RevMatch: An Efficient and Robust Decision Model for Collaborative Malware Detection

Carol Fung, Disney Lam, and Raouf Boutaba
Virginia Commonwealth University and University of Waterloo

Outline

- Introduction
- Related Work
- RevMatch Model
- Evaluation
- Conclusion

Introduction

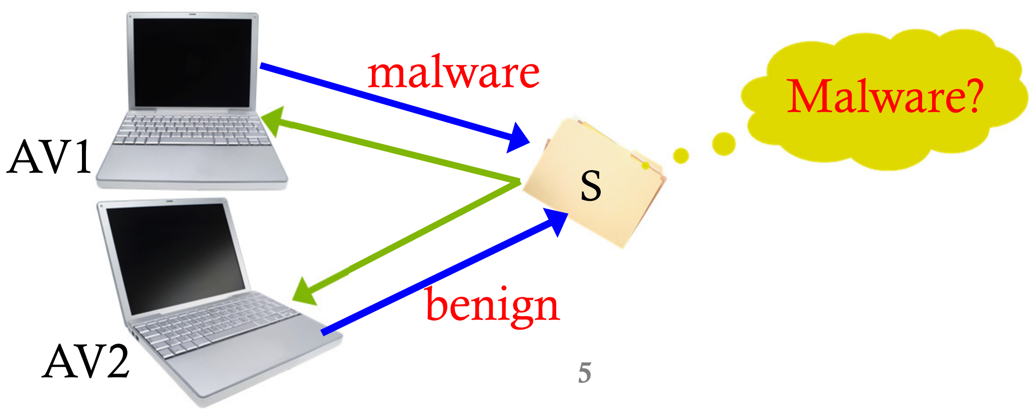
- Millions of new unique malware instances appear every year
- 560 million victims per year (2012)
- Annual economy lost US $110 billion (2012)
- Malware consequences:
  - Botnets (BredoLab, conficker, etc.)
  - Attack others, such as spamming and DDoS attacks
  - Spamhaus attack (2013)

Collaborative Malware Detection

- Anti-virus software (AVs) are commonly used for malware detection
  - Signature-based, behavior-based, heuristic-based, and reputation-based
- Most AV vendors do not share knowledge with each other
- Collaborative malware detection allows and encourages anti-viruses to share knowledge to improve accuracy
  - E.g., CloudAV
  - Challenge: Collaborative decision model

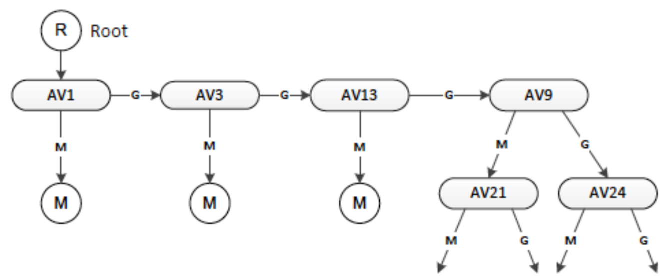
Problem Statement

- A suspicious file S is sent to multiple AVs for scanning
- Collected results are either malware (1) or benign-ware (0) from each AV
- Given that we have the detection results of some malware scanners on a set of known malware and benign-ware, we need to decide whether the file S is malware or benign-ware?



Related Work

- Static Threshold
  - Simple average compared to a fixed threshold
- Weighted Average
  - Weighted average compared to a fixed threshold
- Decision Tree
  - Machine learning approach



- Bayesian Decision
  - Compute probability of malware and optimal decision based on cost of false positive and false negative
  - The assumption is that all AVs are independent

RevMatch Model

- Check the labeled history to find the number of malware M(y) and benign ware G(y) with the same scanning results
  - y is the scanning results vector from all AVs
- If $M(y)+G(y) \geq \tau$
  - We raise malware alarm if

$$\frac{M(y)}{G(y)} \geq \frac{C_{fp} M P_G}{C_{fn} G P_M} = \theta$$

#malware in history

#benign ware in history

Cost of false positive

Cost of false negative

Prior probability of benign ware

Prior probability of malware

Decision Model (con.)

