Supplemental Material to Discovering Signals from Web Sources to Predict Cyber Attacks

September 23, 2019

1 Cyber Terms

The following list of cyber related terms was used in this study: 0day, 0-day, account, acrobat, adobe, blackmail, botnet, breach, breaches, cpe, crypto, cve, databreach, ddos, dhcp, dns, exploit, exploits, explorer, extortion, hack, hacker, hijacked, hijacking, intel, ios, iot, linux, malware, malwares, microsoft, ms16-, ms17-. oracle, password, phishing, ransomware, ransomwares, rootkit, trojan, trojans, udp, usb, vpn, vulnerabilities, vulnerability, win7, windows. windows7, zeroday.

2 ARIMAX Plots

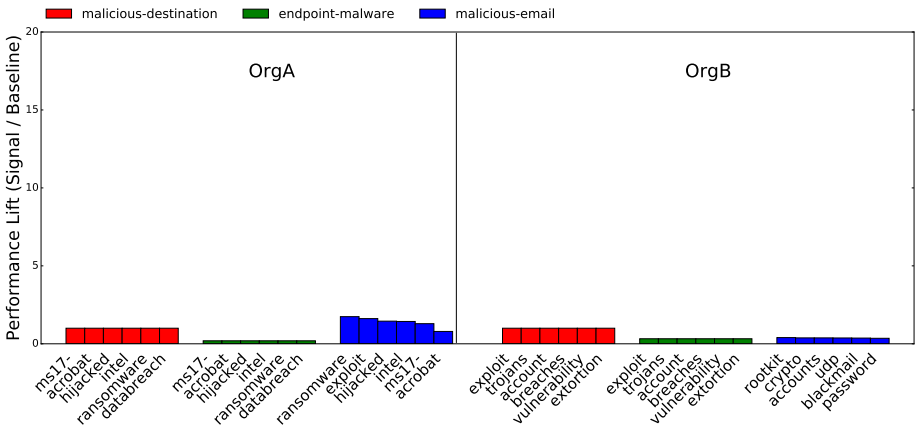

Figure 1: Weekly ARIMAX F1 performance of Twitter signals.

3 Mean RMSE Lifts

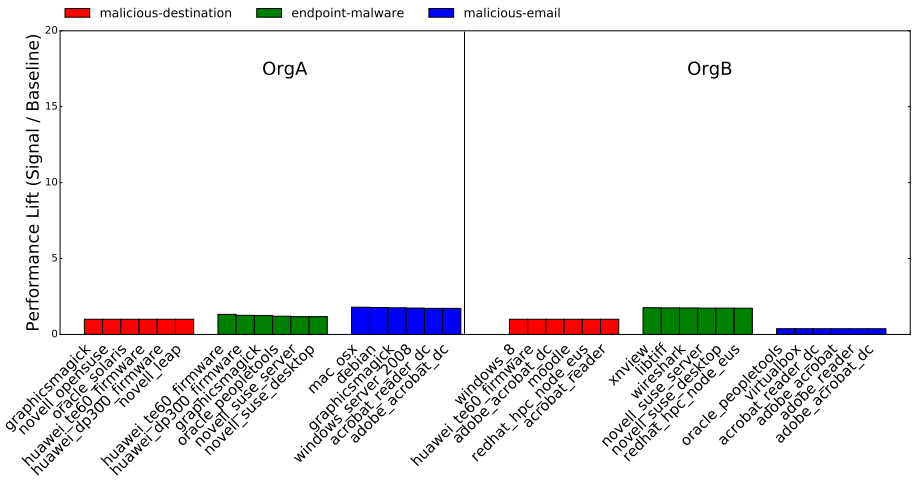

Figure 2: Weekly ARIMAX F1 performance of vulnerability signals.

Table 1: Mean RMSE Lift of GRU model trained on a single signal across seven months for an individual endpoint. Evaluated on OrgA.

