Using Probabilistic Generative Models for Ranking Risks of Android Apps
Peng, Gates, Li, Qi, et. al.
presented: CCS ‘12

Problem:
• Android relies on the User to make security relevant decisions regarding permissions during installation
• In Android, permissions are difficult to understand and often ignored

Goal:
• Create a principled method to calculate the risk of apps, that is...
  • Simple to understand
  • Monotonic
  • Ranks Malware generally as High Risk

Method:
• Use Probabilistic Generative Models
• Train on large amount of unlabeled data
• Create an expectation, measure distance from the expectation to create risk score

Models Explored:
• Naïve Bayes
• Naïve Bayes with Informative Prior
• Mixture of Naïve Bayes
• Hierarchical Mixture of Naïve Bayes

Data
~325,000 apps from Google Play in Feb2012
~400 malware apps
Extract Permission Requests as Features

Monotonic Property:
Naïve Bayes with Informative Prior
Monotonic
Hierarchical Mixture of Naïve Bayes
Not Monotonic

Risk: Malware vs Market Apps

Conclusion:
• Naïve Bayes with Prior is suggested
  • Performance + Simplicity + Monotonic