Large Scale Malware Analysis 2017/18

Introductory information

Bojan Kolosnjaji
Mohammad Norouzian

Chair of IT Security
TU München

Today’s agenda

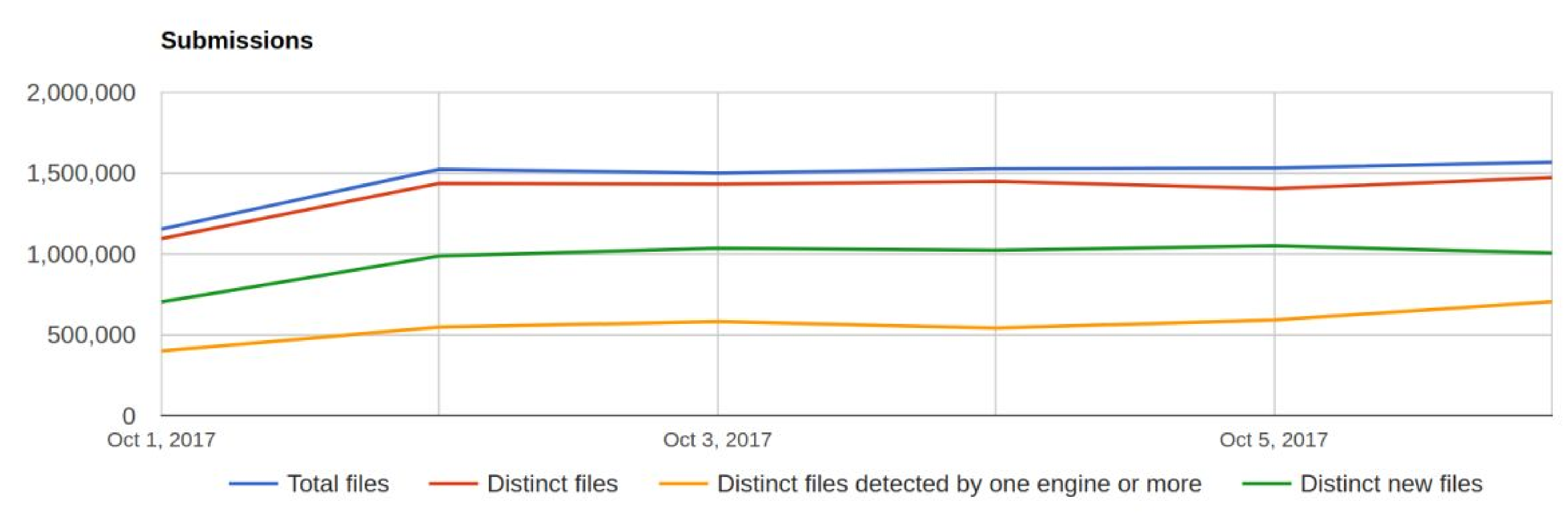
1) Introduction to the research area
2) Seminar instructions
    a) Deliverables
    b) Grading
    c) FAQ

