Cyber Security and Artificial Intelligence

KP CHOW
CENTER FOR INFORMATION SECURITY AND CRYPTOGRAPHY
UNIVERSITY OF HONG KONG (HKU)

JUNE 2018

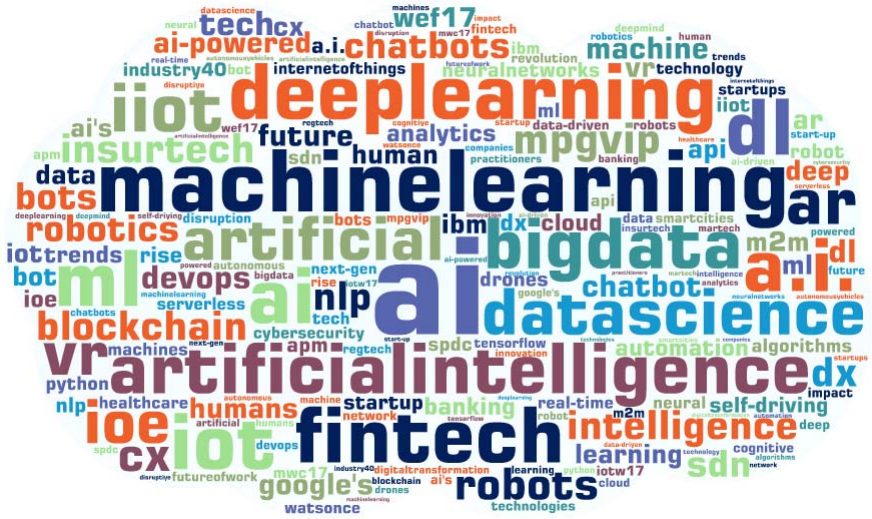


Artificial Intelligence (AI) in 1980s

- Reasoning: Logics, automated reasoning
- Expert systems
- Natural language processing: parsing and semantic analysis
- Machine learning: multi-value logics
- Computer vision: model based approach
- Searching: game tree search

