UNIVERSITÀ DEGLI STUDI DI PADOVA

Simone Milani
Room 216
DEI A

Phone: 049 827 7641
E-mail:
simone.milani@dei.unipd.it

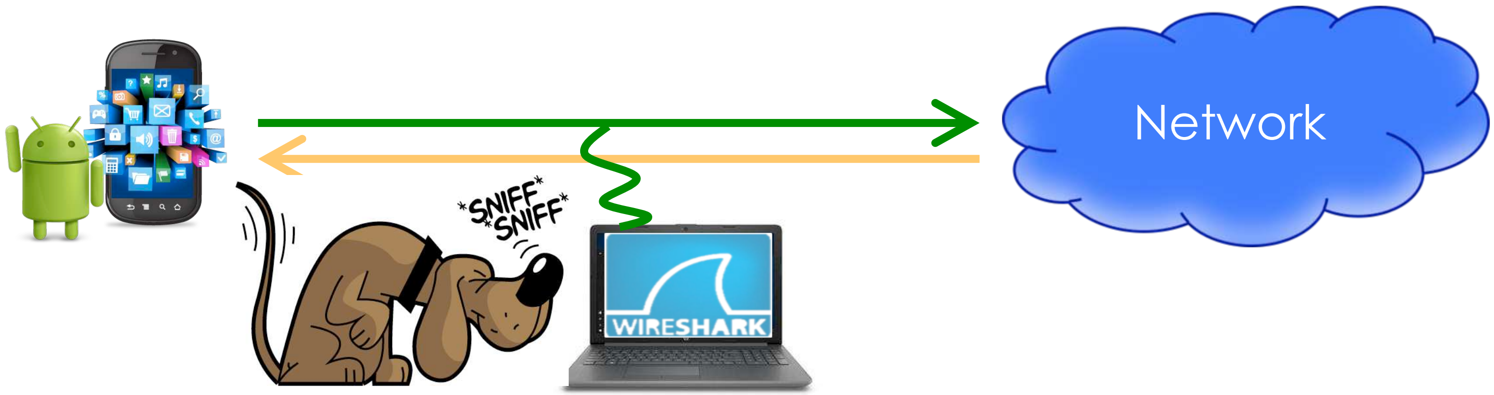
Digital Forensics

A.A. 2018/2019

Lab experience n.1 – Network forensics

Scenario 1: traffic classification

- Let us assume we are sniffing the packets from an android phone
- Monitoring encrypted traffic: only packet lengths are available

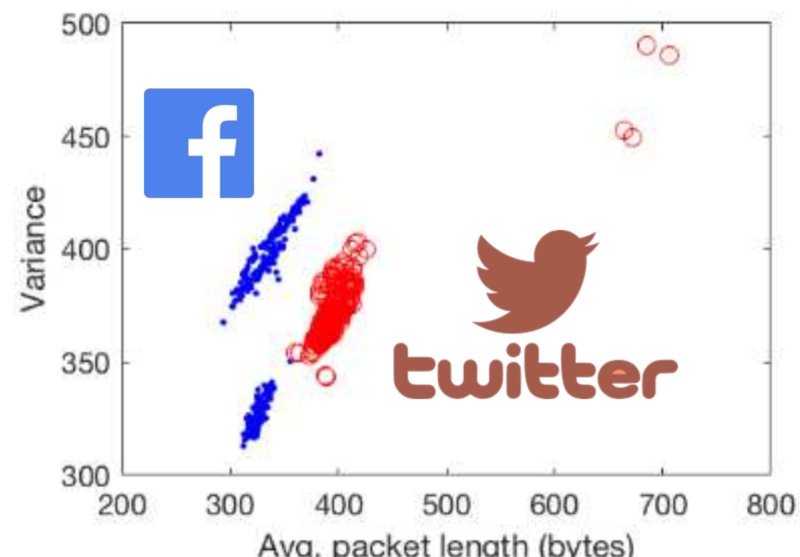


Features: since packets are encrypted we need some high level feature

e.g. discriminating between Twitter and Facebook, average packet length and their variance.

$$x_1 = L = \frac{1}{N}\sum_i l(p_i)$$

$$x_2 = \sigma_L = \sqrt{\frac{1}{N}\sum_i \left(l(p_i) - \overline{L}\right)^2}$$



PCAP files

PCAP files

Stores information about sniffed streams (tools: libpcap)

