Luigi Gallo

Tutor: Prof. Alessio Botta

XXXIV Cycle - I year presentation

Identifying cyber threats and fraud in the context of large companies

Template

- CONTENT
  - Cover
  - Your background
    - Graduation MS, DIETI group, cooperations (mostly written)
    - Type of fellowship
  - Your problem
    - Specific (1 minutes)
  - Your research activity (3 minutes)
    - idea, methodology, developments, expected results, validation
  - Your products
    - List and mention
  - Next years
    - I year credits (table, mark in red if discrepancies occurs with PhD web site table)
    - Specific objects(say)
    - Table for training (expected credits) no words

My Background

Graduation MS - Computer Engineering

Università di Napoli 'Federico II'

Thesis : A system for network anomaly detection with Big Data analysis techniques.

Research Group: TRAFFIC (part of the larger COMICS).

Cooperation with the Cyber Security Lab at Telecom Italia Lab (TIM S.p.A)

Motivations (1/2)

The growth of dependence on cyberspace offers

- new opportunities
- new business scenarios
- new applications

but also 

- new threats
- new vulnerabilities
- new motives for cybercrime

Motivations (2/2)

Machine learning techniques, big data technologies, social engineering etc. work (mainly) on the attacker's side!!

The defense is forced to chase. It is important to investigate more about their use for cyber security purposes in several scenarios. ML-based commercial solutions for defense are not yet consolidated.

For example: Email is currently one of the most used channels for making cyber attacks and very often companies fall victim of financial fraud.

Research Activity (1/3)

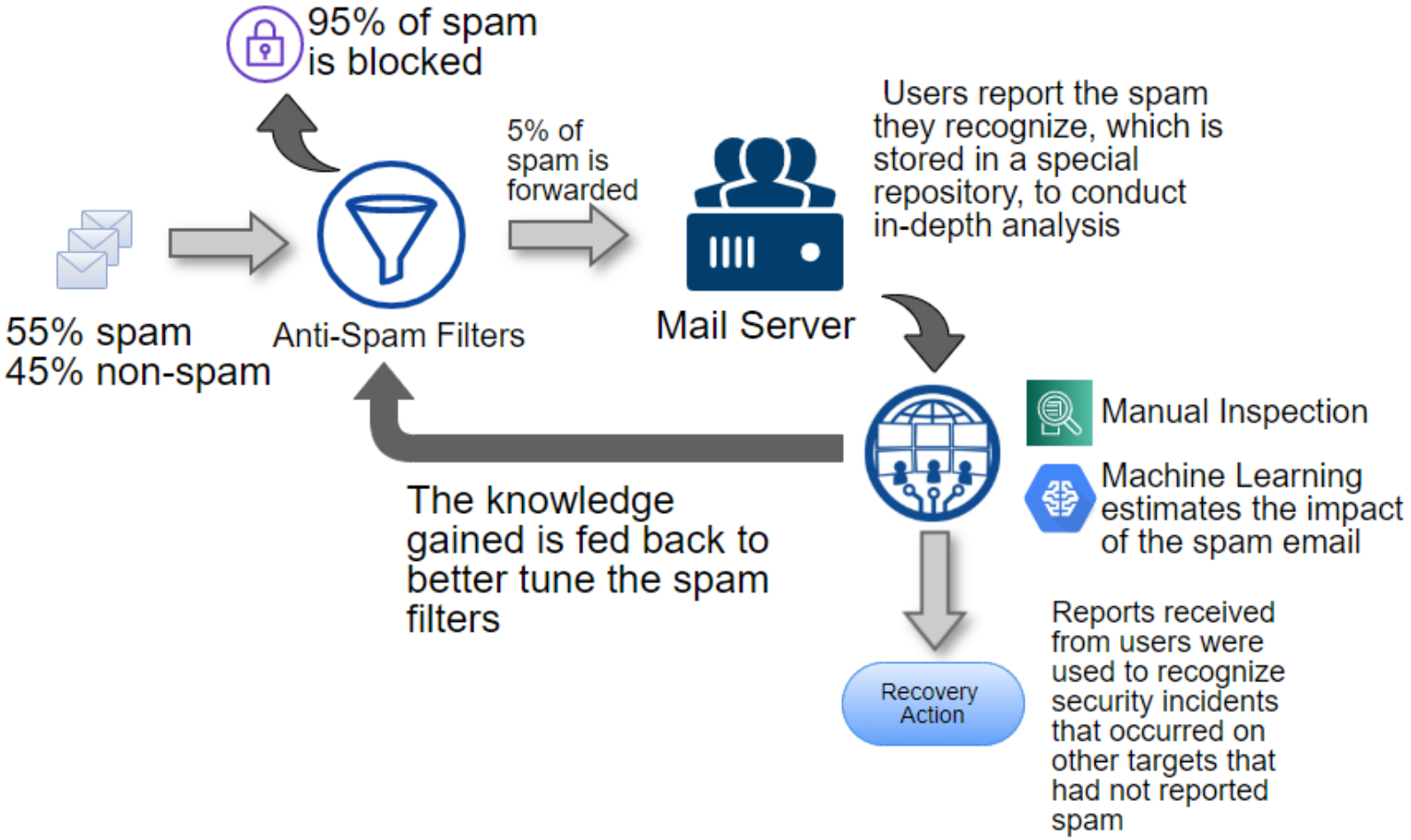
- Study of the literature: books, journals, conferences
- Study of the existing techniques for spam and phishing defense

Recap: anti-spam filters still have major lacks and are easily cheated by evasion attack and poisoning attack.