Introduction to the research area

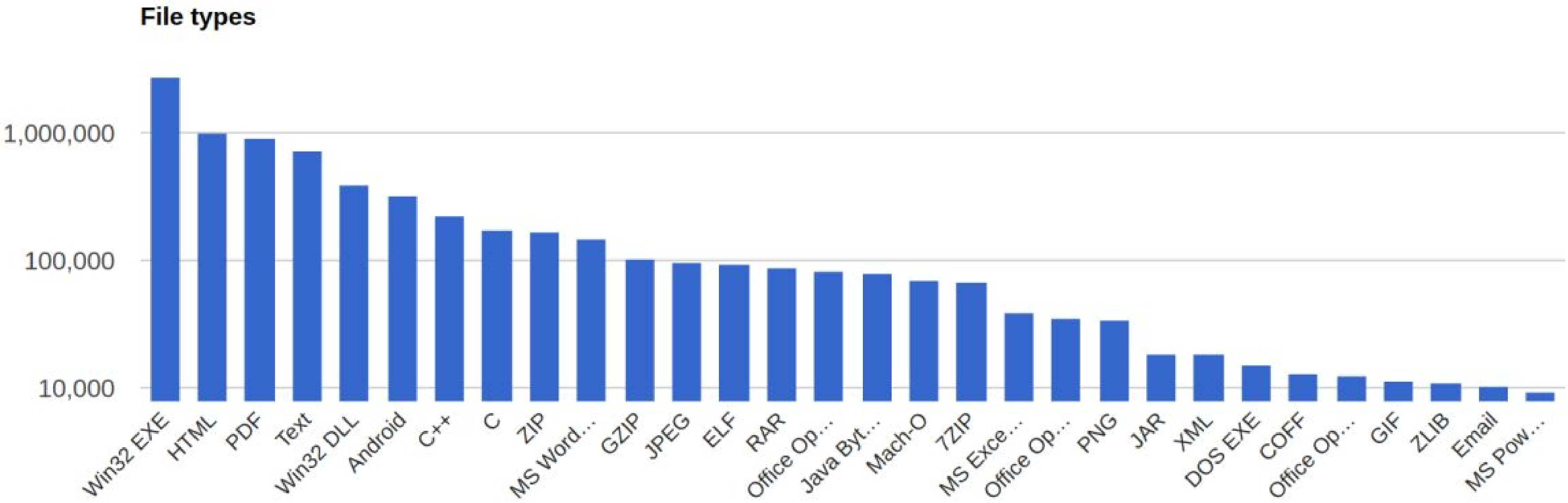
Symantec Threat Report

- Attackers Are **Moving Fast**, Defenders Are Not
- Attackers are Streamlining and **Upgrading their Techniques**, While Companies Struggle to Fight Old Tactics
- Cyberattackers Are **Leapfrogging Defenses** in Ways Companies Lack Insight to Anticipate
- Malware Used In Mass Attacks is **Increasing** and **Adapts**

Submissions to VirusTotal (1)



Submissions to VirusTotal (2)



File types

Research and practical challenges

- How to gather, store, process data?
- How to gain the new insight from the data?
- How to use this insight to improve cybersecurity?

Many researchers have been tackling this problem.

First ideas stem from the early 2000’s

Data Mining Approaches for Intrusion detection (1998)

Modeling system calls for intrusion detection with dynamic window sizes (2001)

Intrusion detection using neural networks and support vector Machines (2002)

Network traffic anomaly detection using packet bytes (2003)

...

However more and more papers

BotGrep: Finding P2P Bots with Structured Graph Analysis (2010)

Polonium: Tera-Scale Graph Mining and Inference for Malware Detection (2011)

“Andromaly”: a behavioral malware detection framework for android devices (2011)

