Stories and narratives in safety engineering

Catherine Menon*, Austen Rainer†

University of Hertfordshire*, Queens University Belfast†

Abstract *The use of stories and narrative is widespread throughout safety engineering, from "war stories" to use cases In this paper we consider the effectiveness of stories in modelling safety-critical systems and challenges. We present a discussion of how aspects of a story such as characterisation, narrative arc and setting can affect the extent to which it adequately illuminates a software engineering problem.*

1 Introduction

Storytelling is one of the oldest forms of human communication, from the narratives conveyed by primitive art and cave painting, to Greco-Roman myths, to the earliest known written story, the epic of Gilgamesh (Kovacs, 1989). Stories have been used to entertain, to moralise, to inform and – perhaps above all – to illuminate.

It is in this last capacity that we see the most significant use of stories in Science, Technology, Engineering and Mathematics (STEM) subjects, particularly engineering. Stories are used as abstractions of specific engineering problems: the Travelling Salesman (Robinson, 1949), the Dining Philosophers (Hoare, 1978), the Byzantine Generals (Lamport, 1982). In some cases the story becomes better known than its origin (Hoare's coinage of the phrase "dining philosophers", for example, is so strongly associated with the deadlock problem that Dijkstra's original "Dining Quintuple" (Dijkstra, 1987) is now largely forgotten – as are the previous usages of the term "dining philosophers" themselves (Acland, 1841), (Martineau, 1838)). Stories have a particular value in the pedagogy of computer science, as they contain elements of narrative and character which students can recognise, and with which they identify. This initial sense of familiarity has been shown to decrease the perceived difficulty of subsequently understanding unfamiliar concepts such as algorithms, code and systems thinking (Parham-Mocello, Ernst and Erwig, 2019).

Although stories are an integral part of the way in which we teach, discuss and represent complex systems, historically there appears to have been very little thought given by engineers to the construction of *effective* stories. In this, of course, we refer only to fictional constructions and not the use of "war stories" or summaries drawn from accident reports. Fictional stories are unique in that their every property – characters, settings, plots – can be tailored to an effective representation of an abstract problem, and yet in many cases we still find that the properties of a constructed story are in conflict with the engineering problem which it seeks to illuminate.

In this paper we consider the effectiveness of stories in reasoning about and modelling safety-critical systems and situations. We present a discussion of how aspects of a story such as narrative, characters and plot can affect the extent to which it adequately illuminates a software engineering problem. We illustrate this with examples drawn from two specific stories representing underlying engineering problems: the Byzantine Generals Problem (Lamport, 1983) and the Dining Philosophers Problem (Hoare, 1978).

2 Stories and story characteristics

Much work already exists on the properties of stories (Yorke, 2014), the process of writing (King, 2000), (Lamott, 1994) and the philosophy behind storytelling as a concept (Bradbury, 1992). From these, we can extract some fundamental characteristics of a successful story.

In this paper we will use the term *primary characteristics* to refer to the essential properties which make a narrative a story (rather than an unrelated series of sentences): the presence of a protagonist, the presence of a desire or aim possessed by the protagonist, an obstacle (antagonist) leading to conflict, and a resolution in alignment with these properties. To be effective, a story must not only contain these elements, but they must be identifiable and understood by the reader, hence the question of who is the target readership of a story becomes of paramount importance, and we return to this in Section 6. It may be stated that – outside of certain genres such as experimental fiction, which are unlikely to be relevant to engineering problems – possession of these primary characteristics is an essential requirement for all stories.

We will use the term *secondary characteristics,* by contrast, to refer to those aspects of a given specific story which correspond to the specific plot, setting and characters of a given story. For example, the secondary characteristics of the Byzantine Generals Problem (Section 3) include a city under siege, an unspecified number of generals, some traitors etc. The secondary characteristics of the Dining Philosophers Problem (Section 4) include some stubborn and hungry philosophers, a single dining room and a limited supply of silverware; we suggest the conflict may safely be left as an exercise for the reader.

