Abstract:
Machine learning is playing an important and growing role in cyber security. A significant barrier to the development of machine learning applications to a greater extent is the lack of availability of high-quality cyber security datasets. To effectively address this problem, several questions need to be answered. Where should we source our data? Must it all come from real machines experiencing real cyber-attacks, or can some of it be synthetically generated? What types of data should we be collecting? How can we objectively evaluate which data should be collected? And what types of preprocessing must be done for this data to be useful? In this seminar, we will explore answers to these and similar questions.

Bio:
James Halvorsen is a Ph.D. student in Computer Science at Washington State University. His research focuses on the cross sections between machine learning and cyber security, emphasizing generative machine learning techniques and intrusion detection systems.

cyser.wsu.edu