Sharing and Automation for Privacy Preserving Attack Neutralization
(H2020 833418)

COMBINING ANOMALY DETECTION MODELS FOR MORE RELIABLE ATTACK DETECTION

F-Secure Artificial Intelligence Center of Excellence

Dmitriy Komashinskiy

F-Secure

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 833418.

INTRODUCTION

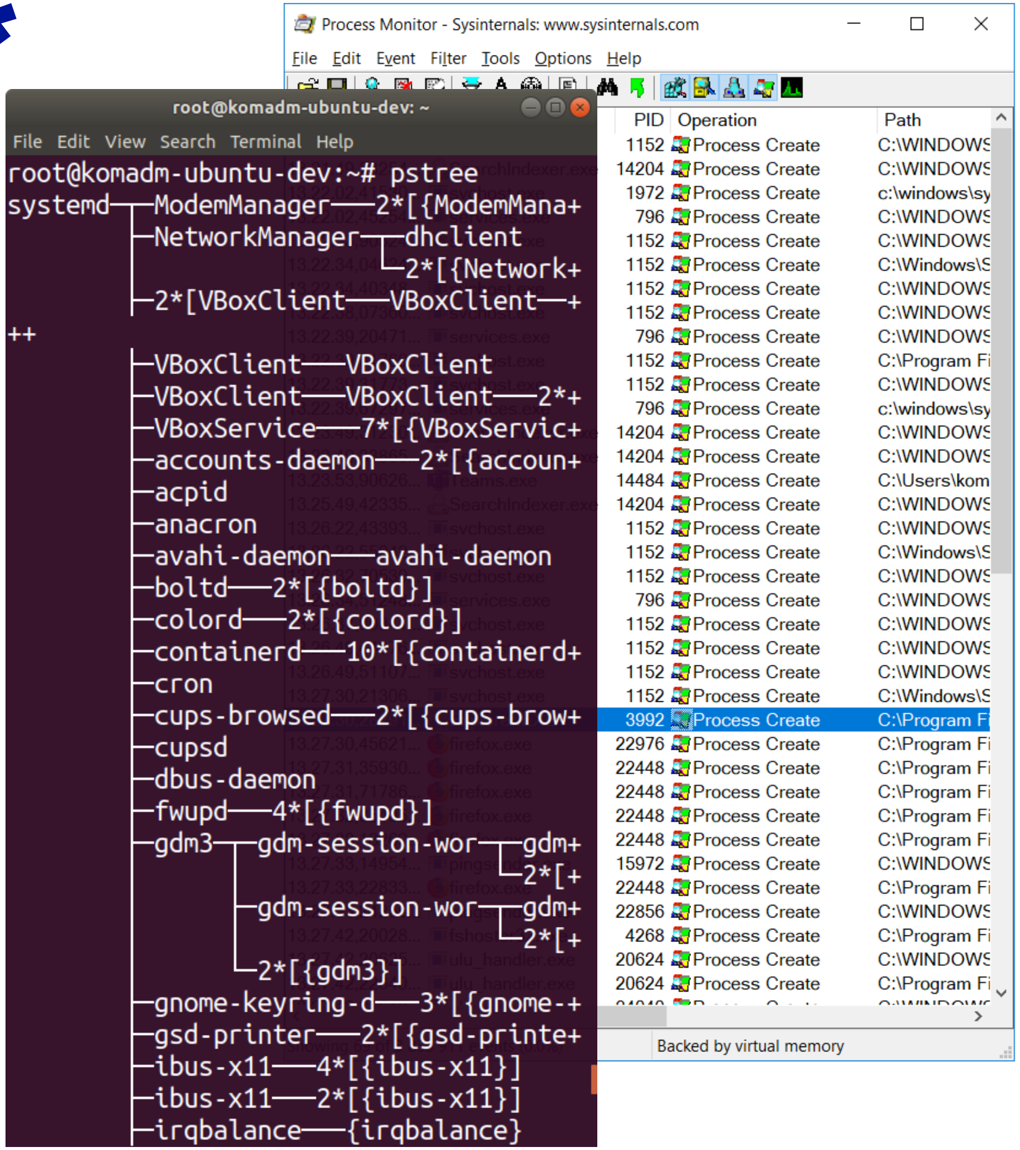
- EDR (Endpoint Detection and Response) client software (a.k.a. sensors) heavily relies on various system call / event data collection mechanisms to collect comprehensive behavioral data from endpoints;
- EDR backend data processing pipelines therefore must deal with enormous volumes of data. Various approaches exist to address this challenge, like sensor-side or BE-side data deduplication / aggregation, whitelisting, misuse / novelty detection logic etc.
- The abovementioned data reduction techniques are proven to be effective, but they leave an open question about how to find a reasonable tradeoff between the unavoidable data loss and the EDR protection's QoS; namely what needs to be done to keep EDR performance, scalability and fidelity in a balanced state.