For a story to be effective, secondary characteristics must be consistent with each other and the assumed or explicit rules of the story's universe. That is, the characters should be of a kind which might credibly be found in that specific setting, and the plot should consist of events and choices which might credibly be made by these characters. Secondary characteristics are, in general, more varied than primary characteristics: we can find a story effective even where the characters act in surprising ways, or find themselves in an unusual setting – in fact, it is arguable from a position of literary criticism that a story without some element of surprise is the poorer for it. More pragmatically, in stories which represent an underlying engineering problem, an unusual protagonist (e.g. a general who dislikes killing), or an unexpected aspect of the setting (e.g. a world where safe passage is always given to messengers) can still be effective provided sufficient explanation and narrative effort is expended on ensuring the readers comprehend the "surprise" (Alwitt, 2002).

3 Byzantine Generals Problem

In this section we will consider a well-known story within the safety community: the Byzantine Generals Problem (BGP). This has the dubious virtue of representing a real flaw in some safety-critical systems: there has been at least one occurrence of a real-life Byzantine Generals Problem in the system failure of the Airbus A330 in 2020 (TTSB, 2021).

There are several presentations of the BGP, but in this paper we take what is perhaps the most commonly-cited within the safety engineering community: that described in (Lamport, 1982):

"We imagine that several divisions of the Byzantine army are camped outside an enemy city, each division commanded by its own general. The generals can communicate with one another only by messenger. After observing the enemy, they must decide upon a common plan of action. However, some of the generals may be traitors, trying to prevent the loyal generals from reaching agreement. The generals must have an algorithm to guarantee that the following two conditions are met:

- BGP1: All loyal generals decide on the same plan of action […]
- BGP2: A small number of traitors cannot cause the loyal generals to adopt a bad plan" (Lamport, 1982)

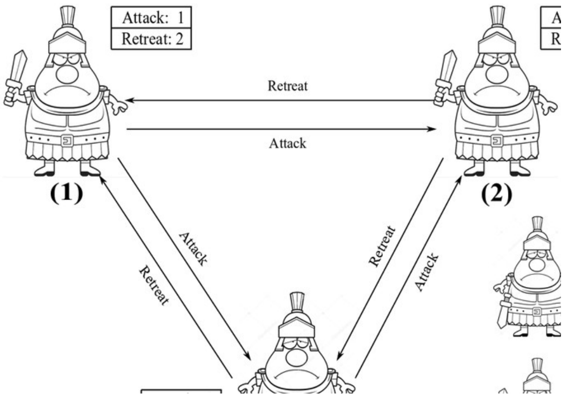

Fig. 1. Byzantine Generals Problem (Salimitari, 2020)

This is certainly a well-understood problem in safety engineering, and accurately represents the underlying challenge of communication between independent components of a system, some of which may not be reliable. Nevertheless, there are some amendments which could be made to the BGP as presented above to improve its accessibility as a story, and promote a deeper understanding of the problem outside of the safety engineering community.

We gave the BGP story to a group of professional writers, and after a single reading, a straw poll established that they presumed the following as part of the story:

Primary story characteristics:

- The generals want to agree on a method to defeat the city (e.g. by assault, sabotage, siege etc.)
- The traitors are trying to stop the generals defeating the city
- The generals will either defeat the city, be defeated in the attempt, or give up and (temporarily?) retreat

Secondary story characteristics:

- The generals are willing to incur some losses to defeat the city
- Messengers might be waylaid or put in danger by the traitors or the city

We emphasise that the writers were asked to react to this only as a story, rather than as a representation of an engineering problem. Moreover, it is also important to note that these readers do not have an engineering background or (as we established) prior knowledge of the BGP. They are therefore reading this story in the capacity of members of the lay public with particular experience in stories and their interpretation. This may go some way towards explaining some of the inconsistencies between the BGP as an engineering problem and the primary / secondary story characteristics deduced above.

Given these caveats, there are nevertheless some inconsistencies which make it clear that the BGP could be improved as a story, and hence as a means of communication about an engineering problem. In particular, none of the writers identified that that the generals' motivation is only to reach agreement, rather than to attack the city, and similarly they failed to identify that the traitors are trying to prevent agreement rather than defend the city. These misunderstandings of the story are discussed in more detail below.

3.2 BGP primary characteristics

A fundamental issue with the BGP – from a story rather than an engineering perspective – is the protagonists (the loyal generals) lack a desire consistent with what we know of their characters and the narrative. The Byzantine generals don't specifically desire to attack the city but instead merely want to reach agreement on what they should do next. While irreproachable as a representation of complex system communication, this choice is inconsistent with the characterisation of the generals (given the lack of any other information, readers assume generals are concerned primarily with conducting war and seeking victory rather than with anxiously ensuring that everybody agrees with everybody else).

