QoS Profile Provisioning in 5GC using Naive Bayes Technique

WiSig Networks

Introduction

Traditional technique provide only 50-60% ACCURACY

This technique uses supervised ML algorithm

WiSig Networks

Objects and Discriminators

Application of classification scheme requires Parametrization of objects to be classified.

Using these parameters the classifier allocates object to class.

Hence object describing parameters are called discriminators

Fundamental object classified:(IP of pair hosts,Protocol Type)

WiSig Networks

Discriminators Used

Flow duration
TCP port
Packet inter arrival time
Payload size
Effective Bandwidth based upon Entropy
Fourier Transform of packet arrival time

WiSig Networks

Traffic Categories

BULK
DATABASE
INTERACTIVE
MAIL
SERVICES
WWW
P2P
ATTACK
GAMES
MULTIMEDIA

WiSig Networks

Machine Learning Classification

Deterministic classification

Assigns data points to mutually exclusive classes

Probabilistic classification

probabilistic classification methods classify data by assigning it with probabilities of belonging to each class of interest.

Reason for Probabilistic classification

Identify similar characteristic of flow

Tractable, Well documented and understood

Robust to measurement error

Allows for supervised training with pre-classified traffic

WiSig Networks

Bayes theorem

$$X = (x_1, x_2, x_3, \ldots.., x_n)$$

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)}$$

p(y) is the prior probability

p(X) is the evidence

Naive assumption

Independence among the features

$$P(y|x_1,...,x_n) = \frac{P(x_1|y)P(x_2|y)...P(x_n|y)P(y)}{P(x_1)P(x_2)...P(x_n)}$$

$$P(y|x_1,...,x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i|y)}{P(x_1)P(x_2)...P(x_n)}$$

$$P(y|x_1,...,x_n) \propto P(y)\prod_{i=1}^{n} P(x_i|y)$$

p(y) is class probability

p(xi/y) is conditional probability

WiSig Networks

Naive Bayesian Classifier

Data sample x = (x1,..........,xn) realization of X = {X1,.........,Xn} such that Xi is a random variable and each Xi is described by m attributes {A1,........,Am}(discriminators)

Assuming k known classes of interest Let C = {c1,........ck} all set of known classes Bayesian rule is given by

$$p(c_j \mid y) = \frac{p(c_j)f(y \mid c_j)}{\sum_{c_j} p(c_j)f(y \mid c_j)}$$

WiSig Networks

Naive Bayes:Gaussian Estimation

The main goal of naive bayes is to estimate $f(\cdot \mid c_j), j = 1, \ldots, k.$

Consider two probabilities for classes c1 and c2

$$p(c_1) = \frac{n_{c1}}{n},$$
$$p(c_2) = \frac{n_{c2}}{n}.$$

WiSig Networks

Naive Bayes:Gaussian Estimation

The normal gaussian distribution is given by:

$$f_{A_1|c_1}(x;\mu_1,\sigma_1^2) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$

$$f_{A_1|c_2}(x;\mu_2,\sigma_2^2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}.$$

WiSig Networks

Naive Bayes:Gaussian Estimation

$$
\begin{aligned}
\hat{\mu}_1 &= \sum_{x_i : C(x_i) = c_1} \frac{x_i}{n_{c_1}}, \\
\hat{\mu}_2 &= \sum_{x_i : C(x_i) = c_2} \frac{x_i}{n_{c_2}}, \\
\hat{\sigma}_1^2 &= \sum_{x_i : C(x_i) = c_1} \frac{(x_i - \hat{\mu}_1)^2}{n_{c_1} - 1}, \\
\hat{\sigma}_2^2 &= \sum_{x_i : C(x_i) = c_2} \frac{(x_i - \hat{\mu}_2)^2}{n_{c_2} - 1}.
\end{aligned}
$$

$$
p(c_i \mid y) = \frac{f_{A_1 \mid c_i}(y; \hat{\mu}_i, \hat{\sigma}_i^2) \times p(c_i)}{N},
$$

