APPROACHING OVERLOAD: DIAGNOSIS AND RESPONSE TO ANOMALIES IN COMPLEX AND AUTOMATED SYSTEMS

Marisa Grayson, [1,2]
[1)] *The Ohio State University, United States*
[2)] *Mile Two, LLC, United States*

Abstract
Web production software systems operate at an unprecedented scale today, requiring extensive automation to develop and maintain services. The systems are designed to regularly adapt to dynamic load to avoid the consequences of overloading portions of the network. As the software systems scale and complexity grows, it becomes more difficult to observe, model, and track how the systems function and malfunction. Anomalies inevitably arise, challenging incident responders or SREs to recognize and understand unusual behaviors as they plan and execute interventions to mitigate or resolve the threat of service outages.

A study of four real cases reveals the interplay between the human and machine agents when problems disrupt the system. The analysis of the incidents directly links the cascade of disturbances below the line of representation (e.g. computer interfaces, monitoring tools) with the cognitive work of Site Reliability Engineers. The Above the Line / Below the Line Framework (ABL) changes the perspective in reviewing the cases post mortem in the tradition of Cognitive Systems Engineering and Resilience Engineering. The case study demonstrates specific and general patterns for complications to incident management in complex web operation systems, as well as directions for designing better tooling to support future, resilient work.

Keywords: anomaly response, saturation, software systems, resilience

1. INTRODUCTION

The incredible scale and dynamism of web operation production software systems requires extensive automation to maintain and continuously update today. As the software systems scale to provide valued services, the systems become more complex as a network with extensive interdependencies. The functional components and subsystems are built to adapt to changing loads regularly, in order to avoid overload or mitigate the consequences of overloading portions of the network. Nevertheless, problems arise that go beyond these resources and capabilities, which pushes functional parts of the system toward overload and necessitates the software engineers to resiliently intervene. As with many other domains, such as energy systems, space systems, and anesthetic management during surgery [1,2], anomaly response has been recognized as a critical function of web production software systems [3].

Several prior research efforts have focused on web production software systems and the adaptations made by the people involved [3,4]. One of the defining characteristics separating

ISBN: 978-91-88898-41-8
DOI: https://doi.org/10.15626/rea8.13

software systems from other domains is that the monitored process is entirely digital. Theoretically, all data in the system is available for collection and analysis, which makes the domain ideal as a natural laboratory. Furthermore, there are a variety of probes and representations that may rely on the phenomenon they are measuring, creating the potential for strange loops [5]. The automated systems often have hidden interconnections and primitive event capture reporting. The diverse group of engineers is highly distributed, though their interactions are recorded via chat logs, which was used for this talk's research.

The human-machine joint cognitive system can be described by the Above the Line / Below the Line framework (ABL) [4,6]. The line of representation is both the visual stimulus portraying the abstract processes and the interface for users' interactions directing the underlying automation. Below the line is a myriad of software layers accomplishing different functional tasks. The software engineers and users are above the line and can never directly interact with the coded automation, but together work to mitigate the risk of saturation and ensure productive activity [7]. The extensive accessibility of data allows for a detailed process trace of interactions between software engineers as they attempt to maintain and diagnose the failing system in a corpus study. The talk describes the high-level takeaways from the thesis research on the insights into the cognitive work of anomaly response challenged by the autonomy and complexity of web operations production systems.

2. METHODOLOGY

The cases were donated by multiple contributors of the SNAFU Catchers Consortium, a group of industry leaders and researchers focused on understanding and coping with the complexity involved in the operation of critical digital services [8]. The initial set was narrowed after discussions with domain experts from the consortium that were anchored with cognitive probes relevant to the research question. The discussions relied on techniques from knowledge elicitation and critical incident methods [9,10]. The selected subset of four cases featured unique qualities related to saturation and difficult anomaly response. Chat log files were gathered from post mortem records as the primary data source for each of the cases. The chat logs were either from IRC logs or from Slack logs, depending on the main communication technology used at the time. Each communication channel had a separate chat log that was integrated into the qualitative analysis tool developed by Adaptive Capacity Labs, Churchkey [11].

