Chap2: Machine Learning Applications in Security

January 24, 2023

Devesh C Jinwala,
Professor, SVNIT and Adjunct Prof., CSE, IIT Jammu
Department of Computer Science and Engineering,
Sardar Vallabhbhai National Institute of Technology, SURAT

- Introduction to Internet architecture. Applications of machine learning to network security. **Overview of real-world case studies** viz. Intrusion Detection System Approaches (Signature-Based Approach, Anomaly-Based Approach), Intrusion Prevention, Phishing Detection, Privacy Preservation, Spam Detection, Risk Assessment, Malware Detection. Adversarial Machine Learning. Supervised learning examples: Spam filtering, phishing. Unsupervised learning examples: Anomaly detection. [2 hours]

Security Basics: Briefly

ITU-T X.805 - Security architectural elements

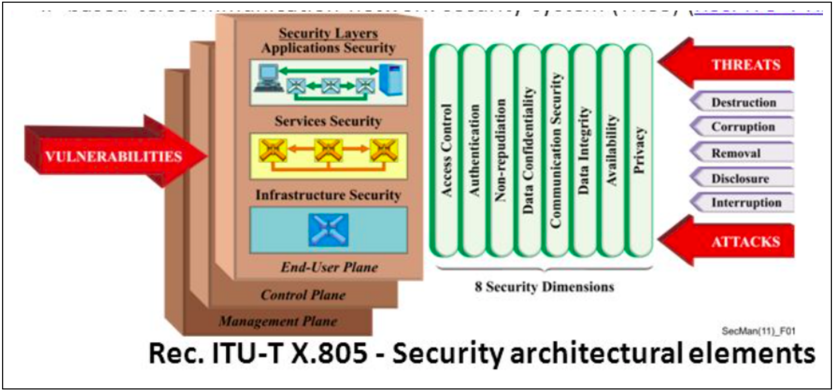

Figure: ITU-T X.805 - Security architectural elements[1]

ITU-T Security architectural elements...

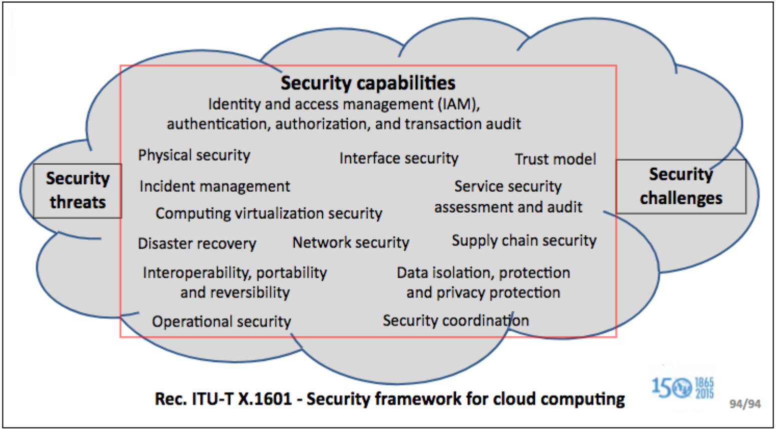

Figure: ITU-T X. 1601 - Security Framework for Cloud computing[2]

ITU-T Security architectural elements...

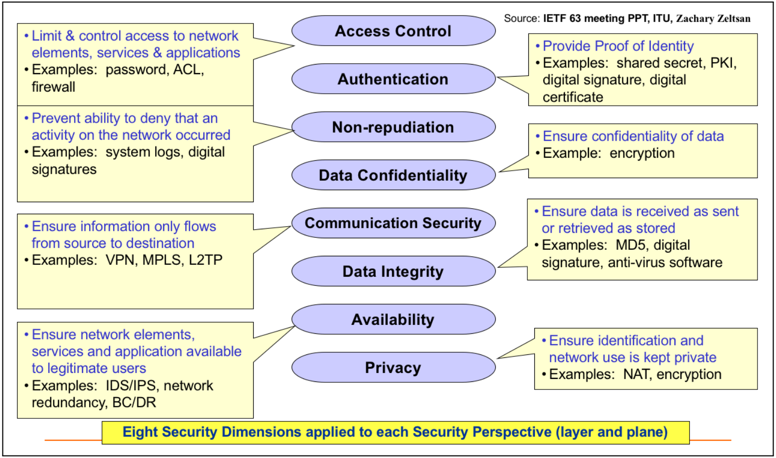

Figure: ITUT Security Dimensions

Cyber Threat Landscape

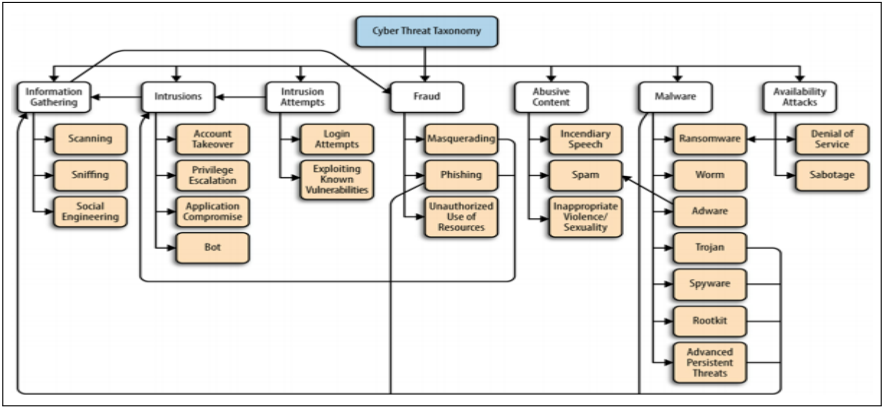

Figure: Cyber Threats Landscape

FigSrc: ML4Sec, Oreilly, Anderson et. al.

Approaches to devise security mechanisms ?

Conventional approaches to cyber defense

- address various cybersecurity threats, including viruses, Trojans, worms, spam, and botnets.

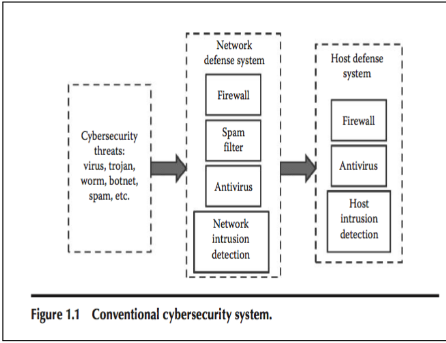

Figure: Conventional cybersecurity system

FigSrc: DM and ML in Sec, Sumit Dua and Xian Du, CRC PRess, 201

Conventional approaches to cyber defense

- address various cybersecurity threats, including viruses, Trojans, worms, spam, and botnets.
- These cybersecurity systems combat cybersecurity threats at two levels to provide

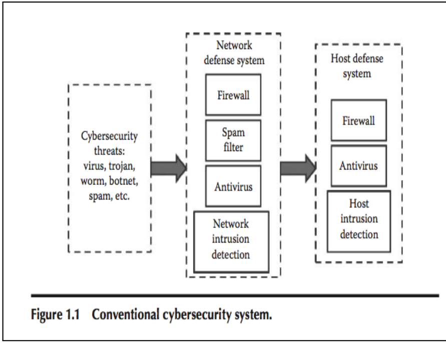

Figure: Conventional cybersecurity system

FigSrc: DM and ML in Sec, Sumit Dua and Xian Du, CRC PRess, 201

Conventional approaches to cyber defense

- address various cybersecurity threats, including viruses, Trojans, worms, spam, and botnets.
- These cybersecurity systems combat cybersecurity threats at two levels to provide
  - network-based defenses

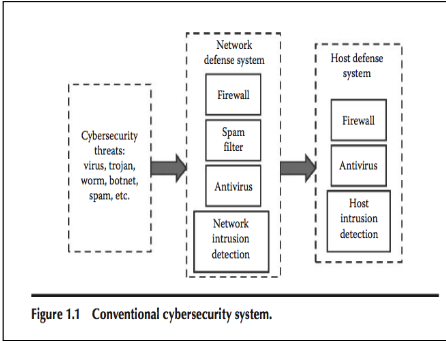

Figure: Conventional cybersecurity system

FigSrc: DM and ML in Sec, Sumit Dua and Xian Du, CRC PRess, 201

Conventional approaches to cyber defense

Conventional approaches to cyber defense

- address various cybersecurity threats, including viruses, Trojans, worms, spam, and botnets.
- These cybersecurity systems combat cybersecurity threats at two levels to provide
  - network-based defenses
  - host-based defenses.

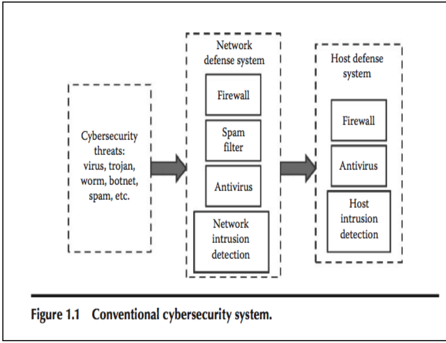

Figure: Conventional cybersecurity system

FigSrc: DM and ML in Sec, Sumit Dua and Xian Du, CRC PRess, 201

Conventional approaches to cyber defense

- address various cybersecurity threats, including viruses, Trojans, worms, spam, and botnets.
- These cybersecurity systems combat cybersecurity threats at two levels to provide
  - network-based defenses
  - host-based defenses.
- and conventionally consist of mechanisms designed in

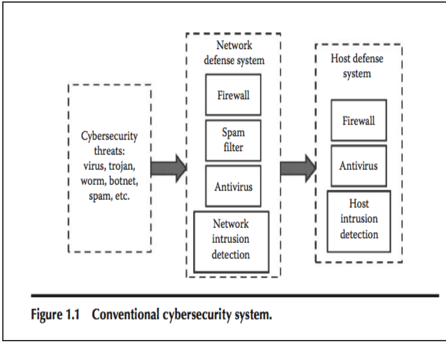

Figure: Conventional cybersecurity system

FigSrc: DM and ML in Sec, Sumit Dua and Xian Du, CRC PRess, 201

Conventional approaches to cyber defense

- address various cybersecurity threats, including viruses, Trojans, worms, spam, and botnets.
- These cybersecurity systems combat cybersecurity threats at two levels to provide
    - network-based defenses
    - host-based defenses.
- and conventionally consist of mechanisms designed in
    - fire-walls,

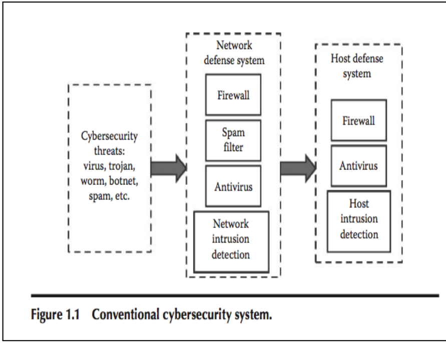

Figure: Conventional cybersecurity system

FigSrc: DM and ML in Sec, Sumit Dua and Xian Du, CRC PRess, 201

Conventional approaches to cyber defense

- address various cybersecurity threats, including viruses, Trojans, worms, spam, and botnets.
- These cybersecurity systems combat cybersecurity threats at two levels to provide
  - network-based defenses
  - host-based defenses.
- and conventionally consist of mechanisms designed in
  - fire-walls,
  - collective authentication portals and filtering routers

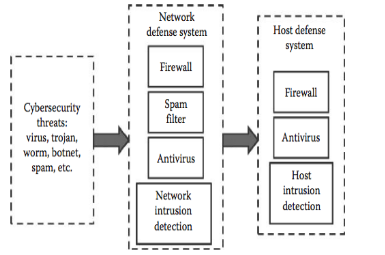

Figure 1.1 Conventional cybersecurity system.

Figure: Conventional cybersecurity system

FigSrc: DM and ML in Sec, Sumit Dua and Xian Du, CRC PRess, 201

Conventional approaches to cyber defense

- address various cybersecurity threats, including viruses, Trojans, worms, spam, and botnets.
- These cybersecurity systems combat cybersecurity threats at two levels to provide
  - network-based defenses
  - host-based defenses.
- and conventionally consist of mechanisms designed in
  - fire-walls,
  - collective authentication portals and filtering routers
  - encryption systems,

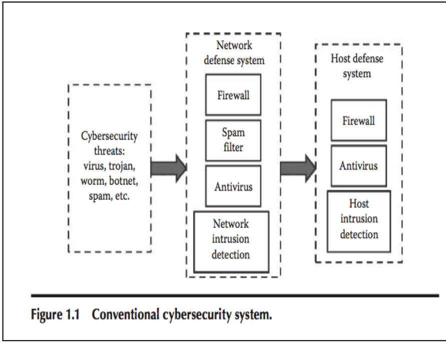

Figure: Conventional cybersecurity system

FigSrc: DM and ML in Sec, Sumit Dua and Xian Du, CRC PRess, 201

Conventional approaches to cyber defense

Conventional approaches to cyber defense

- address various cybersecurity threats, including viruses, Trojans, worms, spam, and botnets.
- These cybersecurity systems combat cybersecurity threats at two levels to provide
  - network-based defenses
  - host-based defenses.
- and conventionally consist of mechanisms designed in
  - fire-walls,
  - collective authentication portals and filtering routers
  - encryption systems,
  - network servers

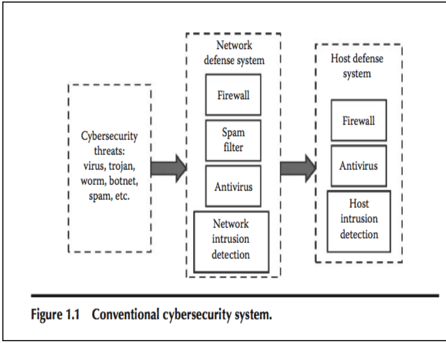

Figure: Conventional cybersecurity system

FigSrc: DM and ML in Sec, Sumit Dua and Xian Du, CRC PRess, 201

Conventional approaches to cyber defense

Conventional approaches to cyber defense

- address various cybersecurity threats, including viruses, Trojans, worms, spam, and botnets.
- These cybersecurity systems combat cybersecurity threats at two levels to provide
  - network-based defenses
  - host-based defenses.
- and conventionally consist of mechanisms designed in
  - fire-walls,
  - collective authentication portals and filtering routers
  - encryption systems,
  - network servers
- that monitor, track, block viruses & other malicious cyber attacks.

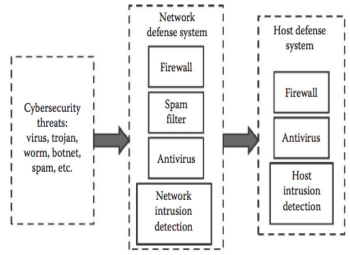

Figure 1.1 Conventional cybersecurity system.

Figure: Conventional cybersecurity system

FigSrc: DM and ML in Sec, Sumit Dua and Xian Du, CRC PRess, 201

Reactive approach to Security

Thus, conventional approaches to cyber defense mechanisms

- create a protective shield for cyberinfrastructure.

Reactive approach to Security

Thus, conventional approaches to cyber defense mechanisms

- create a protective shield for cyberinfrastructure.
- However, the vulnerabilities of these methods, themselves are ubiquitous in applications.

Reactive approach to Security

Thus, conventional approaches to cyber defense mechanisms

- create a protective shield for cyberinfrastructure.
- However, the vulnerabilities of these methods, themselves are ubiquitous in applications.
- In addition, these reactive security software are vulnerable because of

Reactive approach to Security

Thus, conventional approaches to cyber defense mechanisms

- create a protective shield for cyberinfrastructure.
- However, the vulnerabilities of these methods, themselves are ubiquitous in applications.
- In addition, these reactive security software are vulnerable because of
  - flawed design and flawed implementation of software/network infrastructure

Reactive approach to Security

Thus, conventional approaches to cyber defense mechanisms

- create a protective shield for cyberinfrastructure.
- However, the vulnerabilities of these methods, themselves are ubiquitous in applications.
- In addition, these reactive security software are vulnerable because of
  - flawed design and flawed implementation of software/network infrastructure
  - are based on primarily finding and fixing known security problems after they've been exploited by filtering dangerous input.