WiSig Networks

Problems with gaussian estimator

Different discriminators are clearly not independent

Assumptions of normality of discriminator is inaccurate

Naive Bayes:Kernel Estimation

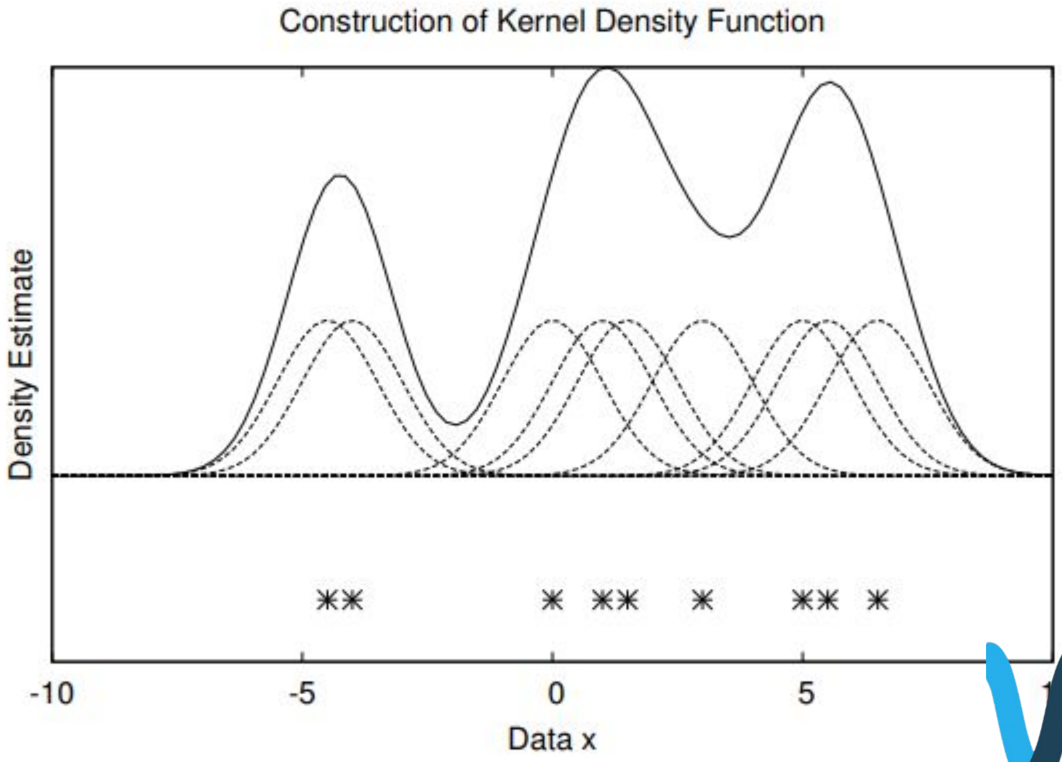
Addresses the problem of approximating every discriminator by a Gaussian distribution

$$\hat{f}(t \mid c_j) = \frac{1}{n_{c_j} h} \sum_{x_i : C(x_i) = c_j} K\left(\frac{t - x_i}{h}\right)$$

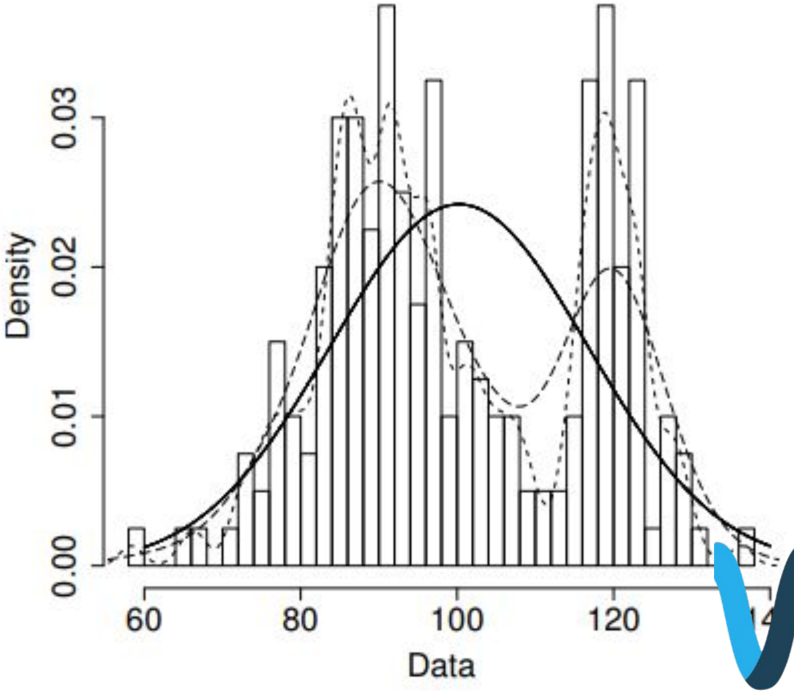
Where h is Bandwidth or smoothing factor measured my mean integrated square method

$$MISE(\hat{f}) = E\left[\int \left(\hat{f}(t) - f(t)\right)^2 dt\right]$$

Naive Bayes:Kernel Estimation



Comparison of Gaussian and Kernel density



Algorithmic complexity

	NAÏVE BAYES		KERNEL	
Operation	Time	Space	Time	Space
Train on n cases	$O(nm)$	$O(m)$	$O(nm)$	$O(nm)$
Test on q cases	$O(qm)$		$O(qnm)$	

Discriminator selection and Dimensionality reduction

Important for removing irrelevant and redundant discriminators.