AI Today

- Neural network and deep learning
- Computer vision
- NLP
- Game playing

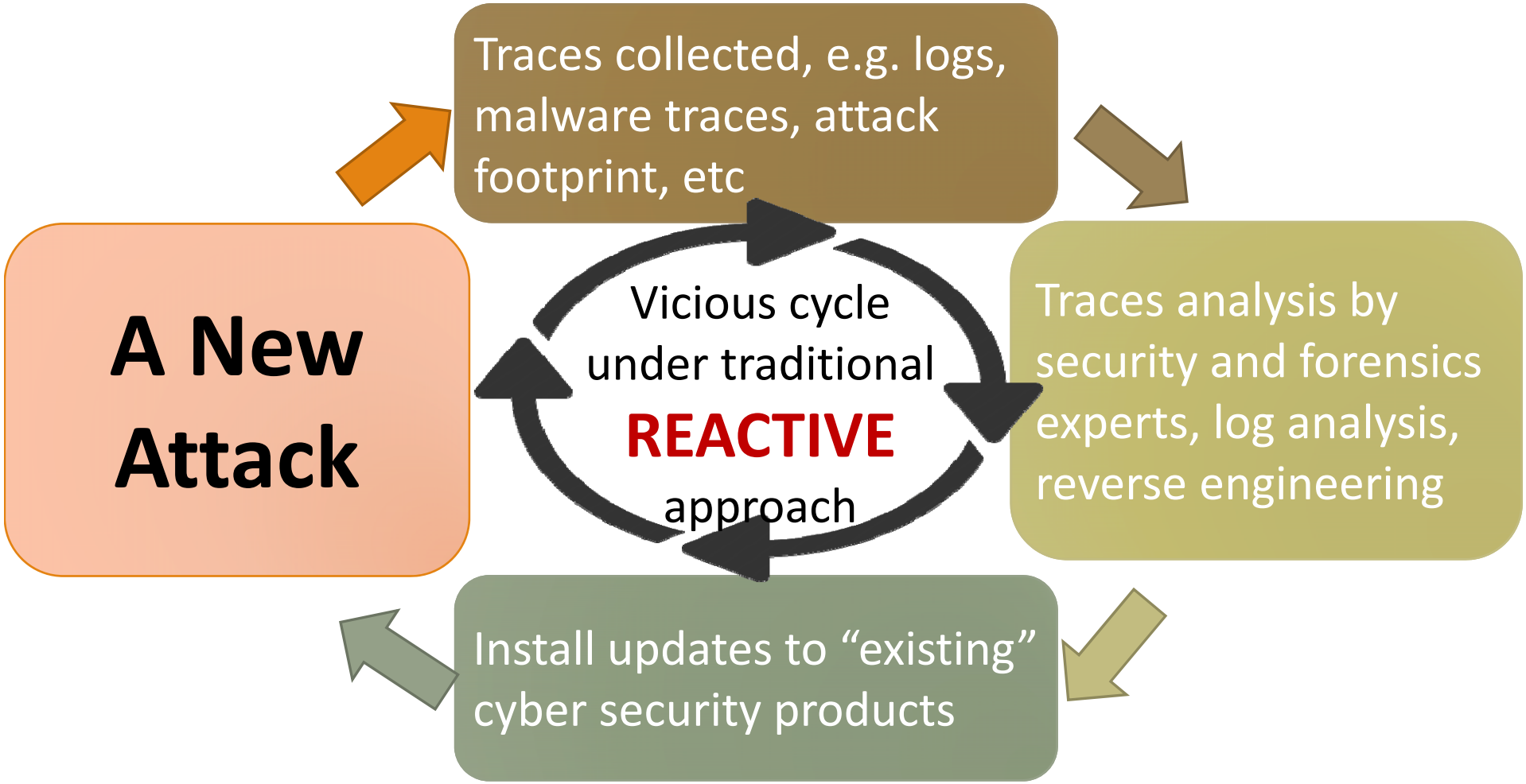



Applications in cyber security?

Major Cyber Security Issues

- New types of malware, ransomware, …
- New types of vulnerabilities, zero days, …
- New devices, IoT devices
- New “approaches” to attack, e.g. using IoT devices as bots
- New …

Traditional Reactive Approach in Cyber Security



Can AI help?

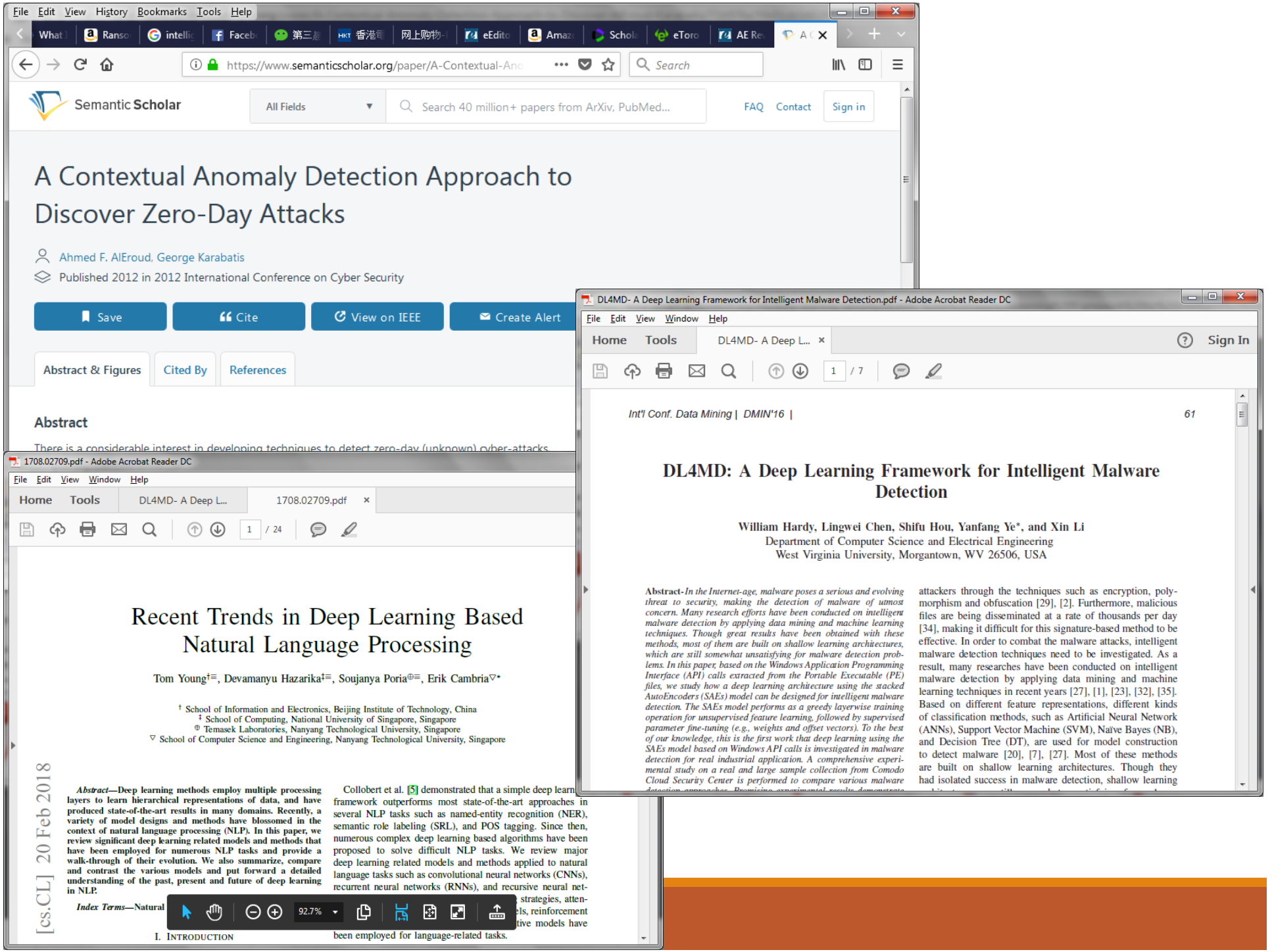
"The good news, Dave, is that the computer's passed the Turing test. The bad news is that you've failed. "