Reactive approach to Security

Thus, conventional approaches to cyber defense mechanisms

- create a protective shield for cyberinfrastructure.
- However, the vulnerabilities of these methods, themselves are ubiquitous in applications.
- In addition, these reactive security software are vulnerable because of
  - flawed design and flawed implementation of software/network infrastructure
  - are based on primarily finding and fixing known security problems after they've been exploited by filtering dangerous input.
  - follows network-centric approach based on penetrate and patch and input filtering

Reactive approach to Security

Thus, conventional approaches to cyber defense mechanisms

- create a protective shield for cyberinfrastructure.
- However, the vulnerabilities of these methods, themselves are ubiquitous in applications.
- In addition, these reactive security software are vulnerable because of
  - flawed design and flawed implementation of software/network infrastructure
  - are based on primarily finding and fixing known security problems after they've been exploited by filtering dangerous input.
  - follows network-centric approach based on penetrate and patch and input filtering

Reactive approach to Security

Thus, conventional approaches to cyber defense mechanisms

- create a protective shield for cyberinfrastructure.
- However, the vulnerabilities of these methods, themselves are ubiquitous in applications.
- In addition, these reactive security software are vulnerable because of
  - flawed design and flawed implementation of software/network infrastructure
  - are based on primarily finding and fixing known security problems after they've been exploited by filtering dangerous input.
  - follows network-centric approach based on penetrate and patch and input filtering

Two critical shortcomings

- the security of the application depends completely on the robustness of the fortress wall of protections that surround it

Reactive approach to Security

Thus, conventional approaches to cyber defense mechanisms

- create a protective shield for cyberinfrastructure.
- However, the vulnerabilities of these methods, themselves are ubiquitous in applications.
- In addition, these reactive security software are vulnerable because of
  - flawed design and flawed implementation of software/network infrastructure
  - are based on primarily finding and fixing known security problems after they've been exploited by filtering dangerous input.
  - follows network-centric approach based on penetrate and patch and input filtering

Two critical shortcomings

- the security of the application depends completely on the robustness of the fortress wall of protections that surround it
- the defense in depth protections themselves are vulnerable - are likely to harbor exploitable development faults and other weaknesses.

Shortcomings of Reactive approaches...

Thus, these software-intensive security systems,

- themselves, could become the first target of te nation-state adversaries, terrorists, and criminals target these

Shortcomings of Reactive approaches...

Thus, these software-intensive security systems,

- themselves, could become the first target of te nation-state adversaries, terrorists, and criminals target these
- are not attack-resistant or attack-resilient enough to prevent the more determined efforts from succeeding.

Shortcomings of Reactive approaches...

Thus, these software-intensive security systems,

- themselves, could become the first target of te nation-state adversaries, terrorists, and criminals target these
- are not attack-resistant or attack-resilient enough to prevent the more determined efforts from succeeding.
- do not directly address the insecurity of the application-level software.

Shortcomings of Reactive approaches...

Thus, these software-intensive security systems,

- themselves, could become the first target of te nation-state adversaries, terrorists, and criminals target these
- are not attack-resistant or attack-resilient enough to prevent the more determined efforts from succeeding.
- do not directly address the insecurity of the application-level software.
- the implication is revealed in these software taking binary approach to protect software - either block everything or allow everything

Shortcomings of Reactive approaches...

Thus, these software-intensive security systems,

- themselves, could become the first target of te nation-state adversaries, terrorists, and criminals target these
- are not attack-resistant or attack-resilient enough to prevent the more determined efforts from succeeding.
- do not directly address the insecurity of the application-level software.
- the implication is revealed in these software taking binary approach to protect software - either block everything or allow everything
- illustrated in the example, next....

- TCP port 80 used for transmitting numerous protocols in Web applications. . .
- How can firewalls be configured to selectively block different application-level protocols ?
- SSL two – way authentication ? SOAP service level authentication ?

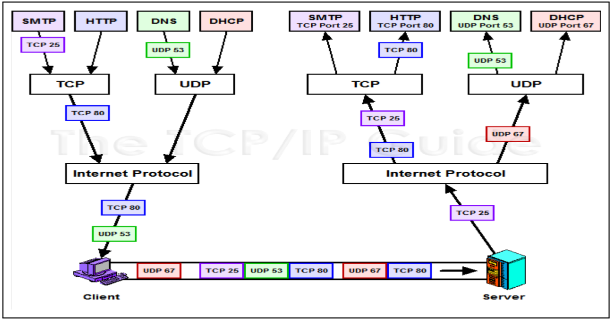

Figure: Nothing

Src: http://www.tcpipguide.com/free/t_TCPIPPortsTransportLayerTCPUDPAddressing-2.htm

Proactive Cyber Security Defense Approaches

Thus, what approach do we need to protect the software applications ?

- We need a proactive approach to build security in....

Proactive Cyber Security Defense Approaches

Thus, what approach do we need to protect the software applications ?

- We need a proactive approach to build security in....
- Proactive approaches anticipate and eliminate vulnerabilities in the cyber system, while remaining prepared to defend effectively and rapidly against attacks.

Proactive Cyber Security Defense Approaches

Thus, what approach do we need to protect the software applications ?

- We need a proactive approach to build security in....
- Proactive approaches anticipate and eliminate vulnerabilities in the cyber system, while remaining prepared to defend effectively and rapidly against attacks.
- To function correctly, proactive security solutions may require one or more of the following

Proactive Cyber Security Defense Approaches

Thus, what approach do we need to protect the software applications ?

- We need a proactive approach to build security in....
- Proactive approaches anticipate and eliminate vulnerabilities in the cyber system, while remaining prepared to defend effectively and rapidly against attacks.
- To function correctly, proactive security solutions may require one or more of the following
  - follow security by design, i.e. incorporate security protection in each phase of SDLC

Proactive Cyber Security Defense Approaches

Thus, what approach do we need to protect the software applications ?

- We need a proactive approach to build security in....
- Proactive approaches anticipate and eliminate vulnerabilities in the cyber system, while remaining prepared to defend effectively and rapidly against attacks.
- To function correctly, proactive security solutions may require one or more of the following
  - follow security by design, i.e. incorporate security protection in each phase of SDLC
  - devise a system capable of security-wise unsafe avoiding programming errors, using static analysis

Proactive Cyber Security Defense Approaches

Thus, what approach do we need to protect the software applications ?

- We need a proactive approach to build security in....
- Proactive approaches anticipate and eliminate vulnerabilities in the cyber system, while remaining prepared to defend effectively and rapidly against attacks.
- To function correctly, proactive security solutions may require one or more of the following
  - follow security by design, i.e. incorporate security protection in each phase of SDLC
  - devise a system capable of security-wise unsafe avoiding programming errors, using static analysis
  - instead of designing a program to detect security threats and incidents, relying on a software to build a vulnerabilities detector based on analyzing the patterns of data.

Proactive Cyber Security Defense Approaches

Thus, what approach do we need to protect the software applications ?

- We need a proactive approach to build security in....
- Proactive approaches anticipate and eliminate vulnerabilities in the cyber system, while remaining prepared to defend effectively and rapidly against attacks.
- To function correctly, proactive security solutions may require one or more of the following
  - follow security by design, i.e. incorporate security protection in each phase of SDLC
  - devise a system capable of security-wise unsafe avoiding programming errors, using static analysis
  - instead of designing a program to detect security threats and incidents, relying on a software to build a vulnerabilities detector based on analyzing the patterns of data.
  - What is this a software to build a vulnerabilities detector supposed to do ?

Proactive Cyber Security Defense Approaches

Thus, what approach do we need to protect the software applications ?

- We need a proactive approach to build security in....
- Proactive approaches anticipate and eliminate vulnerabilities in the cyber system, while remaining prepared to defend effectively and rapidly against attacks.
- To function correctly, proactive security solutions may require one or more of the following
  - follow security by design, i.e. incorporate security protection in each phase of SDLC
  - devise a system capable of security-wise unsafe avoiding programming errors, using static analysis
  - instead of designing a program to detect security threats and incidents, relying on a software to build a vulnerabilities detector based on analyzing the patterns of data.
  - What is this a software to build a vulnerabilities detector supposed to do ?
    - it is not enough to know what we want to detect, for we must also know how to detect what we want.

Proactive Cyber Security Defense Approaches

Thus, what approach do we need to protect the software applications ?

- We need a proactive approach to build security in....
- Proactive approaches anticipate and eliminate vulnerabilities in the cyber system, while remaining prepared to defend effectively and rapidly against attacks.
- To function correctly, proactive security solutions may require one or more of the following
  - follow security by design, i.e. incorporate security protection in each phase of SDLC
  - devise a system capable of security-wise unsafe avoiding programming errors, using static analysis
  - instead of designing a program to detect security threats and incidents, relying on a software to build a vulnerabilities detector based on analyzing the patterns of data.
  - What is this a software to build a vulnerabilities detector supposed to do ?
    - it is not enough to know what we want to detect, for we must also know how to detect what we want.
    - e.g. use software to process known benign/known malicious executables - to determine sequences of byte codes unique to the malicious executables.

Proactive Cyber Security Defense Approaches: Summary

To function correctly, proactive security solutions maybe based on one or more of the following :

1. Practice Security by Design - Secure Software Engineering

Proactive Cyber Security Defense Approaches: Summary

To function correctly, proactive security solutions maybe based on one or more of the following :

1. Practice Security by Design - Secure Software Engineering
2. Rigorously check and prevent the security issues due to memory unsafe programming languages - Static Analysis & other such techniques in Secure Software Engineering

Proactive Cyber Security Defense Approaches: Summary

To function correctly, proactive security solutions maybe based on one or more of the following :

1. Practice Security by Design - Secure Software Engineering
2. Rigorously check and prevent the security issues due to memory unsafe programming languages - Static Analysis & other such techniques in Secure Software Engineering
3. Use higher-level adaptive cyber defense systems i.e. those that discover the embedded and lurking cyber intrusions and cyber intrusion techniques

Proactive Cyber Security Defense Approaches: Summary

To function correctly, proactive security solutions maybe based on one or more of the following :

1. Practice Security by Design - Secure Software Engineering
2. Rigorously check and prevent the security issues due to memory unsafe programming languages - Static Analysis & other such techniques in Secure Software Engineering
3. Use higher-level adaptive cyber defense systems i.e. those that discover the embedded and lurking cyber intrusions and cyber intrusion techniques
    - this approach required statistically inclined technology that could deal with a constantly evolving source of abuse - constantly evolving attack mechanisms

Proactive Cyber Security Defense Approaches: Summary

To function correctly, proactive security solutions maybe based on one or more of the following :

1. Practice Security by Design - Secure Software Engineering
2. Rigorously check and prevent the security issues due to memory unsafe programming languages - Static Analysis & other such techniques in Secure Software Engineering
3. Use higher-level adaptive cyber defense systems i.e. those that discover the embedded and lurking cyber intrusions and cyber intrusion techniques
   - this approach required statistically inclined technology that could deal with a constantly evolving source of abuse - constantly evolving attack mechanisms

Proactive Cyber Security Defense Approaches: Summary

To function correctly, proactive security solutions maybe based on one or more of the following :

1. Practice Security by Design - Secure Software Engineering
2. Rigorously check and prevent the security issues due to memory unsafe programming languages - Static Analysis & other such techniques in Secure Software Engineering
3. Use higher-level adaptive cyber defense systems i.e. those that discover the embedded and lurking cyber intrusions and cyber intrusion techniques
    - this approach required statistically inclined technology that could deal with a constantly evolving source of abuse - constantly evolving attack mechanisms

Our focus here is on the last approach of the above.....one that is based on the use of ML in Security.

ML-based Proactive Cyber Security Defense: A testimonial

Note that the field of computer and network security encompasses an enormous range of threats, mechanisms and domains viz. to name a few.....

- intrusion detection, web application security,

ML-based Proactive Cyber Security Defense: A testimonial

Note that the field of computer and network security encompasses an enormous range of threats, mechanisms and domains viz. to name a few.....

- intrusion detection, web application security,
- malware analysis, social network security,

ML-based Proactive Cyber Security Defense: A testimonial

Note that the field of computer and network security encompasses an enormous range of threats, mechanisms and domains viz. to name a few.....

- intrusion detection, web application security,
- malware analysis, social network security,
- advanced persistent threats, applied cryptography,

ML-based Proactive Cyber Security Defense: A testimonial

Note that the field of computer and network security encompasses an enormous range of threats, mechanisms and domains viz. to name a few.....

- intrusion detection, web application security,
- malware analysis, social network security,
- advanced persistent threats, applied cryptography,
- ... ...

ML-based Proactive Cyber Security Defense: A testimonial

Note that the field of computer and network security encompasses an enormous range of threats, mechanisms and domains viz. to name a few.....

- intrusion detection, web application security,
- malware analysis, social network security,
- advanced persistent threats, applied cryptography,
- ... ...

ML-based Proactive Cyber Security Defense: A testimonial

Note that the field of computer and network security encompasses an enormous range of threats, mechanisms and domains viz. to name a few.....

- intrusion detection, web application security,
- malware analysis, social network security,
- advanced persistent threats, applied cryptography,
- ... ...

Amongst this spectrum of different security software, spam email filtering has emerged as a dominant security software.

ML-based Proactive Cyber Security Defense: A testimonial

Note that the field of computer and network security encompasses an enormous range of threats, mechanisms and domains viz. to name a few.....

- intrusion detection, web application security,
- malware analysis, social network security,
- advanced persistent threats, applied cryptography,
- ... ...

Amongst this spectrum of different security software, spam email filtering has emerged as a dominant security software.

Why is it labelled a testimonial of ML-based proactive security software ?

ML-based Proactive Cyber Security Defense: A testimonial...

Spam emails: even today remain a major focus....

- Spam emails filtering is a testimonial of the enormous advance resulting due to ML-based detection as compared to the traditional simplistic spam filtering techniques

ML-based Proactive Cyber Security Defense: A testimonial...

Spam emails: even today remain a major focus....

- Spam emails filtering is a testimonial of the enormous advance resulting due to ML-based detection as compared to the traditional simplistic spam filtering techniques
- What could be the the percentage of spam emails today ?

ML-based Proactive Cyber Security Defense: A testimonial...

Spam emails: even today remain a major focus....

- Spam emails filtering is a testimonial of the enormous advance resulting due to ML-based detection as compared to the traditional simplistic spam filtering techniques
- What could be the the percentage of spam emails today ?
- What percentage of spams are blocked by the modern spam filters ?

ML-based Proactive Cyber Security Defense: A testimonial...

Spam emails: even today remain a major focus....

- Spam emails filtering is a testimonial of the enormous advance resulting due to ML-based detection as compared to the traditional simplistic spam filtering techniques
- What could be the the percentage of spam emails today ?
- What percentage of spams are blocked by the modern spam filters ?
- Why is it so ?

ML-based Proactive Cyber Security Defense: A testimonial...

Spam emails: even today remain a major focus....

- Spam emails filtering is a testimonial of the enormous advance resulting due to ML-based detection as compared to the traditional simplistic spam filtering techniques
- What could be the the percentage of spam emails today ?
- What percentage of spams are blocked by the modern spam filters ?
- Why is it so ?
- Spam filters have evolved a lot: from simple word filtering and email metadata reputation based to being intelligent and adaptive spam filtering

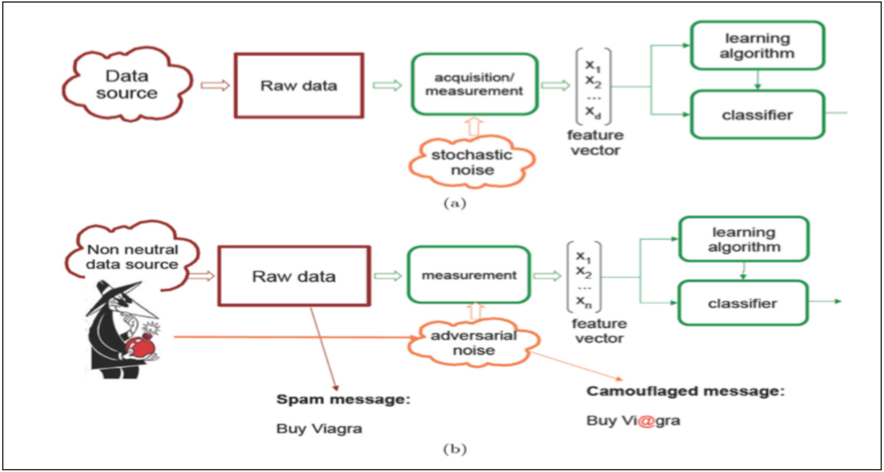

Figure: ML based Spam Filtering

3

Where to apply ML in Security ?
A Case study

Where to apply ML in Security ?