The ability to identify the most important discriminators of the Internet traffic is useful not only because the results will reveal what discriminators are best for traffic classification, but also because classification accuracy can be improved and, a reduction in number of flow discriminators is computationally attractive

Discriminator selection and Dimensionality reduction

A discriminator is irrelevant if it carries no information of interest about class.

A discriminator is redundant if it is highly correlated with other discriminator.

Methods of discrimination:

Filter and wrapper

FCBF:Fast Correlation-Based Filter

In FCBC goodness of discriminator is measured by correlation with the class and other good attributes.

Correlation based on entropy:

The entropy of random variable X taking values in {x1,........,xn}

$$H(X) = -\sum_{x_i} p(x_i) \log_2 p(x_i),$$

FCBF:Fast Correlation-Based Filter

For random variable X and Y

$$H(X \mid Y) = -\sum_{y_j} p(y_j) \sum_{x_i} p(x_i \mid y_j) \log_2 p(x_i \mid y_j),$$

Information gain(measure of correlation between two discriminators

$$IG(X \mid Y) = H(X) - H(X \mid Y).$$

FCBF:Fast Correlation-Based Filter

Symmetric uncertainty

$$SU(X,Y) = 2\left[\frac{IG(X \mid Y)}{H(X)+H(Y)}\right]$$

The FCBF selects good discriminators by two steps:

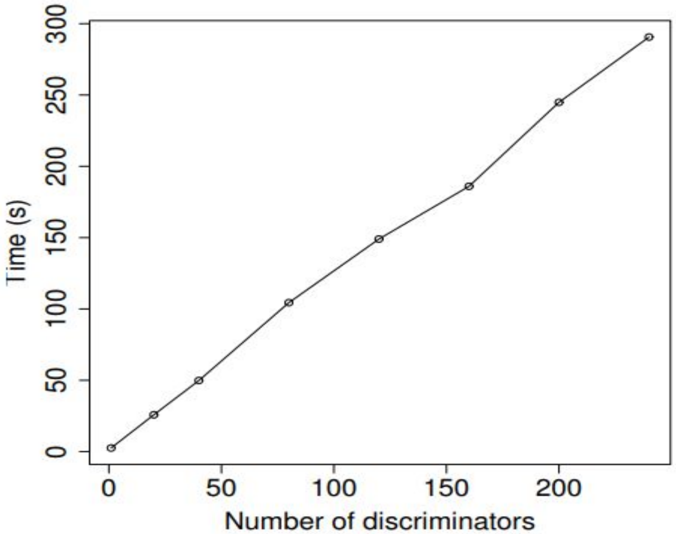
Identifying relevance of discriminator and identifying the redundancy of a feature

Algorithm for best no of discriminators

Algorithm

Experimental results

Time required to train and test data wrt discriminators



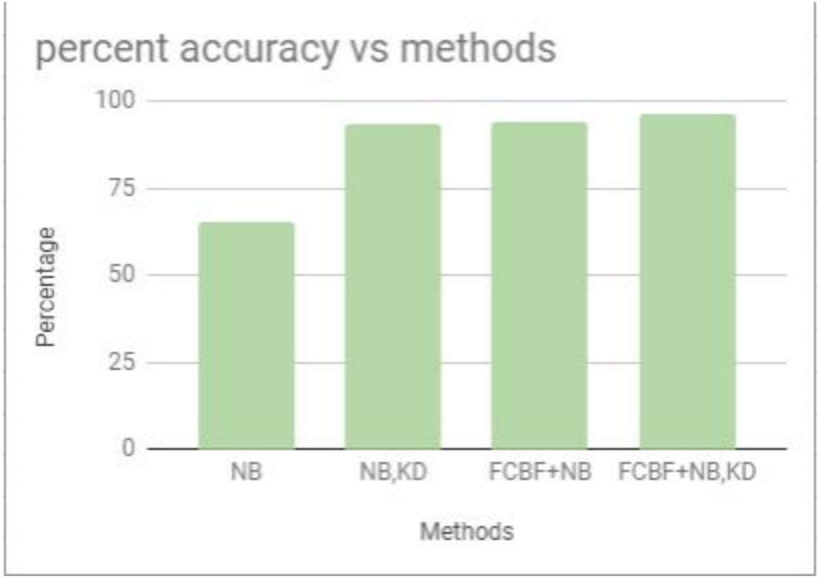
Definitions

Accuracy:Raw count of flows that were classified correctly divided by total no of flows