Can AI help with the "new" problems?

- Automatic detection of new malware?
- Automatic identification of zero day?
- Automatic generation of new protection scheme against new attacks?
- Automatic construction of new defense mechanisms for new devices?
- Automatic analysis and sharing of cyber security intelligence?

Some existing research

- Application of deep learning to automatic analysis of malware
- Application of data mining in behavior analysis to protect against zero days
- Application of natural language processing (NLP) to automatic analysis of cyber security intelligence


File Edit View History Bookmarks Tools Help
https://www.semanticscholar.org/paper/A-Contextual-Ano
Search
Semantic Scholar
All Fields
Search 40 million+ papers from ArXiv, PubMed...
FAQ Contact Sign in
A Contextual Anomaly Detection Approach to Discover Zero-Day Attacks
Ahmed F. AlEroud, George Karabatis
Published 2012 in 2012 International Conference on Cyber Security
Save
Cite
View on IEEE
Create Alert
Abstract & Figures
Cited By
References
Abstract
There is a considerable interest in developing techniques to detect zero-day (unknown) cyber-attacks
DL4MD- A Deep Learning Framework for Intelligent Malware Detection.pdf - Adobe Acrobat Reader DC
File Edit View Window Help
Home Tools
DL4MD- A Deep L...
Sign In
1 / 7
Int'l Conf. Data Mining | DMIN'16 | 61
DL4MD: A Deep Learning Framework for Intelligent Malware Detection
William Hardy, Lingwei Chen, Shifu Hou, Yanfang Ye*, and Xin Li
Department of Computer Science and Electrical Engineering
West Virginia University, Morgantown, WV 26506, USA
Abstract- In the Internet-age, malware poses a serious and evolving threat to security, making the detection of malware of utmost concern. Many research efforts have been conducted on intelligent malware detection by applying data mining and machine learning techniques. Though great results have been obtained with these methods, most of them are built on shallow learning architectures, which are still somewhat unsatisfying for malware detection problems. In this paper, based on the Windows Application Programming Interface (API) calls extracted from the Portable Executable (PE) files, we study how a deep learning architecture using the stacked AutoEncoders (SAEs) model can be designed for intelligent malware detection. The SAEs model performs as a greedy layerwise training operation for unsupervised feature learning, followed by supervised parameter fine-tuning (e.g., weights and offset vectors). To the best of our knowledge, this is the first work that deep learning using the SAEs model based on Windows API calls is investigated in malware detection for real industrial application. A comprehensive experimental study on a real and large sample collection from Comodo Cloud Security Center is performed to compare various malware
attackers through the techniques such as encryption, polymorphism and obfuscation [29], [2]. Furthermore, malicious files are being disseminated at a rate of thousands per day [34], making it difficult for this signature-based method to be effective. In order to combat the malware attacks, intelligent malware detection techniques need to be investigated. As a result, many researches have been conducted on intelligent malware detection by applying data mining and machine learning techniques in recent years [27], [1], [23], [32], [35]. Based on different feature representations, different kinds of classification methods, such as Artificial Neural Network (ANNs), Support Vector Machine (SVM), Naïve Bayes (NB), and Decision Tree (DT), are used for model construction to detect malware [20], [7], [27]. Most of these methods are built on shallow learning architectures. Though they had isolated success in malware detection, shallow learning
1708.02709.pdf - Adobe Acrobat Reader DC
File Edit View Window Help
Home Tools
DL4MD- A Deep L...
1708.02709.pdf
1 / 24
Recent Trends in Deep Learning Based Natural Language Processing
Tom Young[†=], Devamanyu Hazarika[‡=], Soujanya Poria[⊕=], Erik Cambria[▽*]
[†] School of Information and Electronics, Beijing Institute of Technology, China
[‡] School of Computing, National University of Singapore, Singapore
[⊕] Temasek Laboratories, Nanyang Technological University, Singapore
[▽] School of Computer Science and Engineering, Nanyang Technological University, Singapore
[cs.CL] 20 Feb 2018
Abstract—Deep learning methods employ multiple processing layers to learn hierarchical representations of data, and have produced state-of-the-art results in many domains. Recently, a variety of model designs and methods have blossomed in the context of natural language processing (NLP). In this paper, we review significant deep learning related models and methods that have been employed for numerous NLP tasks and provide a walk-through of their evolution. We also summarize, compare and contrast the various models and put forward a detailed understanding of the past, present and future of deep learning in NLP.
Index Terms—Natural
I. Introduction
Collobert et al. [5] demonstrated that a simple deep learn framework outperforms most state-of-the-art approaches in several NLP tasks such as named-entity recognition (NER), semantic role labeling (SRL), and POS tagging. Since then, numerous complex deep learning based algorithms have been proposed to solve difficult NLP tasks. We review major deep learning related models and methods applied to natural language tasks such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and recursive neural net- strategies, atten- els, reinforcement tive models have been employed for language-related tasks.
92.7%