A story to learn from ......

- Let us assume that you graduate and are asked to hold the charge of computer security for your company you work.

Where to apply ML in Security ?

A story to learn from ......

- Let us assume that you graduate and are asked to hold the charge of computer security for your company you work.
- As incharge, you install firewalls, install vaccine in each machine, hold phishing training, ensure secure coding practices, and much more....

Where to apply ML in Security ?

A story to learn from ......

- Let us assume that you graduate and are asked to hold the charge of computer security for your company you work.
- As incharge, you install firewalls, install vaccine in each machine, hold phishing training, ensure secure coding practices, and much more....
- However, what is the CEO of the company concerned with ?

A story to learn from ......

- Let us assume that you graduate and are asked to hold the charge of computer security for your company you work.
- As incharge, you install firewalls, install vaccine in each machine, hold phishing training, ensure secure coding practices, and much more....
- However, what is the CEO of the company concerned with ?
- What system should you design, implement and deploy to serve the purpose ?

Where to apply ML in Security ?

A story to learn from ......

- Let us assume that you graduate and are asked to hold the charge of computer security for your company you work.
- As incharge, you install firewalls, install vaccine in each machine, hold phishing training, ensure secure coding practices, and much more....
- However, what is the CEO of the company concerned with ?
- What system should you design, implement and deploy to serve the purpose ?
- Such a system should check whether...

Where to apply ML in Security ?

A story to learn from ......

- Let us assume that you graduate and are asked to hold the charge of computer security for your company you work.
- As incharge, you install firewalls, install vaccine in each machine, hold phishing training, ensure secure coding practices, and much more....
- However, what is the CEO of the company concerned with ?
- What system should you design, implement and deploy to serve the purpose ?
- Such a system should check whether...
  - For every file sent through the network, does it contain malware?

Where to apply ML in Security ?

A story to learn from ......

- Let us assume that you graduate and are asked to hold the charge of computer security for your company you work.
- As incharge, you install firewalls, install vaccine in each machine, hold phishing training, ensure secure coding practices, and much more....
- However, what is the CEO of the company concerned with ?
- What system should you design, implement and deploy to serve the purpose ?
- Such a system should check whether...
  - For every file sent through the network, does it contain malware?
  - For every login attempt, has someone's password been compromised?

Where to apply ML in Security ?

A story to learn from ......

- Let us assume that you graduate and are asked to hold the charge of computer security for your company you work.
- As incharge, you install firewalls, install vaccine in each machine, hold phishing training, ensure secure coding practices, and much more....
- However, what is the CEO of the company concerned with ?
- What system should you design, implement and deploy to serve the purpose ?
- Such a system should check whether...
  - For every file sent through the network, does it contain malware?
  - For every login attempt, has someone's password been compromised?
  - For every email received, is it a phishing attempt?

Where to apply ML in Security ?

A story to learn from ......

- Let us assume that you graduate and are asked to hold the charge of computer security for your company you work.
- As incharge, you install firewalls, install vaccine in each machine, hold phishing training, ensure secure coding practices, and much more....
- However, what is the CEO of the company concerned with ?
- What system should you design, implement and deploy to serve the purpose ?
- Such a system should check whether...
  - For every file sent through the network, does it contain malware?
  - For every login attempt, has someone's password been compromised?
  - For every email received, is it a phishing attempt?
  - For every request to your servers, is it a denial-of-service (DoS) attack?

Where to apply ML in Security ?

A story to learn from ......

- Let us assume that you graduate and are asked to hold the charge of computer security for your company you work.
- As incharge, you install firewalls, install vaccine in each machine, hold phishing training, ensure secure coding practices, and much more....
- However, what is the CEO of the company concerned with ?
- What system should you design, implement and deploy to serve the purpose ?
- Such a system should check whether...
  - For every file sent through the network, does it contain malware?
  - For every login attempt, has someone's password been compromised?
  - For every email received, is it a phishing attempt?
  - For every request to your servers, is it a denial-of-service (DoS) attack?
  - For every outbound request from your network, is it a bot calling its command-and-control server?

Where to apply ML in Security ?

A story to learn from ......

- Let us assume that you graduate and are asked to hold the charge of computer security for your company you work.
- As incharge, you install firewalls, install vaccine in each machine, hold phishing training, ensure secure coding practices, and much more....
- However, what is the CEO of the company concerned with ?
- What system should you design, implement and deploy to serve the purpose ?
- Such a system should check whether...
  - For every file sent through the network, does it contain malware?
  - For every login attempt, has someone's password been compromised?
  - For every email received, is it a phishing attempt?
  - For every request to your servers, is it a denial-of-service (DoS) attack?
  - For every outbound request from your network, is it a bot calling its command-and-control server?
  - ... ... and many more such functionalities

- What types of basic functionality all these tasks require?

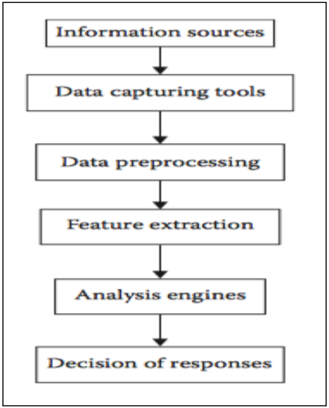

Figure: ML based Adaptive defense system for cybersecurity

[a]

But, note where does this defense system start with ? What is the input?

[a] Summet Dua, Data Mining and Machine Learning in Cybersecurity”

Where to apply ML in Security ?...

- What types of basic functionality all these tasks require?
  - Well, they require classification.......i.e. Classify all events in your network as either malicious or legitimate.

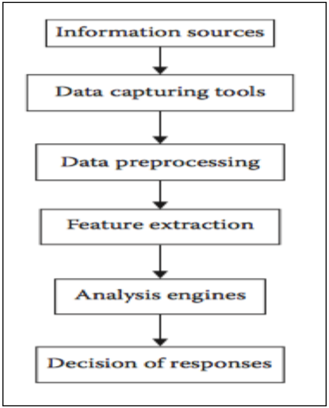

Figure: ML based Adaptive defense system for cybersecurity [a]

But, note where does this defense system start with ? What is the input?

[a] Summet Dua, Data Mining and Machine Learning in Cybersecurity"

Where to apply ML in Security ?...

- What types of basic functionality all these tasks require?
  - Well, they require classification.......i.e. Classify all events in your network as either malicious or legitimate.
- But, the issue is how can we write such a classifier program that "anticipates" and "knows" all the occurrences of malicious or benign network events ?

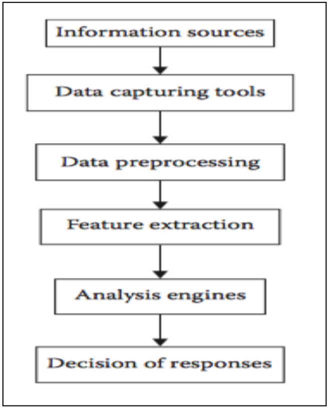

Figure: ML based Adaptive defense system for cybersecurity

[a]

But, note where does this defense system start with ? What is the input?

[a] Summet Dua, Data Mining and Machine Learning in Cybersecurity"

Where to apply ML in Security ?...

- What types of basic functionality all these tasks require?
  - Well, they require classification.......i.e. Classify all events in your network as either malicious or legitimate.
- But, the issue is how can we write such a classifier program that "anticipates" and "knows" all the occurrences of malicious or benign network events ?
- There comes the power of ML to our rescue.....Remember that the power of ML lies in the data

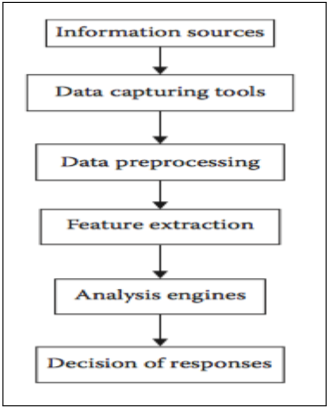

Figure: ML based Adaptive defense system for cybersecurity [a]

But, note where does this defense system start with ? What is the input?

[a] Summet Dua, Data Mining and Machine Learning in Cybersecurity"

Where to apply ML in Security ?...

- What types of basic functionality all these tasks require?
  - Well, they require classification.......i.e. Classify all events in your network as either malicious or legitimate.
- But, the issue is how can we write such a classifier program that "anticipates" and "knows" all the occurrences of malicious or benign network events ?
- There comes the power of ML to our rescue.....Remember that the power of ML lies in the data
- Generically, our ML classifier program shall follow the generic steps as shown here ....

But, note where does this defense system start with ? What is the input?

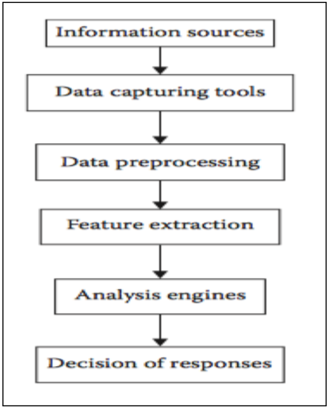

Figure: ML based Adaptive defense system for cybersecurity [a]

[a] Summet Dua, Data Mining and Machine Learning in Cybersecurity"

Where to apply ML in Security ?...

- Information sources: historical logs of binary files, login attempts, emails received, and inbound and outbound requests, attacks in the past, associating the attacks in the past with the events in the logs

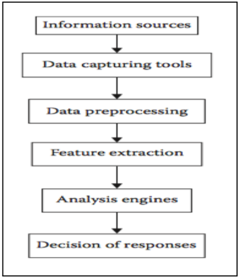

Figure: ML based Adaptive defense system for cybersecurity

[a]

[a] Summet Dua, Data Mining and Machine Learning in Cybersecurity"

- Information sources: historical logs of binary files, login attempts, emails received, and inbound and outbound requests, attacks in the past, associating the attacks in the past with the events in the logs
- Data capturing tools: Libpcap for Linux®, Solaris BSM for SUN®, and Winpcap for Windows®, capture events from the audit trails of resource information sources (e.g., network) etc......

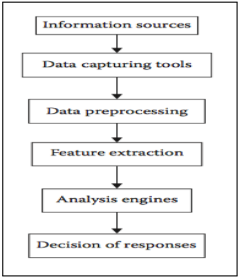

Figure: ML based Adaptive defense system for cybersecurity [a]

[a] Summet Dua, Data Mining and Machine Learning in Cybersecurity"

Where to apply ML in Security ?...

- Information sources: historical logs of binary files, login attempts, emails received, and inbound and outbound requests, attacks in the past, associating the attacks in the past with the events in the logs
- Data capturing tools: Libpcap for Linux®, Solaris BSM for SUN®, and Winpcap for Windows®, capture events from the audit trails of resource information sources (e.g., network) etc......
- Data-preprocessing module: filters out the attacks for which good signatures have been learned.

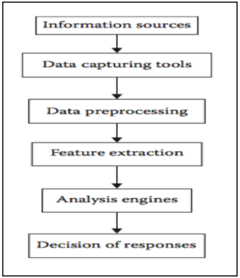

Figure: ML based Adaptive defense system for cybersecurity

[a]

[a] Summet Dua, Data Mining and Machine Learning in Cybersecurity"

Where to apply ML in Security ?...

- Information sources: historical logs of binary files, login attempts, emails received, and inbound and outbound requests, attacks in the past, associating the attacks in the past with the events in the logs
- Data capturing tools: Libpcap for Linux®, Solaris BSM for SUN®, and Winpcap for Windows®, capture events from the audit trails of resource information sources (e.g., network) etc......
- Data-preprocessing module: filters out the attacks for which good signatures have been learned.
- A feature extractor: derives basic features that are useful in event analysis engines,

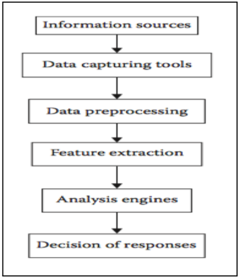

Figure: ML based Adaptive defense system for cybersecurity [a]

[a] Summet Dua, Data Mining and Machine Learning in Cybersecurity"

- Information sources: historical logs of binary files, login attempts, emails received, and inbound and outbound requests, attacks in the past, associating the attacks in the past with the events in the logs
- Data capturing tools: Libpcap for Linux®, Solaris BSM for SUN®, and Winpcap for Windows®, capture events from the audit trails of resource information sources (e.g., network) etc......
- Data-preprocessing module: filters out the attacks for which good signatures have been learned.
- A feature extractor: derives basic features that are useful in event analysis engines,
  - a sequence of system calls,

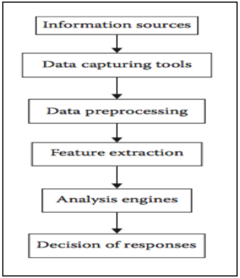

Figure: ML based Adaptive defense system for cybersecurity

[a]

[a] Summet Dua, Data Mining and Machine Learning in Cybersecurity"

Where to apply ML in Security ?...

- Information sources: historical logs of binary files, login attempts, emails received, and inbound and outbound requests, attacks in the past, associating the attacks in the past with the events in the logs
- Data capturing tools: Libpcap for Linux®, Solaris BSM for SUN®, and Winpcap for Windows®, capture events from the audit trails of resource information sources (e.g., network) etc......
- Data-preprocessing module: filters out the attacks for which good signatures have been learned.
- A feature extractor: derives basic features that are useful in event analysis engines,
    - a sequence of system calls,
    - start time,

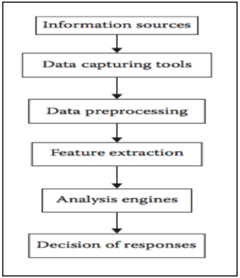

Figure: ML based Adaptive defense system for cybersecurity [a]

[a] Summet Dua, Data Mining and Machine Learning in Cybersecurity"

Where to apply ML in Security ?...

- Information sources: historical logs of binary files, login attempts, emails received, and inbound and outbound requests, attacks in the past, associating the attacks in the past with the events in the logs
- Data capturing tools: Libpcap for Linux®, Solaris BSM for SUN®, and Winpcap for Windows®, capture events from the audit trails of resource information sources (e.g., network) etc......
- Data-preprocessing module: filters out the attacks for which good signatures have been learned.
- A feature extractor: derives basic features that are useful in event analysis engines,
  - a sequence of system calls,
  - start time,
  - duration of a network flow,

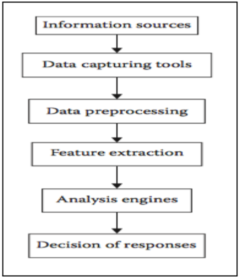

Figure: ML based Adaptive defense system for cybersecurity [a]

[a] Summet Dua, Data Mining and Machine Learning in Cybersecurity"

- Information sources: historical logs of binary files, login attempts, emails received, and inbound and outbound requests, attacks in the past, associating the attacks in the past with the events in the logs
- Data capturing tools: Libpcap for Linux®, Solaris BSM for SUN®, and Winpcap for Windows®, capture events from the audit trails of resource information sources (e.g., network) etc......
- Data-preprocessing module: filters out the attacks for which good signatures have been learned.
- A feature extractor: derives basic features that are useful in event analysis engines,
  - a sequence of system calls,
  - start time,
  - duration of a network flow,
  - source IP and source port,

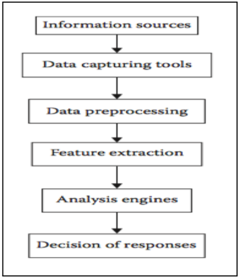

Figure: ML based Adaptive defense system for cybersecurity [a]

[a] Summet Dua, Data Mining and Machine Learning in Cybersecurity"

Where to apply ML in Security ?...

- Information sources: historical logs of binary files, login attempts, emails received, and inbound and outbound requests, attacks in the past, associating the attacks in the past with the events in the logs
- Data capturing tools: Libpcap for Linux®, Solaris BSM for SUN®, and Winpcap for Windows®, capture events from the audit trails of resource information sources (e.g., network) etc......
- Data-preprocessing module: filters out the attacks for which good signatures have been learned.
- A feature extractor: derives basic features that are useful in event analysis engines,
    - a sequence of system calls,
    - start time,
    - duration of a network flow,
    - source IP and source port,
    - destination IP and destination port, protocol,

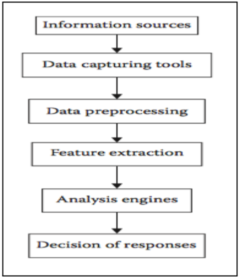

Figure: ML based Adaptive defense system for cybersecurity [a]

[a] Summet Dua, Data Mining and Machine Learning in Cybersecurity"

Where to apply ML in Security ?...