Global Header	Packet Header	Packet data	Packet Header	Packet data	Packet Header	Packet data	…

Global header

guint32 magic_number;	0xa1b2c3d4 0xd4c3b2a1	Defines format and ordering
guint16 version_major;		
guint16 version_minor;		
gint32 thiszone;	*Correction (in sec) between time zone and GMT (UTC)*	
guint32 sigfigs;	*Accuracy of time stamp*	
guint32 snaplen;	"snapshot length" for the capture	
guint32 network;	link-layer header type,	

Packet header

guint32 ts_sec;	timestamp (in sec)
guint32 ts_usec;	timestamp (in microsec)
guint32 incl_len;	Length of stored bytes from packet
guint32 orig_len;	Length (real) of the transmitted packet

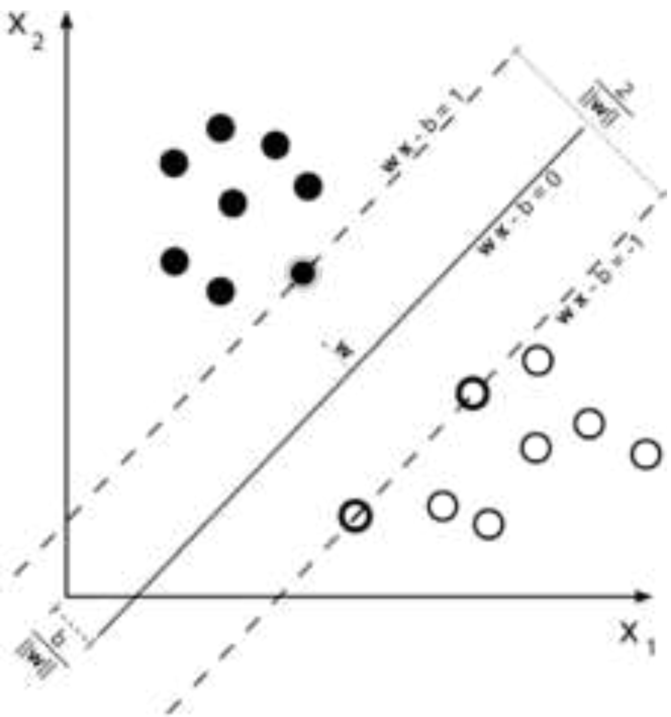
PCAP->CSV files

app	action	seq.	action start	flow nr.	IP dest.	IP dest. resol.	port source	port dest.	flow length	flow start	packets length total
facebook	open facebook	1	1383129102.11	0	x		14146	53	2	XXXX	[79, -185]
facebook	open facebook	1	1383129102.11	1	x		18748	53	2	XXXX	[78, -203]
facebook	open facebook	1	1383129102.11	2	y	a	47559	443	29	XXXX	[250, -1514,...]
facebook	open facebook	1	1383129102.11	3	y	b	42963	443	30	XXXX	[250, -1514,...]
facebook	open facebook	1	1383129102.11	4	x		7633	53	2	XXXX	[76, -183]
facebook	open facebook	1	1383129102.11	5	y	a	33554	443	9	XXXX	[250, -1514,...]
facebook	open facebook	1	1383129102.11	6	y	a	47559	443	2	XXXX	[764, -1014]
facebook	open facebook	1	1383129102.11	7	y	a	42599	443	31	XXXX	[282, -1514, ...]
facebook	open facebook	1	1383129102.11	8	y	b	42963	443	5	XXXX	[92, 92, 93, ...]
facebook	open facebook	1	1383129102.11	9	y	c	49785	443	46	XXXX	[250, -1514, ...]
facebook	open facebook	1	1383129102.11	10	x		39051	53	2	XXXX	[78, -206]
facebook	open facebook	1	1383129102.11	11	y	b	42963	443	7	XXXX	[92, 92, 93,...]
facebook	menu selection	1	1383129213.08	12	y	b	42963	443	7	XXXX	[92, 92, 93, ...]
facebook	menu selection	1	1383129213.08	13	y	b	42963	443	10	XXXX	[92, 92, 93, ...]
facebook	menu selection	1	1383129213.08	14	y	a	33554	443	2	XXXX	[-93, 93]
facebook	menu selection	1	1383129213.08	15	y	a	42599	443	9	XXXX	[917, 1514, ...]
facebook	replacing menu in initial position	1	1383129244.01	16	y	a	47559	443	1	XXXX	[-93]
facebook	replacing menu in initial position	1	1383129244.01	17	y	c	49785	443	1	XXXX	[-93]
facebook	replacing menu in initial position	1	1383129244.01	18	y	b	42963	443	7	XXXX	[92, 92, 93, ...]
facebook	edit search selection	1	1383129275.01	19	y	b	42963	443	7	XXXX	[92, 92, 93,...]