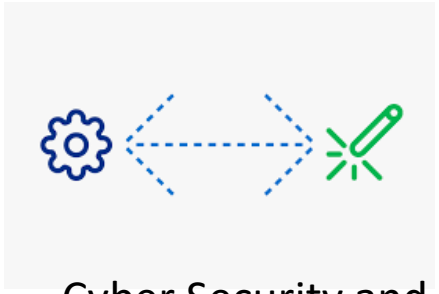
What we are doing now?

- We worked with CISC Ltd to build a Cyber Threat Intelligence Testbed (using open source software)
- Intelligence collection
  - Open source cyber security intelligence through the IaaS (Intelligence-as-a-Service) Platform by CISC Ltd
  - Data from the SHIELDs
  - Any other cyber threat intelligence

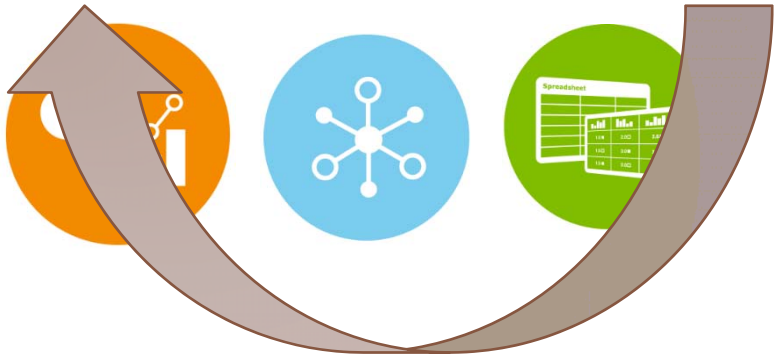
Intelligence Collection

SHIELD Data

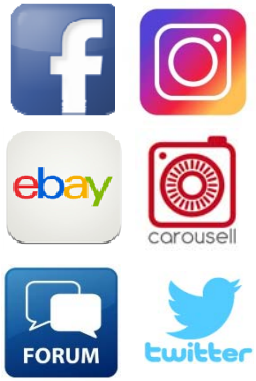
Other sources



Cyber Security and Threat Intelligence



Cyber Security Intelligence



crawl data

iaas

"Intelligence as a Service"

AI and Machine Learning Algos

What types of cyber security intelligence we have?

- Cyber security intelligence from social media and other sharing platforms
  - Text based
  - May have images
  - Large volume
  - Continuous feed
- SHIELD data and others
  - Binary data
  - Unstructured
  - Not human readable

NLP – topic identification and security classification

Deep learning for classification

Thank You

and we prepare to collaborate and share

chow@cs.hku.hk