- Information sources: historical logs of binary files, login attempts, emails received, and inbound and outbound requests, attacks in the past, associating the attacks in the past with the events in the logs
- Data capturing tools: Libpcap for Linux®, Solaris BSM for SUN®, and Winpcap for Windows®, capture events from the audit trails of resource information sources (e.g., network) etc......
- Data-preprocessing module: filters out the attacks for which good signatures have been learned.
- A feature extractor: derives basic features that are useful in event analysis engines,
  - a sequence of system calls,
  - start time,
  - duration of a network flow,
  - source IP and source port,
  - destination IP and destination port, protocol,
  - number of bytes, and number of packets.

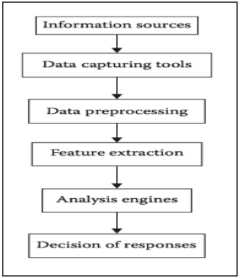

Figure: ML based Adaptive defense system for cybersecurity [a]

[a] Summet Dua, Data Mining and Machine Learning in Cybersecurity"

Where to apply ML in Security ?...

- Analysis engine: would investigate the behavior of the cyberinfrastructure, - that may have (reactive) OR may not have (proactive) appeared before in the record

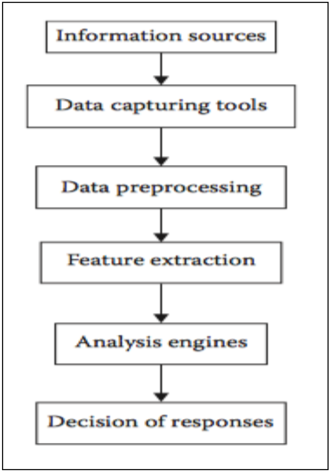

Figure: ML based Adaptive defense system for cybersecurity

Where to apply ML in Security ?...

- Analysis engine: would investigate the behavior of the cyberinfrastructure, - that may have (reactive) OR may not have (proactive) appeared before in the record
- Decision of responses: deployed once a cyber attack is identified

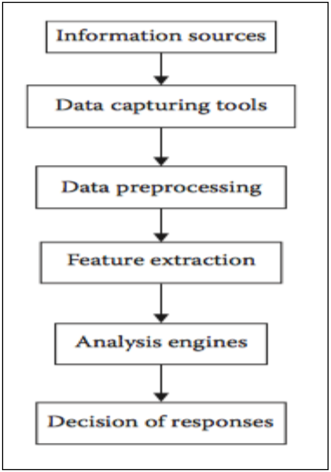

Figure: ML based Adaptive defense system for cybersecurity

- Security Cycle: Protect - Detect - Respond...

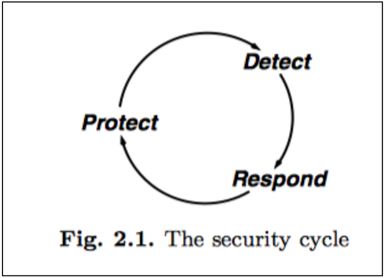

Figure: The Security Cycle

[a]

[a] Summet Dua, Data Mining and Machine Learning in Cybersecurity"

- Security Cycle: Protect - Detect - Respond...
- To which phase of the security cycle, are the ML approaches overviewed earlier applicable, most naturally ?

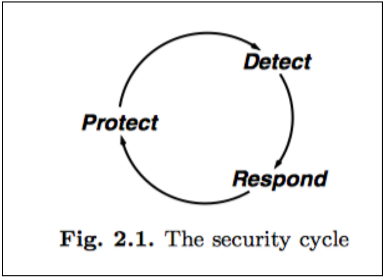

Figure: The Security Cycle

[a]

[a] Summet Dua, Data Mining and Machine Learning in Cybersecurity"

Security Cycle and ML approaches

- Security Cycle: Protect - Detect - Respond...
- To which phase of the security cycle, are the ML approaches overviewed earlier applicable, most naturally ?
- But can one not leverage the power of ML in other phases of the security cycle? We have to investigate further.

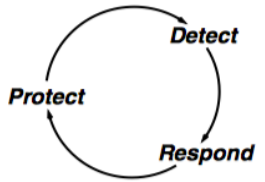

Fig. 2.1. The security cycle

Figure: The Security Cycle [a]

[a] Summet Dua, Data Mining and Machine Learning in Cybersecurity"

Real-World Use cases of Machine Learning in Security

Two broad Use cases of ML Applications in Security

In general, ML's use cases in security can be categorized into two broad categories viz. Pattern Recognition and Anomaly Detection.

Two broad Use cases of ML Applications in Security

In general, ML's use cases in security can be categorized into two broad categories viz. Pattern Recognition and Anomaly Detection.

Pattern Recognition

- the goal here is to discover explicit or latent characteristics hidden in the data.

Two broad Use cases of ML Applications in Security

In general, ML's use cases in security can be categorized into two broad categories viz. Pattern Recognition and Anomaly Detection.

Pattern Recognition

- the goal here is to discover explicit or latent characteristics hidden in the data.
- these characteristics, when distilled into feature sets, can be used to teach an algorithm to recognize other forms of the data that exhibit the same set of characteristics.

Two broad Use cases of ML Applications in Security

In general, ML's use cases in security can be categorized into two broad categories viz. Pattern Recognition and Anomaly Detection.

Pattern Recognition

- the goal here is to discover explicit or latent characteristics hidden in the data.
- these characteristics, when distilled into feature sets, can be used to teach an algorithm to recognize other forms of the data that exhibit the same set of characteristics.
  - e.g. Spam Detection: spam typically has a largely predictable set of characteristics,

Two broad Use cases of ML Applications in Security

In general, ML's use cases in security can be categorized into two broad categories viz. Pattern Recognition and Anomaly Detection.

Pattern Recognition

- the goal here is to discover explicit or latent characteristics hidden in the data.
- these characteristics, when distilled into feature sets, can be used to teach an algorithm to recognize other forms of the data that exhibit the same set of characteristics.
  - e.g. Spam Detection: spam typically has a largely predictable set of characteristics,
  - an algorithm can be trained to recognize those characteristics as a pattern by which to classify emails.

Anomaly Detection:

- the broader goal here also is the knowledge discovery, but is based on identifying anomalies.

Figure: An Anomaly

3

Two broad Use cases of ML Applications in Security...

Anomaly Detection:

- the broader goal here also is the knowledge discovery, but is based on identifying anomalies.
  - an anomaly is a value that deviates from the norm considerably enough to be regarded as a rare exception

Figure: An Anomaly

3

Two broad Use cases of ML Applications in Security...

- Anomaly Detection: *....continued*

- Anomaly Detection: *....continued*
  - here, learning specific patterns that exist within certain subsets of the data are identified and a notion of normality that describes most (say, more than 95%) of a given dataset, is established

- Anomaly Detection: *....continued*
  - here, learning specific patterns that exist within certain subsets of the data are identified and a notion of normality that describes most (say, more than 95%) of a given dataset, is established
  - thereafter, deviations from this normality of any sort will be detected as anomalies.

Two broad Use cases of ML Applications in Security...

- Anomaly Detection: *....continued*
  - here, learning specific patterns that exist within certain subsets of the data are identified and a notion of normality that describes most (say, more than 95%) of a given dataset, is established
  - thereafter, deviations from this normality of any sort will be detected as anomalies.
  - Thus, the process of detection presupposes the establishment of patterns first and then identifying the units violating those patterns.

Two broad Use cases of ML Applications in Security...

- Anomaly Detection: ....*continued*
  - here, learning specific patterns that exist within certain subsets of the data are identified and a notion of normality that describes most (say, more than 95%) of a given dataset, is established
  - thereafter, deviations from this normality of any sort will be detected as anomalies.
  - Thus, the process of detection presupposes the establishment of patterns first and then identifying the units violating those patterns.
  - anomaly detection is challenging, as in most cases, the meaning of anomalies is ambiguous. Why is it so?

Why is the meaning of anomalies is ambiguous, in most cases ?

- An anomaly from any perspective considerably different from the dataset - is usually termed a global anomaly.

Why is the meaning of anomalies is ambiguous, in most cases ?

- An anomaly from any perspective considerably different from the dataset - is usually termed a global anomaly.
  - It is easily identifiable, both automatically and manually.

Two broad Use cases of ML Applications in Security...

Why is the meaning of anomalies is ambiguous, in most cases ?

- An anomaly from any perspective considerably different from the dataset - is usually termed a global anomaly.
    - It is easily identifiable, both automatically and manually.
- But the problem arises with local anomalies - those where the values differ only insignificantly from the primary dataset

Two broad Use cases of ML Applications in Security...

Why is the meaning of anomalies is ambiguous, in most cases ?

- An anomaly from any perspective considerably different from the dataset - is usually termed a global anomaly.
  - It is easily identifiable, both automatically and manually.
- But the problem arises with local anomalies - those where the values differ only insignificantly from the primary dataset
  - even micro-clusters of deviant data standing quite close to the general dataset fall in this category

Two broad Use cases of ML Applications in Security...

Why is the meaning of anomalies is ambiguous, in most cases ?

- An anomaly from any perspective considerably different from the dataset - is usually termed a global anomaly.
    - It is easily identifiable, both automatically and manually.
- But the problem arises with local anomalies - those where the values differ only insignificantly from the primary dataset
    - even micro-clusters of deviant data standing quite close to the general dataset fall in this category

Two broad Use cases of ML Applications in Security...

Why is the meaning of anomalies is ambiguous, in most cases ?

- An anomaly from any perspective considerably different from the dataset - is usually termed a global anomaly.
    - It is easily identifiable, both automatically and manually.
- But the problem arises with local anomalies - those where the values differ only insignificantly from the primary dataset
    - even micro-clusters of deviant data standing quite close to the general dataset fall in this category

Applications of anomaly detection:

- Fraud detection (insurance, banking),

Two broad Use cases of ML Applications in Security...

Why is the meaning of anomalies is ambiguous, in most cases ?

- An anomaly from any perspective considerably different from the dataset - is usually termed a global anomaly.
  - It is easily identifiable, both automatically and manually.
- But the problem arises with local anomalies - those where the values differ only insignificantly from the primary dataset
  - even micro-clusters of deviant data standing quite close to the general dataset fall in this category

Applications of anomaly detection:

- Fraud detection (insurance, banking),
- intrusion detection (computer networks, national surveillance),

Two broad Use cases of ML Applications in Security...

Why is the meaning of anomalies is ambiguous, in most cases ?

- An anomaly from any perspective considerably different from the dataset - is usually termed a global anomaly.
  - It is easily identifiable, both automatically and manually.
- But the problem arises with local anomalies - those where the values differ only insignificantly from the primary dataset
  - even micro-clusters of deviant data standing quite close to the general dataset fall in this category

Applications of anomaly detection:

- Fraud detection (insurance, banking),
- intrusion detection (computer networks, national surveillance),
- Medical informatics (diagnosis, disorder detection),

Two broad Use cases of ML Applications in Security...

Why is the meaning of anomalies is ambiguous, in most cases ?

- An anomaly from any perspective considerably different from the dataset - is usually termed a global anomaly.
  - It is easily identifiable, both automatically and manually.
- But the problem arises with local anomalies - those where the values differ only insignificantly from the primary dataset
  - even micro-clusters of deviant data standing quite close to the general dataset fall in this category

Applications of anomaly detection:

- Fraud detection (insurance, banking),
- intrusion detection (computer networks, national surveillance),
- Medical informatics (diagnosis, disorder detection),
- Fault/damage detection (commerce, industry)

Two broad Use cases of ML Applications in Security...

Why is the meaning of anomalies is ambiguous, in most cases ?

- An anomaly from any perspective considerably different from the dataset - is usually termed a global anomaly.
  - It is easily identifiable, both automatically and manually.
- But the problem arises with local anomalies - those where the values differ only insignificantly from the primary dataset
  - even micro-clusters of deviant data standing quite close to the general dataset fall in this category

Applications of anomaly detection:

- Fraud detection (insurance, banking),
- intrusion detection (computer networks, national surveillance),
- Medical informatics (diagnosis, disorder detection),
- Fault/damage detection (commerce, industry)

Two broad Use cases of ML Applications in Security...

Why is the meaning of anomalies is ambiguous, in most cases ?

- An anomaly from any perspective considerably different from the dataset - is usually termed a global anomaly.
  - It is easily identifiable, both automatically and manually.
- But the problem arises with local anomalies - those where the values differ only insignificantly from the primary dataset
  - even micro-clusters of deviant data standing quite close to the general dataset fall in this category

Applications of anomaly detection:

- Fraud detection (insurance, banking),
- intrusion detection (computer networks, national surveillance),
- Medical informatics (diagnosis, disorder detection),
- Fault/damage detection (commerce, industry)

How do the Pattern Recognition and Anomaly Detection differ ?

Two broad Use cases of ML: How do they differ?

How do the Pattern Recognition and Anomaly Detection differ ?

- First, pattern recognition focusses on identifying similarities, whereas anomaly detection focusses on tracking similarities to identify outliers.

.......*continued*

Two broad Use cases of ML: How do they differ?

How do the Pattern Recognition and Anomaly Detection differ ?

- First, pattern recognition focusses on identifying similarities, whereas anomaly detection focusses on tracking similarities to identify outliers.
- That is, patterns extracted through pattern recognition must be strictly derived from the observed data used to train the algorithm.

.......*continued*

Two broad Use cases of ML: How do they differ?

How do the Pattern Recognition and Anomaly Detection differ ?

- First, pattern recognition focusses on identifying similarities, whereas anomaly detection focusses on tracking similarities to identify outliers.
- That is, patterns extracted through pattern recognition must be strictly derived from the observed data used to train the algorithm.
- On the other hand, in anomaly detection there can be an infinite number of anomalous patterns as anomalies including even those derived from hypothetical data that do not exist in the training or testing datasets.

.......*continued*

Two broad Use cases of ML: How do they differ?

How do the Pattern Recognition and Anomaly Detection differ ?

- First, pattern recognition focusses on identifying similarities, whereas anomaly detection focusses on tracking similarities to identify outliers.
- That is, patterns extracted through pattern recognition must be strictly derived from the observed data used to train the algorithm.
- On the other hand, in anomaly detection there can be an infinite number of anomalous patterns as anomalies including even those derived from hypothetical data that do not exist in the training or testing datasets.
- Can we consider Spam Detection to be an example of anomaly detection?

.......*continued*

Two broad Use cases of ML: How do they differ?

How do the Pattern Recognition and Anomaly Detection differ ?

- First, pattern recognition focusses on identifying similarities, whereas anomaly detection focusses on tracking similarities to identify outliers.
- That is, patterns extracted through pattern recognition must be strictly derived from the observed data used to train the algorithm.
- On the other hand, in anomaly detection there can be an infinite number of anomalous patterns as anomalies including even those derived from hypothetical data that do not exist in the training or testing datasets.
- Can we consider Spam Detection to be an example of anomaly detection?
  - One should think what is easier and natural out of (a) finding similarities between spam messages (b) finding similarities within the broad set of normal traffic ? That answers the issue.

.......*continued*

Two broad Use cases of ML: How do they differ?

How do the Pattern Recognition and Anomaly Detection differ ?

- First, pattern recognition focusses on identifying similarities, whereas anomaly detection focusses on tracking similarities to identify outliers.
- That is, patterns extracted through pattern recognition must be strictly derived from the observed data used to train the algorithm.
- On the other hand, in anomaly detection there can be an infinite number of anomalous patterns as anomalies including even those derived from hypothetical data that do not exist in the training or testing datasets.
- Can we consider Spam Detection to be an example of anomaly detection?
  - One should think what is easier and natural out of (a) finding similarities between spam messages (b) finding similarities within the broad set of normal traffic ? That answers the issue.
- Similarly, can we consider Malware detection and botnet detection to be examples of pattern recognition ?

.......*continued*

Two broad Use cases of ML: How do they differ?

How do the Pattern Recognition and Anomaly Detection differ ?

