Today’s agenda

1) Introduction to the research area

2) Seminar instructions
   a) Deliverables
   b) Grading
   c) FAQ
Introduction to the research area
Learning in adversarial environment

- Problem considered in the research community at least since early 2000s
- With the hype over machine learning (deep learning) the problem gains importance
- Adversarial perturbations studied in vision, text, malware...
Adversarial attacks

- Attack goals
  - Destroy system functionality
  - Change the model output
  - Reveal confidential information

- Attacker knowledge
  - Perfect Knowledge (white-box attack)
  - Limited Knowledge (gray-box attack)
  - Zero Knowledge (black-box attack)

- Attacker capability
  - Make queries to the model
  - Change training data
  - Change validation/test data
Adversarial attacks

● Evasion attacks - modify test input to get a desired classifier decision

● Poisoning attacks - modify training data to change the resulting model (either just make it less accurate or enable particular properties (backdoor))

● Privacy attacks - learn about sensitive data from the model based on queries
Adversarial attacks

- CIA Triad - from classical Infosec

Confidentiality  Integrity  Availability
Attack and defense

● Attack optimization -> find minimal perturbation that achieves the goal
  
  \((\min g(x) , \text{s.t.} \ d(x, x') < d_{\text{max}} (x, x'))\)

● Defense optimization -> find an approach that either:
  ◦ Rejects the attack points as outliers or is
  ◦ Robust to attack points
Defense strategies

- Kerckhoffs principle - do not rely on obscurity
- Assess security against various levels of adversary’s knowledge and capability
- Arms race
Adversarial Attack - discoveries

Roli, Biggio, “Wild Patterns: Half-day Tutorial on Adversarial Machine Learning”, IJCAI
Adversarial Learning

- Attacked methods
  - Linear regression
  - SVM
  - Neural Networks
  - Random Forests
  - ...

Adversarial Learning

- Data under modification
  - Computer Vision
  - Natural Language Processing
  - Speech
  - Malware
  - ...

Vision papers best known...


But also other domains...


Windows Defender: Rise of the machine (learning)

Windows Defender harnesses the power of machine learning, contributing to making Windows 10 Microsoft’s most secure client operating system and providing increased protection against security threats facing consumers and commercial enterprises today.
Computer malware can often evade antivirus security software if the author changes a few lines of code or designs the program to automatically mutate before each new infection.

Artificial neural networks, trained to recognize the characteristics of malicious code by looking at millions of examples of malware and normal malware files, could perhaps offer a far better way to catch such nefarious code. An approach known as deep learning, which involves
Some current work at Chair I20

- Generating adversarial binaries to evade neural network malware detectors based on raw code
Seminar instructions
Our goal

- Get an **overview** of the academic state-of-the-art
- Extend your **knowledge** in machine learning, learn to look at it from the **security** standpoint
- Get a feeling on **how to evaluate risk**
What to deliver?

- 1 presentation
  - 45 minutes + 15 minutes of discussion
  - Present the topic, use all papers

- 1 report
  - 14 Pages LNCS (look up LNCS template in Latex)
Presentation (start on 14.05.)
Presentation

Needs to be:

- Correct
- Complete
- Comprehensible
Presentation - Correct

- Present information from the paper correctly
- Don’t speculate without a reason or proof
- Don’t claim something you cannot explain well
Presentation - Complete

- Explain all key points of the papers

- Be careful about time constraints and distribution

- Convey information without leaving out important insight
Presentation - Comprehensible

- Speak loud and clear
- Think about the audience - fellow students
- Motivate the audience for discussion
- Don’t fight your audience, answer all questions friendly
A PPT is just a presentation aid. It should not be a paper in its own right and your bullet points should be, if possible, less than a line long. Specifically, keep bullet points short by making them clear and concise. Do not be afraid from using incomplete sentences or phrases. In reality, this is the preferred method because it helps to highlight the points you are making during your talk.

This is because having lots of text on your slides makes it difficult to understand the point you are trying to make. Furthermore, your audience will end up reading the text and ignoring you.
Presentation - Emphasis

- ONLY USE CAPITALIZATION WHEN NEEDED
- Use color sparingly
- Only bold key words and phrases
- Light text on a dark background is bad
wow
such color
much text
not comparable
meaning not clear
WHAT DOES THIS MEAN?
Presentation - Structure

- Introduction to the topic
- Present all papers together
  - Introduction
  - Main Points
  - Back up arguments
  - Conclusions (key takeaways)
Presentation - Audience

- Read papers, or at least abstracts, prior to each presentation day
- Listen carefully, write down questions
- Ask questions, comment
- Active participation is appreciated!
Presentation - Grading

● Presentation skills
  ○ General organization, use of slides
  ○ Language, slide text and graphics
  ○ Pace, use of time

● Subject-related competence
  ○ Subject knowledge
  ○ Staying on topic
  ○ Identifying interesting/important points
Report (deadline 09.06.2019.)
Report

- 14 Pages LNCS
- Summarize key points of all papers together - not an easy task
- Use a typical paper structure:

Abstract -> Introduction -> Methodology -> Results -> Discussion -> Conclusion
Report - Abstract

- Summarize the paper
  - Introduction to the problem
  - How was the problem solved? Methodology
  - Short insight in the results
  - What is the impact of the papers?
Report - Introduction

- Describe the context
- What is the preexisting work?
- What does the preexisting work lack?
- How do the papers close the gap?
Report - Methodology

- Describe the mechanisms used to tackle the existing problem
- Lead the reader through the problem solving procedures
- Give arguments for the choice of methods
Report - Results

- Give an overview of the important results
- Add tables, graphs... if you have space
- Shortly comment on the figures
- Avoid phrases like: It is obvious from this graph that...
Report - Discussion

- What do the results actually tell us?
- Compare the results with related work
- What are the limitations of the paper results?
- How can the limitations be addressed?
Report - Conclusion

- Summary of the main findings in 3-4 sentences
- What are the most interesting results?
- What is the impact of the papers?
Report - Grading

- Paper organization
- Language and grammar
- Subject knowledge
- Ability to summarize
- Proper bibliography and citations (!)
How to do your research

- Seminar - (kind of) simulation of scientific research
- Try to be independent, but also ask questions
FAQ

● Allowed to miss a presentation day? Yes, if you have a very good reason.
  ○ Examples of good reason: health issues, schedule clashes at the Uni
  ○ Examples of bad reason: HiWi work, homework, football training, bad mood

● Can I set a meeting if I have problems with my papers?
  ○ Yes, but try to do as much as you can yourself.