Session of packets

NB: > 0 outgoing
<0 incoming

Binary linear separator



Given a set of points

$$(\mathbf{x}_1, y_1),\ (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)$$

Feature array

Binary label: either -1 or 1

The hyperplane that separates the points of each class is

$$\mathbf{w} \cdot \mathbf{x} - b = 0$$

Introducing a margin we have

$$\mathbf{w} \cdot \mathbf{x} - b = 1 \qquad \mathbf{w} \cdot \mathbf{x} - b = -1$$

The distance between the two margin is $\frac{2}{\|\mathbf{w}\|^2}$ which implies that robustness is increased if ||**w**|| is small.

The two classes can be defined as

$$\mathbf{w} \cdot \mathbf{x}_i - b \leq -1 \quad \text{for } y_i=-1$$

$$\mathbf{w} \cdot \mathbf{x}_i - b \geq 1 \quad \text{for } y_i=1$$

$$y_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1$$

Kernel-based classification

It is possible to have a representation of **w** in the space so defined

$$\mathbf{w} = \sum_i \alpha_i y_i \varphi(\mathbf{x}_i)$$

Note that $\mathbf{w}\varphi(\mathbf{x}) = \sum_i \alpha_i y_i \varphi(\mathbf{x}_i)\varphi(\mathbf{x}) = \sum_i \alpha_i y_i k(\mathbf{x}_i, \mathbf{x})$

In the linear case: $\mathbf{w} = \sum_i c_i y_i \mathbf{x}_i$

Commands to be used

Label 1:feat1 2:feat2 ...
Label 1:feat1 2:feat2 ...
Label 1:feat1 2:feat2 ...
Label 1:feat1 2:feat2 ...
Label 1:feat1 2:feat2 ...
.....

Create a .mat file to be read by libsvm

```
>>write_svm_file(mat1,mat2,'train.mat');
```

Create a classifier

```
$svm-train [options] train.mat classifier.mod
```

`train.mat:` training set file
`classifier.mod:` file with classifier parameters (support vector, etc.)

[Options]

`-t <n>` kernel function ('0': linear, '2': RBF)
`-v <n>` cross-validation; <n>: number of partitions
`-e <n>` tolerance for termination
`-b [0|1]` probability estimates

Predict input values

```
$svm-predict test.mat classifier.mod  output.txt
```

`output.txt:` classified labels
`test.mat:` training set file

Compute separating hyperplane

Compute hyperplane with linear kernels

```
% Compute line parameters
w = sum(vector(:,1).*vector(:,2:3));

% Compute slope intercept form
m = w(1)/w(2); % slope
q = rho/w(2); % intercept
```

Support vectors

$c_{ii} \bullet y_i$

$$\mathbf{w} = \sum_i c_i y_i \mathbf{x}_i$$

Find line equation

$$k(\mathbf{x}_i, \mathbf{x}) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}\|)$$