- First, pattern recognition focusses on identifying similarities, whereas anomaly detection focusses on tracking similarities to identify outliers.
- That is, patterns extracted through pattern recognition must be strictly derived from the observed data used to train the algorithm.
- On the other hand, in anomaly detection there can be an infinite number of anomalous patterns as anomalies including even those derived from hypothetical data that do not exist in the training or testing datasets.
- Can we consider Spam Detection to be an example of anomaly detection?
  - One should think what is easier and natural out of (a) finding similarities between spam messages (b) finding similarities within the broad set of normal traffic ? That answers the issue.
- Similarly, can we consider Malware detection and botnet detection to be examples of pattern recognition ?
  - Polymorphic behaviour by the attackers, fuzzing and ML to counter fuzzing

.......*continued*

Two broad Use cases of ML: How do they differ?...

How do the Pattern Recognition and Anomaly Detection differ ?

- On the similar lines as before, can we consider User authentication and user behavior analysis to be examples of anomaly detection?

.......*[Source:Clarence, Freeman et al]*

How do the Pattern Recognition and Anomaly Detection differ ?

- On the similar lines as before, can we consider User authentication and user behavior analysis to be examples of anomaly detection?
  - Is the threat model clearly known OR is it not clearly known ?

.......*[Source:Clarence, Freeman et al]*

Two broad Use cases of ML: How do they differ?...

How do the Pattern Recognition and Anomaly Detection differ ?

- On the similar lines as before, can we consider User authentication and user behavior analysis to be examples of anomaly detection?
  - Is the threat model clearly known OR is it not clearly known ?
- Can we consider Network Outlier detection OR malicious URL detection to be the examples of anomaly detection?

.......*[Source:Clarence, Freeman et al]*

Two broad Use cases of ML: How do they differ?...

How do the Pattern Recognition and Anomaly Detection differ ?

- On the similar lines as before, can we consider User authentication and user behavior analysis to be examples of anomaly detection?
  - Is the threat model clearly known OR is it not clearly known ?
- Can we consider Network Outlier detection OR malicious URL detection to be the examples of anomaly detection?
  - is a hypothetical question...

.......*[Source:Clarence, Freeman et al]*

Two broad Use cases of ML: How do they differ?...

How do the Pattern Recognition and Anomaly Detection differ ?

- On the similar lines as before, can we consider User authentication and user behavior analysis to be examples of anomaly detection?
  - Is the threat model clearly known OR is it not clearly known ?
- Can we consider Network Outlier detection OR malicious URL detection to be the examples of anomaly detection?
  - is a hypothetical question...
- Lastly, can we consider User Access Control to be using anomaly detection as the basis?

.......*[Source:Clarence, Freeman et al]*

Two broad Use cases of ML: How do they differ?...

How do the Pattern Recognition and Anomaly Detection differ ?

- On the similar lines as before, can we consider User authentication and user behavior analysis to be examples of anomaly detection?
  - Is the threat model clearly known OR is it not clearly known ?
- Can we consider Network Outlier detection OR malicious URL detection to be the examples of anomaly detection?
  - is a hypothetical question...
- Lastly, can we consider User Access Control to be using anomaly detection as the basis?
  - An example of access control in a hospital's patient record storage system

.......*[Source:Clarence, Freeman et al]*

Two broad Use cases of ML: How do they differ?...

Pattern Recognition pros and cons

- No false positives

Anomaly detection pros and cons

Source: *https://www.linkedin.com/pulse/anomaly-detection-vs-pattern-recognition-dennis-zimmer/*

Two broad Use cases of ML: How do they differ?...

Pattern Recognition pros and cons

- No false positives
- Instant results (time to value)

Anomaly detection pros and cons

Source: *https://www.linkedin.com/pulse/anomaly-detection-vs-pattern-recognition-dennis-zimmer/*

Two broad Use cases of ML: How do they differ?...

Pattern Recognition pros and cons

- No false positives
- Instant results (time to value)
- No manual root cause analysis needed as root cause is already known most of the time due to the pattern

Anomaly detection pros and cons

Source: *https://www.linkedin.com/pulse/anomaly-detection-vs-pattern-recognition-dennis-zimmer/*

Two broad Use cases of ML: How do they differ?...

Pattern Recognition pros and cons

- No false positives
- Instant results (time to value)
- No manual root cause analysis needed as root cause is already known most of the time due to the pattern
- No training necessary

Anomaly detection pros and cons

Source: *https://www.linkedin.com/pulse/anomaly-detection-vs-pattern-recognition-dennis-zimmer/*

Two broad Use cases of ML: How do they differ?...

Pattern Recognition pros and cons

- No false positives
- Instant results (time to value)
- No manual root cause analysis needed as root cause is already known most of the time due to the pattern
- No training necessary
- But, doesn't find anomalies in a dynamic way, only finds exact patterns (no dynamic behavior)

Anomaly detection pros and cons

Source: *https://www.linkedin.com/pulse/anomaly-detection-vs-pattern-recognition-dennis-zimmer/*

Two broad Use cases of ML: How do they differ?...

Pattern Recognition pros and cons

- No false positives
- Instant results (time to value)
- No manual root cause analysis needed as root cause is already known most of the time due to the pattern
- No training necessary
- But, doesn't find anomalies in a dynamic way, only finds exact patterns (no dynamic behavior)
- Needs high expertise when defining patterns

Anomaly detection pros and cons

Source: *https://www.linkedin.com/pulse/anomaly-detection-vs-pattern-recognition-dennis-zimmer/*

Two broad Use cases of ML: How do they differ?...

Pattern Recognition pros and cons

- No false positives
- Instant results (time to value)
- No manual root cause analysis needed as root cause is already known most of the time due to the pattern
- No training necessary
- But, doesn't find anomalies in a dynamic way, only finds exact patterns (no dynamic behavior)
- Needs high expertise when defining patterns

Anomaly detection pros and cons

- Finds uncommon behavior in the environment potentially uncovering bad

Source: *https://www.linkedin.com/pulse/anomaly-detection-vs-pattern-recognition-dennis-zimmer/*

Two broad Use cases of ML: How do they differ?...

Pattern Recognition pros and cons

- No false positives
- Instant results (time to value)
- No manual root cause analysis needed as root cause is already known most of the time due to the pattern
- No training necessary
- But, doesn't find anomalies in a dynamic way, only finds exact patterns (no dynamic behavior)
- Needs high expertise when defining patterns

Anomaly detection pros and cons

- Finds uncommon behavior in the environment potentially uncovering bad
- Acts dynamic and can be trained

Source: *https://www.linkedin.com/pulse/anomaly-detection-vs-pattern-recognition-dennis-zimmer/*

Two broad Use cases of ML: How do they differ?...

Pattern Recognition pros and cons

- No false positives
- Instant results (time to value)
- No manual root cause analysis needed as root cause is already known most of the time due to the pattern
- No training necessary
- But, doesn't find anomalies in a dynamic way, only finds exact patterns (no dynamic behavior)
- Needs high expertise when defining patterns

Anomaly detection pros and cons

- Finds uncommon behavior in the environment potentially uncovering bad
- Acts dynamic and can be trained
- Is not limited to a specific set of rules

Source: *https://www.linkedin.com/pulse/anomaly-detection-vs-pattern-recognition-dennis-zimmer/*

Two broad Use cases of ML: How do they differ?...

Pattern Recognition pros and cons

- No false positives
- Instant results (time to value)
- No manual root cause analysis needed as root cause is already known most of the time due to the pattern
- No training necessary
- But, doesn't find anomalies in a dynamic way, only finds exact patterns (no dynamic behavior)
- Needs high expertise when defining patterns

Anomaly detection pros and cons

- Finds uncommon behavior in the environment potentially uncovering bad
- Acts dynamic and can be trained
- Is not limited to a specific set of rules
- But, can only find symptoms, not the root cause

Source: *https://www.linkedin.com/pulse/anomaly-detection-vs-pattern-recognition-dennis-zimmer/*

Two broad Use cases of ML: How do they differ?...

Pattern Recognition pros and cons

- No false positives
- Instant results (time to value)
- No manual root cause analysis needed as root cause is already known most of the time due to the pattern
- No training necessary
- But, doesn't find anomalies in a dynamic way, only finds exact patterns (no dynamic behavior)
- Needs high expertise when defining patterns

Anomaly detection pros and cons

- Finds uncommon behavior in the environment potentially uncovering bad
- Acts dynamic and can be trained
- Is not limited to a specific set of rules
- But, can only find symptoms, not the root cause
- Cannot locate issues that don't have symptoms, which is most

Source: *https://www.linkedin.com/pulse/anomaly-detection-vs-pattern-recognition-dennis-zimmer/*

Two broad Use cases of ML: How do they differ?...

Pattern Recognition pros and cons

- No false positives
- Instant results (time to value)
- No manual root cause analysis needed as root cause is already known most of the time due to the pattern
- No training necessary
- But, doesn't find anomalies in a dynamic way, only finds exact patterns (no dynamic behavior)
- Needs high expertise when defining patterns

Anomaly detection pros and cons

- Finds uncommon behavior in the environment potentially uncovering bad
- Acts dynamic and can be trained
- Is not limited to a specific set of rules
- But, can only find symptoms, not the root cause
- Cannot locate issues that don't have symptoms, which is most
- Baseline is only as good as the setup

Source: *https://www.linkedin.com/pulse/anomaly-detection-vs-pattern-recognition-dennis-zimmer/*

A Tutorial

- Consider a case: if all buyers of an e-shop pay Rs 2500 on average for a pair of shoes, and some client pays Rs 25000 for the same purchase, is it an anomaly detection or a pattern recognition problem ?

A Tutorial

- Consider a case: if all buyers of an e-shop pay Rs 2500 on average for a pair of shoes, and some client pays Rs 25000 for the same purchase, is it an anomaly detection or a pattern recognition problem ?
- Consider a case: if an insurer receives an average MRI check for Rs2500/- from patients and suddenly gets a Rs 25000/- check for the same procedure, is it an anomaly detection or a pattern recognition problem ?

A Tutorial

- Consider a case: if all buyers of an e-shop pay Rs 2500 on average for a pair of shoes, and some client pays Rs 25000 for the same purchase, is it an anomaly detection or a pattern recognition problem ?
- Consider a case: if an insurer receives an average MRI check for Rs2500/- from patients and suddenly gets a Rs 25000/- check for the same procedure, is it an anomaly detection or a pattern recognition problem ?
- Consider the following Win-API call sequence viz.

A Tutorial

- Consider a case: if all buyers of an e-shop pay Rs 2500 on average for a pair of shoes, and some client pays Rs 25000 for the same purchase, is it an anomaly detection or a pattern recognition problem ?
- Consider a case: if an insurer receives an average MRI check for Rs2500/- from patients and suddenly gets a Rs 25000/- check for the same procedure, is it an anomaly detection or a pattern recognition problem ?
- Consider the following Win-API call sequence viz.
  - `CreatefileA, CreatePipe, CreateNamedPipeA`;

A Tutorial

- Consider a case: if all buyers of an e-shop pay Rs 2500 on average for a pair of shoes, and some client pays Rs 25000 for the same purchase, is it an anomaly detection or a pattern recognition problem ?
- Consider a case: if an insurer receives an average MRI check for Rs2500/- from patients and suddenly gets a Rs 25000/- check for the same procedure, is it an anomaly detection or a pattern recognition problem ?
- Consider the following Win-API call sequence viz.
  - `CreatefileA, CreatePipe, CreateNamedPipeA`;
  - `OpenFile, OpenFileMappingA`;

A Tutorial

- Consider a case: if all buyers of an e-shop pay Rs 2500 on average for a pair of shoes, and some client pays Rs 25000 for the same purchase, is it an anomaly detection or a pattern recognition problem ?
- Consider a case: if an insurer receives an average MRI check for Rs2500/- from patients and suddenly gets a Rs 25000/- check for the same procedure, is it an anomaly detection or a pattern recognition problem ?
- Consider the following Win-API call sequence viz.
  - `CreatefileA, CreatePipe, CreateNamedPipeA`;
  - `OpenFile, OpenFileMappingA`;
  - `WriteFile, WriteConsoleW, WriteFileEx`;

A Tutorial

- Consider a case: if all buyers of an e-shop pay Rs 2500 on average for a pair of shoes, and some client pays Rs 25000 for the same purchase, is it an anomaly detection or a pattern recognition problem ?
- Consider a case: if an insurer receives an average MRI check for Rs2500/- from patients and suddenly gets a Rs 25000/- check for the same procedure, is it an anomaly detection or a pattern recognition problem ?
- Consider the following Win-API call sequence viz.
  - `CreatefileA`, `CreatePipe`, `CreateNamedPipeA`;
  - `OpenFile`, `OpenFileMappingA`;
  - `WriteFile`, `WriteConsoleW`, `WriteFileEx`;
  - ....and so on, all of which are valid API calls,

A Tutorial

- Consider a case: if all buyers of an e-shop pay Rs 2500 on average for a pair of shoes, and some client pays Rs 25000 for the same purchase, is it an anomaly detection or a pattern recognition problem ?
- Consider a case: if an insurer receives an average MRI check for Rs2500/- from patients and suddenly gets a Rs 25000/- check for the same procedure, is it an anomaly detection or a pattern recognition problem ?
- Consider the following Win-API call sequence viz.
  - `CreatefileA, CreatePipe, CreateNamedPipeA`;
  - `OpenFile, OpenFileMappingA`;
  - `WriteFile, WriteConsoleW, WriteFileEx`;
  - ....and so on, all of which are valid API calls,

A Tutorial

- Consider a case: if all buyers of an e-shop pay Rs 2500 on average for a pair of shoes, and some client pays Rs 25000 for the same purchase, is it an anomaly detection or a pattern recognition problem ?
- Consider a case: if an insurer receives an average MRI check for Rs2500/- from patients and suddenly gets a Rs 25000/- check for the same procedure, is it an anomaly detection or a pattern recognition problem ?
- Consider the following Win-API call sequence viz.
  - `CreatefileA, CreatePipe, CreateNamedPipeA`;
  - `OpenFile, OpenFileMappingA`;
  - `WriteFile, WriteConsoleW, WriteFileEx`;
  - ....and so on, all of which are valid API calls,

  All of these are hashed and a Fuzzy hashcode is computed. Any such Fuzzyhashcode that does not fall in any of these is probably a WORM or a Trojan-Dropper or a Trojan-Downloader Trojan-Spy of a Backdoor. Is it an anomaly detection or a pattern recognition problem ?

- Consider a case: if all buyers of an e-shop pay Rs 2500 on average for a pair of shoes, and some client pays Rs 25000 for the same purchase, is it an anomaly detection or a pattern recognition problem ?
- Consider a case: if an insurer receives an average MRI check for Rs2500/- from patients and suddenly gets a Rs 25000/- check for the same procedure, is it an anomaly detection or a pattern recognition problem ?
- Consider the following Win-API call sequence viz.
  - `CreatefileA, CreatePipe, CreateNamedPipeA`;
  - `OpenFile, OpenFileMappingA`;
  - `WriteFile, WriteConsoleW, WriteFileEx`;
  - ....and so on, all of which are valid API calls,

  All of these are hashed and a Fuzzy hashcode is computed. Any such Fuzzyhashcode that does not fall in any of these is probably a WORM or a Trojan-Dropper or a Trojan-Downloader Trojan-Spy of a Backdoor. Is it an anomaly detection or a pattern recognition problem ?

With this background, we are equipped now to understand the ML Design Paradigms in Security

.......*[Source:Clarence, Freeman et al]*

ML Design Generalization in Security

ML System Design Generalization in Security: Knowledge base

The generalization of ML system designs when applied in security, in terms of their functionalities and positions are as discussed here:

- Knowledge base:
  - is baseline of known normality and/or abnormality, depending on use cases
    - e.g. blacklist(BL),
    - whitelist(WL),
    - watchlist;
    - known malware signatures,
    - system traces, and their families;
    - initial set of malicious web pages;
    - existing security policies or rules, etc.

ML System Design Generalization in Security: Data Sources

The generalization of ML system designs when applied in security, in terms of their functionalities and positions are as discussed here:

- Data Sources:
  - are where relevant data is collected.
  - can be either off-line or live online data feed
  - e.g. malware traces collected after execution(off-line),
  - URL stream(online).