Outcome: the need of entire groups of anti-phishing analysts flooded with spam reports (mostly innocuous).

Idea: a classification that highlights “the needle in the haystack”

Scenario


95% of spam is blocked
5% of spam is forwarded
Users report the spam they recognize, which is stored in a special repository, to conduct in-depth analysis
55% spam
45% non-spam
Anti-Spam Filters
Mail Server
Manual Inspection
Machine Learning estimates the impact of the spam email
The knowledge gained is fed back to better tune the spam filters
Recovery Action
Reports received from users were used to recognize security incidents that occurred on other targets that had not reported spam

Research Activity (2/3)

Collecting Datasets (during one year)

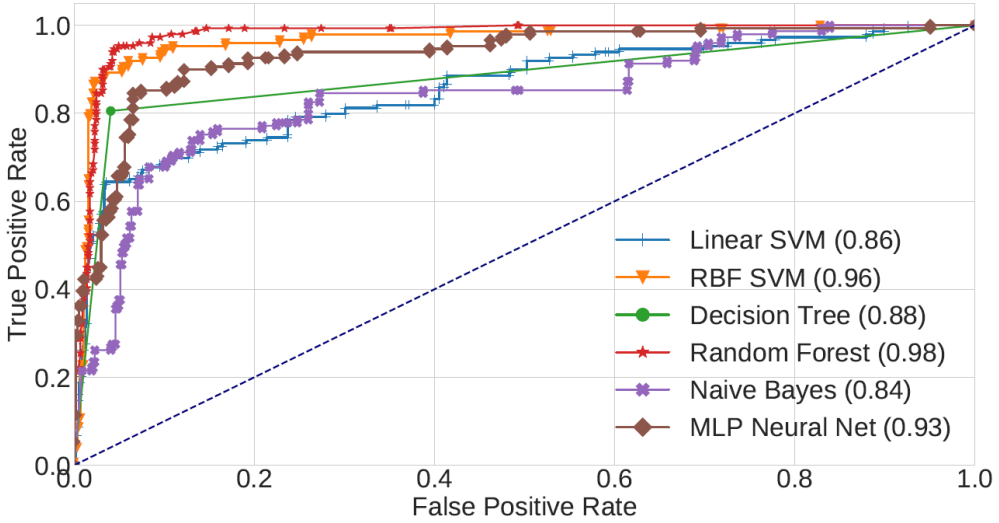
- Spam emails reported by employess of the company (40k samples at the moment)
- Security incidents occurred and recovery actions made by SOC analyst
- User behaviour (collected through awareness campaigns and interviews)
- OSINT

Research Activity (3/3)

Many machine learning algorithms and feature sets have been already tested. Data preparation and Feature Selection are crucial.

Expected results:

- characterization of the phenomenon and its evolution;
- insights about the most Incisive features for an attack to get the target;
- information to develop an automatic system that prioritizes spam reports;
- precise information on where users are weak, so that they can be properly educated.



Validation will take place (is already taking place) directly “on the field”

Other Research Activities

- Anomaly Detection in traffic traces: using Spark for detecting Scans and Dos attacks.
- Security issues of 5G networks.
- Architectures for Robotics applications.

Products

About the main research activity

- *L. Gallo, A. Botta, G. Ventre. Identifying threats in a large company's inbox, Big-DAMA'19: ACM CoNEXT Workshop on Big DAta, Machine Learning and Artificial Intelligence for Data Communication Networks*

About other research activities

- *A. Botta, L. Gallo, G. Ventre. Cloud, Fog, and Dew Robotics: architectures for next generation applications, IEEE Mobile Cloud 2019*
- Poster at : *A. Affinito, A. Botta, L. Gallo, M. Garofalo, G. Ventre. SPADA: SPark-based AnomalyDetection Ace, ACM CoNEXT 2018*
- Conference Paper (Under Review): *A. Affinito, A. Botta, L. Gallo, M. Garofalo, G. Ventre. Spark-Based Port and Net Scan Detection, ACM SAC Conference 2020*

Next Year

Study:

- The cognitive and psychological aspects exploited by spammers to deceive recipients, in order to expand the feature set and improve performance.
- Deep learning approaches, Natural language processing and Sentiment Analysis.

Production:

- Journal papers integrating the conference papers listed above

	Credits year 1								Credits year 2								Credits year 3									
		1	2	3	4	5	6			1	2	3	4	5	6			1	2	3	4	5	6			
	Estimated	bimonth	bimonth	bimonth	bimonth	bimonth	bimonth	Summary	Estimated	bimonth	bimonth	bimonth	bimonth	bimonth	bimonth	Summary	Estimated	bimonth	bimonth	bimonth	bimonth	bimonth	bimonth	Summary	Total	Check
Modules	20	1,6	0	3	0	6	5	16	14							0								0	16	30-70
Seminars	5	0,4	0	0,4	1,5	0	0	2,3	6							0								0	2,3	10-30
Research	35	8	10	6,6	8,5	4	5	42	40							0								0	42	80-140
	60	10	10	10	10	10	10	60	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60	180