The situated work from the cases was analyzed by a process tracing method [12]. Over several iterations, the protocols were analyzed by applying a lightweight coding scheme from the tenets of anomaly response theory and macrocognitive functions [2,13,14]. Several key processes that were focused on include 1) events, 2) hypothesis generation, 3) model revisions, 4) interventions, and 5) stance [15]. Specifically, hypothesis generation is an important part of anomaly response, particularly between distributed problem solvers [2,3]. They communicate active theories to provide direction for diagnostic search, as well as broadening the hypothesis exploration space with multiple perspectives' contributions [16]. The evolution of the hypothesis space has been marked, which created one type of visual artifact from the study. The other main representation showcased the different responses to overload above and below the line, described in the results section.

3. RESULTS

One method of visualization was used to connect signals crossing the line of representation to the developing hypotheses of the responders. The collective hypothesis space above the line was created from the shared hypotheses in the chat logs and are relative to the line of certainty,

which is an ambiguous zone separating tentative plans from ideas that were acted upon as shown in Figure 1, top portion. The cases are re-represented to focus on the evolving mindsets of the people in response to the various anomalies and consequences of overload in the system. The different encodings are portrayed in time and connected to the hypotheses made by the responders to showcase the components of the anomaly response process present in each case.

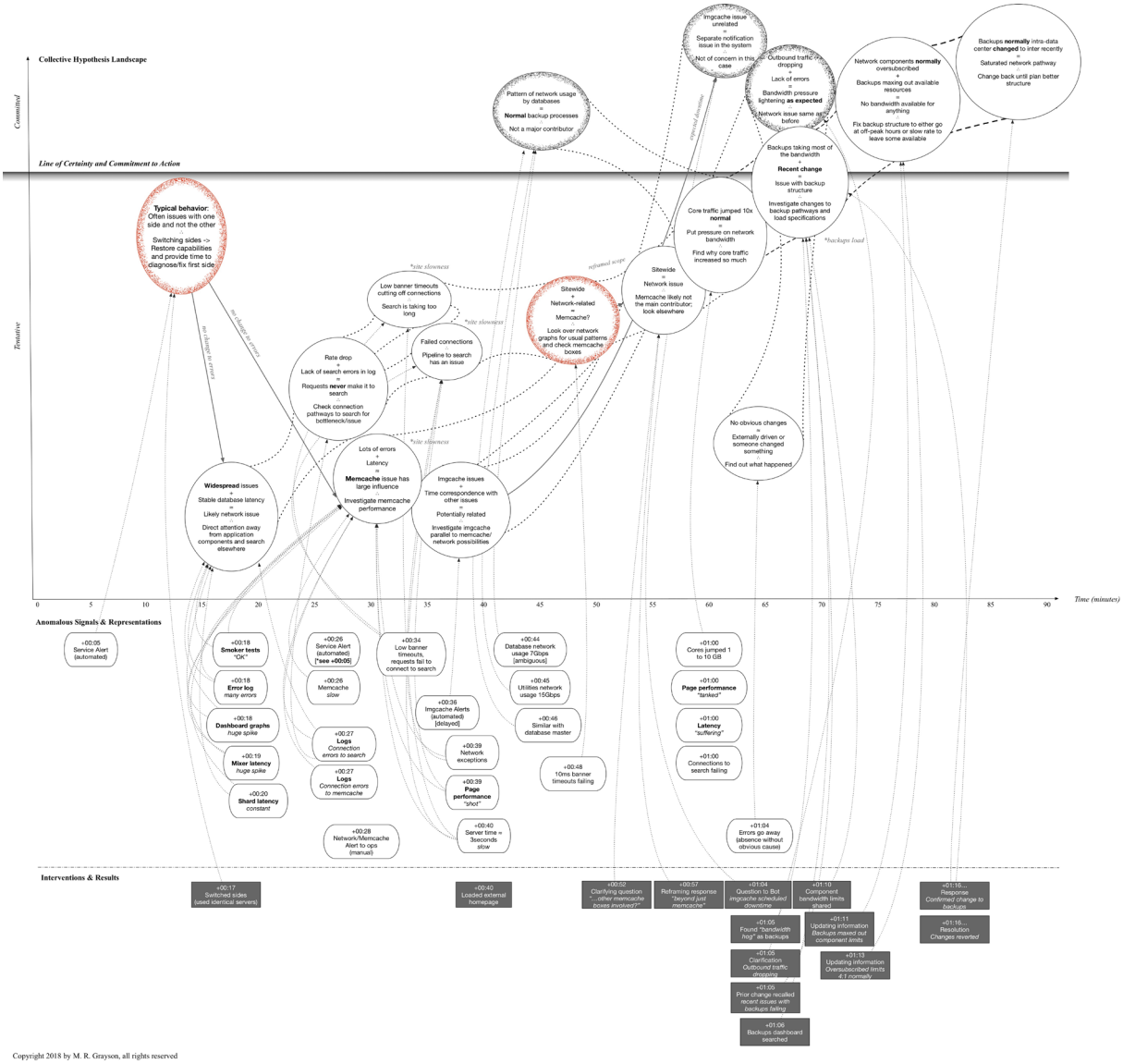

Figure 1. Hypothesis exploration chart with changing hypotheses (top portion) connecting to anomalous signals (middle) and interventions (bottom) noted across the case timeline.

In each of the cases, there was a noticeable appearance of saturation in various forms with mitigations above and below the line. Systems can breakdown when their capacity for maneuver is exhausted as effects cascade throughout the network [17].

This pattern plays a key role in each of the cases, though the manifestation and signals depend on the case's unique circumstances. The four broad strategies to manage bottleneck and overload are 1) shed load, 2) reduce thoroughness, 3) recruit more resources, and 4) shift work in time to lower workload periods [1]. The former two strategies are more tactical with lightening immediate load and freeing resources. The latter are more strategic defenses that

focus on broader resource allocation and management. All four responses were demonstrated by both the human responders and the automation in the corpus as the joint cognitive system was challenged by various disturbances, as shown in Figure 2.