- What if $M(y)+G(y) < \tau$?
  - We perform feedback relaxation: move the feedback from least competent AVs until the number matching samples exceeds τ

- Therefore, we need to sort the level of competence of all participating AVs
  - We use the metrics of 1-FN-FP=TP-FP

Example

$\tau=2$
$\theta=1$

$y=\{1,1,0\}$

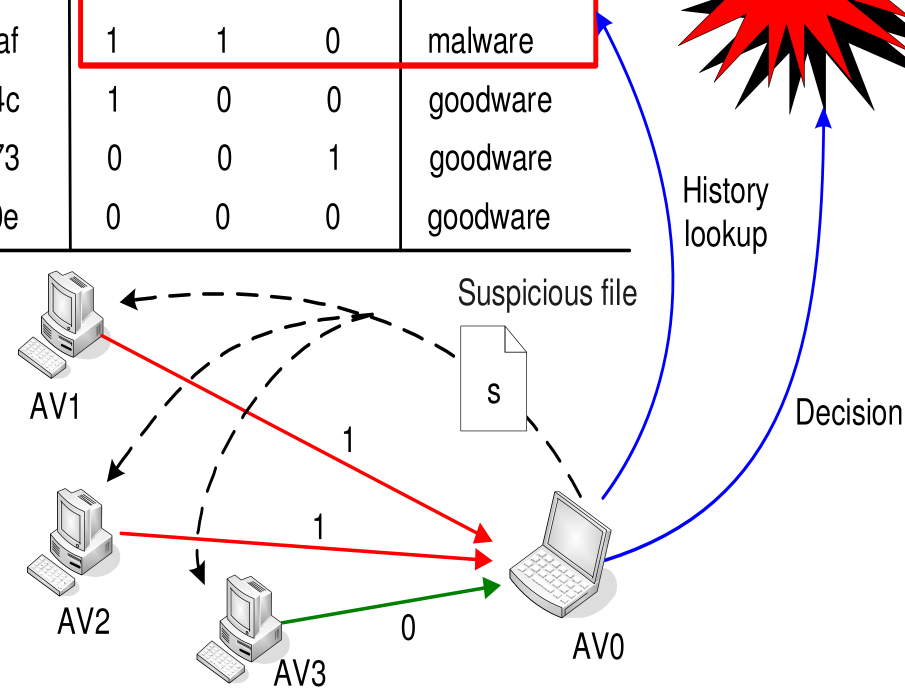
$M(y)=2$
$G(y)=0$

$M(y)/G(y)\geq\theta$

Raise Alarm!

Labeled history for AV0

Digest	Feedback Set			Ground Truth
	AV1	AV2	AV3	
df73	1	1	1	malware
48c2	1	1	0	malware
9faf	1	1	0	malware
3a4c	1	0	0	goodware
3473	0	0	1	goodware
cc0e	0	0	0	goodware



History Maintenance

- Use files with ground truth to obtain labeled history
- Detection results where the ground truth are revealed later can also be used as labeled history
- Enforce minimum time gap Δt for history updates with the same detection results
  - E.g., if the last update of{1,0,0,malware} is at time 0 then {1,0,0,malware} at time Δt-1 will not be recorded in history
  - Prevent from manipulated history poisoning

Evaluation Data Set

DATA SETS

Dataset ID	Dataset description	Samples	Year	Malware alarm rate
S1	Old malware	58,730	2008–2009	84.8%
S2	New malware	29,413	2011–2012	59.5%
S3	Hybrid malware	50,000	2009–2012	69.7%
S4	Goodware (SourceForge)	56,023	2012	0.3%
S5	Goodware (Manual)	944	2012	7.9%
S6	Hybrid Goodware	5,000	2012	1.6%

