# Interactive Process Tree Analysis: Exploring the Behaviour of Processes with Visual Analytics for Security Operators

Robert-Carl Rapp* University of Stuttgart
Christoph Müller † University of Stuttgart
Franziska Becker ‡ University of Stuttgart
Paolo Palumbo§ F-Secure Corporation
Thomas Ertl¶ University of Stuttgart

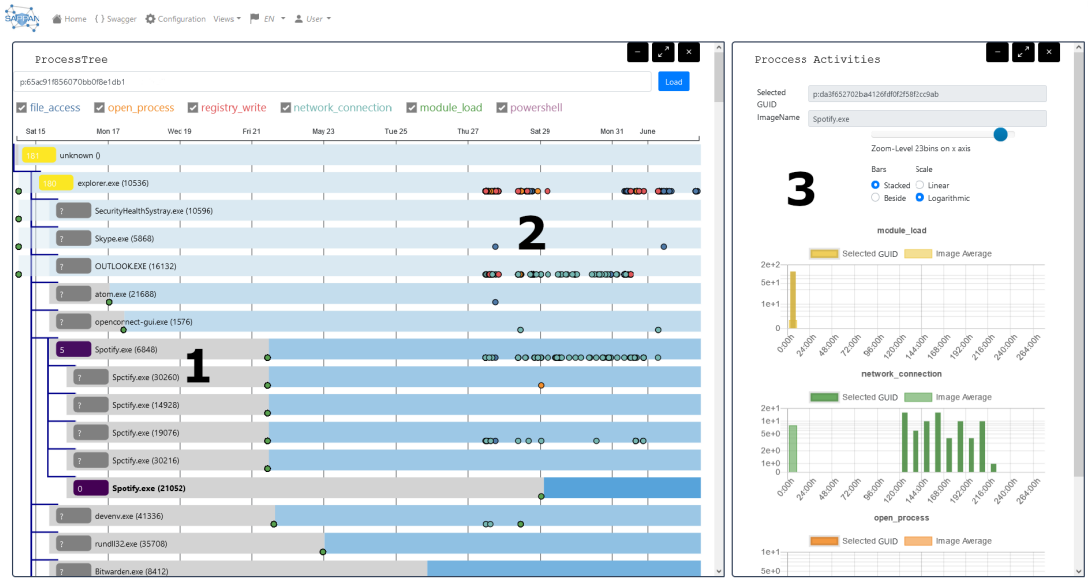


Figure 1: Our proposed process tree visualisation combines the hierarchical depiction (1) of the parent-child relationship with a temporal dimension that allows analysts to quickly understand when child processes have been spawned. The coloured dots (2) on the "process bar" of each process indicate different kinds of events like for example registry accesses coloured in red that the endpoint sensor recorded for the process. A histogram view (3) of these events for the same process image enables analysts to visually judge whether an event is a typical behaviour of the process at a specific point in time or an unusual one.

## Abstract

Despite constant efforts to improve automation for IT security incidents, analysts are often confronted with numerous alerts and have to make sure that they do not miss the most critical of them. The analysts need to quickly decide based on a plethora of yet incomplete information. This information often includes a tree of parent and child processes in real-world scenarios. We present an augmented visualisation of such a process tree, which not only shows the static hierarchy as previous ones do, but also conveys the temporal relation between processes, thus allowing for investigating the hierarchy and time perspective of the process tree at the same time. Furthermore, it makes additional process-related events collected by endpoint sensors accessible for a more complete view on process behaviour.

