Network Anomaly Detection

Mohammad Reza Norouzian
Chair of IT Security (I20)
Department of Informatics
Technische Universität München

January 24, 2017
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 effectively
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?*
Intrusion Detection System

Master Seminar
Summer Semester, 2017
Goals of the Seminar

• Learn how IDSs detect malicious activities
• Another look at NIDSs with High Cost of Errors
• How to address the challenges in NIDS
• Use machine learning to solve some challenges
  • Detection
  • Analysis
  • Making conclusions, countermeasures
Research Areas

• Network Intrusion Detection System (NIDS)
  • Anomaly based
  • Hybrid based

• Feature Selection

• Industrial control system (ICS) intrusion detection
  • e.g. APT attacks: Stuxnet
Prerequisites

- MSc students of Informatics or similar
- Basics of IT security
- Machine Learning/Data Mining – very beneficial
- English speaking skills :)

15
Student Assignments

- The structure will be determined in group of 3 or 4 students
- Report + Presentation
- Students have to write reports about the papers they’ve got
  - You have to write the report on your own words, direct copy and paste will be determined as plagiarism
- Each student has to present core and approved papers
- List of papers will be provided later via website
- Grade:
  - Report (70%) + Presentation (30%)
- Report:
  - Core papers + Student’s searched papers
- Students search papers has to approved by instructor
How to Apply?

- Attend the Kick-Off
- Send a short CV to:
  - norouzian@sec.in.tum.de until 08.02.17
- Register on the matching system
  - look up http://docmatching.in.tum.de/
- If you cannot use the matching system for some reason, let me know!
• For any questions, ask now or contact me later:
  • norouzian@sec.in.tum.de
Topics + Groups

Industrial Intrusion Detection
SCADA
ICS
4 + 3 persons

Non-Industrial Intrusion Detection
Anomaly based
Hybrid based
4 + 3 persons

All groups have to present “Anomaly Detection: A Survey“ in 15 minuets on 04.05.17