ML System Design Generalization in Security: Training data

The generalization of ML system designs when applied in security, in terms of their functionalities and positions are as discussed here:

- Training data: are labelled data which are fed to classifiers in training. A few examples are as follows:
    - standard research datasets,
    - new data(mostly from industry) labeled by human,
    - synthetic datasets, or a mix.

The generalization of ML system designs when applied in security, in terms of their functionalities and positions are as discussed here:

- Pre-processor and feature extractor: construct features from data sources
  - URL aggregators,
  - graph representations,
  - SMTP header extractions,
  - n-gram model builders.

ML Paradigms in Security Problems

ML Paradigms in Security Problems

- Here, we discuss different ways of classifying the security problems, vis-a-vis the well-understood ML paradigms.

ML Paradigms in Security Problems

- Here, we discuss different ways of classifying the security problems, vis-a-vis the well-understood ML paradigms.
- the security problems

ML Paradigms in Security Problems

- Here, we discuss different ways of classifying the security problems, vis-a-vis the well-understood ML paradigms.
- the security problems
  - concern with the details about the attacker type, means of attack, and purpose of attack

ML Paradigms in Security Problems

- Here, we discuss different ways of classifying the security problems, vis-a-vis the well-understood ML paradigms.
- the security problems
    - concern with the details about the attacker type, means of attack, and purpose of attack
- the ML paradigms

ML Paradigms in Security Problems

- Here, we discuss different ways of classifying the security problems, vis-a-vis the well-understood ML paradigms.
- the security problems
  - concern with the details about the attacker type, means of attack, and purpose of attack
- the ML paradigms
  - concern with the details about the semi-supervised, the supervised, the unsupervised, the Human-in-the-loop(HITL) learning and the Game Theory(GT)-

ML Paradigms: The Security Problems

The attacker types

- Passive attackers: make no attempt to evade detections; their behaviors fit into descriptions of the threat models.

ML Paradigms: The Security Problems

The attacker types

- Passive attackers: make no attempt to evade detections; their behaviors fit into descriptions of the threat models.
- Semi-aggressive attackers: have knowledge of the detectors, and only attempt to evade detections.

The attacker types

- Passive attackers: make no attempt to evade detections; their behaviors fit into descriptions of the threat models.
- Semi-aggressive attackers: have knowledge of the detectors, and only attempt to evade detections.
- Active attackers

The attacker types

- Passive attackers: make no attempt to evade detections; their behaviors fit into descriptions of the threat models.
- Semi-aggressive attackers: have knowledge of the detectors, and only attempt to evade detections.
- Active attackers
    - these do not only have knowledge of the detectors and attempt to evade detection

ML Paradigms: The Security Problems

The attacker types

- Passive attackers: make no attempt to evade detections; their behaviors fit into descriptions of the threat models.
- Semi-aggressive attackers: have knowledge of the detectors, and only attempt to evade detections.
- Active attackers
  - these do not only have knowledge of the detectors and attempt to evade detection
  - but also actively try to poison, mislead, or thwart detection.

ML Paradigms: The Security Problems

The attacker types

- Passive attackers: make no attempt to evade detections; their behaviors fit into descriptions of the threat models.
- Semi-aggressive attackers: have knowledge of the detectors, and only attempt to evade detections.
- Active attackers
  - these do not only have knowledge of the detectors and attempt to evade detection
  - but also actively try to poison, mislead, or thwart detection.
- In addition, the attacks can be on either confidentiality, (availability), (integrity)

The ML paradigms for the above security problems

The ML paradigms

- Supervised learning: use labeled data for training.

The ML paradigms for the above security problems

The ML paradigms

- Supervised learning: use labeled data for training.
- Semi-supervised learning: use both labeled and unlabeled data for training.

The ML paradigms

- Supervised learning: use labeled data for training.
- Semi-supervised learning: use both labeled and unlabeled data for training.
- Unsupervised learning: has no labeled data available for training.

The ML paradigms for the above security problems

The ML paradigms

- Supervised learning: use labeled data for training.
- Semi-supervised learning: use both labeled and unlabeled data for training.
- Unsupervised learning: has no labeled data available for training.
- Human-in-the-loop(HITL) learning: incorporates active human feedback to algorithm's decisions into the knowledge base and/or algorithms.

The ML paradigms for the above security problems

The ML paradigms

- Supervised learning: use labeled data for training.
- Semi-supervised learning: use both labeled and unlabeled data for training.
- Unsupervised learning: has no labeled data available for training.
- Human-in-the-loop(HITL) learning: incorporates active human feedback to algorithm's decisions into the knowledge base and/or algorithms.
- Game Theory(GT)-based learning:

The ML paradigms for the above security problems

The ML paradigms

- Supervised learning: use labeled data for training.
- Semi-supervised learning: use both labeled and unlabeled data for training.
- Unsupervised learning: has no labeled data available for training.
- Human-in-the-loop(HITL) learning: incorporates active human feedback to algorithm's decisions into the knowledge base and/or algorithms.
- Game Theory(GT)-based learning:
  - considers learning as a series of strategic interactions between the model learner and actors with conflicting goals.

The ML paradigms for the above security problems

The ML paradigms

- Supervised learning: use labeled data for training.
- Semi-supervised learning: use both labeled and unlabeled data for training.
- Unsupervised learning: has no labeled data available for training.
- Human-in-the-loop(HITL) learning: incorporates active human feedback to algorithm's decisions into the knowledge base and/or algorithms.
- Game Theory(GT)-based learning:
  - considers learning as a series of strategic interactions between the model learner and actors with conflicting goals.
  - the actors can be data generators, feature generators, chaotic human actors, or a combination.

The ML paradigms for the security problems: Observations

Following are the general observations on the current state-of-the-art in ML security research

1. the majority of research in different security domains use supervised learning to deal with passive or semi-aggressive attackers.

The ML paradigms for the security problems: Observations

Following are the general observations on the current state-of-the-art in ML security research

1. the majority of research in different security domains use supervised learning to deal with passive or semi-aggressive attackers.
    - however, since the labeled data for training - is not always viable or easy to obtain - it is interesting research direction to explore semi-supervised and unsupervised learning approaches.

The ML paradigms for the security problems: Observations

Following are the general observations on the current state-of-the-art in ML security research

1. the majority of research in different security domains use supervised learning to deal with passive or semi-aggressive attackers.
   - however, since the labeled data for training - is not always viable or easy to obtain - it is interesting research direction to explore semi-supervised and unsupervised learning approaches.
2. in addition, many ML applications in security assume that training and testing data come from the same distribution.

The ML paradigms for the security problems: Observations

Following are the general observations on the current state-of-the-art in ML security research

1. the majority of research in different security domains use supervised learning to deal with passive or semi-aggressive attackers.
   - however, since the labeled data for training - is not always viable or easy to obtain - it is interesting research direction to explore semi-supervised and unsupervised learning approaches.
2. in addition, many ML applications in security assume that training and testing data come from the same distribution.
   - however, in the real world, it is highly unlikely that data are stationary.

The ML paradigms for the security problems: Observations

Following are the general observations on the current state-of-the-art in ML security research

1. the majority of research in different security domains use supervised learning to deal with passive or semi-aggressive attackers.
   - however, since the labeled data for training - is not always viable or easy to obtain - it is interesting research direction to explore semi-supervised and unsupervised learning approaches.
2. in addition, many ML applications in security assume that training and testing data come from the same distribution.
   - however, in the real world, it is highly unlikely that data are stationary.
   - in fact, the data could very well be generated by an adversarial data generator producing training and/or testing data sets

The ML paradigms for the security problems: Observations

Following are the general observations on the current state-of-the-art in ML security research

1. the majority of research in different security domains use supervised learning to deal with passive or semi-aggressive attackers.
   - however, since the labeled data for training - is not always viable or easy to obtain - it is interesting research direction to explore semi-supervised and unsupervised learning approaches.
2. in addition, many ML applications in security assume that training and testing data come from the same distribution.
   - however, in the real world, it is highly unlikely that data are stationary.
   - in fact, the data could very well be generated by an adversarial data generator producing training and/or testing data sets
   - hence, GT-based learning approaches and HITL learning system designs should be explored more

The ML paradigms for the security problems: Observations

Following are the general observations on the current state-of-the-art in ML security research

1. the majority of research in different security domains use supervised learning to deal with passive or semi-aggressive attackers.
   - however, since the labeled data for training - is not always viable or easy to obtain - it is interesting research direction to explore semi-supervised and unsupervised learning approaches.
2. in addition, many ML applications in security assume that training and testing data come from the same distribution.
   - however, in the real world, it is highly unlikely that data are stationary.
   - in fact, the data could very well be generated by an adversarial data generator producing training and/or testing data sets
   - hence, GT-based learning approaches and HITL learning system designs should be explored more
   - this helps design more efficient security defense mechanisms that could deal with active and unpredictable adversaries.

ML Applications in Security

We have now set the stage for overviewing the applications of ML in security.....

A Taxonomy of ML Applications in Security

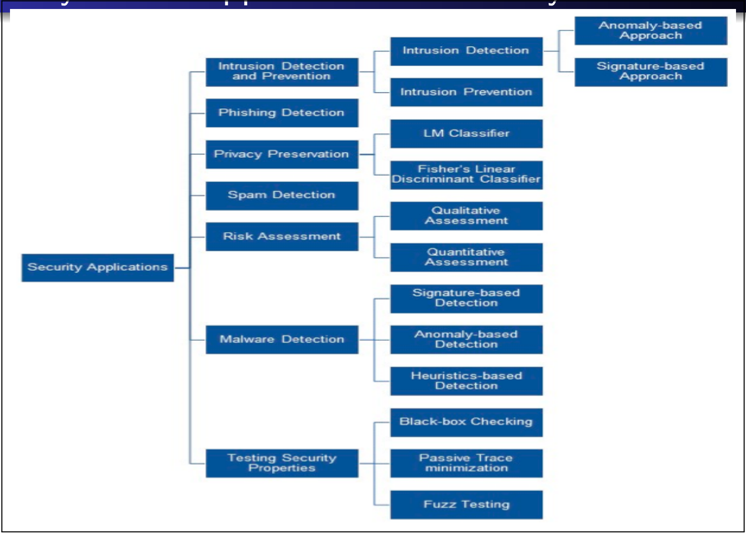

Figure: A Taxonomy of ML Applications in Security

Rutvij Jhaveri et al, 2019

ML Applications in Network Security

Machine Learning Applications in Network Security

We discuss use cases of ML applications in Network Security in the following areas:

- Machine Learning for Network Protection

Machine Learning Applications in Network Security

We discuss use cases of ML applications in Network Security in the following areas:

- Machine Learning for Network Protection
- Machine Learning for Endpoint Protection

Machine Learning Applications in Network Security

We discuss use cases of ML applications in Network Security in the following areas:

- Machine Learning for Network Protection
- Machine Learning for Endpoint Protection
- Machine Learning for Application Security

We discuss use cases of ML applications in Network Security in the following areas:

- Machine Learning for Network Protection
- Machine Learning for Endpoint Protection
- Machine Learning for Application Security
- Machine Learning for User Behavior Analytics

Machine Learning Applications in Network Security

We discuss use cases of ML applications in Network Security in the following areas:

- Machine Learning for Network Protection
- Machine Learning for Endpoint Protection
- Machine Learning for Application Security
- Machine Learning for User Behavior Analytics
- Machine Learning for Process Behavior Analytics

Machine Learning Applications in Network Security

We discuss use cases of ML applications in Network Security in the following areas:

- Machine Learning for Network Protection
- Machine Learning for Endpoint Protection
- Machine Learning for Application Security
- Machine Learning for User Behavior Analytics
- Machine Learning for Process Behavior Analytics
- Adversarial Machine Learning

ML Applications in Network Security

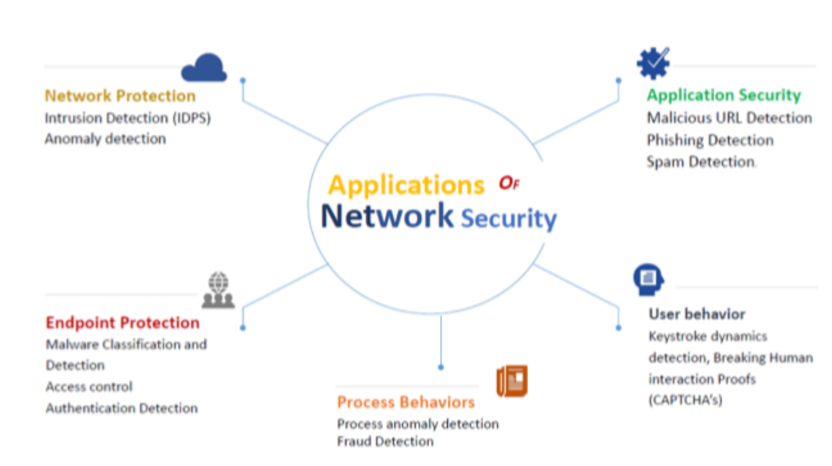

Figure: ML Applications in Network Security

1

[1] Olakunle Ibitoye et al, 2020

Machine Learning for Network Protection

- To monitor events dynamically in a computer network or system, the Intrusion Detection Systems (IDS) are applied.

Machine Learning for Network Protection

- To monitor events dynamically in a computer network or system, the Intrusion Detection Systems (IDS) are applied.
- typically, a preemptive approach that identifies potential threats with the help of ML classifiers and respond to prevent misuse.

Machine Learning for Network Protection

- To monitor events dynamically in a computer network or system, the Intrusion Detection Systems (IDS) are applied.
- typically, a preemptive approach that identifies potential threats with the help of ML classifiers and respond to prevent misuse.
- Two basic types of IDS viz. signature-based and anomaly-based

Network Intrusion Detection Systems

- The IDS is placed along the network boundary or between the network and the server.

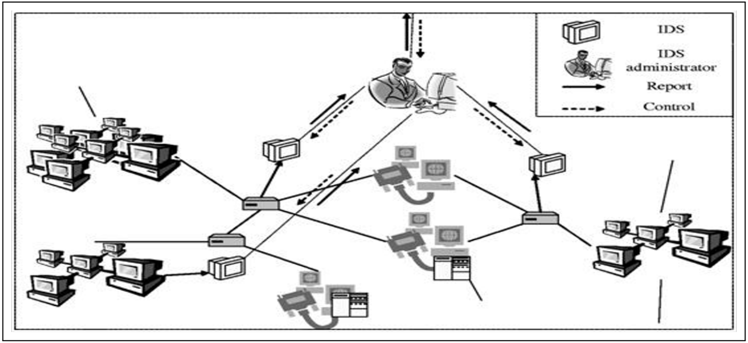

Figure: Network IDSs

1

[1]Ref: https://www.elprocus.com/basic-intrusion-detection-system/

Network Intrusion Detection Systems

- The IDS is placed along the network boundary or between the network and the server.
- the system monitors continuously the traffic on individual networks or subnets by comparing it with the known attacks in the library.

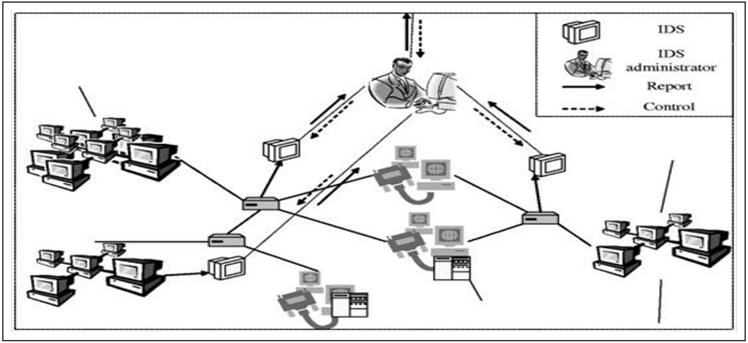

Figure: Network IDSs [1]

[1] Ref: https://www.elprocus.com/basic-intrusion-detection-system/

Host-based Intrusion Detection Systems

- work on individual operating systems where the incoming and outgoing of packets are constantly monitored and the auditing of system files is done...