**Index Terms:** Human-centred computing—Visualization—Visualization application domains—Visual analytics; Security and privacy—Human and societal aspecty of security and privacy—Usability in security and privacy—

*e-mail: robert.rapp@vis.uni-stuttgart.de
†e-mail: christoph.mueller@visus.uni-stuttgart.de
‡e-mail: franziska.becker@vis.uni-stuttgart.de
§e-mail: paolo.palumbo@f-secure.com
¶e-mail: thomas.ertl@vis.uni-stuttgart.de

## 1 Introduction

Cybersecurity is becoming constantly more important, but also more complex, as more corporate assets and processes become digital ones. Using data from sensors on endpoint devices, companies try not only to catch known malicious activities but also to detect previously unknown attacks from behavioural analysis. The latter is subject to a great degree of uncertainty, which requires human security analysts to decide whether an automated alert is a false-positive or an actual incident. For that, analysts often rely on the process hierarchy to follow the traces of a potential attack. This hierarchy reveals which process is the parent of which other process and is similar to what advanced process management tools like *Sysinternals Process Explorer* [8] from Microsoft, *htop* and *Process Hacker* use. These tools rely on established techniques to represent hierarchies [5], mostly node-link diagrams [2] or indented tree plots, to provide a nearly real-time view of the system's process hierarchy. In contrast to that, the process trees we are working with are reconstructed from endpoint sensor data of a commercial managed-security product after the data passed the incident detection pipeline. While our process tree, therefore, is not a real-time view of the current system state, it is more than a mere snapshot of any point in time as we have the history of process creation at our disposal.

Albeit the availability of the temporal dimension of the process, the visualisation the analysts are working with is a conventional depiction of the hierarchy, which sparked the desire for the integration of the temporal dimension presented in this poster. To achieve the related goal of combining the hierarchy of grouped elements and time, Burch et al. [1] developed timeline trees, which use a tree

layout for hierarchical information and a timeline to explore events and analyse temporal changes. Furthermore, the process hierarchy uses only one kind of events, namely the creation of a new process. The process tree is a principal tool for analysts and we strive to integrate the other types of events (e.g. access to the registry or to files) that analysts use for detection, but hardly accessible for them, into a unified visualisation. Hassan et al. [6] used such events to obtain a certain degree of automating triage to address the issue of *high false-positive rates* from the survey paper of Kokulu et al. [7]. In contrast, the approach of making raw events accessible directly from the process tree is mainly targeting *slow response speed*.

In summary, we aim at achieving several goals: first and foremost, we want to fulfil the wish of analysts to grasp the temporal development of the process tree. As analysts have limited time to familiarise themselves with radically new visualisations, we secondly want to keep the solution close to the currently used indented tree view. The same reasoning is behind choosing a dashboard-based [3] web application. Finally, our solution makes additional process-related events accessible via the process tree view, which is the central interaction point of analysts with the data.

## 2 Approach

Fig. 1 shows a typical layout of our dashboard, composed of cards that can be flexibly opened, closed and arranged. Analysts can choose visualisations and open them in cards. The process tree on the left represents the process hierarchies and shows the occurrence of events the endpoint sensor recorded for the process on top of the process bar (as coloured dots). These events, which we subsequently call *activities*, mainly comprise process operations, file and registry accesses, network connections and certain system log entries. The histogram view on the right side puts the activities of a process selected from the tree in context with the average activities of the same process image.

Process tree. Based on observations of security operations centre (SOC) analysts' work, we assume that the process tree is often the starting point of analysis as most data from the endpoint sensors are process-centric. For the process tree, we eventually settled for a combination of a of a *Gantt chart-like* [4] horizontal bar chart using colour to indicate activity and process runtime and an indented tree plot that communicates the hierarchy. We, therefore, overloaded the x-axis with two different variables to make interrelated patterns between the temporal behaviour and hierarchical dimensions accessible. Before we proceeded with this approach, we opted using the x-axis purely for the temporal dimension and conveying the hierarchy with an overlaid node-link diagram with curved links. Especially long-running processes like the system root process caused a visualisation that quickly becomes cluttered and makes it challenging to identify connected nodes and required to scroll horizontally. We avoided this cluttering by overloading the x-axis with the multiple of a fixed indentation in the tree plot. However, this indentation width is not likely to represent the time difference between the start of the parent and the child, wherefore we need a visual representation to express this mismatch. Suppose a process starts shortly after its parent – e.g. "Skype.exe" (Fig. 1) – and its bar would need to start before the hierarchical indentation level. In that case, the hierarchy level is conveyed by the start of the fully opaque part of the blue bar. The semi-transparent extension to the left shows the start time. Similarly, for a process starting later than its hierarchy level would mandate, the grey part on the left indicates the hierarchy level and the start of the blue part indicates the start time.

