9]. Moreover, the type of the malicious software can be identified through malware family-level classification and label extrapolation, a concept widely applied [10].

Machine learning algorithms, specifically deep learning networks, are actively used in the process of detecting/classifying malicious software from benign ones [10], [11]. Generally, machine/deep learning networks, thanks to their high performance, are widely used in a wide range of applications, such as health-care [12], finance [13], industry [14], [15], computer-vision [16], and cyber-security [17], [18]. For instance, machine learning theory is leveraged into the process of software graph analysis to build more powerful analysis tools [19]. One such application is exploring IoT malware using both graph analysis and machine learning [9]. These models not only can learn the representative characteristics of the graph, but also can be utilized to build automatic detection system to predict the label of the unseen software. However, the rise in the utilization of deep learning models in security-related

Increase in Malware Incidents



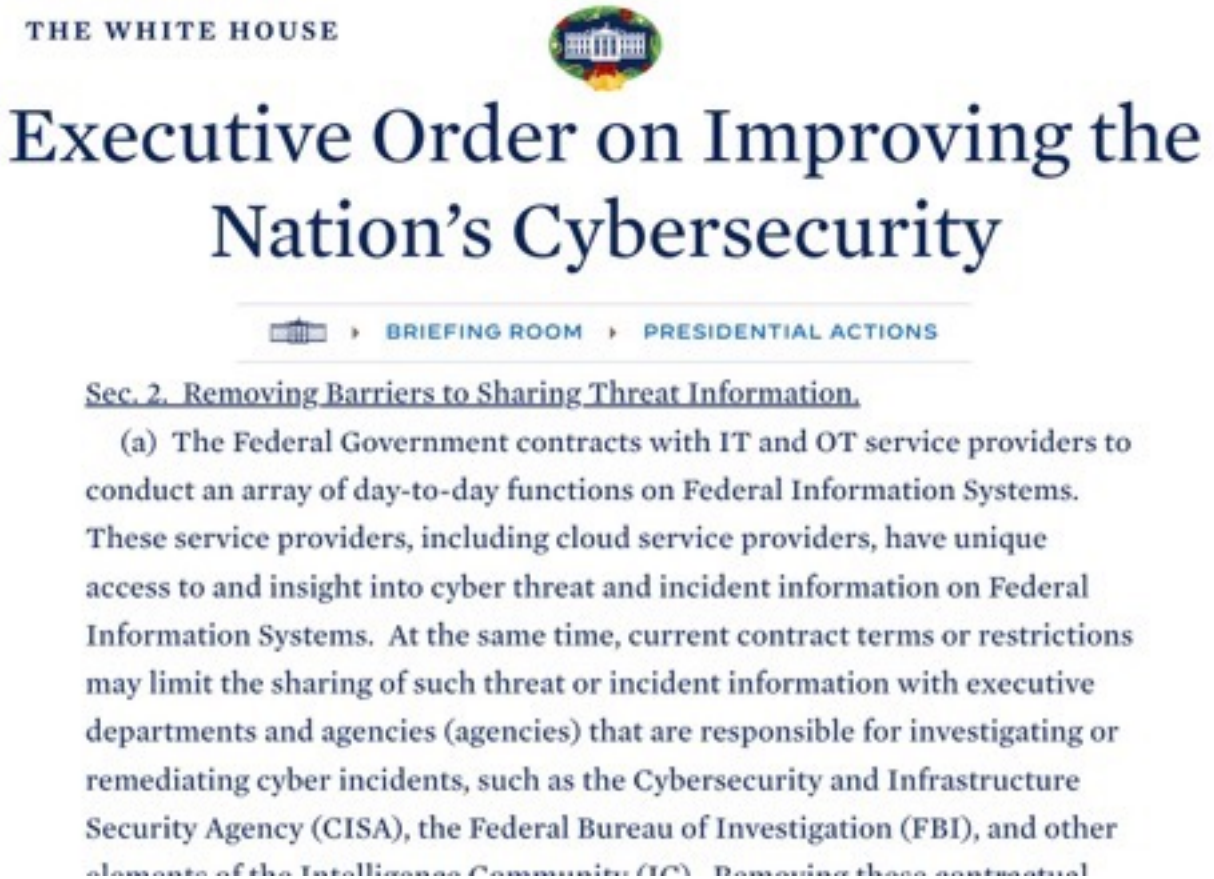
Source: ChatGPT (Microsoft Security Intelligence Report)

Broadening Security Community

- The community has grown in size
- On the other hand, cyber-criminals often adopt newer strategies to evade detection
- We then see the rise of auctioning of these strategies, that at times are sold for millions
- Shared intelligence can help improve shared security stature of all involved parties

Threat Intelligence Sharing

- Security researchers and organizations share their findings/analysis results on the web



THE WHITE HOUSE

Executive Order on Improving the Nation's Cybersecurity

BRIEFING ROOM › PRESIDENTIAL ACTIONS

Sec. 2. Removing Barriers to Sharing Threat Information.