Event Type	Signal	Source	Relative Lift
ep-malware	win_server_2008	vulnerability	1.03
ep-malware	cpe	twitter	1.02
ep-malware	trojan	blogs	1.02
ep-malware	iot	d2web	1.01
ep-malware	hp10_ctr	honeypots	0.96
mal-dest.	iphone_os	vulnerability	1.01
mal-dest.	linux	twitter	0.99
mal-dest.	hijacking	blogs	0.97
mal-dest.	trojans	d2web	0.99
mal-dest.	hp1_ctr	honeypots	1.28
mal-email	fedora_project	vulnerability	1.02
mal-email	breach	twitter	1.02
mal-email	exploits	blogs	1.03
mal-email	exploits	d2web	1.02
mal-email	hp10_clicks	honeypots	1.54

Table 2: Mean RMSE Lift of GRU model trained on a single signal across seven months for an individual endpoint. Evaluated on OrgB.

Event Type	Signal	Source	Relative Lift
ep-malware	imagemagick	vulnerability	0.62
ep-malware	account	twitter	0.72
ep-malware	ios	blogs	0.69
ep-malware	ms17-	d2web	0.68
ep-malware	hp1_impressions	honeypots	1.13
mal-dest.	ntp	vulnerability	0.81
mal-dest.	usb	twitter	0.81
mal-dest.	breaches	blogs	0.81
mal-dest.	iot	d2web	0.81
mal-dest.	hp10_ctr	honeypots	1.00
mal-email	windows_vista	vulnerability	0.51
mal-email	udp	twitter	0.53
mal-email	windows7	blogs	0.54
mal-email	hacker	d2web	0.53
mal-email	hp1_clicks	honeypots	1.02

Table 3: Mean RMSE Lift of GRU model trained on a single signal across 28 weeks for an individual endpoint. Evaluated on OrgA.

Event Type	Signal	Source	Relative Lift
ep-malware	win_server_2008	vulnerability	0.99
ep-malware	blackmail	twitter	1.03
ep-malware	dhcp	blogs	1.03
ep-malware	ms17-	d2web	1.02
ep-malware	hp10_clicks	honeypots	0.95
mal-dest.	iphone_os	vulnerability	0.99
mal-dest.	dhcp	twitter	0.98
mal-dest.	cve	blogs	0.98
mal-dest.	trojans	d2web	0.98
mal-dest.	hp1_ctr	honeypots	1.27
mal-email	redhat_desktop	vulnerability	1.00
mal-email	dhcp	twitter	1.01
mal-email	usb	blogs	1.00
mal-email	zeroday	d2web	1.01
mal-email	hp1_ctr	honeypots	0.97

Table 4: Mean RMSE Lift of GRU model trained on a single signal across 28 weeks for an individual endpoint. Evaluated on OrgB.

Event Type	Signal	Source	Relative Lift
ep-malware	msoft_win_10	vulnerability	1.04
ep-malware	crypto	twitter	1.20
ep-malware	cve	blogs	1.04
ep-malware	ms17-	d2web	1.14
ep-malware	hp10_ctr	honeypots	2.19
mal-dest.	adobe_reader	vulnerability	1.00
mal-dest.	malware	twitter	1.00
mal-dest.	hijacked	d2web	1.00
mal-dest.	intel	blogs	1.01
mal-dest.	hp1_ctr	honeypots	1.28
mal-email	adobe_acrobat	vulnerability	1.06
mal-email	breaches	twitter	1.02
mal-email	ms17-	blogs	1.11
mal-email	0day	d2web	1.08
mal-email	hp1_ctr	honeypots	5.20