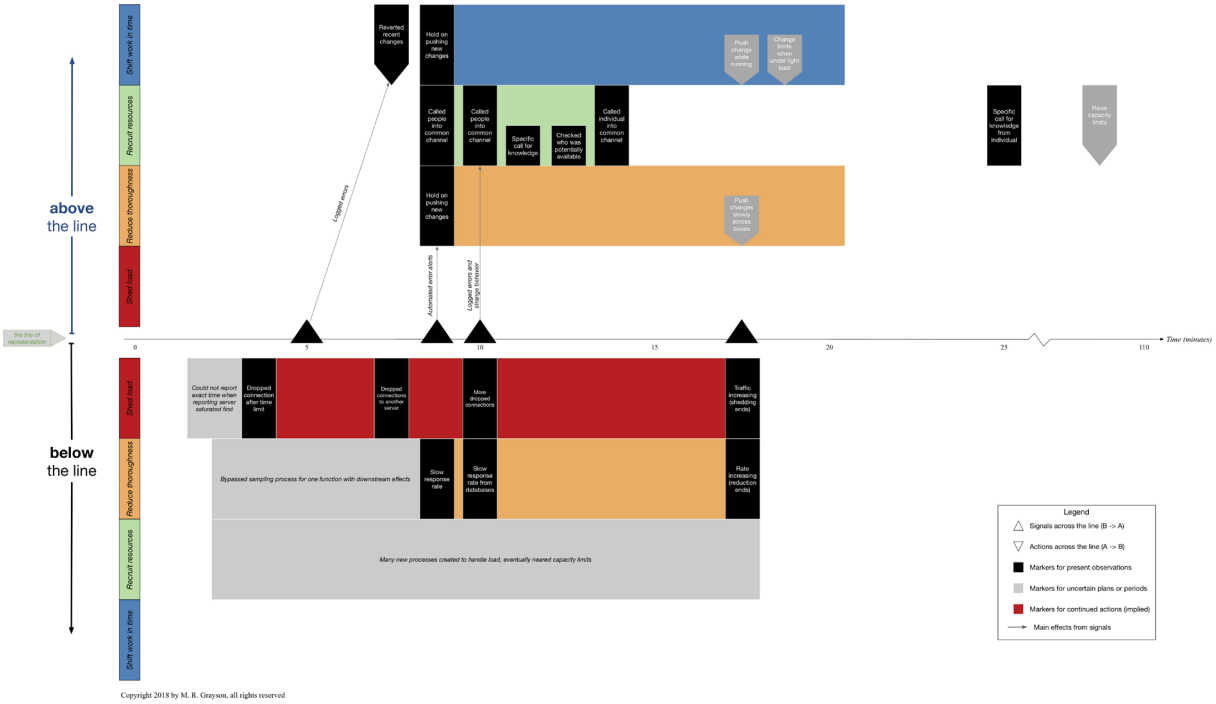

Figure 2. Responses to overload chart showing the four types of responses above and below the line across the case timeline.

A noticeable trend emerges from the results: the automation's responses tended to be more tactical; whereas, the humans responded more strategically. People offered a variety of solutions, which fit into each of the four categories, in contrast to the autonomous system that was mostly limited to the first two or three strategies. Web-based software systems typically have elastic boundaries and few hard physical constraints. However, resource capacities can still be finite when pushed beyond normal expectations. The humans ultimately filled in the gaps in the system, which supports the importance of development accompanying operational monitoring.

Besides the brittleness of the automation, other difficulties arose that challenged responders' abilities to manage the system, including masking, strange loops, concurrent issues, and measurement limitations. The opaque systems restricted observability, degrading the coordination between the human and machine agents [18]. More detail about the specifics from the cases and the extracted patterns can be found in the full thesis document [19]. The cyclical case exploration revealed several core findings about the challenges autonomy and the complexity of web operations places on the cognitive work of anomaly response:

- Despite safeguards, overload occurs, propagates, and is hard to see.
- Mental models have gaps and are updated during anomaly response.
- Network complexity produces effects at a distance and weak representations hinder diagnostic search.

- ABL was an effective framework for analyzing anomaly response.

4. DISCUSSION

The analysis and re-representation of the cases revealed several patterns of insight that point at new directions for improvement. The layers of active automation and processes beyond visible observation hindered human responders' ability to diagnose and respond to anomalies. The current tooling is primarily event driven alerts based on state changes and thresholds with some minor behavioral shifts. The measurements are limited in scope and often leave the functional integration to the engineers. New forms of tooling and monitoring could increase their observability into the dynamics of the automated processes and support their future work. One method could focus on new representations to explore effects at a distance. Another method of promoting this broader hypothesis generation could be to visualize a dual hypothesis space exploration [20]. The reorientation mechanism could have different forms, such as active role in the anomaly response team or a part of a tool monitoring generated hypotheses from common patterns and strategies in communication channels [21,22]. Additional supports for sensemaking could be in improving contrasting sources of data on the same phenomenon of interest. Future research would test the effectiveness of the different support solutions offered in the course of responding to simulated or real incidents.

The Above the Line / Below the Line framework expands on past work by opening the black box and demonstrating saturation below the line in light of crossing signals and distributed observations above. The visualization reframes the cases as the evolving dynamics between the humans and automation in the joint cognitive system. The case study illustrated the beneficial perspective, though it was limited by the chat data from the postmortem without direct data records. This issue could be remedied by gathering the computer logs to augment the below-the-line section, potentially in real-time as the data is already recorded and actively being monitored throughout the system. The ABL framework provides a basis for future work in decision support tools to narrow the time between detection, identification, and possibly sufficient resolution of anomalies.