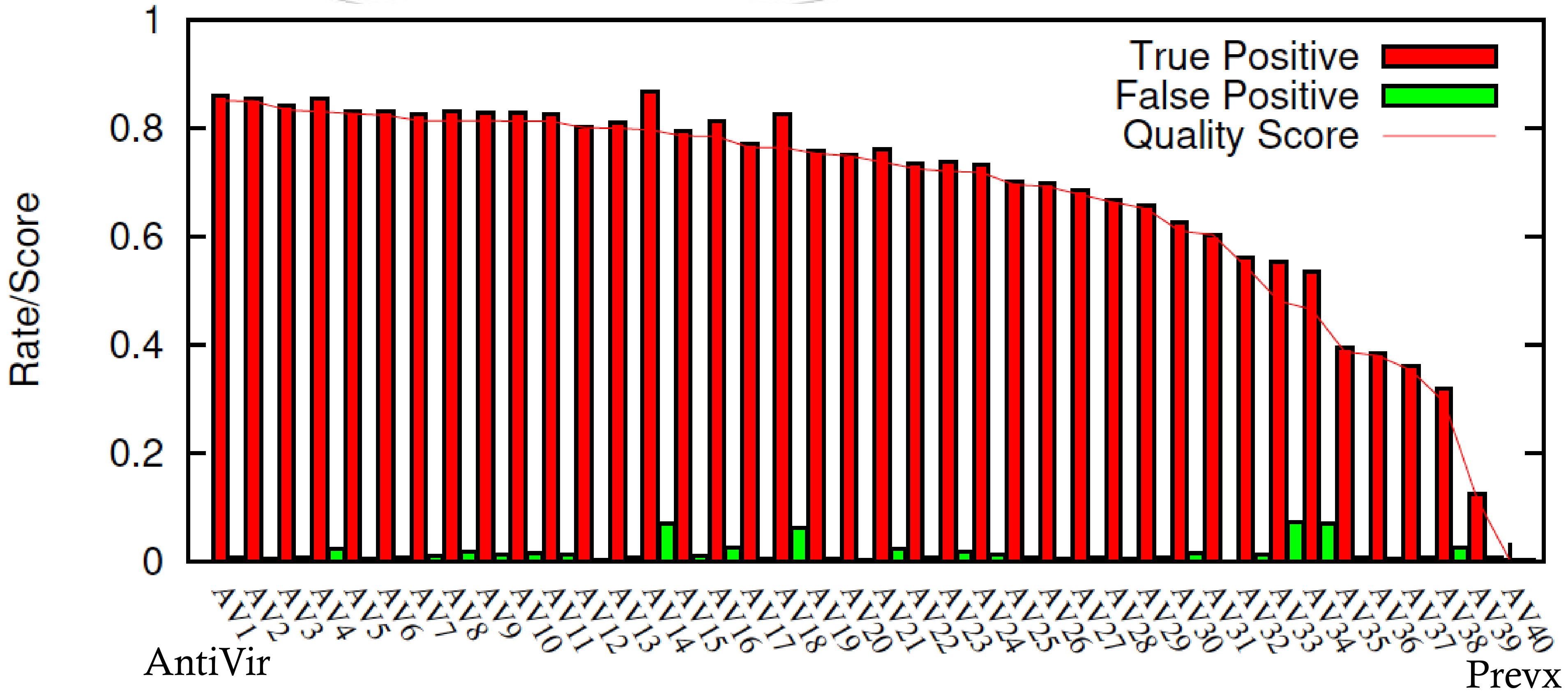
CW-Sandbox and Offensive-computing

List of Anti-viruses

AhnLab-V3	Comodo	Jiangmin	Rising
AntiVir	DrWeb	K7AntiVirus	Sophos
Antiy-AVL	Emsisoft	Kaspersky	SUPERAntiSpyware
Avast	eSafe	McAfee	Symantec
AVG	eTrust-Vet	Microsoft	TheHacker
BitDefender	Fortinet	NOD32Norman	TrendMicro
ByteHero	F-Prot	nProtect	VBA32
CAT-QuickHeal	F-Secure	Panda	VIPRE
ClamAV	GData	PCTools	ViRobot
Commtouch	Ikarus	Prevx	VirusBuster

List of AVs from VirusTotal

Comparison of AVs



Comparison of Accuracy

Method	True Positive TP	False Negative FN	False Positive FP	Quality Score 1-FN-FP
Static Threshold	0.903	0.097	0.022	0.881
Weighted Threshold	0.908	0.092	0.025	0.883
Decision Tree	0.956	0.044	0.077	0.879
Bayesian Decision	0.871	0.129	0.013	0.858
RevMatch	0.927	0.073	0.007	0.920
Best Single AV	0.859	0.141	0.008	0.851