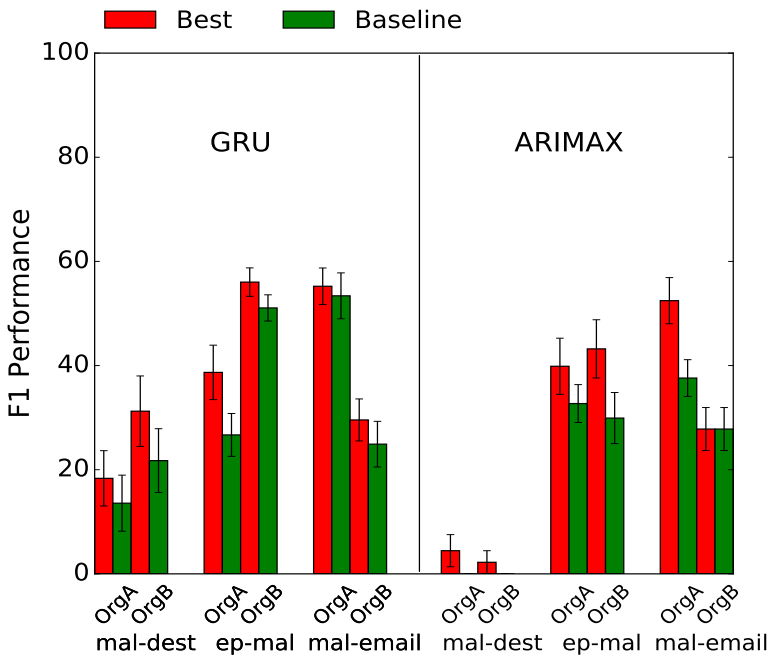

Figure 3: Comparing weekly models trained on the best signal for each configuration against baseline.

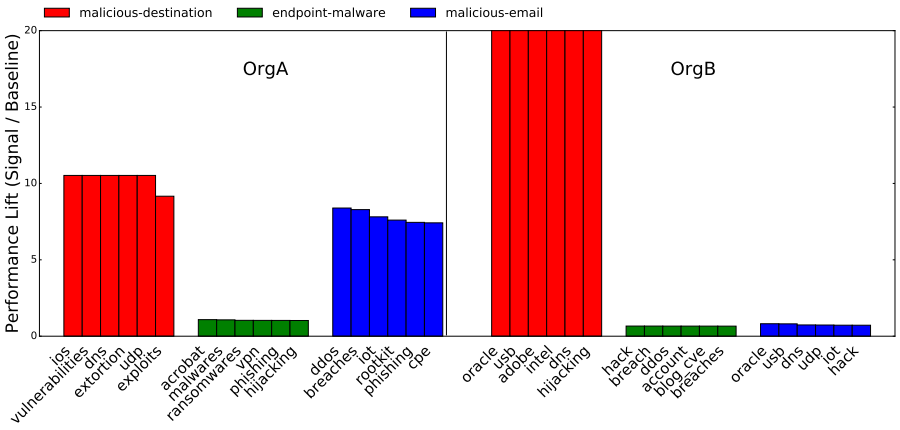

Figure 4: Monthly ARIMAX F1 performance on blog signals.

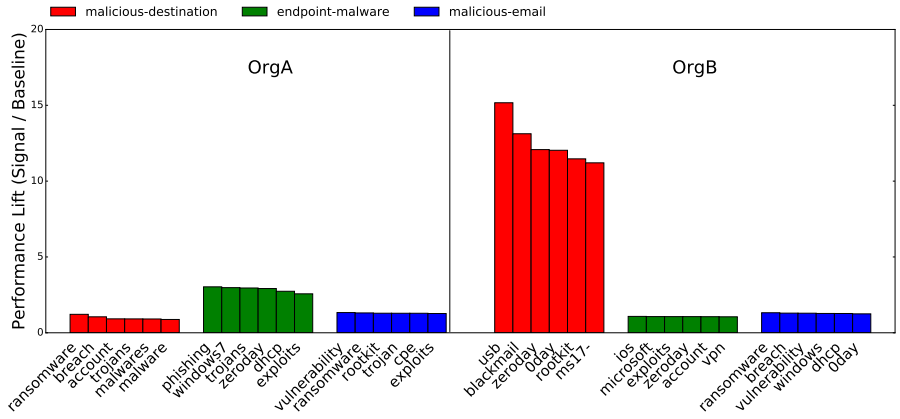

Figure 5: Monthly GRU F1 performance on blog signals.