The talk presents our work-in-progress effort focusing on endpoint anomaly detection facilitating scalable BE side attack detection and response processes for EDR service.

BACKGROUND MODEL*

The Process Launch Distribution model (referred to as PLD) focuses on detecting anomalous process launch events in a computing system;

- Operations in computing systems are carried out by so-called processes, instantiating at run-time software programs and containing their code, resources, activities, etc.
- Processes start each other in various ways, for example, a web browser typically starts a PDF reader to open a PDF file found on the Internet.
- An action of a ***parent process*** starting, or launching, a ***child process*** is called a process launch event.
- Such events can often be used for reliable identification of attempts of cybercriminals to compromise computing systems.



* *Details are available there: https://2020.ares-conference.eu/detailed-program/index.html*

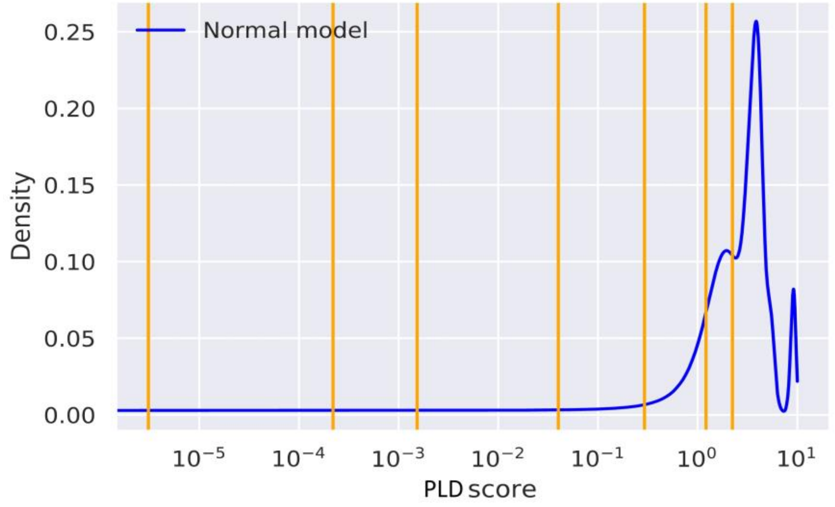
F-Secure

BACKGROUND MODEL: DETAILS

The PLD score is always a non-negative number. The lower the PLD score is, i.e., the closer it is to zero, the more anomalous the process launch event is from the PLD model point of view:

$$Score_{2vars}(x, y) = \frac{\frac{count(x,y)+\alpha}{n_{total}+\beta}}{\frac{count(x)+\alpha}{n_{total}+\beta} \cdot \frac{count(y)+\alpha}{n_{total}+\beta}} \cdot e^{H(Y|x)} \cdot e^{H(X|y)}$$

$$H(A|b) = -\sum_{a \in A} \frac{\frac{count(a,b)+\alpha}{n_{total}+\beta}}{\frac{count(b)+\alpha}{n_{total}+\beta}} \cdot \log \frac{\frac{count(a,b)+\alpha}{n_{total}+\beta}}{\frac{count(b)+\alpha}{n_{total}+\beta}}$$



Additional info: Das, K. and Schneider, J.: Detecting anomalous records in categorical datasets.

F-Secure

BACKGROUND MODEL: DETAILS

The PLD score is always a non-negative number. The lower the PLD score is, i.e., the closer it is to zero, the more anomalous the process launch event is from the PLD model point of view:

$$Score_{2vars}(x, y) = \frac{\frac{count(x,y)+\alpha}{n_{total}+\beta}}{\frac{count(x)+\alpha}{n_{total}+\beta} \cdot \frac{count(y)+\alpha}{n_{total}+\beta}} \cdot e^{H(Y|x)} \cdot e^{H(X|y)}$$

$$H(A|b) = -\sum_{a \in A} \frac{\frac{count(a,b)+\alpha}{n_{total}+\beta}}{\frac{count(b)+\alpha}{n_{total}+\beta}} \cdot \log \frac{\frac{count(a,b)+\alpha}{n_{total}+\beta}}{\frac{count(b)+\alpha}{n_{total}+\beta}}$$

Category	Fraction, %	Interpretation
1	90	9 out of 10 events
10	9	~ 1 per 10 events
20	0.9	~ 1 per 100 events
30	9e-2	~ 1 per 1K events
40	9e-3	~ 1 per 10K events
50	9e-4	~ 1 per 100K events
60	9e-5	~ 1 per 1M events
70	9e-6	~ 1 per 10M events
80	9e-7	~ 1 per 100M events
90	9e-8	~ 1 per 1B events

Additional info: Das, K. and Schneider, J.: Detecting anomalous records in categorical datasets.