Moreover, this desire is inconsistent with their previous assumed actions. The fact that the generals begin the story camping outside the city elicited from all readers the understanding that the generals desired to attack it. As a result, our straw poll readers all misunderstood the story in a way that would mischaracterise the engineering problem it represents: the BGP seeks to establish an algorithm for agreement amongst components, not an algorithm to obtain a particular specified decision outcome.

The second immediate problem with the BGP story is that the antagonists (traitor generals) also lack a desire consistent with their characterisation. They aren't seeking to save the city, but rather, desire only to prevent the loyal generals from agreeing. In this respect they are much more consistent with a "force of chaos" rather than a group of traitors (as an engineering-focused reader would expect, given the underlying engineering problem deals with system failure rather than malicious attack)

The third problem with the assumed primary characteristics of the BGP is the difference between the concept of a "plan of action" within the context of safety-critical system operation, and within the context of storytelling. In the safety-critical systems world, failure of the system components to come to an agreement represents a system failure with potentially catastrophic consequences. In the story world, if the generals fail to come to an agreement the reader's assumption is almost always that they will stay where they are. (They have not agreed to leave, attack, disband or take any other action, and within the story framework they must continue to exist somewhere: there is no "negative space" within the world of the story for them to vanish into).

Our straw poll of readers all interpreted "failure to come to an agreement" as a method of attack on the city. That is, the readers assumed that in the absence of an agreement the generals would stay where they were – and furthermore assumed that that constituted an attack by siege on a city that would eventually run out of resources. The readers therefore considered this a valid fallback option on the part of the generals, an implication which is missing within the safety-critical context. For a safety-critical system, of course, there is no concept of "still doing something" attached to a failure to decide.

We note that these problems stem from the story framing of the BGP, not the representation of the engineering problem. Specifically the setting of a city under siege and the generals as the protagonists lead readers to assumptions which align with stories about war, rather than stories about agreement, negotiation and diplomacy. We suggest that perhaps renaming this particular problem as the Byzantine Diplomats Problem might negate some of these misunderstandings!

3.2 BGP secondary characteristics

There are also some further concerns with the story presentation of the BGP in (Lamport, 1982). In particular, the story and engineering implementation are not separated: there is "story" information that can only be deduced by reading the engineering dissection of the problem. While this is understandable in the context of the original paper, it means that an inexperienced reader or member of the lay public has only an unclear – or worse, a misconceived – idea of the story, and hence the underlying engineering problem.

To take one example, the information given later in (Lamport, 1982) is that the generals must all receive the same information as each other, and moreover, that deciding on the same plan of action implies that the generals must have received the same information. However, the readers' assumptions in Section 3 about the secondary characteristics of the story are already in direct contradiction to this additional information. Specifically, all readers assumed that some generals might vote not to attack (either because it was personally detrimental for their troops, or because they had been given false information by a traitor general), but that the generals as a group could still agree on attacking even given disproportionate costs to a minority of their members. That is, as a group the generals are willing to incur some losses, and even those generals personally facing the losses would accept the plan.

However, if as given in (Lamport, 1982) there can be no such majority vote (i.e. the generals will only agree on a plan if they have been given the same information), then the reader's assumptions must be reassessed and corrected. That is, we are in a different story: the generals must all attack together if they are to have any hope of success, and any general who doesn't want to attack will refuse, thus ruining the plan. This is not the reader's stereotypical assumption of what coordinated army looks like and again we suggest an alternative: The Byzantine Loosely-Allied Tribes?

4 Dining Philosophers

We emphasise that the problems identified above with the BGP relate only to the story presentation, rather than the underlying engineering issue. To illustrate how an effective story can assist in our understanding of software, we next turn to another well-established story: the Dining Philosophers (DP):

"Five philosophers spend their lives thinking and eating. The philosophers share a common dining room where there is a circular table surrounded by five chairs, each belonging to one philosopher. In the centre of the table there is a large bowl of spaghetti, and the table is laid with five forks. On feeling hungry, a philosopher enters the dining room, sits in his own chair, and picks up the fork on the left of his place. Unfortunately, the spaghetti is so tangled that he needs to pick up and use the fork on his right as well, When he has finished, he puts down both forks and leaves the room" (Hoare, 1978).