REFERENCES

[1] Woods,D.D. & Hollnagel,E.(2006a).AutomationSurprise,in. *JointCognitiveSystems Patterns in Cognitive Systems Engineering* (pp. 113-142) Boca Raton: CRC/Taylor & Francis.
[2] Woods, D.D. & Hollnagel, E.(2006b). Anomalyresponse,in *JointCognitiveSystems: Patterns in Cognitive Systems Engineering* (pp. 69-95) Boca Raton: CRC/Taylor & Francis.
[3] Allspaw, J. (2015). Trade-Offs under Pressure: Heuristics and Observations of Teams Resolving Internet Service Outages. Masters Thesis. Lund, Sweden: Lund University.
[4] Woods, D.D., ed. (2017). STELLA: Report from the SNAFU Catchers workshop on coping with complexity. Retrieved from https://snafucatchers.github.io
[5] Hofstadter, D.R. (2007).*Iamastrangeloop.* Basicbooks.
[6] Cook, R.I. and Woods, D.D. (2016). Situationnormal: Allfouledup. Velocity:Web Performance and Operations Conference, New York. Retrieved from https://www.oreilly.com/ideas/situationnormal-all-fouled-up
[7] Allspaw, J. & Cook, R. (2018). SRECognitiveWork. In Blank-Edelman, D.N. Seeking SRE: Conversations About Running Production Systems at Scale (pp. 439-462). O'Reilly Media, Inc.
[8] Aboutus. (2016). Retrieved from https://www.snafucatchers.com/about-us

[9] Flanagan, J.C. (1954). Thecriticalincidenttechnique. *Psychologicalbulletin*, 51(4), 327.
[10] Klein, G. A., Calderwood, R., & Macgregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on systems, man, and cybernetics*, 19(3), 462-472. https://doi.org/10.1109/21.31053
[11] Adaptive Capacity Labs. (2018). Retrieved June 20, 2018, from www.adaptivecapacitylabs.com
[12] Woods, D. D. (1993). Process tracing methods for the study of cognition outside of the experimental psychology laboratory. In Klein GA, Orasanu J, Calderwood R, Zsambok CE, eds. *Decision making in action: Models and methods* (pp. 228-251). Westport, CT: Ablex Publishing.
[13] Cook, R. I. (1998). How complex systems fail. *Cognitive Technologies Laboratory, University of Chicago. Chicago IL.*
[14] Klein, G., Ross, K. G., Moon, B. M., Klein, D. E., Hoffman, R. R., & Hollnagel, E. (2003). Macrocognition. *IEEE intelligent systems*, 18(3), 81-85.
[15] Chow, R., Christoffersen, K., & Woods, D. D. (2000, July). A model of communication in support of distributed anomaly response and replanning. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 44, No. 1, pp. 34-37). Sage CA: Los Angeles, CA: SAGE Publications.
[16] Gettys, C. F. & Fisher, S. D. (1979). Hypothesis plausibility and hypothesis generation. *Organizational Behavior and Human Performance*, 24, 93-110. https://doi.org/10.1016/0030-5073(79)90018-7
[17] Woods, D. D., & Branlat, M. (2011). Basic patterns in how adaptive systems fail. *Resilience engineering in practice*, 127-144.
[18] Klein, G., Feltovich, P. J., Bradshaw, J. M., & Woods, D. D. (2005). Common ground and coordination in joint activity. *Organizational simulation*, 53, 139-184. https://doi.org/10.1002/0471739448.ch6
[19] Grayson, M. R. (2018). *Approaching overload: Diagnosis and response to anomalies in complex and automated production software systems* (Doctoral dissertation, The Ohio State University).
[20] Dunbar, K. (1993). Concept discovery in a scientific domain. *Cognitive Science*, 17(3), 397-434.
[21] Smith, P. J., Miller, T. E., Fraser, J., Smith, J. W., Svirbely, J. R., Rudmann, S., ... & Kennedy, M. (1991). An empirical evaluation of the performance of antibody identification tasks. Transfusion, 31(4), 313-317.
[22] Shute, S. J., & Smith, P. J. (1993). Knowledge-based search tactics. Information Processing & Management, 29(1), 29-45. https://doi.org/10.1016/0306-4573(93)90021-5