BACKGROUND MODEL: EXAMPLE

process_name	child_name	#(parent, child)	#(parent)	#(child)	exp(H(Child|parent))	exp(H(Parent|child))	total	score	category
OneDrive.exe	cmd.exe	0	217199	1153611132	2.942220	47.879398	8.977204e+09	0.000505	60
winlogon.exe	cmd.exe	1053	29784977	1153611132	8.032356	47.879398	8.977204e+09	0.105815	40
browser_broker.exe	cmd.exe	4	39688	1153611132	2.734009	47.879398	8.977204e+09	0.105233	40
rundll32.exe	cmd.exe	2211	18496270	1153611132	1.291744	47.879398	8.977204e+09	0.057535	40
rundll32.exe	CompatTelRunner.exe	551	18496270	18719209	1.291744	2.162630	8.977204e+09	0.039917	40
services.exe	dllhost.exe	378853	940410425	28639235	6.742798	1.078851	8.977204e+09	0.918618	30
explorer.exe	net.exe	477	54675157	19363424	113.828870	2.801295	8.977204e+09	1.290000	30
rundll32.exe	WerFault.exe	1767	18496270	29619574	1.291744	27.233744	8.977204e+09	1.018645	30
dccw.exe	rundll32.exe	0	27	58196745	1.000000	4.354420	8.977204e+09	2.478583	30
services.exe	vmtoolsd.exe	199693	940410425	14993360	6.742798	2.691576	8.977204e+09	2.307462	30

INPUT

OUTPUT

ADDED PLD-LIKE AD MODELS (1)

- Module load distribution model:
  - **1st attribute:** process' file image name;
  - **2nd attribute:** module' file image name;
  - Extra details are available in documentation for LoadLibrary, LoadLibraryEx API functions.

- Open process and open thread distribution models *:
  - **1st attribute:** Actor process' file image name;
  - **2nd attribute:** Target process' file image name;
  - **3rd attribute:** Desired access value;
  - Extra details are available in documentation for OpenProcess, OpenThread API functions.

** Use updated score calculation logic for three input variables.*

ADDED PLD-LIKE AD MODELS (2)

- File Access distribution model *:
  - **1st attribute:** Actor process' file image name;
  - **2nd attribute:** Concatenation of access mode and file extension (e.g. 'READONLY txt', 'MODIFY vbs');
  - **3rd attribute:** top-level directory identifier (e.g. '%temp%', '%user%', '%systemroot%');

- Network access distribution model *:
  - **1st attribute:** Actor process' file image name;
  - **2nd attribute:** Port number (source or destination port for inbound or outbound connection respectively);
  - **3rd attribute:** Domain name of the remote host (if unavailable, IP type / range is used).

** Uses updated score calculation logic for three input variables.*

F-Secure

COMBINING THE DATA: EXAMPLE

New process data snippet (most anomalous)

process_name	process_gpid	child_name	child_gpid	category
OneDrive.exe	p:22b49a3403aff9d8e7c31a16bff53dd8	cmd.exe	p:ef96f9dba333d4d4b330f5f8f8071a52	60
OneDrive.exe	p:3f0d47aab83f7b7209fee90d68a01953	cmd.exe	p:8ad6d9488574052daad266caf4cd5a0a	60
rundll32.exe	p:05fa69b8b8eab6658bfa496ca33e7d2e	cmd.exe	p:9a0e164d5e121b29578a82c09eba08d4	40
rundll32.exe	p:a8b11206236616359b94912233fd53be	CompatTelRunner.exe	p:a168ffd7233e28ba2fbca1960668cb59	40
rundll32.exe	p:a8c90c78bd991e8e22085502d460b211	cmd.exe	p:f9a11a446d2160ce8347263114a4c824	40

Open process data snippet (most anomalous)

process_name	process_gpid	target_name	target_gpid	desired_access	category
powershell.exe	p:dbf8148bafb7e1a4c2a962ae5f78d57d	OneDrive.exe	p:3f0d47aab83f7b7209fee90d68a01953	5242	60
powershell.exe	p:915600ba8adef7ad783c73005e41c45f	OneDrive.exe	p:22b49a3403aff9d8e7c31a16bff53dd8	5242	60
svchost.exe	p:a99f3073a9e0b17b52a67794831e755a	basic_exe winsxs injection.exe	p:17f5de3065810aa687dc7eed1f16bc00	2097151	50
basic_exe.exe	p:47fb2e2a5e242ae2501fba6f68e5bb9a	explorer.exe	p:5bec70f5ef40f352024b80f76c54f84b	2097151	50
svchost.exe	p:b366a528a53c4ead46162172285f35fe	basic_exe.exe	p:47fb2e2a5e242ae2501fba6f68e5bb9a	2097151	50