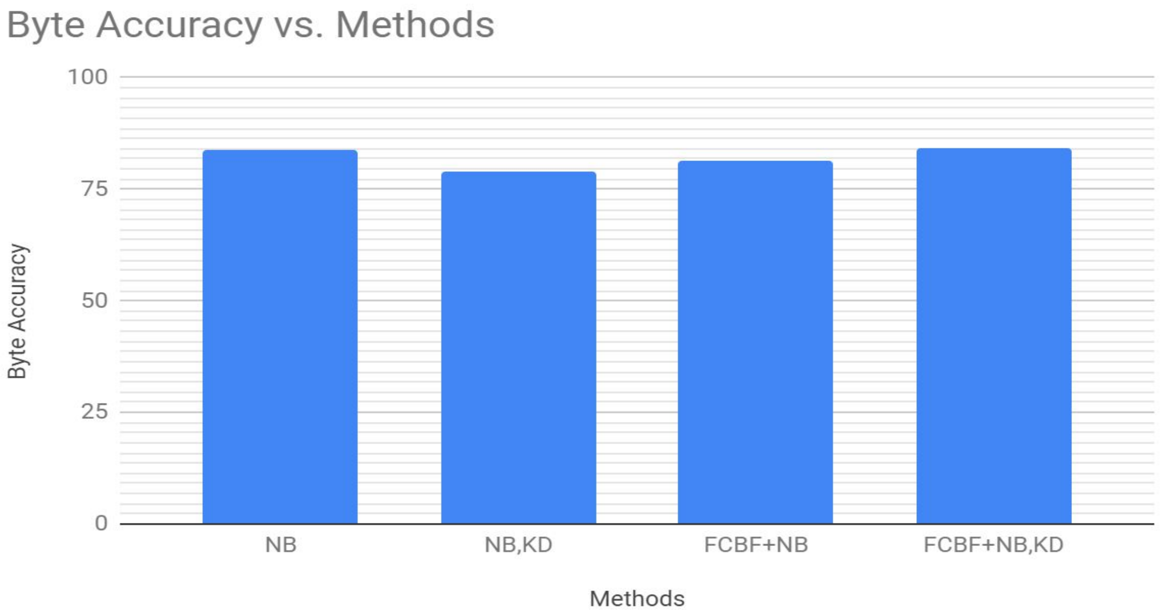
Trust:Indication of how much classification can be trusted

Accuracy by bytes:Raw count of correctly classified flow bytes divided by total no of byte flow

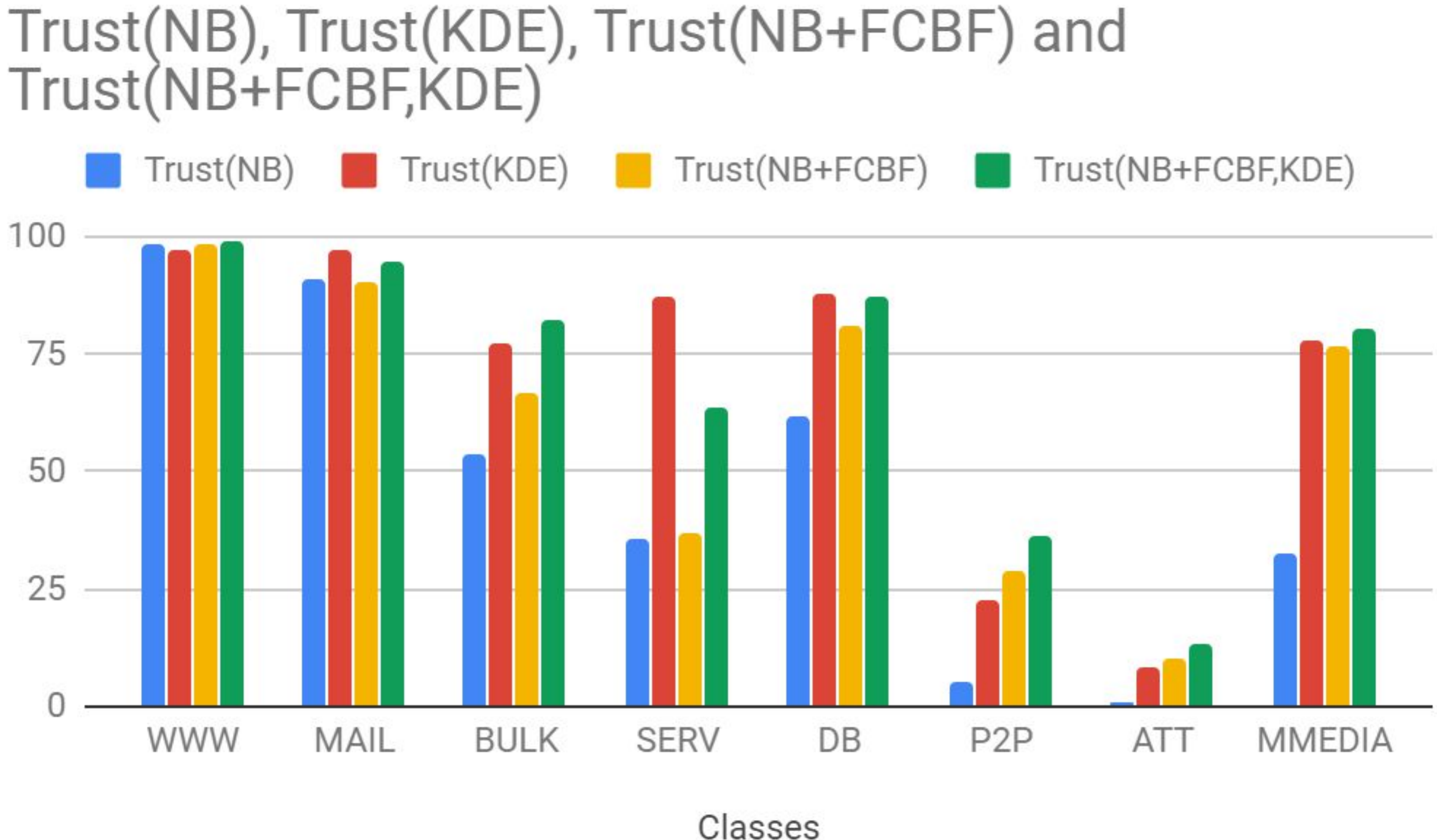
Average percentage of accurately classified flows