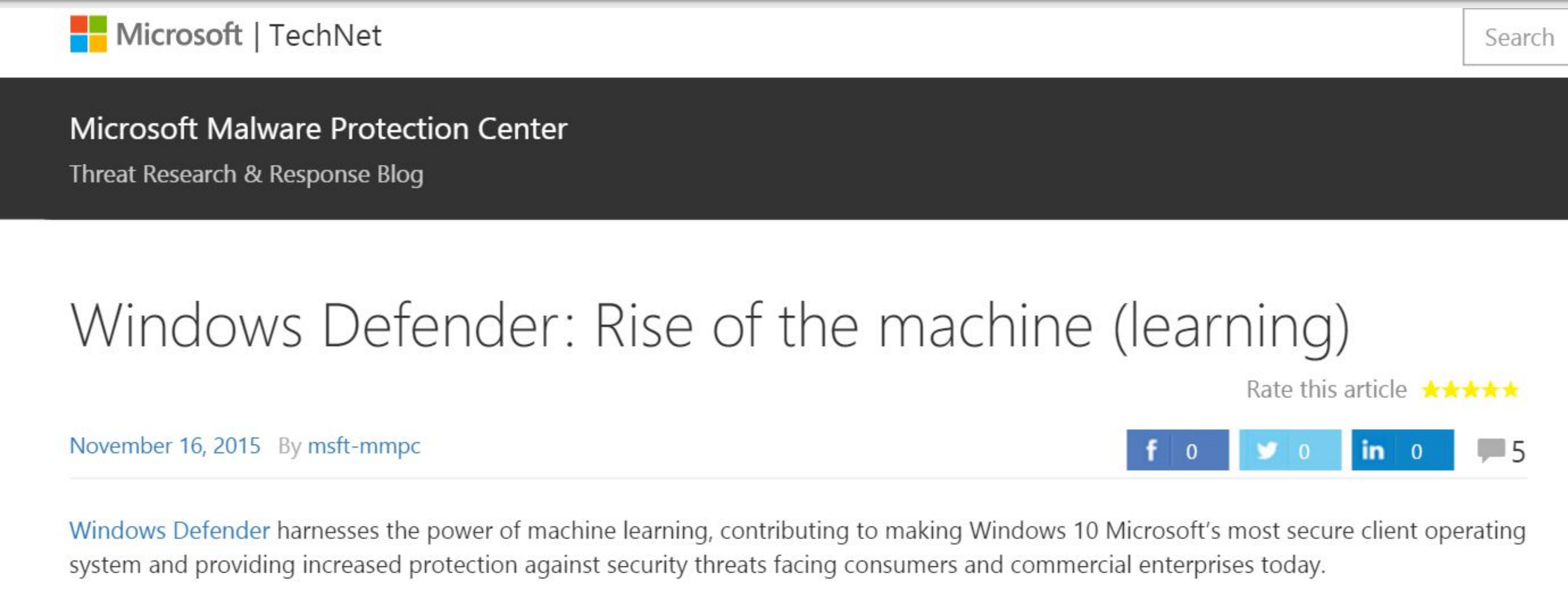
Leveraging String Kernels for Malware Detection (2013)

Recognizing Functions in Binaries using Neural Networks (2015)

Ensemble Learning for Low-level Hardware-Supported Malware Detection (2015)

...

Machine Learning in AV industry



Microsoft Malware Protection Center

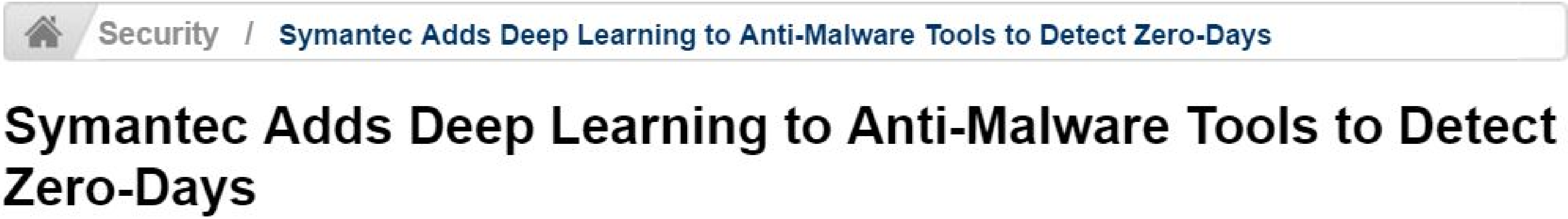
Threat Research & Response Blog

Windows Defender: Rise of the machine (learning)