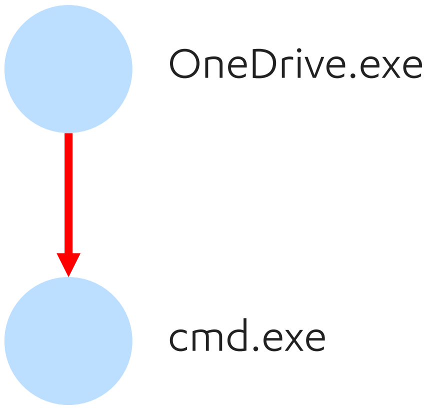
COMBINING THE DATA: EXAMPLE

New process data snippet (most anomalous)

process_name	process_gpid	child_name	child_gpid	category
OneDrive.exe	p:22b49a3403aff9d8e7c31a16bff53dd8	cmd.exe	p:ef96f9dba333d4d4b330f5f8f8071a52	60
OneDrive.exe	p:3f0d47aab83f7b7209fee90d68a01953	cmd.exe	p:8ad6d9488574052daad266caf4cd5a0a	60
rundll32.exe	p:05fa69b8b8eab6658bfa496ca33e7d2e	cmd.exe	p:9a0e164d5e121b29578a82c09eba08d4	40
rundll32.exe	p:a8b11206236616359b94912233fd53be	CompatTelRunner.exe	p:a168ffd7233e28ba2fbca1960668cb59	40
rundll32.exe	p:a8c90c78bd991e8e22085502d460b211	cmd.exe	p:f9a11a446d2160ce8347263114a4c824	40

Open process data snippet (most anomalous)

process_name	process_gpid	target_name	target_gpid	desired_access	category
powershell.exe	p:dbf8148bafb7e1a4c2a962ae5f78d57d	OneDrive.exe	p:3f0d47aab83f7b7209fee90d68a01953	5242	60
powershell.exe	p:915600ba8adef7ad783c73005e41c45f	OneDrive.exe	p:22b49a3403aff9d8e7c31a16bff53dd8	5242	60
svchost.exe	p:a99f3073a9e0b17b52a67794831e755a	basic_exe winsxs injection.exe	p:17f5de3065810aa687dc7eed1f16bc00	2097151	50
basic_exe.exe	p:47fb2e2a5e242ae2501fba6f68e5bb9a	explorer.exe	p:5bec70f5ef40f352024b80f76c54f84b	2097151	50
svchost.exe	p:b366a528a53c4ead46162172285f35fe	basic_exe.exe	p:47fb2e2a5e242ae2501fba6f68e5bb9a	2097151	50



F-Secure

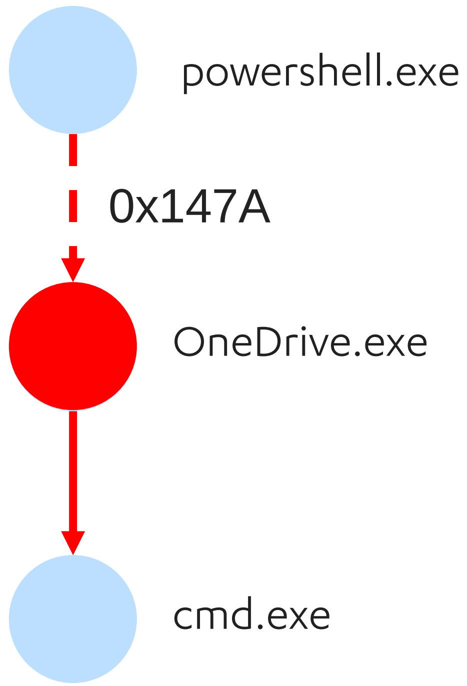
COMBINING THE DATA: EXAMPLE

New process data snippet (most anomalous)

process_name	process_gpid	child_name	child_gpid	category
OneDrive.exe	p:22b49a3403aff9d8e7c31a16bff53dd8	cmd.exe	p:ef96f9dba333d4d4b330f5f8f8071a52	60
OneDrive.exe	p:3f0d47aab83f7b7209fee90d68a01953	cmd.exe	p:8ad6d9488574052daad266caf4cd5a0a	60
rundll32.exe	p:05fa69b8b8eab6658bfa496ca33e7d2e	cmd.exe	p:9a0e164d5e121b29578a82c09eba08d4	40
rundll32.exe	p:a8b11206236616359b94912233fd53be	CompatTelRunner.exe	p:a168ffd7233e28ba2fbca1960668cb59	40
rundll32.exe	p:a8c90c78bd991e8e22085502d460b211	cmd.exe	p:f9a11a446d2160ce8347263114a4c824	40

