to match the malicious applications with signatures which are stored in the virus signature database. Therefore, capabilities of detection mechanisms rely on signature database capabilities and it also relies on detection engine capabilities. The Assumption of having an engine that is capable of extracting and matching signatures will help to skip that section because detection engine and virus signature database are another research topic and both research topics have too many technical challenges inside.

Clustering and signature creation are two different bounded context and two separate research groups could work on each context. However, in this project limitation of time and people, the output of clustering process was used for the signature creation process. In short, the strings and methods processed in the previous clustering section were used to process signature creation.

There are multiple criteria to create a signature but the first concern in this research is the generality of signature to detect all samples of a malware family in the test set and the malware samples in the wild because too strict signatures most probably will cause true negative detection in the wild. This phase starts with scoring strings by different criteria such as string length, characters in the string, whether it is base64 string or not, if there is a space in the beginning or end of the string, strings that has more than 5 uppercase characters, string that starts with lowercase characters and ends with non-alphabetical characters. On the other hand, methods scoring are simply by method's opcode length.

After scoring operation for every single element which is gathered from malicious applications, those elements need to be selected by order to generate the most inclusive signature with the least number of elements. Therefore, in this research 6 elements have been chosen to generate the inclusive signature for a signature family just by starting to choose from string elements and if there are not enough string elements, method elements will be appended to the signature. At the end of the signature creating, there is a quality check operation to ensure that the signature will not cause the high number of false positive detections and will be successful in the wild. The quality of signature is decided by having more than 10 scores in the sum of all elements in the signature.

IV. RESULTS

The clustering mechanism and the signature creation process have successfully applied to a security product to test the results. Clustering success has been compared to Av-Test [8] malware family repository and the research has successfully achieved the %99.2 accuracy of clustering. However that could cause the too strict modelling to achieve this success. Signature creation and detection tests has been rewarded by %99.5 detection results by private test companies.

V. FUTURE WORKS

The processes mentioned above are not fully managed and be capable of processing thousands of files one by one in a day. Because the process needs to handle multiple files at one. Also, another problem of the research is that having more files may increase the process time even exponentially. Therefore the research project needs to be converted a flow process that can process each file separately and keep the processing time in reasonable time.

On the other hand, increasing speed of machine learning development, the research may focus on machine learning techniques to adopt new malicious applications and/or new malicious families easier. That may help to handle features easily instead of using preset multipliers and extraction routines.

The whole process may move to serverless architecture to gather malicious and benign samples and process them in the way that is provided features by the serverless provider. Development of tensorflow [9] and it's technologies have been spread into the serverless architecture to model machine learning procedures and run them in special computer parts named Tensorflow Processing Unit(TPU) [10].

Those technologies may speed up the research process and accuracy of clustering and accuracy of the generating signatures.

REFERENCES

[1] Hamandi, Khodor, et al. "Android SMS malware: Vulnerability and mitigation." Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference on. IEEE, 2013.

[2] Chebyshev, Victor, and Roman Unuchek. "Mobile malware evolution: 2013." Kaspersky Lab ZAO's SecureList 24 (2014).

[3] Wikipedia contributors. (2018, April 26). Ransomware. In Wikipedia, The Free Encyclopedia. Retrieved 21:46, April 29, 2018, from https://en.wikipedia.org/w/index.php?title=Ransomware&oldid=838351381

[4] O'Gorman, G., & McDonald, G. (2012). Ransomware: A growing menace. Symantec Corporation.

[5] Mercaldo, F., Nardone, V., Santone, A., & Visaggio, C. A. (2016, June). Ransomware steals your phone. formal methods rescue it. In International Conference on Formal Techniques for Distributed Objects, Components, and Systems (pp. 212-221). Springer, Cham.

[6] Wijesekera, P., Baokar, A., Hosseini, A., Egelman, S., Wagner, D., & Beznosov, K. (2015, August). Android Permissions Remystified: A Field Study on Contextual Integrity. In USENIX Security Symposium (pp. 499-514).

[7] Mercaldo, F., Nardone, V., Santone, A., & Visaggio, C. A. (2016, June). Ransomware steals your phone. formal methods rescue it. In International Conference on Formal Techniques for Distributed Objects, Components, and Systems (pp. 212-221). Springer, Cham.

[8] GmbH, A. (2018, April 24). AV-TEST – The Independent IT-Security Institute. Retrieved from https://www.av-test.org/en/