DeepSquatting: Learning-based Typosquatting Detection at Deeper Domain Levels

Paolo Piredda[2], Davide Ariu[1,2], Battista Biggio[1,2], Igino Corona[1,2], Luca Piras[1,2], Giorgio Giacinto[1,2], and Fabio Roli[1,2]

1. Pluribus One, Italy – **http://www.pluribus-one.it**
2. PRA Lab – Pattern Recognition and Applications Lab – Dept. of Electrical and Electronic Eng., University of Cagliari, Italy

Pluribus One
seeing one in many

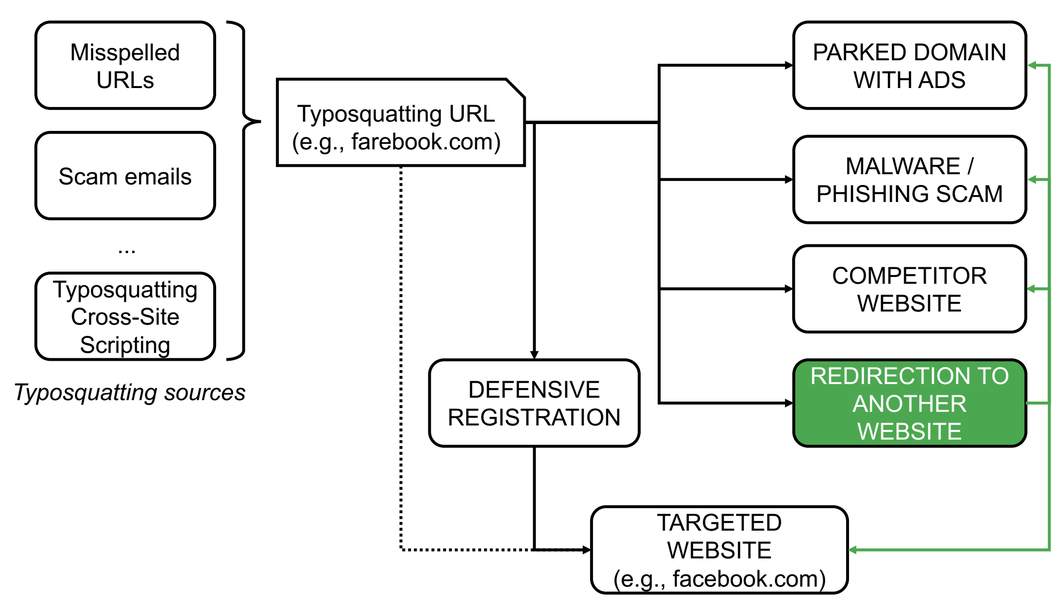
PRA Lab

University
Of Cagliari

Department of Electrical
and Electronic
Engineering

TypoSquatting

- ✓ Domain names registered by cybercriminals to resemble legitimate, reputable, and well-known websites (e.g., farebook.com *vs* facebook.com)
- ✓ **Goal:** to harvest and monetize Internet traffic originally destined to the legitimate, targeted website, exploiting its popularity and potential user typing errors



Related Work

- ✓ Typosquatting domains detected by *generating* all 2LD typo-variants within an *edit distance* of 1 or 2 from legitimate domains
- ✓ Identifying suspicious domains from DNS traffic within an *edit distance* of 1 from legitimate domains
- ✓ **Limitations:** typosquatting domains may contain more than one/two typos, and occur at deeper domain levels (not only at the 2LD)

Our Work

- ✓ Typosquatting domain detection based on passive analysis of DNS traffic at the ISP level
- ✓ **Main idea:** to learn a similarity measure of domain names from data with machine learning
- ✓ We use *n*-grams to detect similar domain names and *typosquatting patterns* (no assumptions on the number of typos) also at lower domain levels (not only at the 2LD)

Legitimate domain "google" vs typosquatting ""gooooooogle"