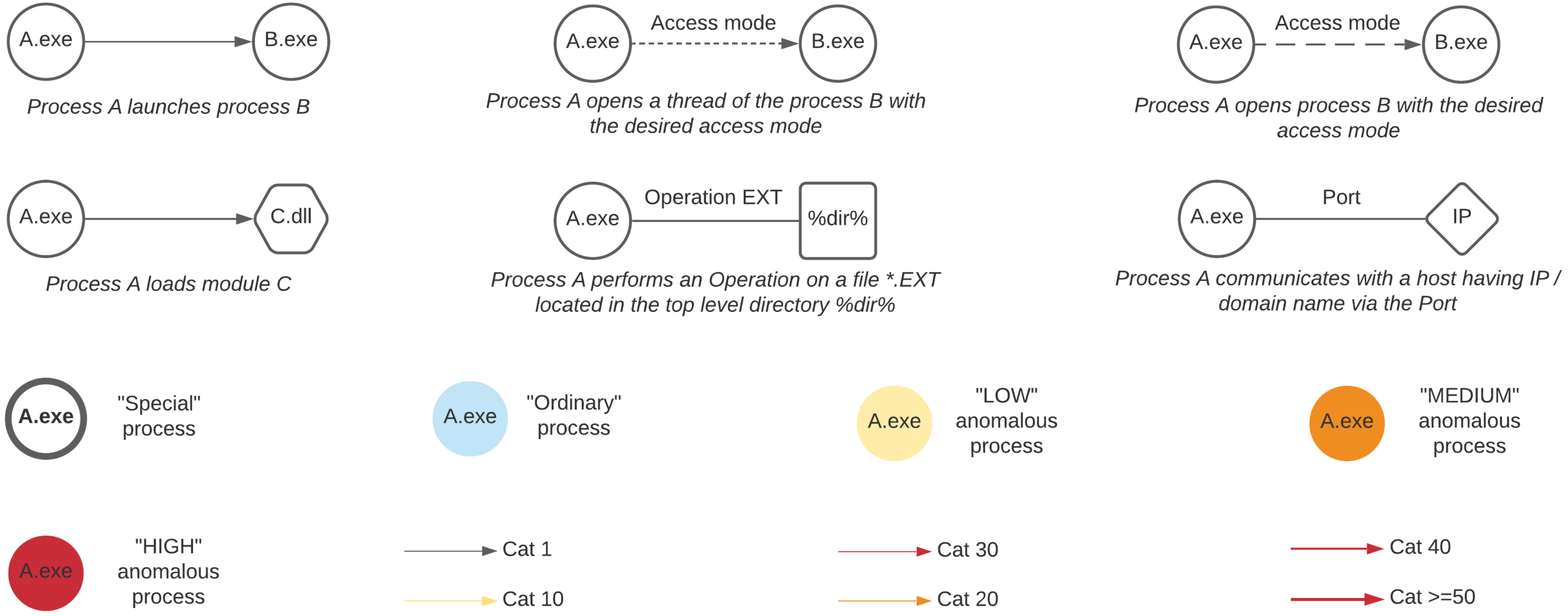
Open process data snippet (most anomalous)

process_name	process_gpid	target_name	target_gpid	desired_access	category
powershell.exe	p:dbf8148bafb7e1a4c2a962ae5f78d57d	OneDrive.exe	p:3f0d47aab83f7b7209fee90d68a01953	5242	60
powershell.exe	p:915600ba8adef7ad783c73005e41c45f	OneDrive.exe	p:22b49a3403aff9d8e7c31a16bff53dd8	5242	60
svchost.exe	p:a99f3073a9e0b17b52a67794831e755a	basic_exe winsxs injection.exe	p:17f5de3065810aa687dc7eed1f16bc00	2097151	50
basic_exe.exe	p:47fb2e2a5e242ae2501fba6f68e5bb9a	explorer.exe	p:5bec70f5ef40f352024b80f76c54f84b	2097151	50
svchost.exe	p:b366a528a53c4ead46162172285f35fe	basic_exe.exe	p:47fb2e2a5e242ae2501fba6f68e5bb9a	2097151	50