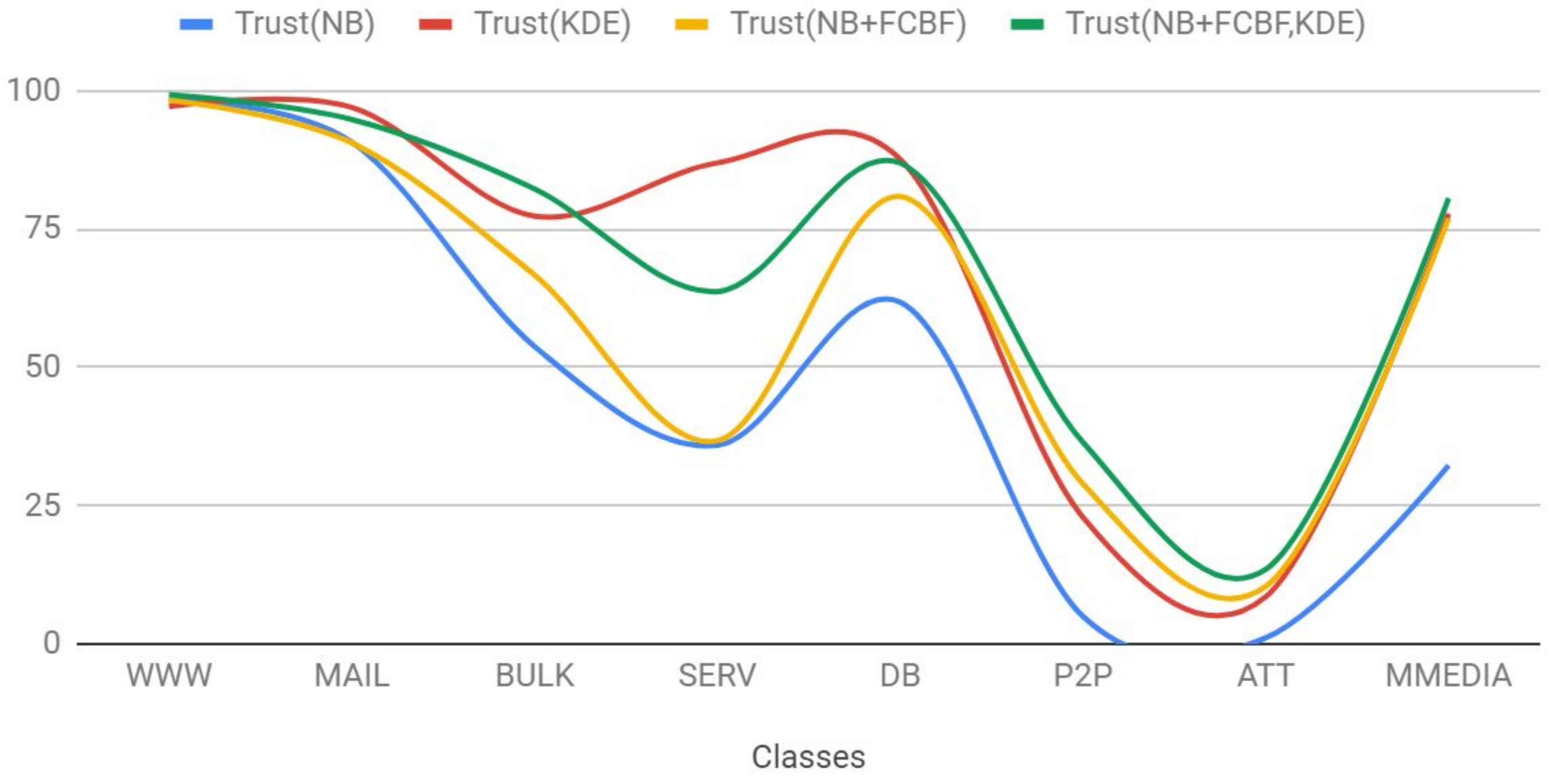
Average percentage of classified bytes by diff methods