Tested on S3 + S6 and 10-fold cross-validation

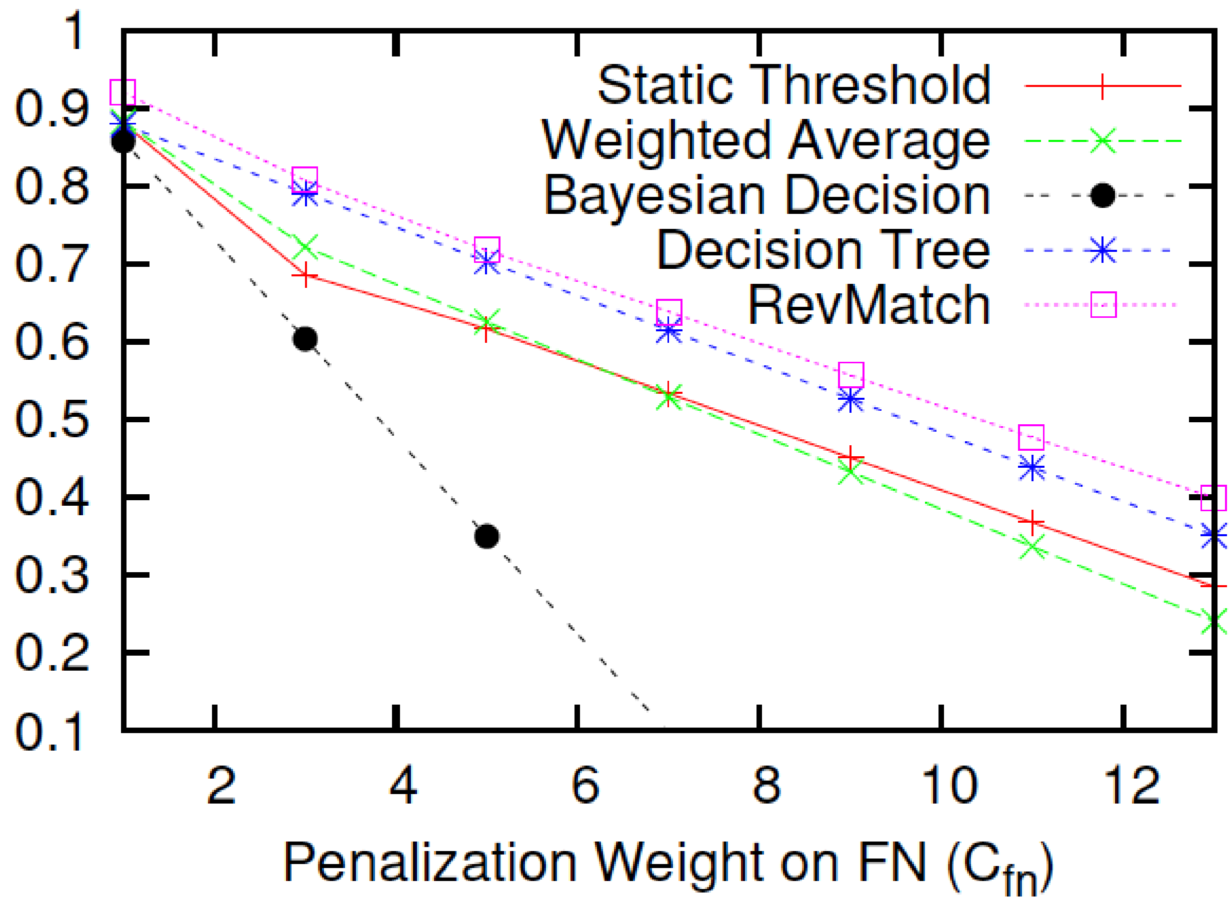


Figure: Quality score of all models with different Cfn

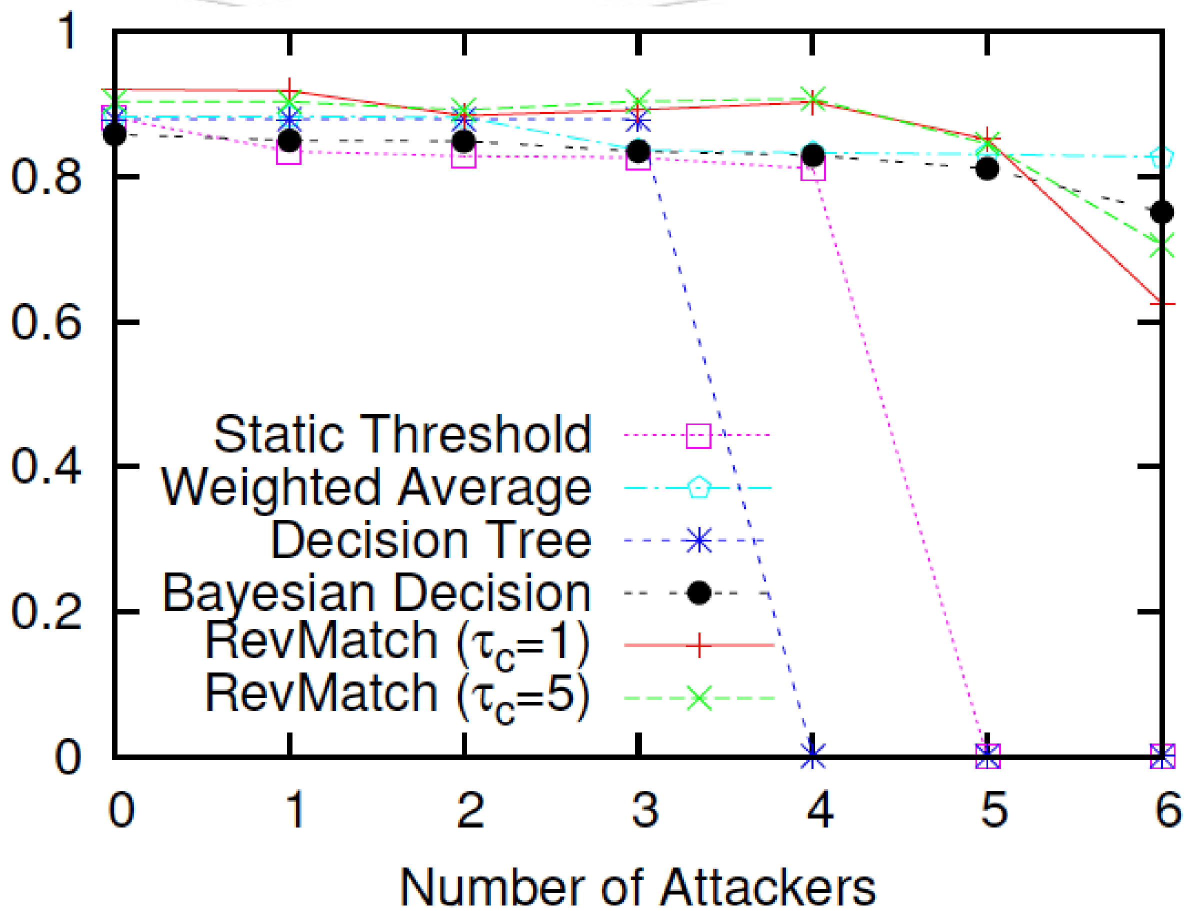


Figure: Quality score versus the number of attackers

Performance Comparison

Decision Model	Decision Quality	Runtime Runtime	Attacker Tolerance	Partial Feedback	Flexi-bility
Static Threshold	medium	fast	4 attackers	no	yes
Weighted Average	medium	fast	5+ attackers	yes	yes
Decision Tree	medium	fast	3 attackers	no	no
Bayesian Decision	low	fast	5+ attackers	yes	yes
RevMatch	high	medium	5+ attackers	yes	yes

Robustness

- History poisoning attack
  - An malicious AV knows a type of zero-day attack and can accurately detect the attack while others cannot
  - The malicious AV creates many malware records where only itself can detect it
  - Afterwards the AV suddenly reports benign-ware to be malware
- Defense
  - Enforce minimum history update gap Δt to prevent from quick history poisoning
  - Files are only sent for scanning if anormalies are detected

Conclusion and Future Work

- Proposed RevMatch: a new decision model for collaborative malware detection
- Proposed evaluation metrics to compare with other models
- Higher accuracy, flexibility, partial feedback tolerance, and robustness against insider attacks
- Improve the feedback relaxation algorithm
- Improve the run-time efficiency

Thank You