F-Secure

USED NOTATION

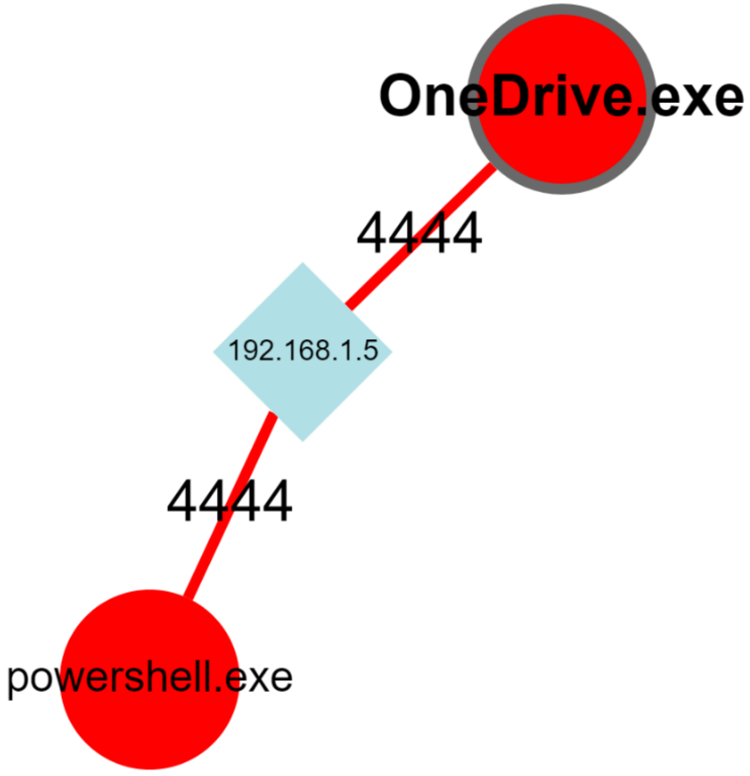


AN ATTACK EXAMPLE STUDY

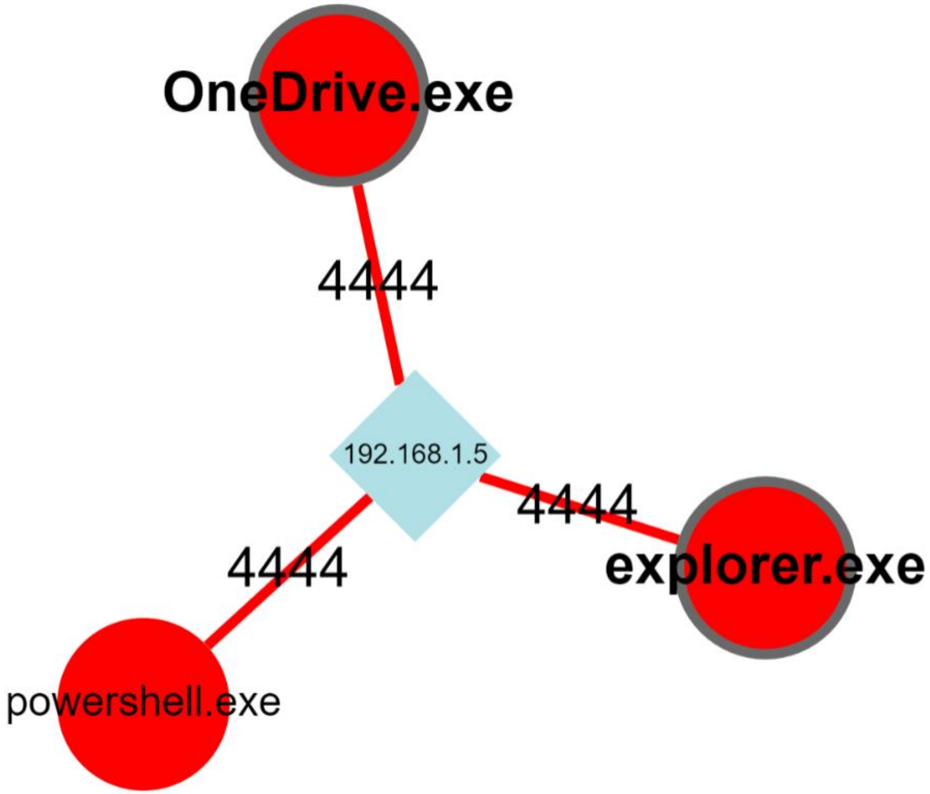
- An elementary data block represents all relevant events (the events that can be assessed by available PLD-like AD models) submitted by a single sensor within 24 hours time interval;
- A limited set of known positive (having confirmed attack traces) data blocks is available for the initial experimentation;
- For every positive data block:
  - All events get categories from corresponding PLD-like models;
  - For every category (starting from the most anomalous one, i.e. from 90) the events get combined either by subject (process ID) or by object (ID of network / file resource).
- Next slides present a typical layout for positive samples: for attacks, anomalous events tend to be connected in the provenance-like graph form.

