REA705 PROJECT STATUS UPDATE

GROUP SSL

GASTON CARVALLO

LOYD RAFOLS

PROJECT TOPIC REMINDER

- Practical introduction to using machine learning for malware detection
- Consists of a theoretical portion and practical portion
- Self contained environment (VMs) or can be done manually using commands

UPDATES FROM PROJECT PROPOSAL PRESENTATION

- Officially integrated into Seneca curriculum
  - Practical components of project to be used as labs for SRT521, introduced in Fall 2021
  - Coordination with Asma to ensure labs are sufficient quality
- Milestones adjusted
  - Practical and technical components moved to top priority to accommodate for new schedule and deadlines
  - Theoretical components moved to REA820 as they are lower priority
  - Final report draft now occupies last ≈ 2 – 4 weeks of REA705

PROJECT PLAN

- Integration with official Seneca curriculum required project plan changes
- REA705 – Labs and Final Report Draft
  - Week 1 to 10 – Perform lab tasks and write lab
  - Week 10 to 13 – Write final report draft
- REA820 – Tutorials and Final Report
  - Week 1 to Study Week – Write tutorials on various fundamental ML concepts
  - Study Week to End of Semester – Write final report

OVERALL PROJECT STATUS – REA705

- 4 milestones to be achieved this semester
  - Represents submission of major deliverables
  - Deliverables include practical components (i.e. labs), and final project report draft
- Overall, ahead of schedule slightly
  - Buffer of about a week

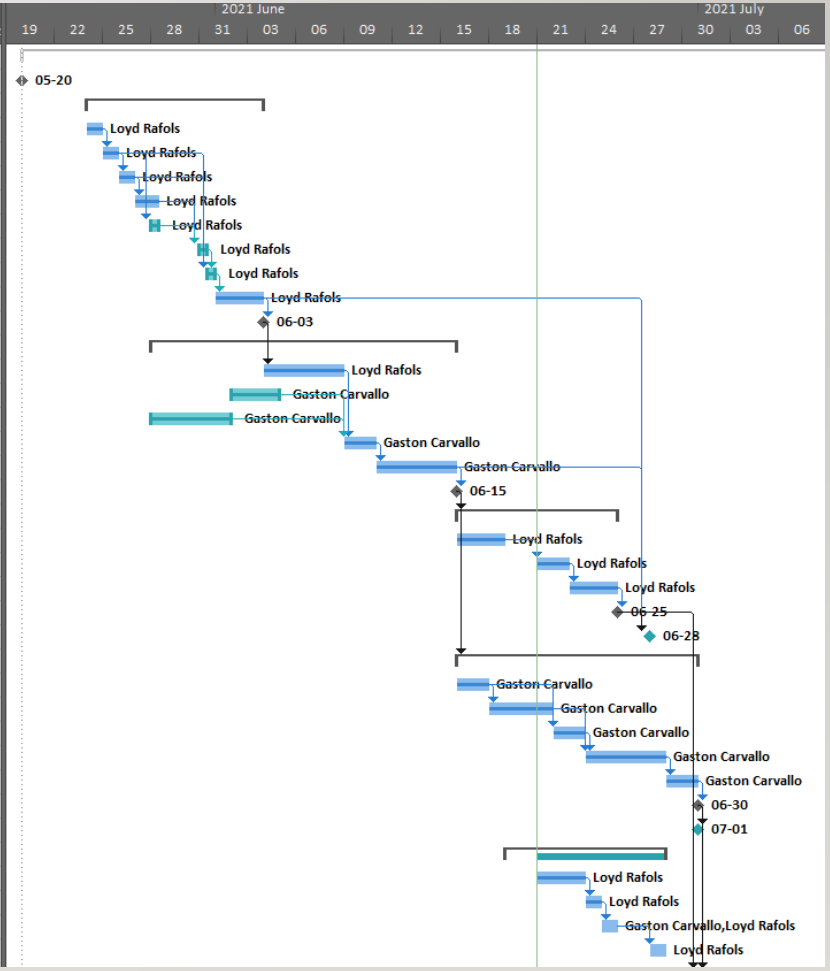
CURRENT PROJECT PROGRESS

- Milestone 1 - Labs 1 to 3 completed
  - All three labs completed ahead of schedule
- Milestone 2 – Lab 4 completed
  - Completed ahead of schedule
- Milestone 3 – Lab 5 and 6 completed
  - To be started, projected to be completed ahead of schedule
- Milestone 4 - Final Report Draft
  - To be started, will have about 3-4 weeks of time to perform work