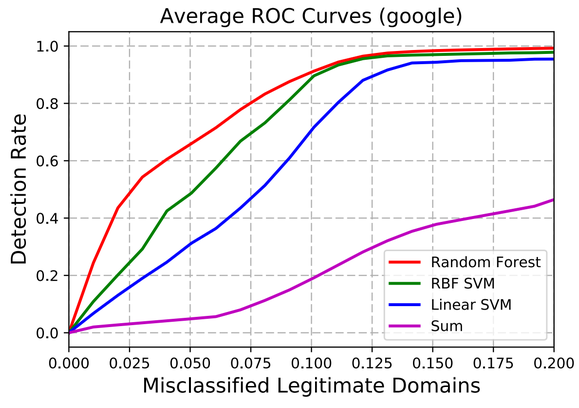
`#google$ vs #gooooooogle$ → #g go oo og gl le e$ vs #g go oo ... oo og gl le e$ → [1111111]`

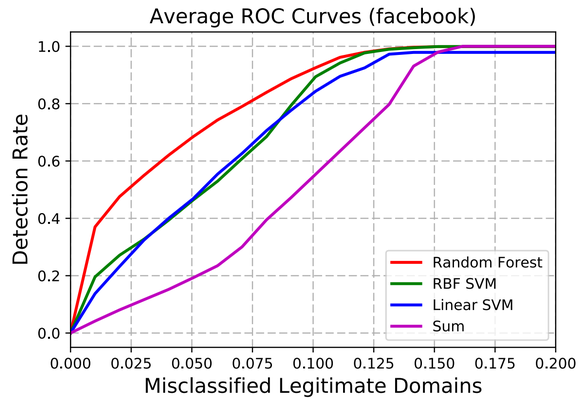
Features are computed by intersecting the n-grams of the typosquatting domain with those of the legitimate domain, at 2LD, 3LD, 4LD, and from 5LD to 10LD, and then concatenated to form a single vector

Machine-learning algorithms (SVM, RBF SVM and Random Forests) trained on this representation

Experiments

- ✓ DNS traffic from large ISP (Aug-Nov 2016)
- ✓ Ground-truth labeling using blacklisting services and manual inspection
- ✓ Typosquatting vs *Google* and *Facebook*





google	DL = 0			DL = 1			DL >1			Overall (DL ≥0)		
	True	*Detected*		*True*	*Detected*		*True*	*Detected*		*True*	*Detected*	
2LD	0	0		576	458	79,5%	412	328	79,6%	988	786	79,6%
3LD	305	162	53,1%	17	5	29,4%	97	63	64,9%	419	230	54,9%
4LD	483	50	10,4%	13	10	76,9%	193	43	22,3%	689	103	14,9%
5LD	161	27	16,8%	0	0		54	31	57,4%	215	58	27,0%
6LD	55	24	43,6%	1	0	0,0%	34	16	47,1%	90	40	44,4%
7LD	17	11	64,7%	0	0		4	1	25,0%	21	12	57,1%
8LD+	8	4	50,0%	0	0		11	7	63,6%	19	11	57,9%
Total										2441	1240	50,8%

facebook	DL = 0			DL = 1			DL >1			Overall (DL ≥0)		
	True	*Detected*		*True*	*Detected*		*True*	*Detected*		*True*	*Detected*	
2LD	0	0		387	314	81,1%	928	134	14,4%	1315	448	34,1%
3LD	347	26	7,5%	16	14	87,5%	334	282	84,4%	697	322	46,2%
4LD	216	69	31,9%	2	0	0,0%	352	342	97,2%	570	411	72,1%
5LD	83	31	37,3%	0	0		7	2	28,6%	90	33	36,7%
6LD	22	0	0,0%	0	0		0	0		22	0	0,0%
7LD	1	1	100,0%	0	0		89	89	100,0%	90	90	100,0%
8LD+	11	5	45,5%	0	0		0	0		11	5	45,5%
Total										2795	1309	46,8%