(a) The Federal Government contracts with IT and OT service providers to conduct an array of day-to-day functions on Federal Information Systems. These service providers, including cloud service providers, have unique access to and insight into cyber threat and incident information on Federal Information Systems. At the same time, current contract terms or restrictions may limit the sharing of such threat or incident information with executive departments and agencies (agencies) that are responsible for investigating or remediating cyber incidents, such as the Cybersecurity and Infrastructure Security Agency (CISA), the Federal Bureau of Investigation (FBI), and other

Open Source Cyber Threat Intelligence

- Intelligence being shared on Social Media

Open Source Cyber Threat Intelligence

Northeastern University
Khoury College of
Computer Sciences

- Intelligence being shared on Social Media

Open Source Cyber Threat Intelligence

Northeastern University
Khoury College of
Computer Sciences

- Intelligence being shared on Social Media

- Feeds on Git
- Security reports/bulletins
- Blog posts

Open Source Cyber Threat Intelligence

Northeastern University
Khoury College of
Computer Sciences

- Intelligence being shared on Social Media

- Feeds on Git
- Security reports/bulletins
- Blog posts

Goals

- Identify sources

Open Source Cyber Threat Intelligence

Northeastern University
Khoury College of
Computer Sciences

- Intelligence being shared on Social Media

- Feeds on Git
- Security reports/bulletins
- Blog posts

Goals

- Identify sources
- Extract the IoCs

Open Source Cyber Threat Intelligence

Northeastern University
Khoury College of
Computer Sciences

- Intelligence being shared on Social Media

- Feeds on Git
- Security reports/bulletins
- Blog posts

Goals

- Identify sources
- Extract the IoCs
- Assign a disclosure date to each IoC

Open Source Cyber Threat Intelligence

Northeastern University
Khoury College of
Computer Sciences

- Intelligence being shared on Social Media

- Feeds on Git
- Security reports/bulletins
- Blog posts

Goals

- Identify sources
- Extract the IoCs
- Assign a disclosure date to each IoC
- Find them on your network

Challenges - Content

Challenges - Content

Northeastern University
Khoury College of
Computer Sciences

- Social Media
  - ❖May contain IoCs as content
  - ❖May contain a URL that contains IoCs
    - ➢Is the domain malicious?! ☹
  - ❖Post date given an idea of disclosure
- Git Repos
  - ❖We have the push date

Challenges - Content

- Social Media
  - ❖May contain IoCs as content
  - ❖May contain a URL that contains IoCs
    - ➢Is the domain malicious?! ☹
  - ❖Post date given an idea of disclosure
- Git Repos
  - ❖We have the push date

Push Date as Disclosure

Northeastern University
Khoury College of
Computer Sciences

```
APPENDIX A: BAZARCALL CAMPAIGN EMAILS - PARTIAL DATA FROM MARCH 2021 THROUGH APRIL 2021

EXAMPLES OF SENDING MAIL SERVERS:

- Received: from 52.151.9[.]80 ([52.151.9[.]80])
- Received: from localhost ([18.209.29[.]210])
- Received: from 4klever[.]com ([87.251.88[.]11])
- Received: from 4room[.]net ([87.251.88[.]13])
- Received: from 51qxct[.]com ([101.36.112[.]175])
```

Challenges - Content

- Social Media
  - ❖May contain IoCs as content
  - ❖May contain a URL that contains IoCs
    - ➢Is the domain malicious?! ☹
  - ❖Post date given an idea of disclosure
- Git Repos
  - ❖We have the push date
  - ❖What is the update frequency?
    - ➢Daily, Weekly, Monthly, Quarterly feeds
    - ➢Frequency may vary within a repository as well!

Challenges - Content

- VirusTotal
  - ❖Gives a historical behavior of a domain/subdomain
  - ❖Latency between OSCTI and IoCs being marked as malicious
    - ➢We often ran into IP addresses / domains found by CTI marked as benign in VirusTotal

- Multiple languages
  - ❖Japanese, Russian, Chinese, etc.