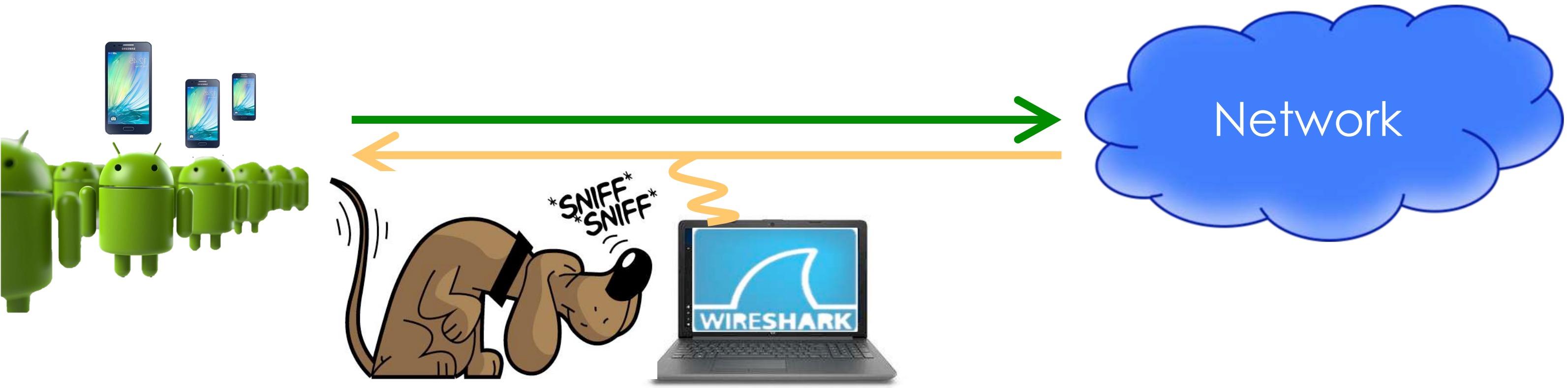
Compute hyperplane with RBF kernels

```
diff_sv_vet=zeros(nbf,length(X0(:)));
for isv=1:nbf
          diff_sv_vet(isv,:)=exp(-1*gamma*mean((ones(length(X0(:)),1)*…
                    vector(isv,2:3)-[X0(:) X1(:)]).^2,2));
          diff_sv_vet(isv,:)=vector(isv,1)*diff_sv_vet(isv,:);
end
val_line=sum(diff_sv_vet,1)-rho;   %line points
iii=find(abs(val_line)<0.05);      %for line points coordinates are close to 0
```

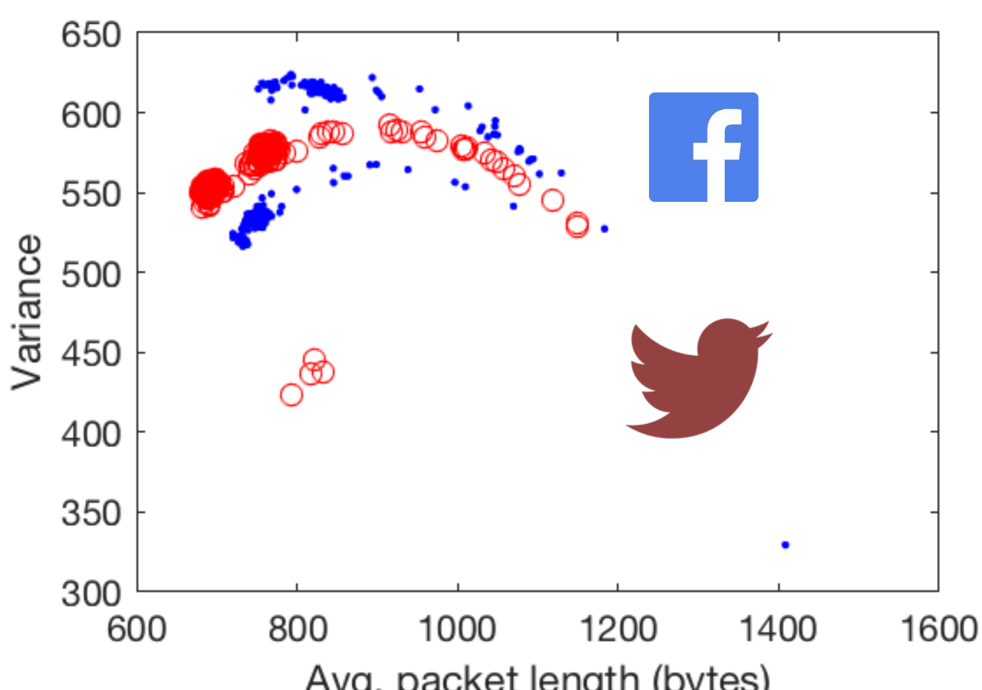
$$y_i k(\mathbf{x}_i, \mathbf{x})$$

$$\mathbf{w}\varphi(\mathbf{x}) = \sum_i c_i y_i k(\mathbf{x}_i, \mathbf{x})$$

Scenario 2: incoming traffic



What happens if we analyze the incoming traffic?



Question 1

① Compute a classifier with linear kernel for scenario 1
② Compute a classifier with linear kernel for scenario 2
③ Is it possible to improve the accuracy using different kernels?
④ What about overfitting?