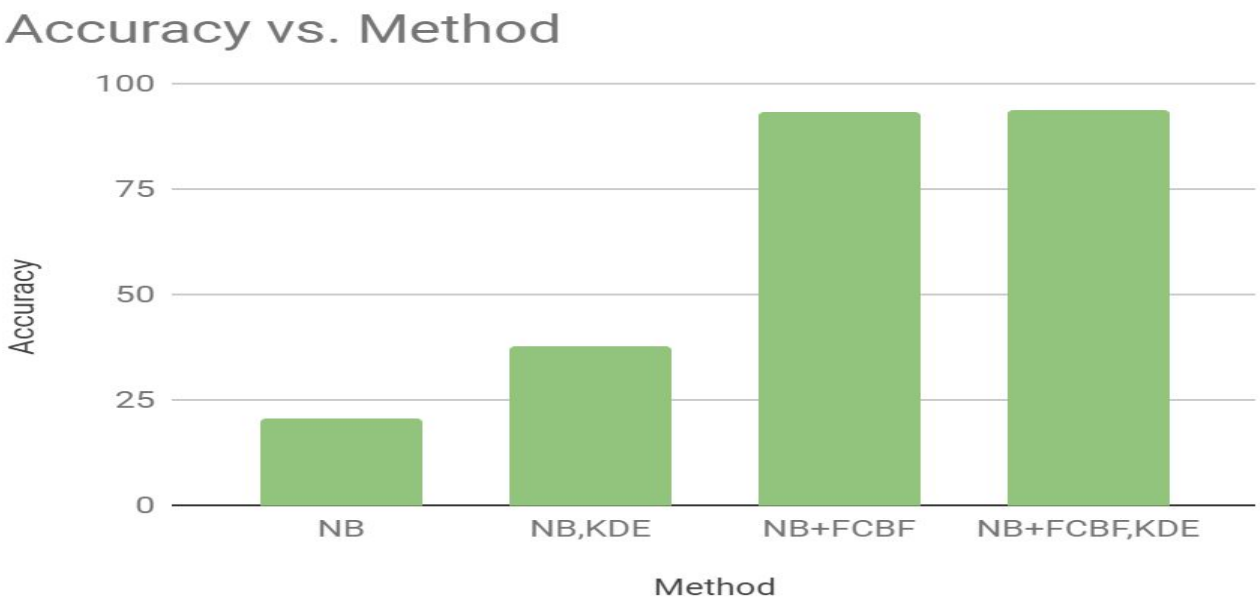
Measure of belief in certain class after all methods



Trust(NB), Trust(KDE), Trust(NB+FCBF) and Trust(NB+FCBF,KDE)