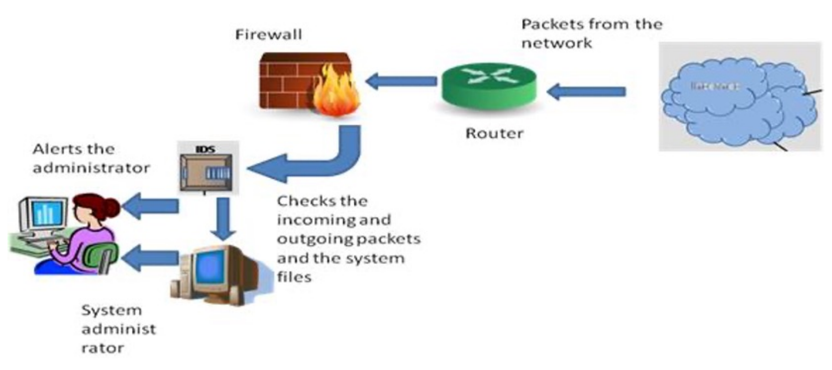

Figure: Host-based IDSs [1]

[1] Ref: https://www.elprocus.com/basic-intrusion-detection-system/

- simply detects the kind of malware operation and issues an alert to the system or network administrator

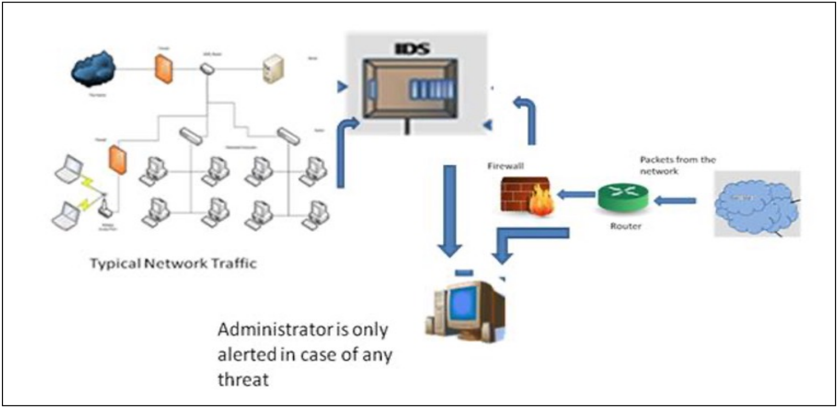

Figure: Passive IDSs

[1] Ref: https://www.elprocus.com/basic-intrusion-detection-system/

Signature-based IDSs

- typically detect attacks based on the repository of attacks signatures

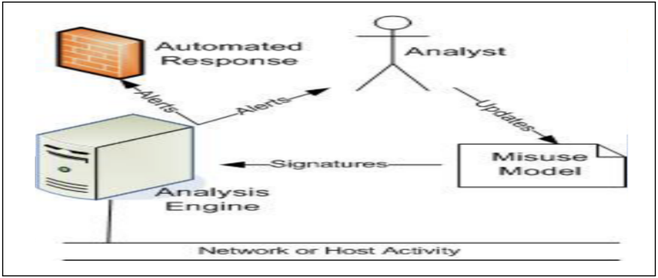

Figure: Signature based IDSs

[1] Ref: Gilbert R. Hendry, Applicability of Clustering to Cyber Intrusion Detection, RIT, NY, USA

Signature-based Intrusion Detection System

Signature-based IDSs

- typically detect attacks based on the repository of attacks signatures
- identifying specific patterns such as malicious instruction sequences or byte sequences - i.e. the signatures

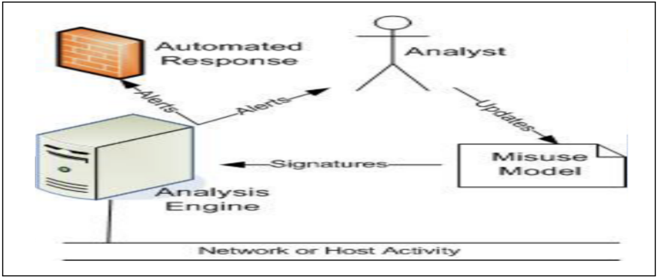

Figure: Signature based IDSs

[1] Ref: Gilbert R. Hendry, Applicability of Clustering to Cyber Intrusion Detection, RIT, NY, USA

Signature-based IDSs

- typically detect attacks based on the repository of attacks signatures
- identifying specific patterns such as malicious instruction sequences or byte sequences - i.e. the signatures
- do not give any false alarm.

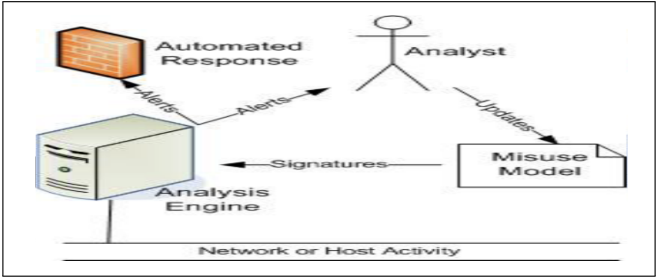

Figure: Signature based IDSs

[1] Ref: Gilbert R. Hendry, Applicability of Clustering to Cyber Intrusion Detection, RIT, NY, USA

Signature-based Intrusion Detection System

Signature-based IDSs

- typically detect attacks based on the repository of attacks signatures
- identifying specific patterns such as malicious instruction sequences or byte sequences - i.e. the signatures
- do not give any false alarm.
- downside: zero-day attacks can easily bypass signature-based IDS.

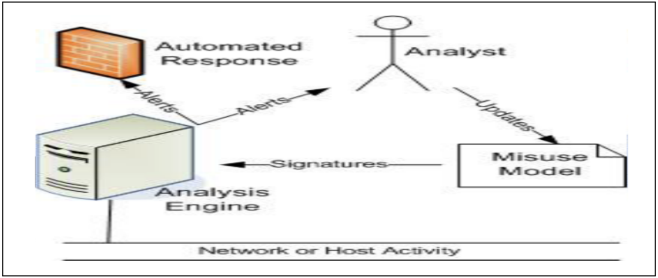

Figure: Signature based IDSs

[1] Ref: Gilbert R. Hendry, Applicability of Clustering to Cyber Intrusion Detection, RIT, NY, USA

Signature-based Intrusion Detection Systems...

Signature-based IDSs ...Common signatures used by such IDSs are as follows:

- Connection attempt from a reserved IP address

.....continued

Signature-based IDSs ...Common signatures used by such IDSs are as follows:

- Connection attempt from a reserved IP address
  - is easily identified by checking the source address field in an IP header.

.....*continued*

Signature-based Intrusion Detection Systems...

Signature-based IDSs ...Common signatures used by such IDSs are as follows:

- Connection attempt from a reserved IP address
  - is easily identified by checking the source address field in an IP header.
- Packet with an illegal TCP flag combination.

.....*continued*

Signature-based Intrusion Detection Systems...

Signature-based IDSs ...Common signatures used by such IDSs are as follows:

- Connection attempt from a reserved IP address
  - is easily identified by checking the source address field in an IP header.
- Packet with an illegal TCP flag combination.
  - can be found by comparing the flags set in a TCP header against known good or bad flag combinations.

.....*continued*

Signature-based Intrusion Detection Systems...

Signature-based IDSs ...Common signatures used by such IDSs are as follows:

- Connection attempt from a reserved IP address
  - is easily identified by checking the source address field in an IP header.
- Packet with an illegal TCP flag combination.
  - can be found by comparing the flags set in a TCP header against known good or bad flag combinations.
- Email containing a particular virus.

.....continued

Signature-based Intrusion Detection Systems...

Signature-based IDSs ...Common signatures used by such IDSs are as follows:

- Connection attempt from a reserved IP address
  - is easily identified by checking the source address field in an IP header.
- Packet with an illegal TCP flag combination.
  - can be found by comparing the flags set in a TCP header against known good or bad flag combinations.
- Email containing a particular virus.
  - the IDS can compare the subject of each email to the subject associated with the virus-laden email, OR

.....*continued*

Signature-based Intrusion Detection Systems...

Signature-based IDSs ...Common signatures used by such IDSs are as follows:

- Connection attempt from a reserved IP address
  - is easily identified by checking the source address field in an IP header.
- Packet with an illegal TCP flag combination.
  - can be found by comparing the flags set in a TCP header against known good or bad flag combinations.
- Email containing a particular virus.
  - the IDS can compare the subject of each email to the subject associated with the virus-laden email, OR
  - it can look for an attachment with a particular name.

.....continued

Signature-based Intrusion Detection Systems...

Signature-based IDSs ...Common signatures used by such IDSs are as follows:

- Connection attempt from a reserved IP address
  - is easily identified by checking the source address field in an IP header.
- Packet with an illegal TCP flag combination.
  - can be found by comparing the flags set in a TCP header against known good or bad flag combinations.
- Email containing a particular virus.
  - the IDS can compare the subject of each email to the subject associated with the virus-laden email, OR
  - it can look for an attachment with a particular name.
- DNS buffer overflow attempt contained in the payload of a query.

.....continued

Signature-based Intrusion Detection Systems...

Signature-based IDSs ...Common signatures used by such IDSs are as follows:

- Connection attempt from a reserved IP address
  - is easily identified by checking the source address field in an IP header.
- Packet with an illegal TCP flag combination.
  - can be found by comparing the flags set in a TCP header against known good or bad flag combinations.
- Email containing a particular virus.
  - the IDS can compare the subject of each email to the subject associated with the virus-laden email, OR
  - it can look for an attachment with a particular name.
- DNS buffer overflow attempt contained in the payload of a query.
  - By parsing the DNS fields and checking the length of each of them, the IDS can identify an attempt to perform a buffer overflow using a DNS field.

.....continued

Signature-based Intrusion Detection Systems...

Signature-based IDSs ...Common signatures used by such IDSs are as follows:

- Connection attempt from a reserved IP address
  - is easily identified by checking the source address field in an IP header.
- Packet with an illegal TCP flag combination.
  - can be found by comparing the flags set in a TCP header against known good or bad flag combinations.
- Email containing a particular virus.
  - the IDS can compare the subject of each email to the subject associated with the virus-laden email, OR
  - it can look for an attachment with a particular name.
- DNS buffer overflow attempt contained in the payload of a query.
  - By parsing the DNS fields and checking the length of each of them, the IDS can identify an attempt to perform a buffer overflow using a DNS field.
  - Another method is to lookfor exploit shell code sequences in the payload.

.....*continued*

Signature-based Intrusion Detection Systems...

Signature-based IDSs ...Common signatures used by such IDSs are as follows:

- Denial of service attack on a POP3 server caused by issuing the same command thousands of times.

Signature-based Intrusion Detection Systems...

Signature-based IDSs ...Common signatures used by such IDSs are as follows:

- Denial of service attack on a POP3 server caused by issuing the same command thousands of times.
    - what could be the probable signature for detection, in this attack ?

Signature-based Intrusion Detection Systems...

Signature-based IDSs ...Common signatures used by such IDSs are as follows:

- Denial of service attack on a POP3 server caused by issuing the same command thousands of times.
  - what could be the probable signature for detection, in this attack ?
    - to keep track of how many times the command is issued and to alert when that number exceeds a certain threshold.

Signature-based Intrusion Detection Systems...

Signature-based IDSs ...Common signatures used by such IDSs are as follows:

- Denial of service attack on a POP3 server caused by issuing the same command thousands of times.
  - what could be the probable signature for detection, in this attack ?
    - to keep track of how many times the command is issued and to alert when that number exceeds a certain threshold.
- File access attack on an FTP server - orchestrated by issuing file and directory commands to it, without first logging in.

Signature-based Intrusion Detection Systems...

Signature-based IDSs ...Common signatures used by such IDSs are as follows:

- Denial of service attack on a POP3 server caused by issuing the same command thousands of times.
  - what could be the probable signature for detection, in this attack ?
    - to keep track of how many times the command is issued and to alert when that number exceeds a certain threshold.
- File access attack on an FTP server - orchestrated by issuing file and directory commands to it, without first logging in.
  - what could be the probable signature for detection, in this attack ?

Signature-based Intrusion Detection Systems...

Signature-based IDSs ...Common signatures used by such IDSs are as follows:

- Denial of service attack on a POP3 server caused by issuing the same command thousands of times.
  - what could be the probable signature for detection, in this attack ?
    - to keep track of how many times the command is issued and to alert when that number exceeds a certain threshold.
- File access attack on an FTP server - orchestrated by issuing file and directory commands to it, without first logging in.
  - what could be the probable signature for detection, in this attack ?
    - a state-tracking signature could be developed which would monitor FTP traffic for a successful login

Signature-based Intrusion Detection Systems...

Signature-based IDSs ...Common signatures used by such IDSs are as follows:

- Denial of service attack on a POP3 server caused by issuing the same command thousands of times.
  - what could be the probable signature for detection, in this attack ?
    - to keep track of how many times the command is issued and to alert when that number exceeds a certain threshold.
- File access attack on an FTP server - orchestrated by issuing file and directory commands to it, without first logging in.
  - what could be the probable signature for detection, in this attack ?
    - a state-tracking signature could be developed which would monitor FTP traffic for a successful login
    - it would alert if certain commands were issued before the user had authenticated properly.

- Anomaly-based IDSs - misuse detection IDSs

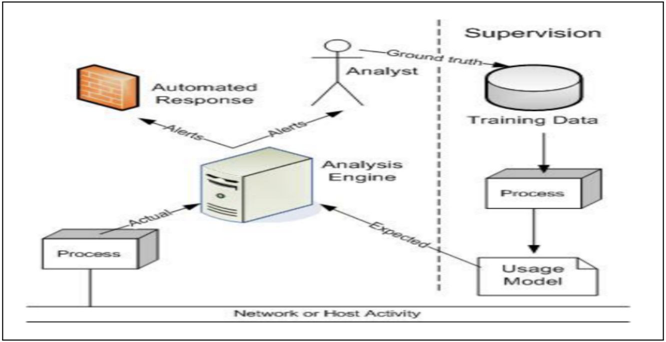

Figure: Anomaly based IDSs

1

[1] Ref: Gilbert R. Hendry, Applicability of Clustering to Cyber Intrusion Detection, RIT, NY, USA

- Anomaly-based IDSs - misuse detection IDSs
  - detect misuse in a network with the help of ML-based classifiers.

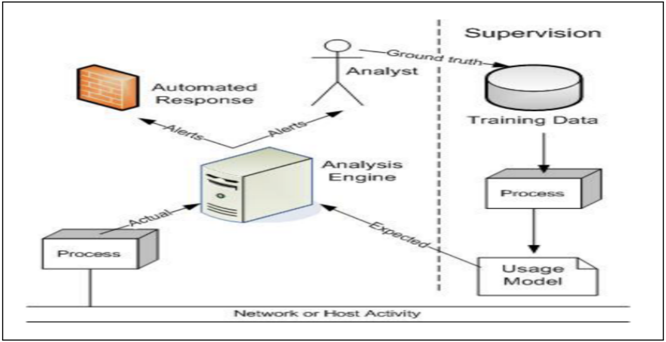

Figure: Anomaly based IDSs

1

[1] Ref: Gilbert R. Hendry, Applicability of Clustering to Cyber Intrusion Detection, RIT, NY, USA

- Anomaly-based IDSs - misuse detection IDSs
  - detect misuse in a network with the help of ML-based classifiers.
  - use ML to detect zero-day attacks and anomalies, too.

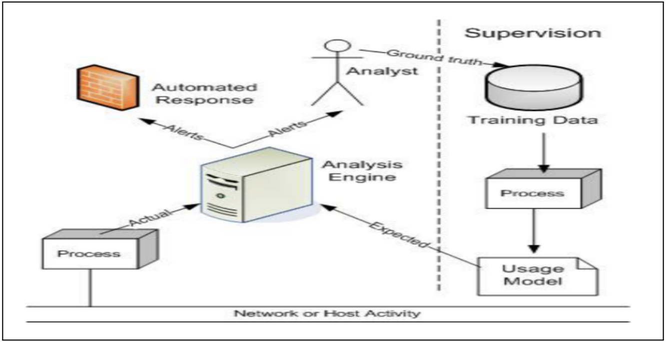

Figure: Anomaly based IDSs

1

[1] Ref: Gilbert R. Hendry, Applicability of Clustering to Cyber Intrusion Detection, RIT, NY, USA

- Anomaly-based IDSs - misuse detection IDSs
  - detect misuse in a network with the help of ML-based classifiers.
  - use ML to detect zero-day attacks and anomalies, too.
  - but, has the tendency to generate a significant number of false positives.