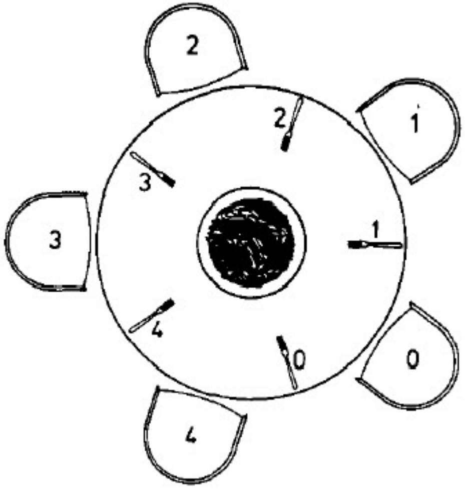

Fig. 2. The Dining Philosophers Problem (Hoare, 1978)

The DP as presented is an essentially absurdist situation: we are asked to accept that the philosophers will willingly starve should they not be able to obtain two forks with which to eat spaghetti. The inadequacy of their table manners notwithstanding, this scenario works as a story because of its internal consistency: a reader expects (stereotypical) philosophers to engage in nonsensical, overly-abstruse debate and to refuse all common-sense solutions. In other words the protagonists have a believable desire within the story world – to eat – believable characterisation, and the resultant conflict is in alignment with both.

It is also worth noting that the presentation of the DP in (Hoare, 1978) is a near-complete separation of implementation and "story". As such, it can be presented to non-technical readers in its entirety, without risking the contradictory assumptions which arise with the BGP.

We do, however, note one aspect of the BGP story presentation which equals – or betters – that of the DP: the story title. The original title of the Chinese Generals Problem (Lamport, 2021) is not only problematic from the perspective of racial prejudice, but it does not give readers any insight into why these generals are acting in a counter-intuitive and highly constrained manner. The translation to Byzantine generals, who might – in common with the rest of the Byzantine Empire – be expected to operate in a labyrinthine atmosphere of rules, constraints, plots and betrayal, accurately conveys the flavour of the story.

5 Discussion

As we emphasise in both Section 3 and Section 4, the critique of the stories presented here does not imply any criticism of the complexity, relevance or "worth" of the underlying engineering problems. Nevertheless, if stories are intended to convey an understanding of these problems – particularly to the lay public – it is worth examining why some are more effective than others.

Returning to the BGP of Section 3, we refer again to the observation in (Lamport, 2021) that the original problem related to Chinese generals. It is unclear from a modern position why these generals should be specifically Chinese – or, indeed, any other specific nationality, except Byzantine. Stories which appeal to stereotypes in this way often founder when such stereotypes become outdated or unknown. It is interesting that both the BGP in its current form and the DP (Section 4) hark back to relatively uncontroversial "ancient history" stereotypes of behaviour, rather than modern.

We suggest that it might also be instructive to consider whether the chosen story says something to readers about the importance of the problem. In general, the story chosen can encourage or discourage particular assumptions, or focus the reader's attention on a particular outcome. For example, an argument can be variously described as a war (with an implication of conflict, a winner and a loser) or as a dance (with an implication of decorum, collaboration and diplomacy). The construction and regulatory acceptance of a safety case argument is perhaps more akin to the latter in theory, and the former in practice!

In general, stories which postulate a potentially fatal outcome for one or more characters tend to receive more traction in safety discourse. The Trolley Problem (Foot, 1967) is an example: although this story largely does not capture the primary concerns of safety engineers, there are arguably few such engineers who have not referred to this during a discussion of autonomous vehicle safety. Similarly, the BGP is phrased as a story about "dramatic", high-worth events: war, invasion and treaties. We speculate that the story may have received less traction within the engineering community if the characters were of a very different kind (The Byzantine Schoolgirls Problem?).

6 Conclusions and further work

We acknowledge that seasoned engineers are unlikely to rely on the story of an engineering problem to provide them with their full understanding of it. However, safety-critical engineering is a discipline which relies on communication and public understanding of risk. Emerging technologies such as autonomous systems will require a greater degree of public acceptance, understanding and willing engagement (Information Commissioners Office and Turing Institute, 2020), and adequate communication – including in the form of accessible stories – is a necessary first step towards that.