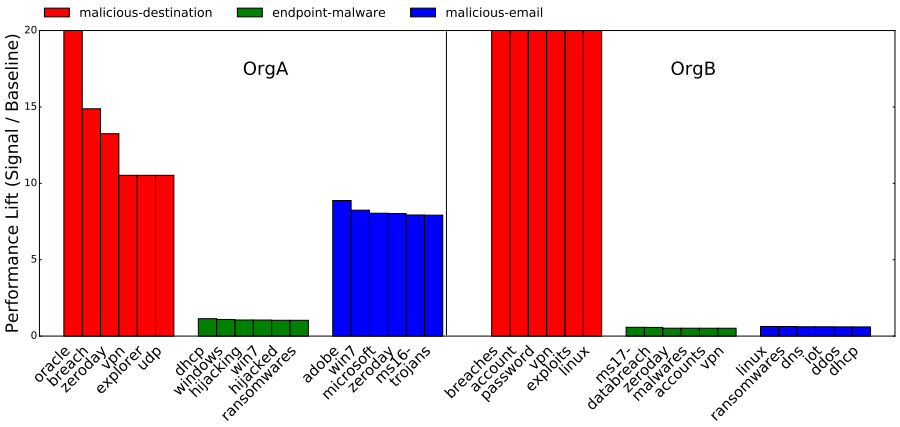

Figure 6: Monthly ARIMAX F1 performance on d2web signals.

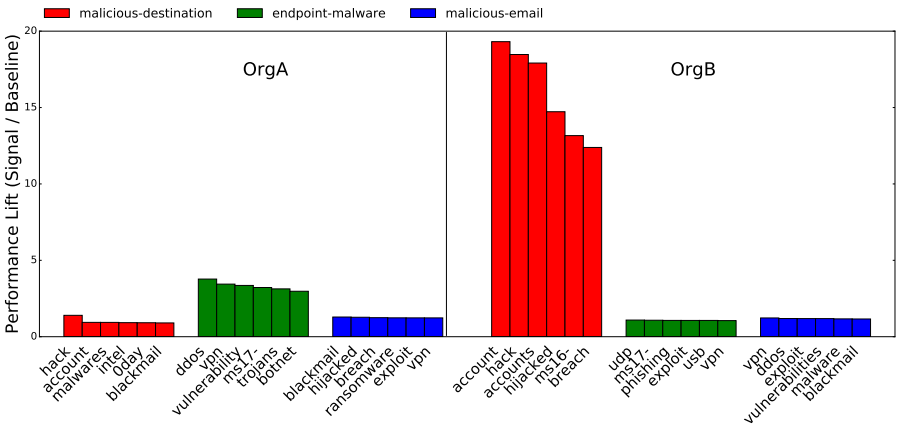

Figure 7: Monthly GRU F1 performance on d2web signals.

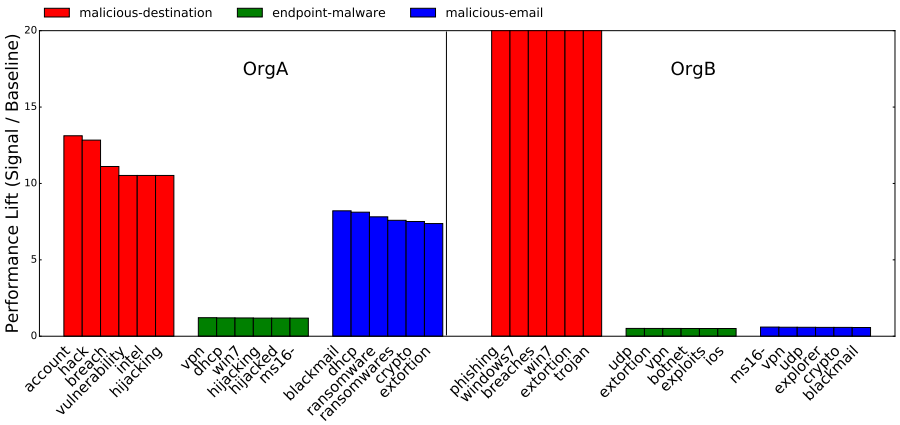

Figure 8: Monthly ARIMAX F1 performance on Twitter signals.

