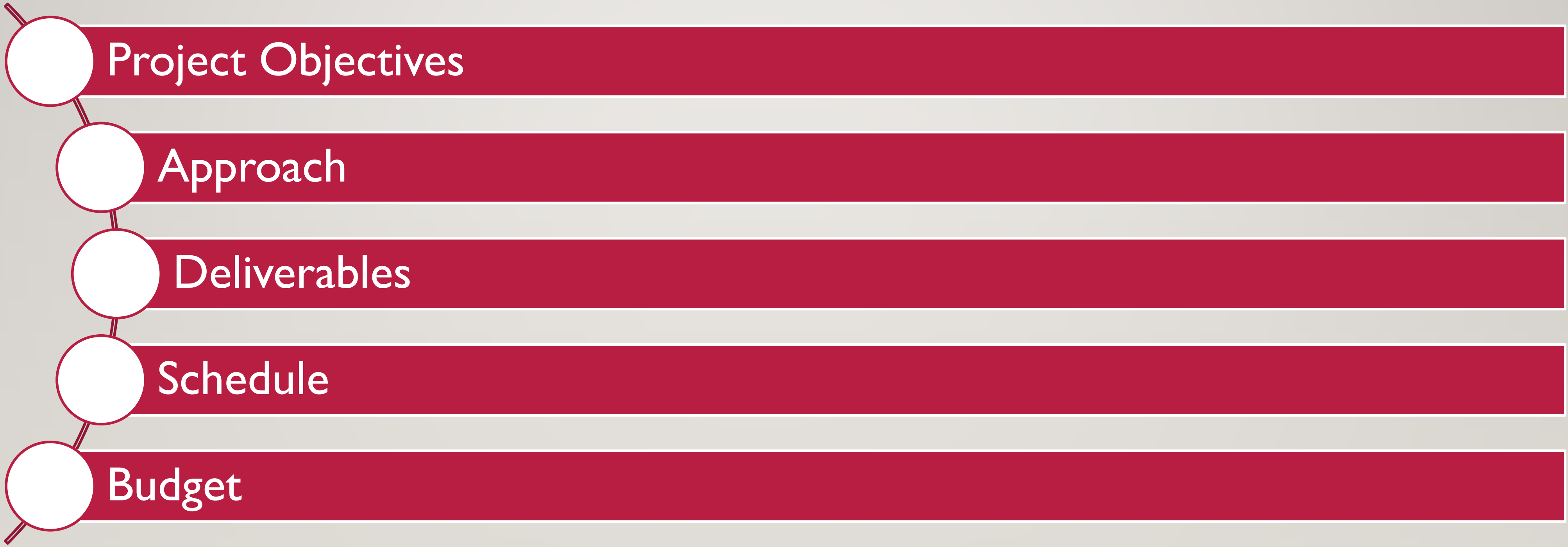
A PRACTICAL INTRODUCTION TO APPLYING MACHINE LEARNING TO MALWARE DETECTION

GROUP SSL:

GASTON CARVALLO

LOYD RAFOLS

AGENDA



- Project Objectives
- Approach
- Deliverables
- Schedule
- Budget

RESEARCH TOPIC

Our Topic was

- Malware detection using machine learning

Our Solution was

- Hybrid model using both network and host features

PROJECT JUSTIFICATION

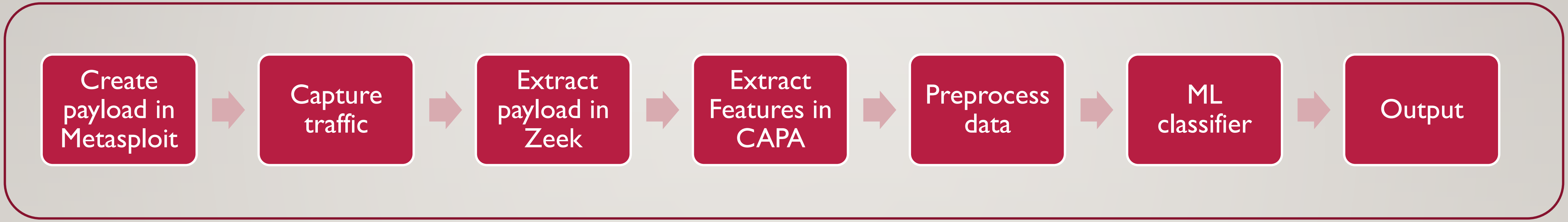
Students Knowledge

Industry Trend

PROJECT OBJECTIVE

- Focus on learning
- Create our own data
- Output to a log

PROPOSED WORKFLOW



APPROACH TO TECHNICAL IMPLEMENTATION

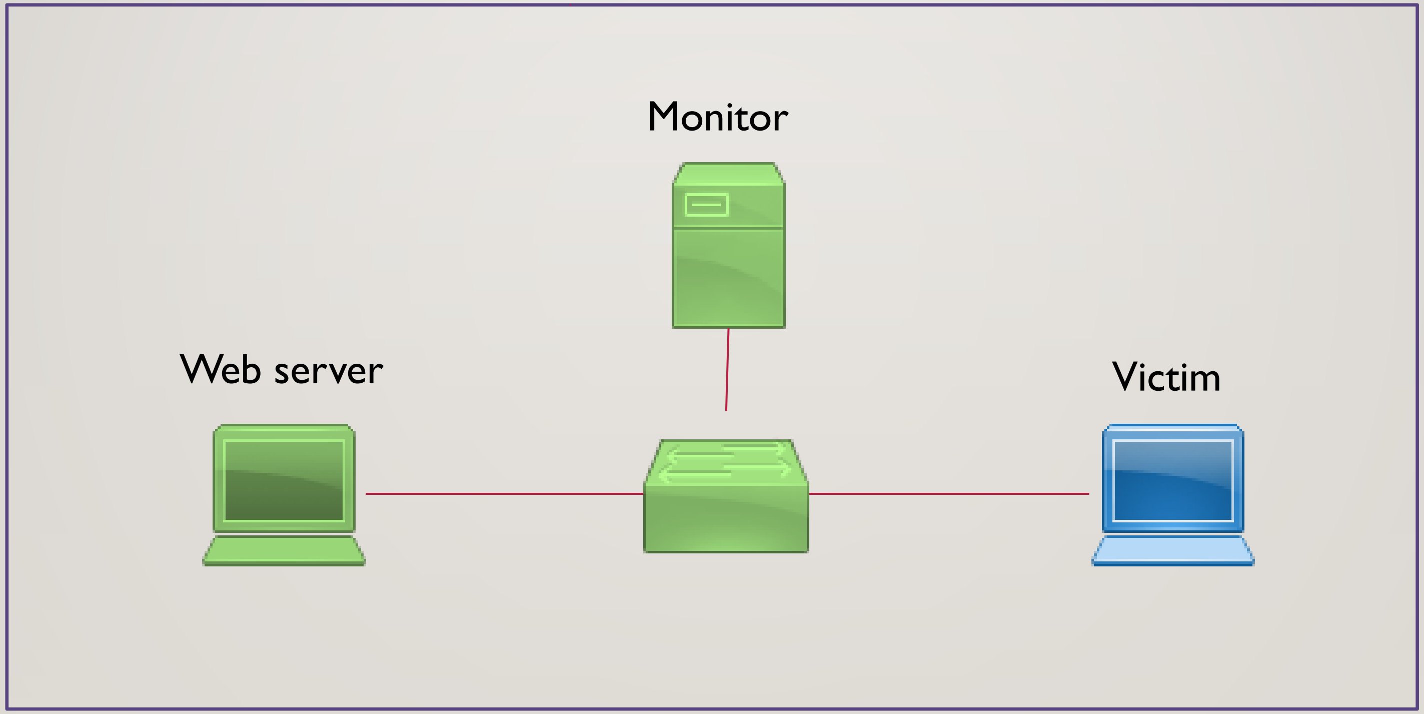
Deploy testbed

Create dataset

Train and validate classifiers

Implement Classifier

TESTBED



DATASET

- Metasploit reverse shell payloads targeting windows
- Benign files
- Automated process

MACHINE LEARNING

- Extract features through CAPA
- Preprocess data in python
- Scikit-learn library
- Implement live classifier

PROJECT MANAGEMENT INFORMATION

Phases, Deliverables, Scheduling, Budgeting, Feasibility

PROJECT MANAGEMENT INFORMATION

Phases, Deliverables, Scheduling, Budgeting, Feasibility

PROJECT PHASES

- Four different objective milestones broken down into phases
  - Dataset creation
  - Setting up and configuring machine learning model
  - Creating education material/configuring learning environment
  - Writing final deliverable report
- Each phase has separate individual tasks to complete

PROJECT DELIVERABLES

- Linux (Ubuntu) OVA file containing learning environment for malicious executable detection using machine learning
- Sample datasets (as part of above)
- Scripts used to implement machine learning infrastructure/sample data generation
- Presentations/lectures for major machine learning workflow steps
- Practical labs for major machine learning workflow steps
- Final companion report

PROJECT SCHEDULE

- 2 semesters of work (REA705 + REA820), 12 weeks each = 24 weeks of work
- Semester 1 (REA705)
  - Dataset creation (Week 1 – Week 6)
  - Feature selection (Week 6 – Week 9)
  - Machine learning model setup and configuration (Week 10 – Week 12)
- Semester 2 (REA820)
  - Creating educational content (Week 1 – Week 3)
  - Drafting final report (Week 4 – Week 8)
  - Compose final report (Week 9 – Week 12)
- Buffers and total float accounted for

PROJECT BUDGET

- Free!
- All software used are generally at least free (most are FOSS)
  - scikit-learn
  - Metasploit Framework
  - Zeek
  - Capa
- Main learning environment is Linux-based, so no need to purchase a license for using most Linux-based OS'

PROJECT FEASIBILITY

- Technical feasibility
  - Technical experience and knowledge to implement project objectives and complete deliverables
- Legal feasibility
  - Software licenses
- Schedule feasibility
  - Project management is key – dependencies not necessarily finish-to-start
  - Parallelization

SUMMARY

- Our project is about “A Practical Introduction to Applying Machine Learning to Malware Detection”
- Proposed workflow, approaches
- Testbed, experimentation
- Project information
  - Phases
  - Deliverables
  - Scheduling
  - Budget
  - Feasibility

THANK YOU

- Group SSL
  - Gaston Carvallo
  - Loyd Rafols
- Topic: A Practical Introduction to Applying Machine Learning to Malware Detection
- Website: https://rea.000109.xyz/
- Website post and slides download: https://rea.000109.xyz/ProposalPresentation/
- Course: REA605 Winter 2021, Mark Shtern