From our preliminary steps in researching how the lay public interpret stories representing engineering problems, we have identified some axioms for constructing and communicating effective stories:

- The story must be an accurate model of the engineering problem. As with all models, omissions are inevitable, and are more tolerable than misrepresentations.
- The story and any engineering solution to the problem should not be mixed in the presentation
- Metaphors and assumptions within the story should be well-understood and, so far as possible, parallel the details of the engineering problem
- The setting and characterisation of the story should not rely on stereotypes which may be misunderstood, particularly where these are used to convey information about the underlying engineering problem
- The story must contain an element of drama, such as the potential for a fatal outcome for one or more characters

As future work, we propose to validate and extend these axioms, moving toward a comprehensive and engineering-focused theory of story construction. Specifically, we propose to examine how these axioms relate to two key safety outputs: safety case reports and accident investigation reports. These must both provide a compelling, credible story which is sufficient to convince the reader that the system is adequately safe in its proposed context of use (safety case report), or that the sequence of events leading to an accident has been comprehensively analysed (accident investigation report).

We also propose to explore how information can be transmitted in the opposite direction: that is, how a reader's reaction to an inconsistent narrative, characterisation or setting can be used to identify those aspects of the underlying engineering problem which might be inadequately specified. One specific approach to this would be to empirically investigate the transformation of the BGP story from the story through safety requirements, software design, code, and then execution and evaluation. As part of this process we will seek to investigate how variances in the detail of the story affect the transformations between lifecycle artefacts, and consequently affect our modelling of safety-critical systems. We propose to use this investigation to create a re-telling of the BGP, in a form which addresses some of the potential drawbacks of the current story.

References

Acland, A. (1841) A Letter to the Right Reverend Fathers In God. Ancient and Modern Ways of Charity, British Critic and Quarterly Review, Vol 29, J.G.F & J. Rivington.

Alwitt, L. (2002) Maintaining Attention to a Narrative Event, Advances in Psychology Research, vol. 18, pp. 99–114.

Bradbury, R. (1992), Zen in the Art of Writing, Bantam.

Dijkstra, E. EWD-1000 (1987) E.W. Dijkstra Archive, https://www.cs.utexas.edu/users/EWD/ewd10xx/EWD1000.PDF

Foot, P. (1967) The Problem of Abortion and the Doctrine of the Double Effect, Oxford Review, Vol 5.

Hoare, C.A.R. (1978) Communicating Sequential Processes, Communications of the ACM, Vol 21, Issue 8.

Information Commissioners Office, Turing Institute (2020) Explaining Decisions Made With AI. https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf, accessed October 2021.

King, S. (2000), On Writing: A Memoir of the Craft, Scribner.

Kovacs, M. (1989) The Epic of Gilgamesh, Stanford University Press.

Lamott, A. (1994) Bird by Bird, Anchor.

Lamport, L., Shostak, R., Pease, M. (1982) The Byzantine Generals Problem. ACM Transactions on Programming Languages and Systems, Vol 4, 382-401.

Lamport, L. (2021) My Writings http://lamport.azurewebsites.net/pubs/pubs.html#trans, accessed October 2021.

Martineau, H. (1838) Retrospect of Western Travel, Saunders & Otley.

Parham-Mocello, J., Ernst, S., Erwig, M. (2019) Story Programming: Explaining Computer Science Before Coding, Proceedings of the ACM Special Interest Group on Computer Science Edcation Technical Symposium.

Robinson, J. (1949) On the Hamiltonian Game (A Traveling Salesman Problem), Rand Corporation RM-303, https://www.rand.org/pubs/research_memoranda/RM303.html, accessed October 2021.

Salimitari, M., Chatterjee, M., Fallah, Y. (2020). A Survey on Consensus Methods in Blockchain for Resource-constrained IoT Networks. Internet of Things, Vol 11.

Taiwan Transportation Safety Board (2021), Major Transportation Occurrence Final Report: Airbus A330, TTSB-AOR-21-09-21

Yorke, J. (2014), Into The Woods: How Stories Work and Why We Tell them, Penguin.