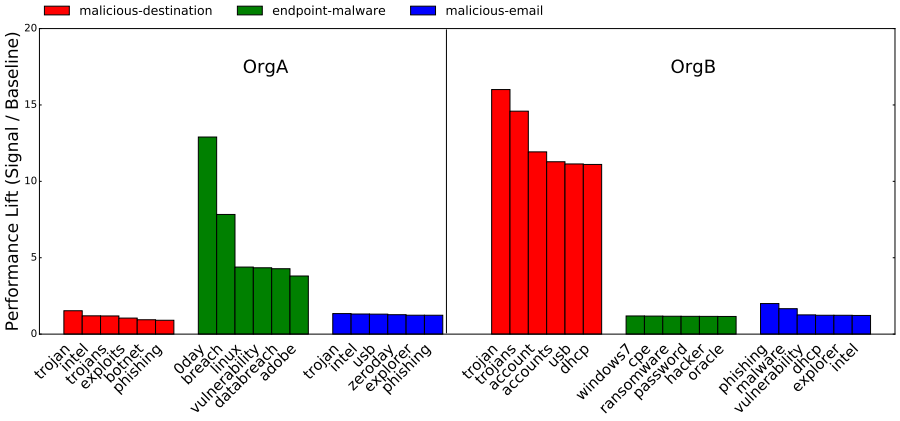

Figure 9: Monthly GRU F1 performance on Twitter signals.

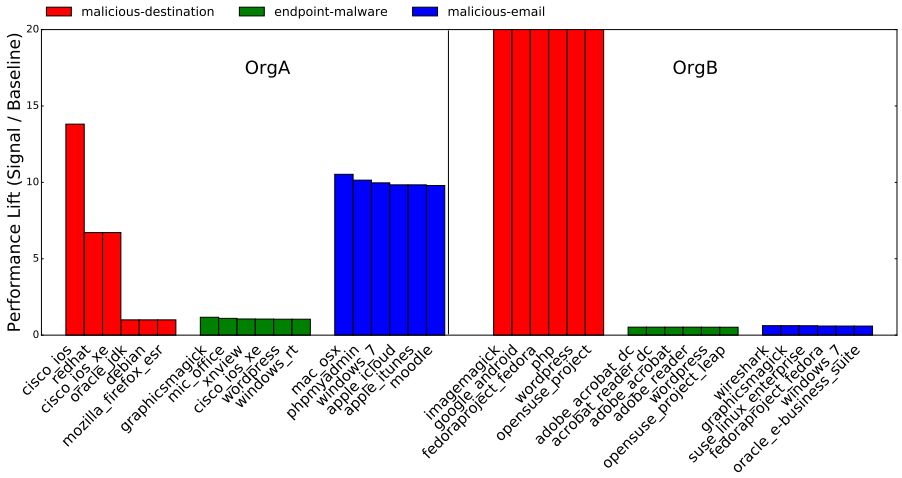

Figure 10: Monthly ARIMAX F1 performance on vulnerability signals.