November 16, 2015 By msft-mmpc

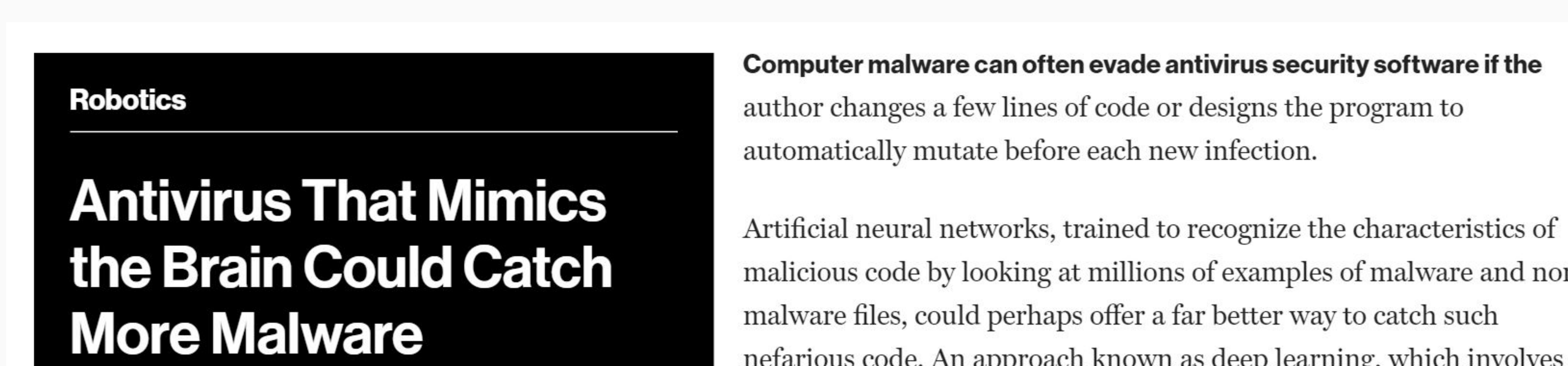
Windows Defender harnesses the power of machine learning, contributing to making Windows 10 Microsoft's most secure client operating system and providing increased protection against security threats facing consumers and commercial enterprises today.

Even deep learning...



Security / Symantec Adds Deep Learning to Anti-Malware Tools to Detect Zero-Days

Symantec Adds Deep Learning to Anti-Malware Tools to Detect Zero-Days



Robotics

Antivirus That Mimics the Brain Could Catch More Malware

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 non malware files, could perhaps offer a far better way to catch such nefarious code. An approach known as deep learning, which involves

Still many problems to tackle

- **Feature extraction** - we have lots of data, but what is important?
- **Scalability** - how to maintain our model?
- **Evasion** in sandboxes
- **Adversarial** environment, **mimicry** attacks

Seminar instructions

Our goal

- Get an **overview** of the academic state-of-the-art
- Extend your **knowledge** in security and/or machine learning
- Get a feeling on **how to apply** machine learning, big data in the security context

Topics from different context

- Malware Detection
- Analyzing Network Data
- Malicious Web Pages and Domains
- Evasion and Poisoning

What to deliver?

- 1 presentation
  - 20 minutes + 10 minutes of discussion

- 1 report
  - 14 Pages LNCS (look up LNCS template in Latex)

