

Introduction for Seminar: Intrusion Detection Systems

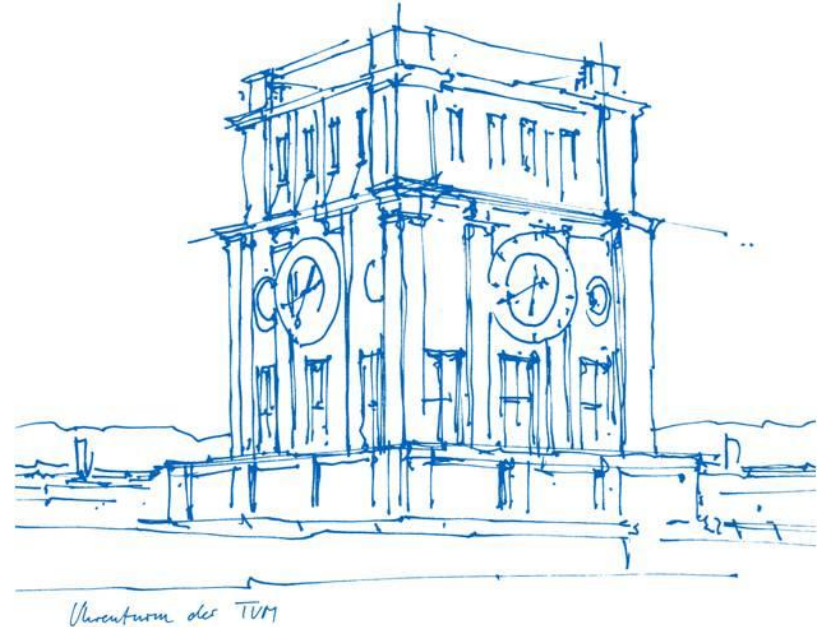
Mohammad Reza Norouzian

Technische Universität München

Fakultät für Informatik

Lehrstuhl für IT Sicherheit

17.04.2018



Outline



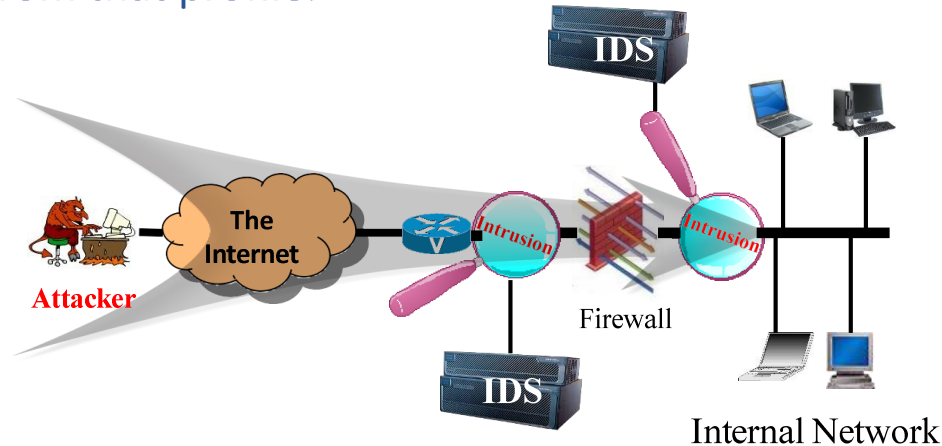
- Introduction to research area
- Student assignments
- Grading
- Time table
- Presentation guidance
- Report guidance
- FAQ

What's an Intrusion?

- Successful attack is usually (but not always) associated with an **access control violation**
 - A buffer overflow has been exploited, and now attack code is being executed inside a legitimate program
 - Outsider gained access to a protected resource
 - A program or file has been modified
 - System is not behaving “as it should”
- The goal of an intrusion detection system (IDS) is to detect that bad things are happening (intrusion)
 - Just as they start happening (hope so)
 - How is this different from a firewall?

Intrusion detection styles

- Misuse detection: precise descriptions of known malicious behavior.
- Anomaly detection: have a notion of normal activity and flag deviations from that profile.



- *Specification-based detection: defining allowed types of activity in order to flag any other activity as forbidden.

Detection Styles in Actual Deployments

- Striking imbalance deployments:
 - Almost exclusively only misuse detectors in use
 - Detect signatures (characteristic byte sequences)
- Question:
 - However, anomaly detection is extremely appealing (in the literatures)
 - Promises to find novel attacks w/o anticipating specifics
 - Machine learning works so well in other domains
 - But it's hard to find any machine learning NIDS in real-world deployments, why?