Process activity histograms. While the coloured dots on the bars directly convey the activities of a process, it is hard to judge whether this time series is suspicious or not. Therefore, we give the option to further investigate this series in a separate histogram view on the dashboard (cf. Fig. 1-3). This view shows two overlaid histograms per activity, one for the occurrences in the selected process and the other for the average occurrences for all processes of the same process image (Fig. 1, right). As time is relative to the start of each process, these bar charts immediately show whether the selected process performed activities at similar points in its life cycle. While this might not immediately point to suspicious behaviour, particularly in interactive applications having activities triggered by user interaction, it still helps to identify spikes of activities as probably usual. We support linear and logarithmic scaling and interactive zooming to counter the conflict between the histogram covering large ranges and that in IT security, a single activity might be the malicious one an analyst is looking for.

## 3 Discussion

We presented a visualisation for historical process trees used in security operations centres to further investigate automated alerts before making critical decisions. The solution meets the design goals of integrating well with the web-based dashboard solutions currently used by analysts and heavily borrowing from the current depictions based on indented tree plots, therefore being familiar to analysts. At the same time, the analysts can now see the temporal sequence of processes being started in the same visualisation. Although our solution overloads the x-axis for hierarchy and time, the proposed solution for this issue is easy to understand for our expert. We consider at least once explained to analysts that they understand the representation but we need to validate this with analysts. Furthermore, we integrate the full variety of per-process activities collected by the endpoint sensors in the process tree, adding the possibility to open specific activities in a detail view as the raw events per se are of limited use. This view allows analysts to compare the average occurrences of events with the specific ones of the selected process, thus providing the opportunity to identify the unusual accumulation of activities. Overall, we expect our visualisation to constitute a quality-of-life improvement for SOC operators during the triage of incidents, but validating the approach in real-life scenarios is part of our future work. This will also address the issue that pseudonymisation of the data currently used prevents us from deeply integrating frequently used security applications such as *Cyberchef* or *VirusTotal* to enrich process data with further information.


## Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 833418.


## References


[1] M. Burch, F. Beck, and S. Diehl. Timeline trees. In S. Levialdi, editor, *Proceedings of the working conference on Advanced visual interfaces*, page 75, 2008.

[2] Christian Wojner. Procdot's home, 16.03.2021. `https://www.procdot.com/index.htm`.

[3] S. Few. *Information dashboard design: Displaying data for at-a-glance monitoring*. Analytics Press, Burlingame, Calif., 2 ed. edition, 2013.

[4] H. L. Gantt. Work, wages and profit. *The Engineering Magazine*, 1910.

[5] M. Graham and J. Kennedy. A survey of multiple tree visualisation. *Information Visualization*, 9(4):235–252, 2010.

[6] W. U. Hassan, S. Guo, D. Li, Z. Chen, K. Jee, Z. Li, and A. Bates. Nodoze: Combatting threat alert fatigue with automated provenance triage. In *Proceedings 2019 Network and Distributed System Security Symposium*, 2019.

[7] F. B. Kokulu, A. Soneji, T. Bao, Y. Shoshitaishvili, Z. Zhao, A. Doupé, and G.-J. Ahn. Matched and mismatched SOCs. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, pages 1955–1970, 11062019.

[8] M. Russinovich. Process explorer - windows sysinternals, 09.06.2021. `https://docs.microsoft.com/de-de/sysinternals/downloads/process-explorer`.