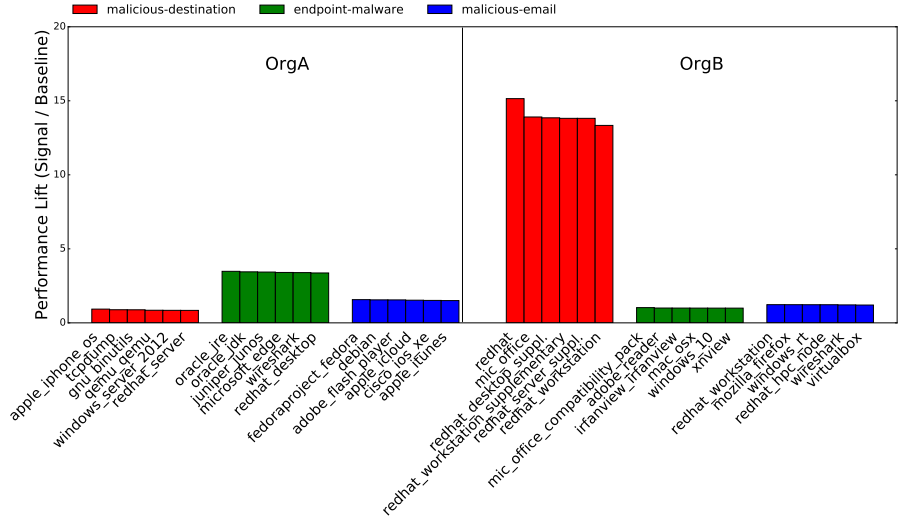

Figure 11: Monthly GRU F1 performance on vulnerability signals.

