Title	Network Forensics of Ransomware and Sensor Integrity in Machine Learning; Toward Building Robust Systems against Cyberattacks
Author(s)	Kurniawan, Ade
Citation	大阪大学, 2024, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/96223
rights	
Note	やむを得ない事由があると学位審査研究科が承認したため、全文に代えてその内容の要約を公開しています。全文のご利用をご希望の場合は、<a href=”https://www.library.osaka-u.ac.jp/thesis/#closed”>大阪大学の博士論文について</a>をご参照ください。

Form 3

Abstract of Thesis

Name	(Ade Kurniawan)
Title	Network Forensics of Ransomware and Sensor Integrity in Machine Learning; Toward Building Robust Systems against Cyberattacks （ランサムウェアに対するネットワーク・フォレンジックと機械学習におけるセンサー情報の完全性 ーサイバーアタックに対するロバストなシステムの構築に向けてー)

Abstract of Thesis

As the information technologies become important infrastructures, cyberattacks have become major concerns. These attacks not only compromise information system integrity but can also lead to substantial economic losses, reputation damage, and physical safety risks. Therefore, the information systems should be more robust against such cyberattacks. To make the system more robust against attacks, it is important to comprehend how existing attacks emerge and spread by analyzing them in detail by reconstructing events of cyberattacks. Based on the knowledge obtained by the analysis of the existing attacks, we can make the system more robust against cyberattacks. In this thesis, we first analyze ransomware-based cyberattacks using network forensics. We introduce a methodology to reconstruct the event chain of such attacks from packet capture data, which helps identify infected hosts and their routes of infection. We apply this method to the case of the CERBER ransomware. As a result, we found the event chain related to this ransomware; the user’ s search on bing.com led them to www.homeimprovement.com, a website compromised by cybercriminals. We also found that they used a pseudoDarkleech script to redirect visitors to a server that deployed the RIG Exploit Kit, resulting in the download of CERBER Ransomware. While the initial parts of the thesis focused on traditional information systems such as servers, PCs, and smartphones, we also extended our analysis to systems based on machine learning models. With the widespread adoption of machine-learning and sensor technologies, these models have become new targets for cyberattacks. Therefore, protecting these models is crucial to ensuring robust defense mechanisms against such evolving cyber threats. Adversarial examples (AEs) are one of the largest vulnerabilities of the machine learning models. In this attacks, an adversary generates inputs that causes a machine learning system to generate incorrect outputs. Especially in the system using multiple sensors, some sensors may be vulnerable and can be used to generate AEs. Even if an attacker can attack a machine learning model by using only a small part of sensors, the vulnerabilities of sensors have a significant risk. However, the impact of hacking a small part of the sensor has not been discussed thus far. Therefore, we also discuss the impact of a small part of sensors on the machine learning models and demonstrate that the attacker can change the output of the ML models using multiple sensors if the attacker can manipulate the values from a part of sensors in this thesis. We call this attack sensor-based AEs. We performed experiments using the human activity recognition model with three sensor devices attached to the chest, wrist, and ankle of a user, and demonstrate that attacks are possible by hacking one of the sensor devices. Then, we also discuss the countermeasure against sensor-based AEs. One of the approach to protecting the system from the sensor-based AEs is to protect all sensor devices. However, it is difficult to protect all sensor devices, because the risk of the existence of the vulnerable sensor devices increases as the number of sensor devices increases. Therefore, we need a method to protect machine learning models even if a part of sensors are compromised by the attacker. Therefore, in this thesis, we propose a new countermeasure focusing on the sensorbased AEs. This method detects sensor-based AEs and the sensors used by the attackers by checking the inconsistency of the output of the machine learning model obtained by changing the features used by the model. By detecting the sensors used by the attackers, we can check and replace them. Our method is based on the features of the sensor-based AEs that the attacker cannot avoid; the output of the machine learning model is altered when the values from the sensors used by the attacker are incorporated. We evaluated our method using a human activity recognition model with sensors attached to the user’ s chest, wrist, and ankle. We demonstrate that our method can accurately detect sensors used by the attacker and achieves an average Recall of Detection of 0.92, and the average Precision of Detection is 0.72.

様式7

論文審査の結果の要旨及び担当者

氏　名　（　Ade Kurniawan　）			
論文審査担当者		(職)	氏　名
	主 査	教授	村田 正幸
	副 査	教授	渡辺 尚
	副 査	教授	長谷川 亨
	副 査	教授	山口 弘純
	副 査	教授	下西 英之

論文審査の結果の要旨

As the information technologies become important infrastructures, cyberattacks have become major concerns. These attacks not only compromise information system integrity but can also lead to substantial economic losses, reputation damage, and physical safety risks. Therefore, the information systems should be more robust against such cyberattacks. To make the system more robust against attacks, it is important to comprehend how existing attacks emerge and spread by analyzing them in detail by reconstructing events of cyberattacks. Based on the knowledge obtained by the analysis of the existing attacks, the system is expected to make more robust against cyberattacks.

The thesis first analyzes ransomware-based cyberattacks using network forensics. A methodology is introduced to reconstruct the event chain of such attacks from packet capture data, which helps identify infected hosts and their routes of infection. It is then applied to the case of the CERBER ransomware. As a result, the event chain related to this ransomware is first found; the user's search on bing.com led them to www.homeimprovement.com, a website compromised by cybercriminals. It is also found that they used a pseudo-Darkleech script to redirect visitors to a server that deployed the RIG Exploit Kit, resulting in the download of CERBER Ransomware. While the initial parts of the thesis focused on traditional information systems such as servers, PCs, and smartphones, it is extended to systems based on machine learning (ML) models.

Adversarial examples (AEs) are one of the largest vulnerabilities of the machine learning models. In these attacks, an adversary generates inputs that causes a machine learning system to generate incorrect outputs. Especially in the system using multiple sensors, some sensors may be vulnerable and can be used to generate AEs. Even if an attacker can attack a machine learning model by using only a small part of sensors, the vulnerabilities of sensors have a significant risk. However, the impact of hacking a small part of the sensor has not been discussed thus far. Therefore, the impact of a small part of sensors on the machine learning models is discussed and it is demonstrated that the attacker can change the output of the ML models using multiple sensors if the attacker can manipulate the values from a part of sensors in this thesis. This kind of the attack is called sensor-based AEs. Experiments are performed using the human activity recognition model with three sensor devices attached to the chest, wrist, and ankle of a user, and demonstrate that attacks are possible by hacking one of the sensor devices. Then, the countermeasure against sensor-based AEs is discussed, and a new method should be introduced to protect machine learning models even if a part of sensors is compromised by the attacker.

Accordingly, a new countermeasure focusing on the sensor-based AEs is proposed. This method detects sensor-based AEs and the sensors used by the attackers by checking the inconsistency of the output of the machine learning model obtained by changing the features used by the model. By detecting the sensors used by the attackers, those can be checked and replaced. The proposed method is based on the features of the sensor-based AEs that the attacker cannot avoid; the output of the machine learning model is altered when the values from the sensors used by the attacker are incorporated. The proposed method is evaluated using a human activity recognition model with sensors attached to the user's chest, wrist, and ankle. It is demonstrated that the method can accurately detect sensors used by the attacker and achieves an average Recall of Detection of 0.92, and the average Precision of Detection is 0.72.

As described above, this thesis has produced useful research results for network forensics of ransomware and sensor integrity in machine learning toward building robust systems against cyberattacks. Therefore, it is recognized as a valuable thesis for the degree of Doctor of Philosophy (Information Science).