Chatter: Order-based Features for Malware Classification

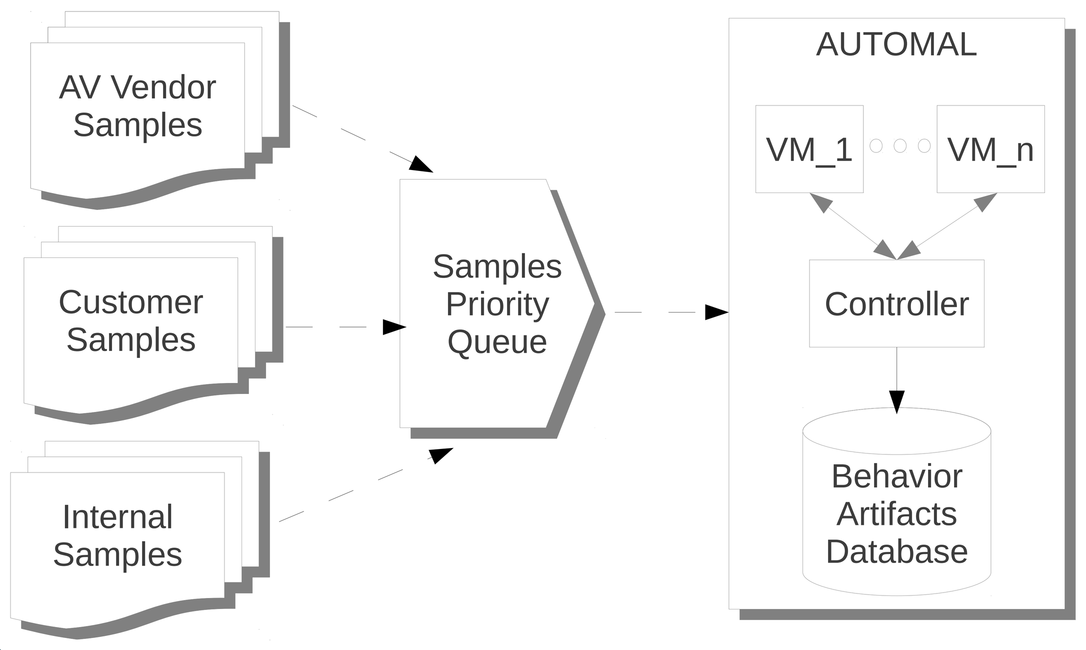
Aziz Mohaisen[1], Omar Alrawi [2], Andrew West[1], Allison Mankin[1], and Trevor Tonn[1]
[1]Verisign Labs, VA, USA [2]Qatar Foundation
First IEEE Conference on Communications and Network Security. Oct. 14-16, 2013, Washington, DC, USA.

Contribution

- A system for malware classification using cheap-to-obtain features; the features used in our system rely on the order in which artifacts associated with malware samples happen during the execution of a sample.
- An evaluation of the system's performance using several families of malware samples.

Behavior-based Analaysis

- Executes malware samples (binaries) in a sandboxed environment
- Artifacts are collected and used to fingerprint various malware families: memory, file system, registry, and network.
- Using those artifacts, researchers created features vectors and used machine learning algorithms for classification and clustering of malware samples.
- Often times, expensive algorithms: require co-residence and variety of features.



Design Goals and Objectives

- Cost effective: no deep features
- Less invasive: ideally can be implemented as an outside observer unit.
- Generalizable and multiple purpose: can be used for malware as well as other malicious activities characterization.
- Evolvable to address behavior changes: easy to tune to address malware circumvention mechanisms.
- Accurate: to meet operational standards.