MS PROJECT OVERVIEW

GANTT CHART

	Task Mode	Task Name	Duration	Start	Finish	Predecessors	Resource Names
0		**REA705 Implementation**	**59 days**	**Thu 21-05-20**	**Tue 21-08-10**		
1	✓	Project Kickoff	0 days	Thu 21-05-20	Thu 21-05-20		
2	✓	**Lab 1: Setup Environment**	**9 days**	**Mon 21-05-24**	**Thu 21-06-03**		
3	✓	Deploy Kali (attacker)	1 day	Mon 21-05-24	Mon 21-05-24		Loyd Rafols
4	✓	Deploy Windows MS3 (victim)	1 day	Tue 21-05-25	Tue 21-05-25	3	Loyd Rafols
5	✓	Deploy and configure Ubuntu monitoring server	1 day	Wed 21-05-26	Wed 21-05-26	4	Loyd Rafols
6	✓	Configure Zeek to extract files	1.5 days	Thu 21-05-27	Fri 21-05-28	5	Loyd Rafols
7	✓	Install and configure ELK stack	0.5 days	Fri 21-05-28	Fri 21-05-28	4	Loyd Rafols
8	✓	Configure promiscous interface	0.5 days	Mon 21-05-31	Mon 21-05-31	6,7	Loyd Rafols
9	✓	Validate network connectivity and configuration	0.5 days	Mon 21-05-31	Mon 21-05-31	4,5,8	Loyd Rafols
10	✓	Create lab 1: Setup environment	3 days	Tue 21-06-01	Thu 21-06-03	9	Loyd Rafols
11	✓	Lab 1 completed	0 days	Thu 21-06-03	Thu 21-06-03	10	
12	✓	**Lab 2: Create dataset**	**13 days**	**Fri 21-05-28**	**Tue 21-06-15**		
13	✓	Gather malicious samples	3 days	Fri 21-06-04	Tue 21-06-08	11	Loyd Rafols
14	✓	Gather Benign samples	3 days	Wed 21-06-02	Fri 21-06-04		Gaston Carvallo
15	✓	Create scripts to generate dataset	3 days	Fri 21-05-28	Tue 21-06-01		Gaston Carvallo
16	✓	Generate dataset	2 days	Wed 21-06-09	Thu 21-06-10	13,14,15	Gaston Carvallo
17	✓	Create lab 2: Data Collection and feature extraction	3 days	Fri 21-06-11	Tue 21-06-15	16	Gaston Carvallo
18	✓	Lab 2 completed	0 days	Tue 21-06-15	Tue 21-06-15	17	
19	✓	**Lab 3: Classification using random forest**	**8 days**	**Wed 21-06-16**	**Fri 21-06-25**	**18**	
20	✓	Data pre-processing	3 days	Wed 21-06-16	Fri 21-06-18		Loyd Rafols
21	✓	Train and validate ML using random forest	2 days	Mon 21-06-21	Tue 21-06-22	20	Loyd Rafols
22	✓	Create lab 3: Classification using random forest	3 days	Wed 21-06-23	Fri 21-06-25	21	Loyd Rafols
23	✓	Lab 3 completed	0 days	Fri 21-06-25	Fri 21-06-25	22	
24	✓	Milestone 1: labs 1 - 3	0 days	Mon 21-06-28	Mon 21-06-28	10,17,23	
25	✓	**Lab 4: Classification using C4.5 and Naïve Bayes**	**11 days**	**Wed 21-06-16**	**Wed 21-06-30**	**18**	
26	✓	Data pre-processing	2 days	Wed 21-06-16	Thu 21-06-17		Gaston Carvallo
27	✓	Train and validate ML using C4.5	2 days	Fri 21-06-18	Mon 21-06-21	26	Gaston Carvallo
28	✓	Train and validate ML using Naïve Bayes	2 days	Tue 21-06-22	Wed 21-06-23	26	Gaston Carvallo
29	✓	Create Lab 4	3 days	Thu 21-06-24	Mon 21-06-28	27,28	Gaston Carvallo
30	✓	Revise Lab 4	2 days	Tue 21-06-29	Wed 21-06-30	29	Gaston Carvallo
31	✓	Lab 4 completed	0 days	Wed 21-06-30	Wed 21-06-30	30	
32	✓	Milestone 2: Lab 4	0 days	Thu 21-07-01	Thu 21-07-01	31	
33		**Project Status Update Presentation**	**7 days**	**Sat 21-06-19**	**Mon 21-06-28**		
34	✓	Create slide contents	3 days	Mon 21-06-21	Wed 21-06-23		Loyd Rafols
35	✓	Prettify slides	1 day	Thu 21-06-24	Thu 21-06-24	34	Loyd Rafols
36		Record voiceovers	1 day	Fri 21-06-25	Fri 21-06-25	35	Gaston Carvallo,Loy
37		Compose video and submit	1 day	Mon 21-06-28	Mon 21-06-28	36	Loyd Rafols



SUMMARY

- Project Topic: A Practical Introduction to Applying Machine Learning to Malware Detection
- Project plan changes due to integration into Seneca curriculum
- 4 milestones to be achieved in REA705, currently a few days ahead of schedule
  - Practical components (labs) and final report draft

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/REA705ProjectStatusUpdate
- Course: REA705 Summer 2021, Mark Shtern & Mohammed Alani