Extracting Intel From Content

- We have the IoCs now.
  - ❖Are they reliable?
  - ❖Can they be directly used for triaging?
- Noise
  - ❖Private addresses
  - ❖Google DNS
  - ❖Intelligence sharing platforms (Pastebin, Emerging Threats)

Temporal Inconsistency

Northeastern University
Khoury College of
Computer Sciences

pastebin.com

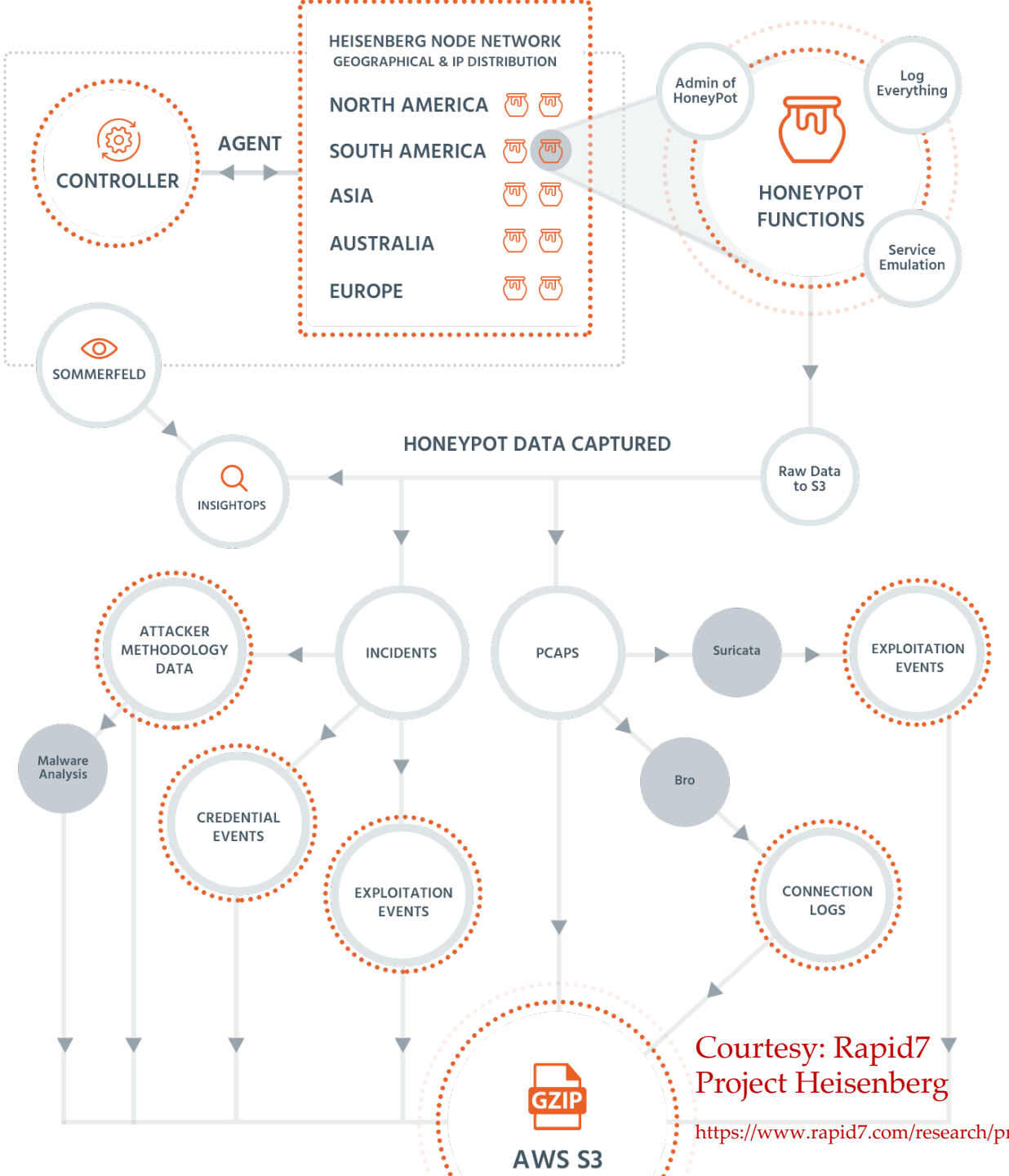
Date resolved	Detections	Resolver	IP
2022-04-13	0 / 96	VirusTotal	172.67.34.170
2020-09-21	0 / 96	VirusTotal	104.23.99.190
2020-09-21	0 / 96	VirusTotal	104.23.98.190
2020-02-18	2 / 96	VirusTotal	104.20.67.143
2020-02-16	1 / 96	VirusTotal	104.20.68.143

Achieving Actionable Intelligence

- IP addresses are volatile
- The industry still follows the 30 day rule
  - ❖Is fine as a conservative approach to restrict access
  - ❖But, leads to loss of revenue
  - ❖And, it might not be reliable
- Possible approach
  - ❖Identify disclosure date – *Min(disclosure dates)*
  - ❖Profile each ASN or subnet and relate it latency Δ
    - ➢Δ would define how long since disclosure can it be applied for triaging

Honeypot

Northeastern University
Khoury College of
Computer Sciences



Courtesy: Rapid7
Project Heisenberg

https://www.rapid7.com/research/project-heisenberg/

Thank you!

Get in touch:

Afsah Anwar

afsahanwar@gmail.com

https://www.afsah.org/