Expansion of Features

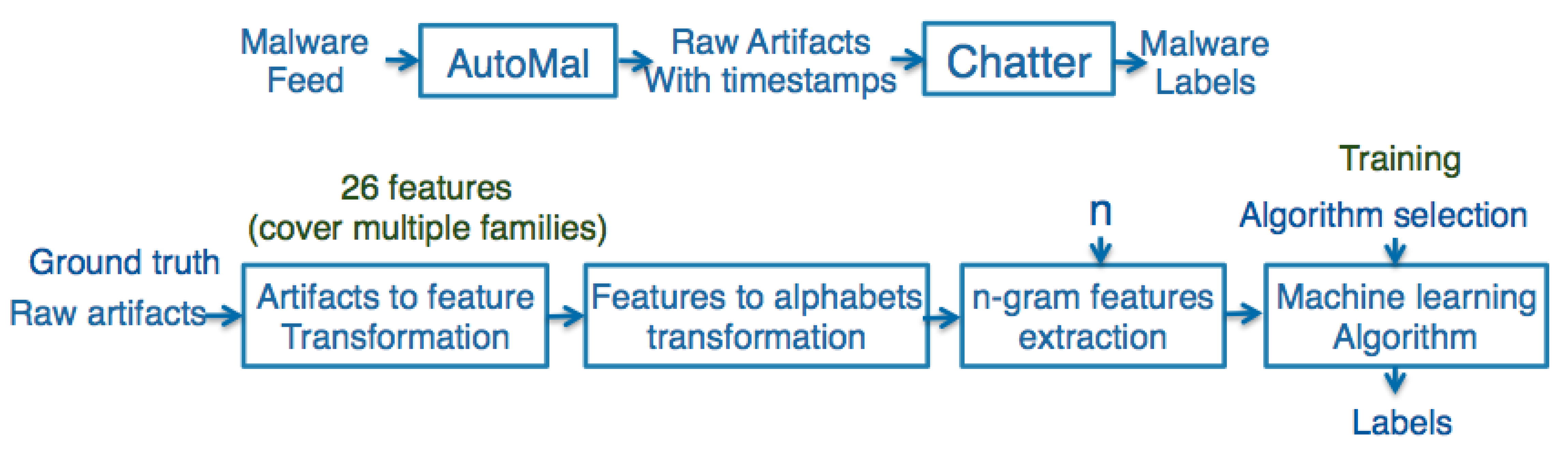
- Not all combination of features are used
- Ideally, for k characters and n length of a feature, there exist k^n possible features. The number explodes quickly as both parameters grow
- We use a condensed representation: we use non-zero features across the multiple samples. The number of features is reduced to 0.0005% when $n = 6$.
- Reduced features not only support efficient representation but fast of operation when using machine learning algorithms

n value	1	2	3	4	5	6	7	8
Zeus	24	102	250	481	943	1690	2638	3794
Darkness	24	103	243	461	875	1503	2266	3149
SRAT	25	105	247	460	877	1536	2337	3300

References

[1] Aziz Mohaisen et al., Chatter: Behavior and Order-based Features for Malware Classification In *Technical Report*, VeriSign Labs, October 2013

Design and Example



- Protocol type (TCP, UDP, RAW)
- port (53, 80, 443, 8080,10, 8000, other)
- Request and response size (in and out; quartiles; total of 8 features).

- Fixed length of alphabets (26)
- Variable length of behavior profiles (200 ~ 1000 characters)

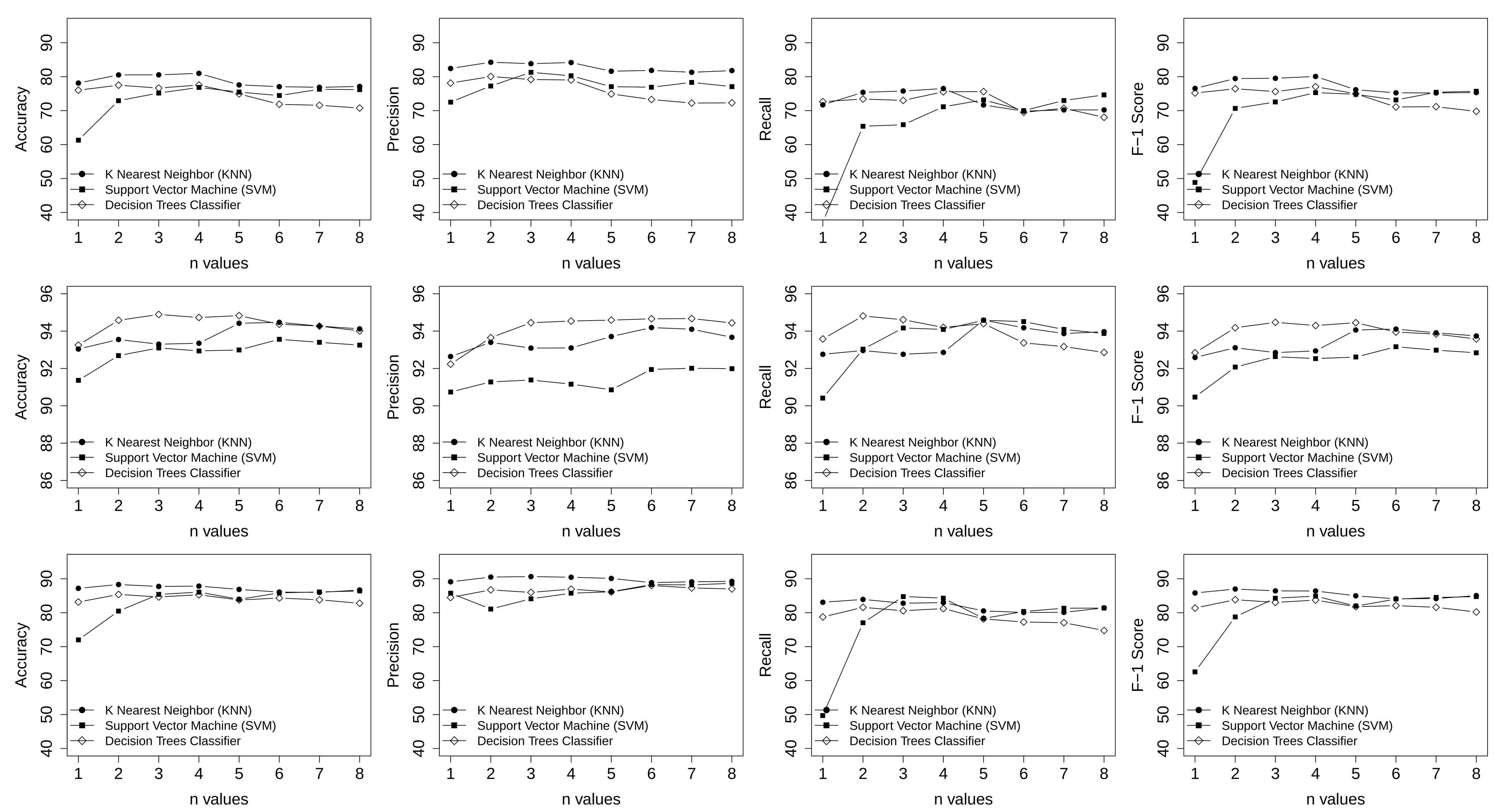
- Variable # of features (24-2000)
- Dense representation