Misuse Detection (Signature-Based)

- Set of **rules** defining a behavioral signature likely to be associated with attack of a certain type
 - **Example: buffer overflow**
 - A setuid program spawns a shell with certain arguments
 - A network packet has lots of NOPs in it
 - Very long argument to a string function
 - **Example: SYN flooding (denial of service)**
 - Large number of SYN packets without ACKs coming back
 - ...or is this simply a poor network connection?
- Attack signatures are usually very specific and may miss variants of known attacks
 - **Why not make signatures more general?**

Anomaly Detection

- Originally introduced by Dorothy Denning in 1987
 - Assumption: attacks exhibit characteristics NOT observed for normal usage
 - Propose: host-based IDS
 - Host-level system building per-user profiles of activity
 - E.g., login frequency, session duration, resource consumption
- Machine learning (ML):
 - Training: trained with reference input to “learn” its specifics
 - Supervised or Unsupervised
 - Test: deployed on previously unseen input for the actual detection process

Anomaly Detection Cont'd

- Define a **profile** describing “normal” behavior
 - Works best for “small”, well-defined systems single program rather than huge multi-user OS
- Profile may be statistical
 - Build it manually (this is hard)
 - Use machine learning and data mining techniques
 - Log system activities for a while, then “train” IDS to recognize normal and abnormal patterns
 - Risk: attacker trains IDS to accept his activity as normal - adversarial learning
 - Daily low-volume port scan may train IDS to accept port scans

Machine Learning in Other Domains

- Examples (for comparison):
 - Amazon/Netflix – product recommendation
 - OCR (optical character recognition) systems
 - Natural language translation
 - Spam detection
- Claim: the task of finding attacks is fundamentally different from other applications
 - Making it significantly harder for us to employ ML

Machine Learning in Intrusion Detection



- Some well-known problems:
 - High false positive rate
 - Lack of (attack-free) training data
 - Attackers can try to evade detection
- Goal:
 - Using anomaly detection effectively in the real world operational environments (for network intrusion detection)

Challenges of Using Machine Learning



- Outlier Detection
- Lack of Training Data
- High Cost of Errors
- Semantic Gap (interpretation of results)
- Diversity of Network Traffic
- Difficulties with Evaluation
 - Training Data
 - Evasion Risk
- *So, What could be the solutions?*

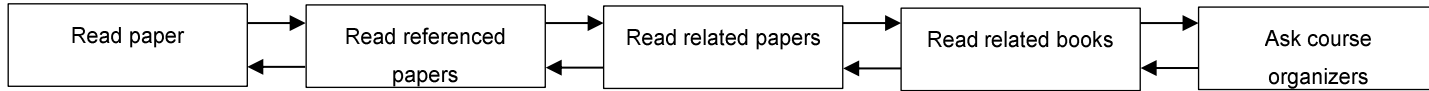
Student Assignments

- Pick 2 papers from the list (course homepage¹), or propose papers
 - You cannot select both papers from the Survey category!
- 1 paper used for a seminar presentation (20' + 10' discussion)
- Write a report about both papers (14 pages LNCS format)
- What to deliver:
 - 1 Presentation
 - 1 Report

1- <https://www.sec.in.tum.de/i20/teaching/ss2018/intrusion-detection-systems>

How to do your research

- Seminar - (kind of) simulation of scientific research



- Try to be independent, but also ask questions

Grading

- Grading consist of different parameters:
- Report (60%)
- Presentation (30%)
- Participation and discussion (10%)
 - Almost neglected!

Time Table

- 29.01.18 – Kick-off Meeting
- 17.04.18 – Introduction, Rules and Division of papers
- 05.06.18 – Students Presentation
- 12.06.18 – Students Presentation
- 19.06.18 – Students Presentation
- 26.06.18 – Students Presentation
- 03.07.18 – Students Presentation
- 10.07.18 – Students Presentation

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 paper
- 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

Presentation - Structure



- Introduction to the topic
- Present paper
 - Introduction
 - Main Point
 - 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 13.07.2018)

Report



- 14 Pages LNCS - fit both papers, describe them separately
- Summarize key points of both papers - 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 paper?

Report - Introduction

- Describe the context
- What is the preexisting work?
- What does the preexisting work lack?
- How does this paper 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?
- How can the limitations be addressed?

Report - Conclusion

- Summary of the paper in 3-4 sentences
- What are the most interesting results?
- What is the impact of the paper?

Report - Grading

- Paper organization
- Language and grammar
- Subject knowledge
- Ability to summarize
- Proper bibliography and citations

- 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.

1. Sommer, R., & Paxson, V. (2010). Outside the Closed World: On Using Machine Learning for Network Intrusion Detection. *2010 IEEE Symposium on Security and Privacy*, 0(May), 305–316.
2. Bhuyan, M. H., Bhattacharyya, D. K., & Kalita, J. K. (2014). Network Anomaly Detection: Methods, Systems and Tools. *Communications Surveys & Tutorials, IEEE*, 16(1), 303–336.
3. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(September), 1–58.
4. Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. *IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA 2009*, (Cisda), 1–6.
5. García-Teodoro, P., Díaz-Verdejo, J., Maciá-Fernández, G., & Vázquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. *Computers & Security*, 28, 18–28.