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

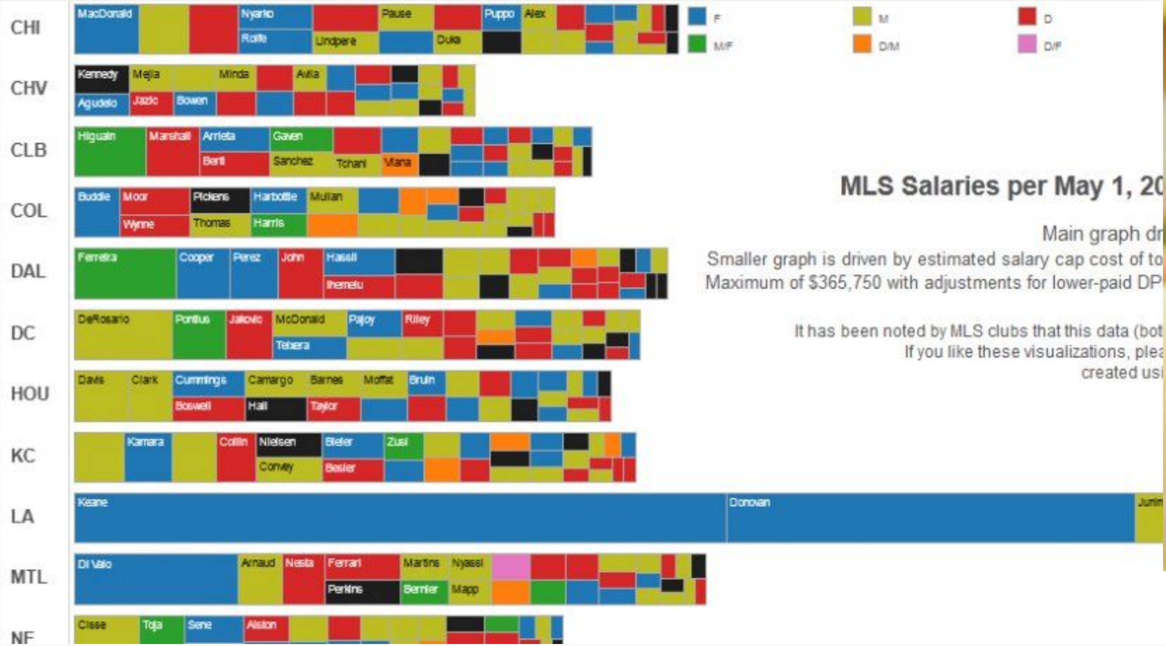
Presentation - Concise Text

- 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

Presentation - Pictures



MLS Salaries per May 1, 20

Main graph dr

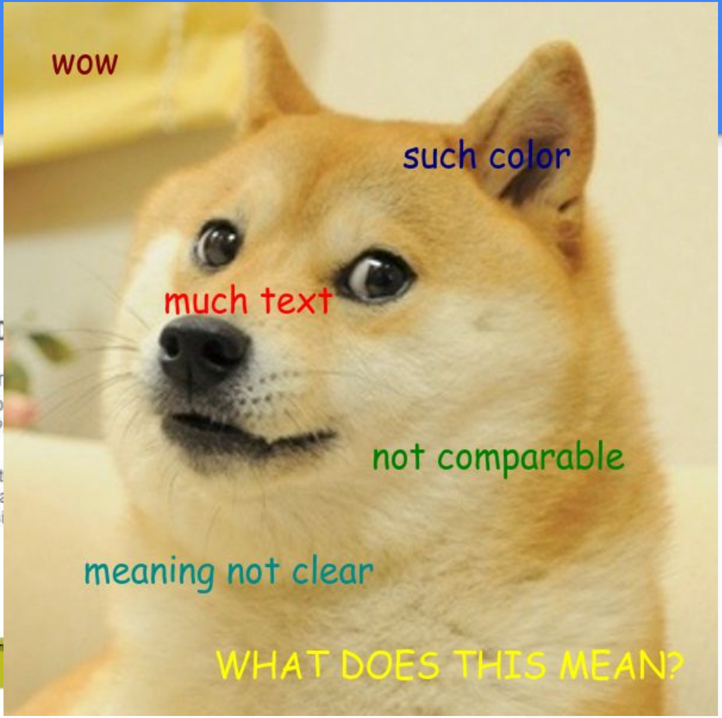
Smaller graph is driven by estimated salary cap cost of to

Maximum of $365,750 with adjustments for lower-paid DP

It has been noted by MLS clubs that this data (bot

If you like these visualizations, plea

created usi



Presentation - Structure

- 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 (more important)
  - Subject knowledge
  - Staying on topic
  - Identifying interesting/important points

Report (deadline 15.02.2018.)

Report

- 14 Pages LNCS - fit both papers
- Summarize key points of both papers - not an easy task
- If the papers are related, try to compare them in one section at the end
- 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?
    - 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 procedure
- 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

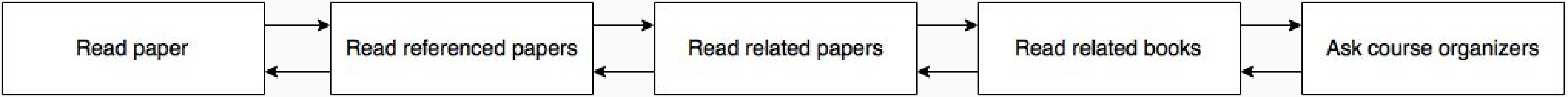
- Summary of the paper in 3-4 sentences
- What are the most interesting results?
- What is the impact of the paper?

Report - Grading

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

How to do your research

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



Read paper

Read referenced papers

Read related papers

Read related books

Ask course organizers

FAQ

- Allowed to miss 1 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.