F-Secure

THRESHOLDS: 70-90



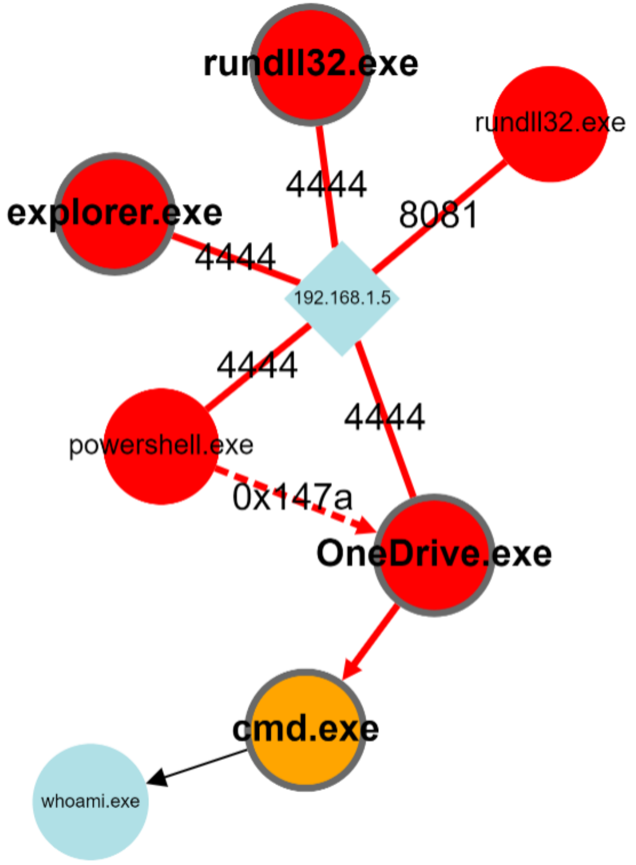
Threshold: 90



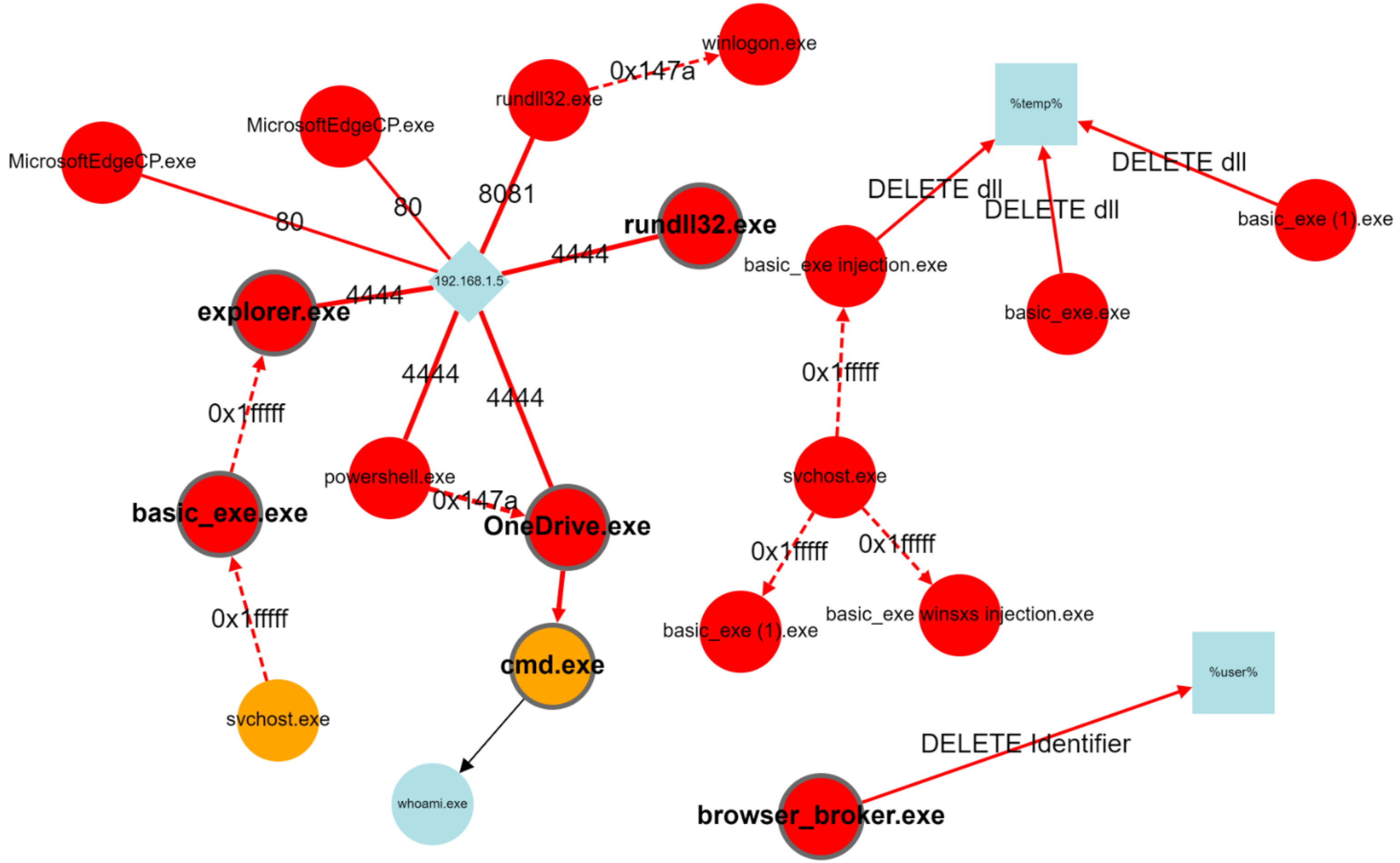
Thresholds: 70, 80

F-Secure

THRESHOLDS: 50-60



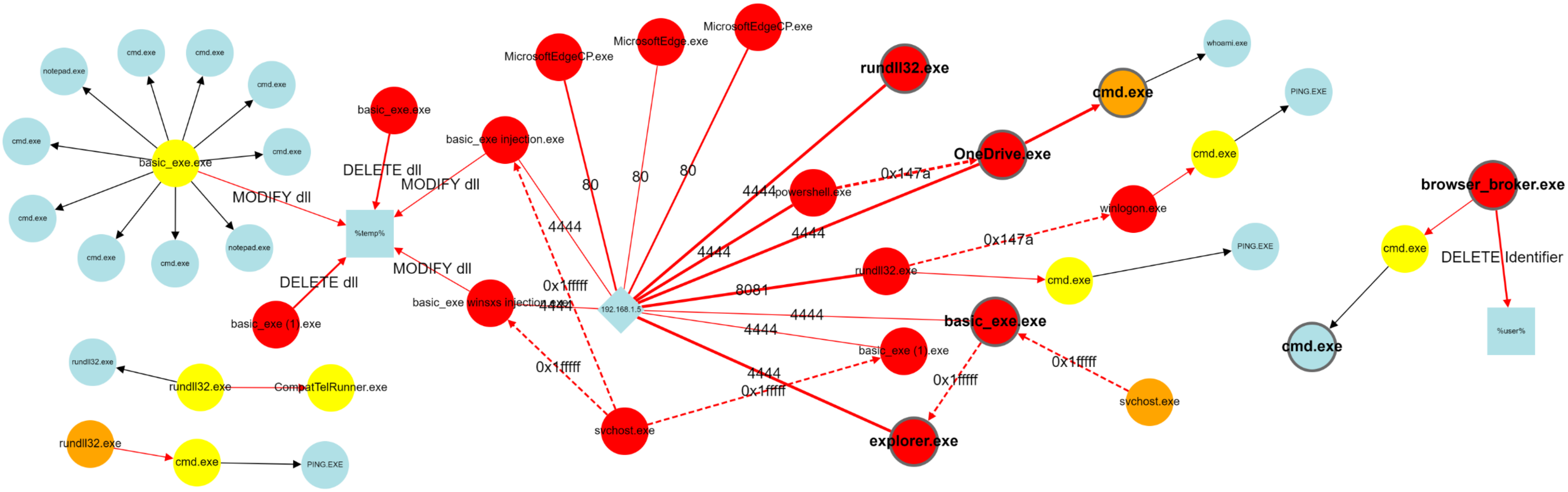
Threshold: 60



Threshold: 50

F-Secure

THRESHOLD: 40



CONCLUSIONS AND FUTURE WORK

Summary:

- Positive feedback from security analysts (the first realistic use case is to apply obtained data structures for deeper investigation of initially confirmed incidents);
- On-sensor data selection and aggregation, ready for prioritized decision making;

Open questions:

- Missing validation: definition of false positives / negatives, labelled data.
- Detection of ”low and slow” attack patterns;
- Decreasing models’ memory footprints;
- Applicability of additional PLD-like models (memory, registry, etc.)

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 833418.

USED AND USEFUL REFERENCES

- Das, K. and Schneider, J., 2007, August. Detecting anomalous records in categorical datasets. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 220-229).
- Braun, U.J., Shinnar, A. and Seltzer, M.I., 2008. Securing provenance. In *Proceedings of the 3rd USENIX Workshop on Hot Topics in Security (HotSec'08)*. USENIX Association.
- Milajerdi, S.M., Gjomemo, R., Eshete, B., Sekar, R. and Venkatakrishnan, V.N., 2019, May. Holmes: real-time apt detection through correlation of suspicious information flows. In *2019 IEEE Symposium on Security and Privacy (SP)* (pp. 1137-1152). IEEE.
- Hassan, W.U., Guo, S., Li, D., Chen, Z., Jee, K., Li, Z. and Bates, A., 2019, February. Nodoze: Combatting threat alert fatigue with automated provenance triage. In *Network and Distributed Systems Security Symposium*.
- Han, X., Pasquier, T., Bates, A., Mickens, J. and Seltzer, M., 2020. Unicorn: Runtime provenance-based detector for advanced persistent threats. *arXiv preprint arXiv:2001.01525*.

F-Secure