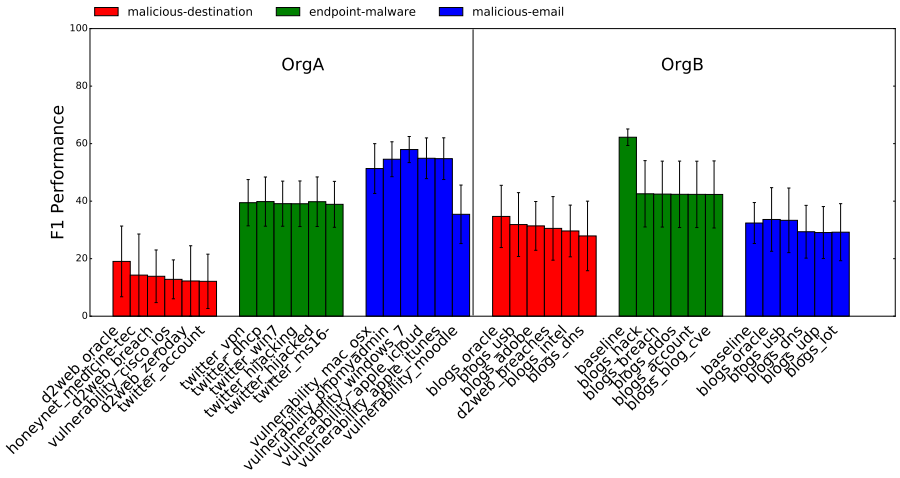

Figure 12: F1 Performance of best signals for ARIMAX Monthly

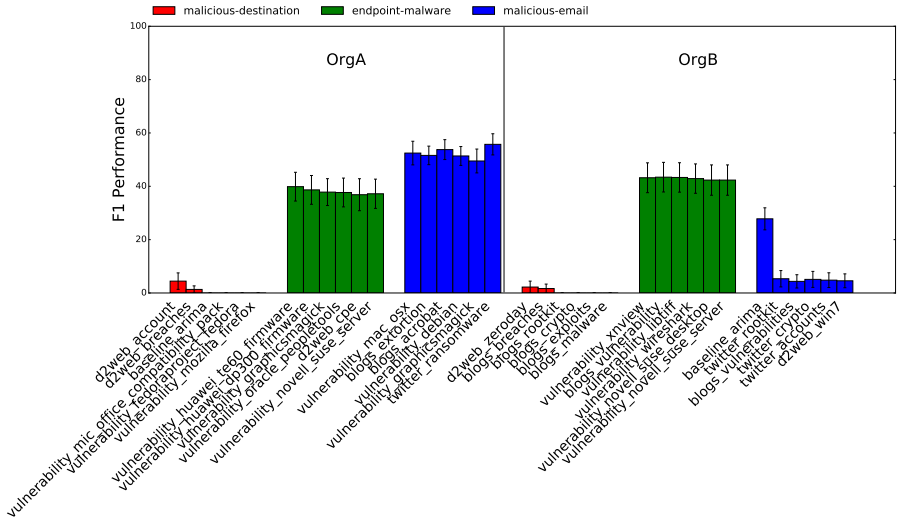

Figure 13: F1 Performance of best signals for ARIMAX Weekly

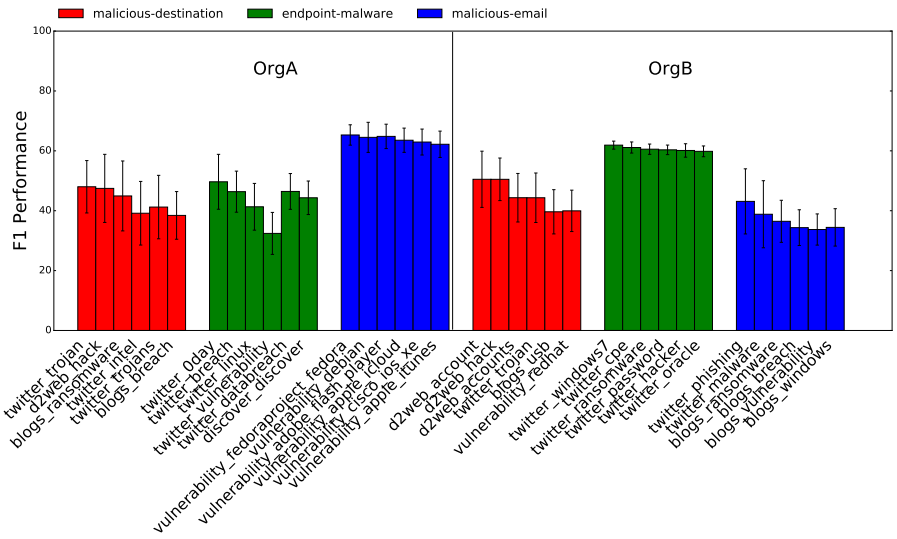

Figure 14: F1 Performance of best signals for GRU Monthly

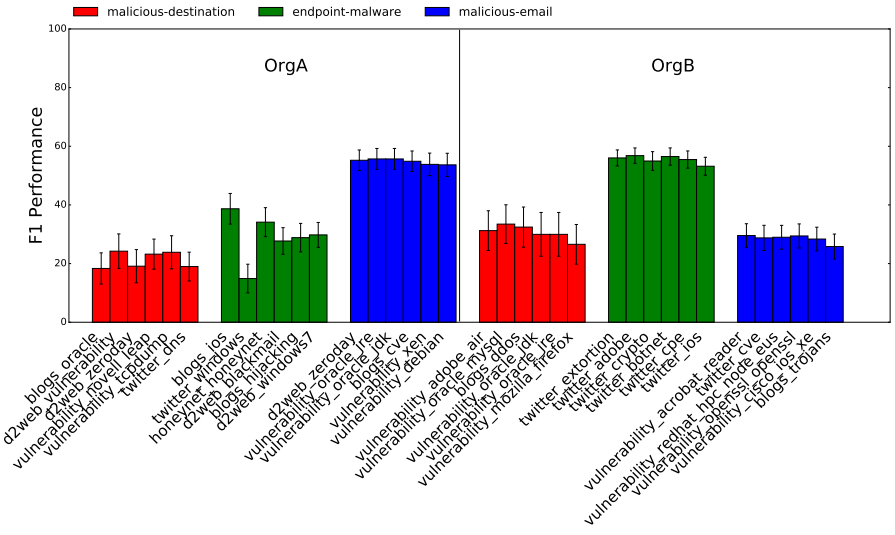

Figure 15: F1 Performance of best signals for GRU Weekly