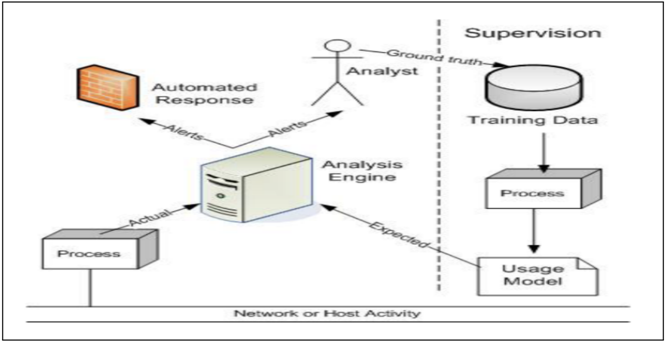

Figure: Anomaly based IDSs

1

[1] Ref: Gilbert R. Hendry, Applicability of Clustering to Cyber Intrusion Detection, RIT, NY, USA

ML Paradigms for Endpoint Protection

Machine Learning for Endpoint Protection

The main uses cases under this category are as follows:

- Malware detection

Machine Learning for Endpoint Protection

The main uses cases under this category are as follows:

- Malware detection
- Automatic Analysis of Malware Behavior

Machine Learning for Endpoint Protection

The main uses cases under this category are as follows:

- Malware detection
- Automatic Analysis of Malware Behavior

Machine Learning for Endpoint Protection

The main uses cases under this category are as follows:

- Malware detection
- Automatic Analysis of Malware Behavior
- Malware detection includes detection on varied platforms viz. workstations, servers, cloud instances, and mobile devices.

Machine Learning for Endpoint Protection

The main uses cases under this category are as follows:

- Malware detection
- Automatic Analysis of Malware Behavior
- Malware detection includes detection on varied platforms viz. workstations, servers, cloud instances, and mobile devices.
    - the goal is to detect and identify malicious activities caused by malware.

Machine Learning for Endpoint Protection

The main uses cases under this category are as follows:

- Malware detection
- Automatic Analysis of Malware Behavior
- Malware detection includes detection on varied platforms viz. workstations, servers, cloud instances, and mobile devices.
  - the goal is to detect and identify malicious activities caused by malware.
  - the common conventional approach used for MD is signature-based malware detection.

Machine Learning for Endpoint Protection

The main uses cases under this category are as follows:

- Malware detection
- Automatic Analysis of Malware Behavior
- Malware detection includes detection on varied platforms viz. workstations, servers, cloud instances, and mobile devices.
  - the goal is to detect and identify malicious activities caused by malware.
  - the common conventional approach used for MD is signature-based malware detection.
  - however, ML can cope with this increase and can detect and discover underlying patterns.

Machine Learning for Endpoint Protection

The main uses cases under this category are as follows:

- Malware detection
- Automatic Analysis of Malware Behavior
- Malware detection includes detection on varied platforms viz. workstations, servers, cloud instances, and mobile devices.
    - the goal is to detect and identify malicious activities caused by malware.
    - the common conventional approach used for MD is signature-based malware detection.
    - however, ML can cope with this increase and can detect and discover underlying patterns.
- Why use ML for malware detection ?

Machine Learning for Endpoint Protection

The main uses cases under this category are as follows:

- Malware detection
- Automatic Analysis of Malware Behavior
- Malware detection includes detection on varied platforms viz. workstations, servers, cloud instances, and mobile devices.
  - the goal is to detect and identify malicious activities caused by malware.
  - the common conventional approach used for MD is signature-based malware detection.
  - however, ML can cope with this increase and can detect and discover underlying patterns.
- Why use ML for malware detection ?
  - dealing with unknown malware or zero-day malware

Machine Learning for Endpoint Protection

The main uses cases under this category are as follows:

- Malware detection
- Automatic Analysis of Malware Behavior
- Malware detection includes detection on varied platforms viz. workstations, servers, cloud instances, and mobile devices.
  - the goal is to detect and identify malicious activities caused by malware.
  - the common conventional approach used for MD is signature-based malware detection.
  - however, ML can cope with this increase and can detect and discover underlying patterns.
- Why use ML for malware detection ?
  - dealing with unknown malware or zero-day malware
  - with the increase in the variety of malware activities, the need for automatic detection

Machine Learning for Endpoint Protection...

Some of the common research excursions proposed include

- using clustering similar malware behaviors into classes,

Machine Learning for Endpoint Protection...

Some of the common research excursions proposed include

- using clustering similar malware behaviors into classes,
- using different machine learning algorithms for classification such as

Machine Learning for Endpoint Protection...

Some of the common research excursions proposed include

- using clustering similar malware behaviors into classes,
- using different machine learning algorithms for classification such as
    - Logistic Regression e.g. to predict the next system call for executable process and compare it with real ones. Why ?

Machine Learning for Endpoint Protection...

Some of the common research excursions proposed include

- using clustering similar malware behaviors into classes,
- using different machine learning algorithms for classification such as
  - Logistic Regression e.g. to predict the next system call for executable process and compare it with real ones. Why ?
  - Random Forest, Naive Bayes, Random Tree, Sequential Minimal Optimization (SMO) - to classify programs into malware, OR spyware OR ransomware etc.

Machine Learning for Endpoint Protection...

Some of the common research excursions proposed include

- using clustering similar malware behaviors into classes,
- using different machine learning algorithms for classification such as
  - Logistic Regression e.g. to predict the next system call for executable process and compare it with real ones. Why ?
  - Random Forest, Naive Bayes, Random Tree, Sequential Minimal Optimization (SMO) - to classify programs into malware, OR spyware OR ransomware etc.
- Classification of newly retrieved malware samples into a predefined set of malware classes by inspecting system call sequences

Machine Learning for Endpoint Protection...

Some of the common research excursions proposed include

- using clustering similar malware behaviors into classes,
- using different machine learning algorithms for classification such as
  - Logistic Regression e.g. to predict the next system call for executable process and compare it with real ones. Why ?
  - Random Forest, Naive Bayes, Random Tree, Sequential Minimal Optimization (SMO) - to classify programs into malware, OR spyware OR ransomware etc.
- Classification of newly retrieved malware samples into a predefined set of malware classes by inspecting system call sequences
- ... ... and so on.

ML Paradigms for Applications Security

The main uses cases under this category are as follows:

- malicious web attack detection,

Machine Learning for Application Security

The main uses cases under this category are as follows:

- malicious web attack detection,
- phishing detection,

Machine Learning for Application Security

The main uses cases under this category are as follows:

- malicious web attack detection,
- phishing detection,
- spam detection,

Machine Learning for Application Security

The main uses cases under this category are as follows:

- malicious web attack detection,
- phishing detection,
- spam detection,
- malicious URL detection

Machine Learning for Application Security

The main uses cases under this category are as follows:

- malicious web attack detection,
- phishing detection,
- spam detection,
- malicious URL detection

Machine Learning for Application Security

The main uses cases under this category are as follows:

- malicious web attack detection,
- phishing detection,
- spam detection,
- malicious URL detection

The ML tasks involved are

- Classification of phishing emails, and clustering

The main uses cases under this category are as follows:

- malicious web attack detection,
- phishing detection,
- spam detection,
- malicious URL detection

The ML tasks involved are

- Classification of phishing emails, and clustering
- Using Support Vector Machines (SVM), Leave One Model Out, Biased SVM, Neural Networks, Self Organizing Maps (SOMs) and K-Means for clustering,

Machine Learning for Application Security

The main uses cases under this category are as follows:

- malicious web attack detection,
- phishing detection,
- spam detection,
- malicious URL detection

The ML tasks involved are

- Classification of phishing emails, and clustering
- Using Support Vector Machines (SVM), Leave One Model Out, Biased SVM, Neural Networks, Self Organizing Maps (SOMs) and K-Means for clustering,
- Adaptively Detecting Malicious Queries in Web Attacks,

Machine Learning for Application Security

The main uses cases under this category are as follows:

- malicious web attack detection,
- phishing detection,
- spam detection,
- malicious URL detection

The ML tasks involved are

- Classification of phishing emails, and clustering
- Using Support Vector Machines (SVM), Leave One Model Out, Biased SVM, Neural Networks, Self Organizing Maps (SOMs) and K-Means for clustering,
- Adaptively Detecting Malicious Queries in Web Attacks,
- Learning a URL Representation for Malicious URL Detection

Machine Learning for Application Security

The main uses cases under this category are as follows:

- malicious web attack detection,
- phishing detection,
- spam detection,
- malicious URL detection

The ML tasks involved are

- Classification of phishing emails, and clustering
- Using Support Vector Machines (SVM), Leave One Model Out, Biased SVM, Neural Networks, Self Organizing Maps (SOMs) and K-Means for clustering,
- Adaptively Detecting Malicious Queries in Web Attacks,
- Learning a URL Representation for Malicious URL Detection
- ... ... and so on.

ML Paradigms for Behavioral Analytics

User behavior analytics

- is a cybersecurity process that involves analyzing patterns in human behaviour

Machine Learning for User Behavior Analytics

User behavior analytics

- is a cybersecurity process that involves analyzing patterns in human behaviour
- aim is to detect anomalies that could be an indication of fraudulent activities or insider threats.

Machine Learning for User Behavior Analytics

User behavior analytics

- is a cybersecurity process that involves analyzing patterns in human behaviour
- aim is to detect anomalies that could be an indication of fraudulent activities or insider threats.
- anomalies are found in user actions such as unusual login tries, keystrokes pressed etc

Machine Learning for User Behavior Analytics

User behavior analytics

- is a cybersecurity process that involves analyzing patterns in human behaviour
- aim is to detect anomalies that could be an indication of fraudulent activities or insider threats.
- anomalies are found in user actions such as unusual login tries, keystrokes pressed etc
- The research attempts that have been made include the following

Machine Learning for User Behavior Analytics

User behavior analytics

- is a cybersecurity process that involves analyzing patterns in human behaviour
- aim is to detect anomalies that could be an indication of fraudulent activities or insider threats.
- anomalies are found in user actions such as unusual login tries, keystrokes pressed etc
- The research attempts that have been made include the following
    - Authentication with Keystroke Dynamics

Machine Learning for User Behavior Analytics

User behavior analytics

- is a cybersecurity process that involves analyzing patterns in human behaviour
- aim is to detect anomalies that could be an indication of fraudulent activities or insider threats.
- anomalies are found in user actions such as unusual login tries, keystrokes pressed etc
- The research attempts that have been made include the following
  - Authentication with Keystroke Dynamics
    - a Probabilistic Neural Network (PNN) was used to capture the keystroke dynamics and thereby the typing style of a user

User behavior analytics

- is a cybersecurity process that involves analyzing patterns in human behaviour
- aim is to detect anomalies that could be an indication of fraudulent activities or insider threats.
- anomalies are found in user actions such as unusual login tries, keystrokes pressed etc
- The research attempts that have been made include the following
    - Authentication with Keystroke Dynamics
        - a Probabilistic Neural Network (PNN) was used to capture the keystroke dynamics and thereby the typing style of a user
        - eight attributes were used to monitor the enrollment and authentication attempts by a user.

Machine Learning for User Behavior Analytics

User behavior analytics

- is a cybersecurity process that involves analyzing patterns in human behaviour
- aim is to detect anomalies that could be an indication of fraudulent activities or insider threats.
- anomalies are found in user actions such as unusual login tries, keystrokes pressed etc
- The research attempts that have been made include the following
  - Authentication with Keystroke Dynamics
    - a Probabilistic Neural Network (PNN) was used to capture the keystroke dynamics and thereby the typing style of a user
    - eight attributes were used to monitor the enrollment and authentication attempts by a user.
    - an accuracy of 90% was obtained in classifying legitimate users from imposters

Machine Learning for User Behavior Analytics

User behavior analytics

- is a cybersecurity process that involves analyzing patterns in human behaviour
- aim is to detect anomalies that could be an indication of fraudulent activities or insider threats.
- anomalies are found in user actions such as unusual login tries, keystrokes pressed etc
- The research attempts that have been made include the following
  - Authentication with Keystroke Dynamics
    - a Probabilistic Neural Network (PNN) was used to capture the keystroke dynamics and thereby the typing style of a user
    - eight attributes were used to monitor the enrollment and authentication attempts by a user.
    - an accuracy of 90% was obtained in classifying legitimate users from imposters
    - a comparison of the training time between the PNN system and a Multi-Layer Perception Neural Network (MLPNN) showed that the PNN was four times faster.

Machine Learning for User Behavior Analytics

User behavior analytics: a few of the research attempts ...

- Text-based CAPTCHA Strengths and Weaknesses

User behavior analytics: a few of the research attempts ...

- Text-based CAPTCHA Strengths and Weaknesses
  - it was noted in a study that several well known websites still implemented CAPTCHA technologies that have been proven to be vulnerable to cyber attacks

Machine Learning for User Behavior Analytics

User behavior analytics: a few of the research attempts ...

- Text-based CAPTCHA Strengths and Weaknesses
  - it was noted in a study that several well known websites still implemented CAPTCHA technologies that have been proven to be vulnerable to cyber attacks
  - it was shown that only Google and Recaptcha were able to resist to the automated attack.

Machine Learning for User Behavior Analytics

User behavior analytics: a few of the research attempts ...

- Text-based CAPTCHA Strengths and Weaknesses
  - it was noted in a study that several well known websites still implemented CAPTCHA technologies that have been proven to be vulnerable to cyber attacks
  - it was shown that only Google and Recaptcha were able to resist to the automated attack.
  - the study also revealed the need for more robust CAPTCHA designs in most of the widely used schemes.

Some of the other interesting research excursions in User Behavior Analytics are as follows:

- social network (especially on Twitter & MySpace) spam detection that gathers legitimate and spam profiles and feeds them to Support Vector Machine (SVM) model to identify spam.

Machine Learning for User Behavior Analytics...

Some of the other interesting research excursions in User Behavior Analytics are as follows:

- social network (especially on Twitter & MySpace) spam detection that gathers legitimate and spam profiles and feeds them to Support Vector Machine (SVM) model to identify spam.
- determining fraudulent transactions within banking systems, identifying outliers, classifying types of fraud and for clustering various business processes.

Machine Learning for User Behavior Analytics...

Some of the other interesting research excursions in User Behavior Analytics are as follows:

- social network (especially on Twitter & MySpace) spam detection that gathers legitimate and spam profiles and feeds them to Support Vector Machine (SVM) model to identify spam.
- determining fraudulent transactions within banking systems, identifying outliers, classifying types of fraud and for clustering various business processes.
- detecting most of the attacks on Industrial Control System (ICS) using CNN

Machine Learning for User Behavior Analytics...

Some of the other interesting research excursions in User Behavior Analytics are as follows:

- social network (especially on Twitter & MySpace) spam detection that gathers legitimate and spam profiles and feeds them to Support Vector Machine (SVM) model to identify spam.
- determining fraudulent transactions within banking systems, identifying outliers, classifying types of fraud and for clustering various business processes.
- detecting most of the attacks on Industrial Control System (ICS) using CNN
- using Deep Learning techniques for Side-Channel Analysis - a type of attack that attempts to leak information from a system by exploiting some parameters from the physical environment

Machine Learning for User Behavior Analytics...

Some of the other interesting research excursions in User Behavior Analytics are as follows:

- social network (especially on Twitter & MySpace) spam detection that gathers legitimate and spam profiles and feeds them to Support Vector Machine (SVM) model to identify spam.
- determining fraudulent transactions within banking systems, identifying outliers, classifying types of fraud and for clustering various business processes.
- detecting most of the attacks on Industrial Control System (ICS) using CNN
- using Deep Learning techniques for Side-Channel Analysis - a type of attack that attempts to leak information from a system by exploiting some parameters from the physical environment
- detecting credit card frauds

Machine Learning for User Behavior Analytics...

Some of the other interesting research excursions in User Behavior Analytics are as follows:

- social network (especially on Twitter & MySpace) spam detection that gathers legitimate and spam profiles and feeds them to Support Vector Machine (SVM) model to identify spam.
- determining fraudulent transactions within banking systems, identifying outliers, classifying types of fraud and for clustering various business processes.
- detecting most of the attacks on Industrial Control System (ICS) using CNN
- using Deep Learning techniques for Side-Channel Analysis - a type of attack that attempts to leak information from a system by exploiting some parameters from the physical environment
- detecting credit card frauds
- ... ... and so on.

Next ML for Privacy Preservation

B l a n k

Blank

B l a n k

B l a n k

B l a n k