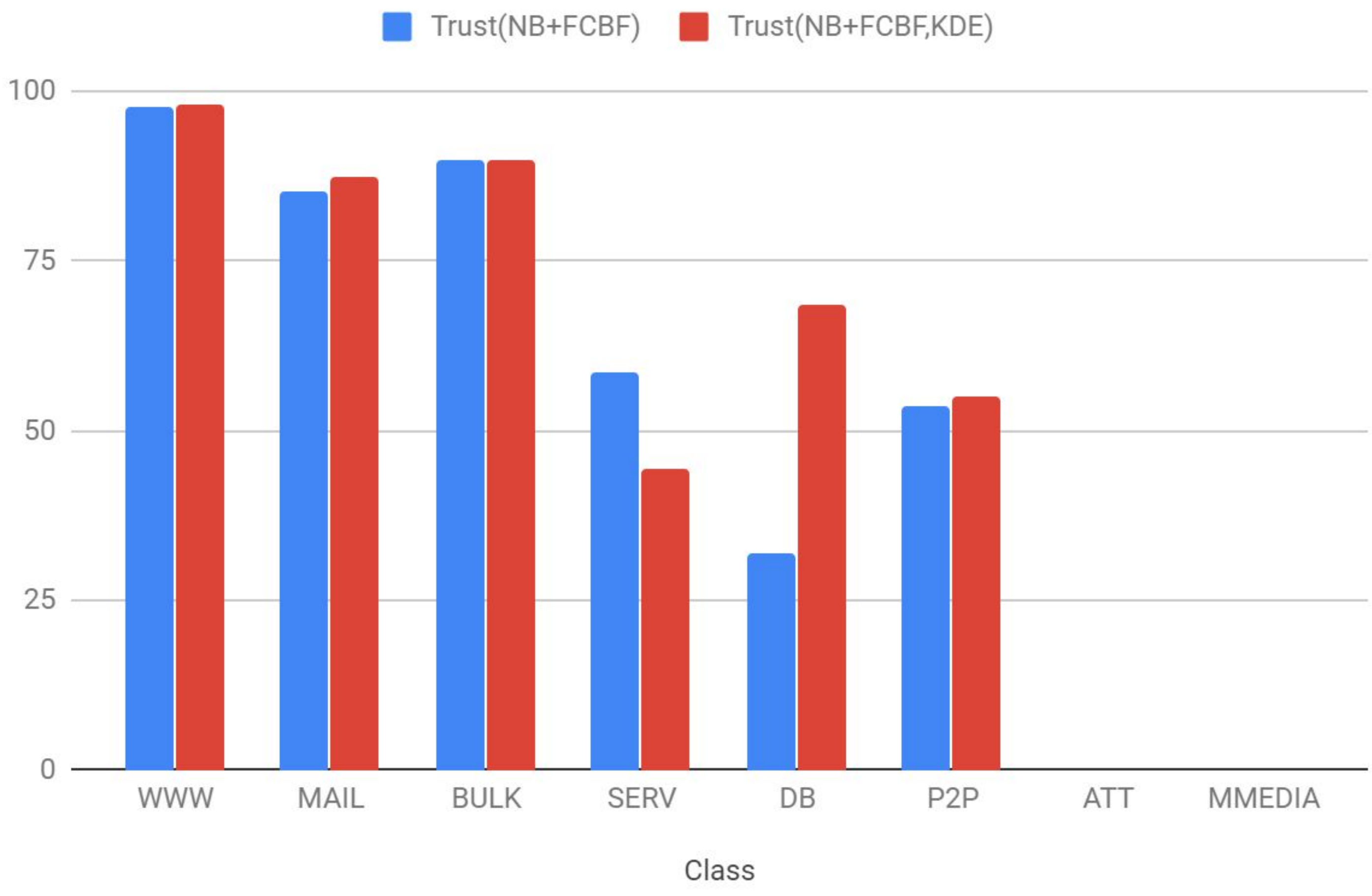


Measured accuracy for future data

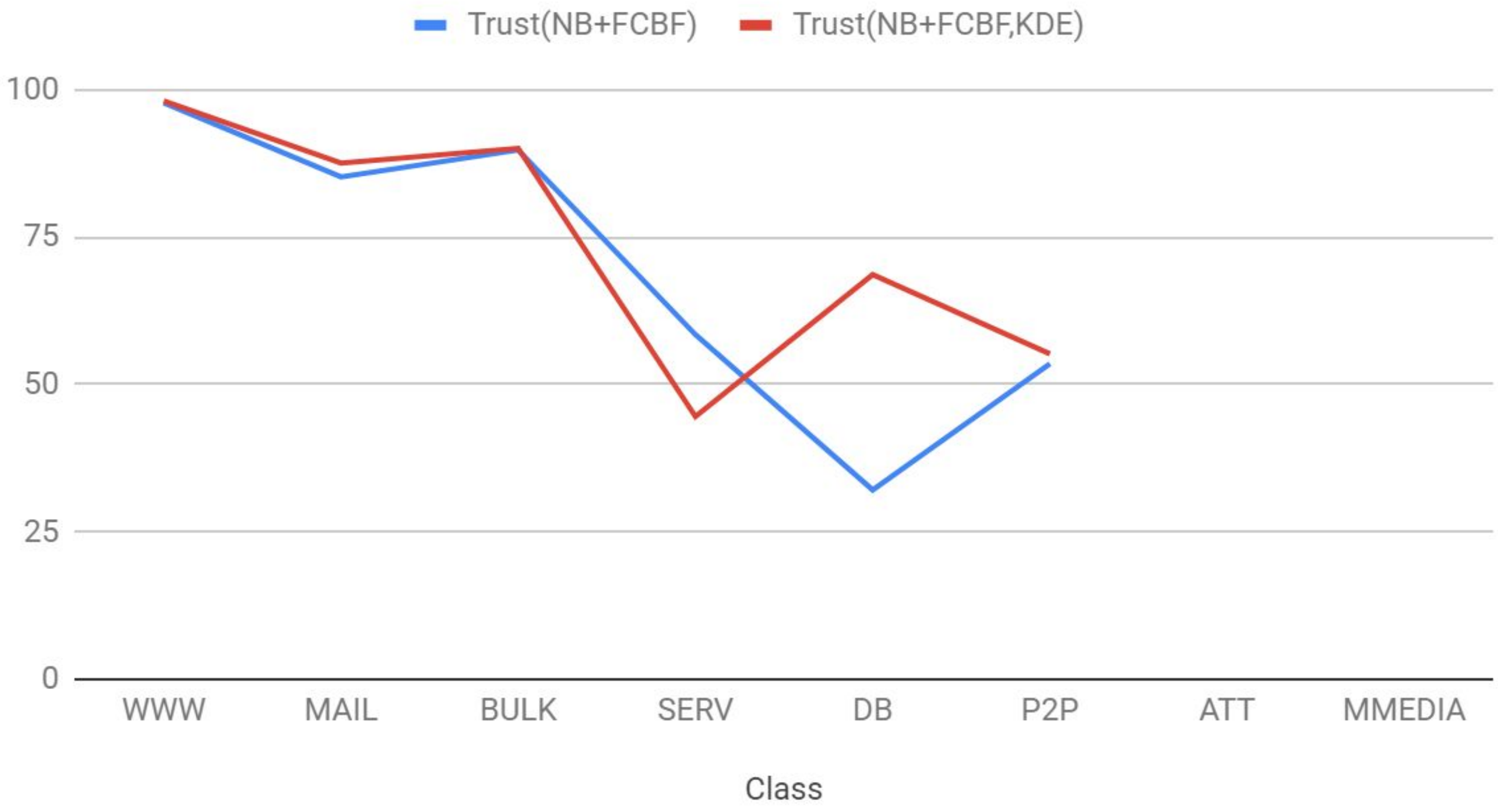


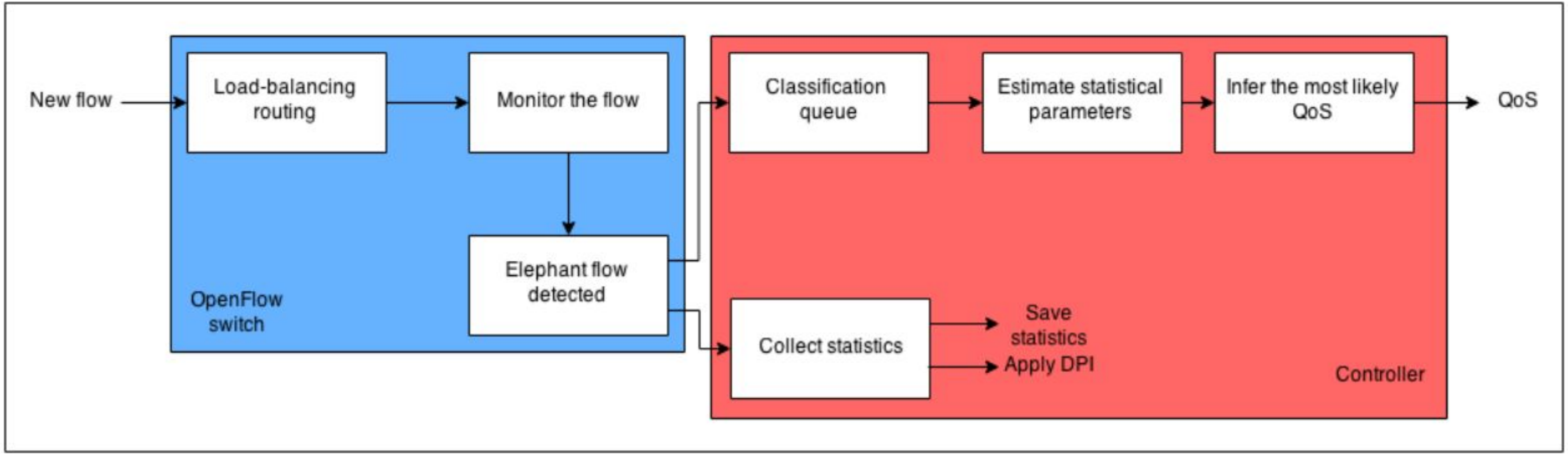
Measurement of belief in certain class after NB+FCBC,(NB+FCBC,KDE)(for later time)

Trust(NB+FCBF) and Trust(NB+FCBF,KDE)



Trust(NB+FCBF) and Trust(NB+FCBF,KDE)




New flow
Load-balancing routing
Monitor the flow
Elephant flow detected
OpenFlow switch
Classification queue
Estimate statistical parameters
Infer the most likely QoS
QoS
Collect statistics
Save statistics
Apply DPI
Controller