Example

T1: Outgoing traffic, UDP connection, Port 53, DNS A record, 25KB, …
T2: Incoming traffic, UDP connection, Port 53, DNS AAAA record, 57KB, …
t1: A0, A3, A6, A8, A10, … n=1 -> (A0, A3, A6, A8, A10), n=2 -> (A0A3, A3A6, A6A8, A8A10, ..)
t2: A1, A3, A6, A9, A12, … n=1 -> (A1, A3, A6, A9, A10), n=2 -> (A1A3, A3A6, A6A9, A8A10, ..)

Results

- *Datasets:* Zeus (1025 samples, 50.74 avg chars), Darkness (544 samples, 61.47 avg chars), and SRAT (1130 samples, 52.74 avg chars).
- *Evaluation metric:* we use the accuracy, precision, recall, and F1-score.
  - Precision $= \frac{T_p}{T_p+F_p}$
  - Recall $= \frac{T_p}{T_p+F_n}$
  - Accuracy $= \frac{T_p+T_n}{T_p+T_n+F_p+F_n}$
  - F1 score $= 2 \times \frac{\text{precision}\times\text{recall}}{\text{precision}+\text{recall}}$
- *Parameters and settings:* n used for the n-gram is in the range of 1 to 8. A random set with equal size is used against the given family. Only network features are used (26 of them across multiple families). The table below is for augmented experiment with file-system artifacts. (below: chatter; above: +fs, order: darkness, zeus, SRAT).



	n-grams	1				4				8			
	Algorithms	P	R	A	F1	P	R	A	F1	P	R	A	F1
Zeus	k-NN	80.79	79.68	81.48	79.97	79.07	83.90	82.25	81.35	78.29	78.17	79.64	78.09
	SVM	67.41	82.67	72.69	73.92	75.96	80.47	78.67	77.84	80.41	82.87	82.45	81.50
	Decision Trees	80.14	80.90	81.74	80.42	81.13	81.82	82.67	81.35	80.82	82.82	83.02	81.77
Dark.	k-NN	76.22	73.13	76.08	74.56	80.40	71.52	77.70	75.57	71.38	69.58	71.65	70.20
	SVM	76.82	32.38	62.24	45.05	78.18	71.32	76.45	74.35	76.62	76.36	77.22	76.27
	Decision Trees	80.45	72.56	78.20	76.07	81.75	72.89	79.04	76.93	80.50	68.37	76.39	73.59
SRAT	k-NN	81.38	76.78	82.78	78.45	83.87	81.83	85.51	81.95	83.99	74.28	82.93	78.16
	SVM	76.88	65.43	75.88	69.55	83.70	82.94	86.23	83.03	85.68	80.86	86.33	82.71
	Decision Trees	85.16	81.11	86.44	82.60	88.28	81.65	88.01	84.45	86.13	78.92	85.54	81.85