# Computational Representations of High Profile International Cyber Incidents

Roger Hurwitz
Computer Science and Artificial Intelligence Laboratory (CSAIL)
Massachusetts Institute of Technology
rhhu@csail.mit.edu

Patrick Winston
Computer Science and Artificial Intelligence Laboratory (CSAIL)
Massachusetts Institute of Technology
phw@csail.mit.edu

Paper presented to the panel "**Multi-Disciplinary Methods for Cyberspace Research**" at the Annual Meeting of the International Studies Association
Montreal, Quebec, Canada, March 16, 2011

This work is funded by the Office of Naval Research under award number N00014-09-1-0597. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the Office of Naval Research.

Several high profile incidents have shaped both popular and government understanding of international cyber conflicts. One of the most iconic is the distributed denial of service attack (DDoS) on Estonian government, media and financial sites in April-May, 2007. The attack by "hacktivists" in Russia, perhaps supported by the Russian government, was a response to symbolic and legal moves by the Estonian government to expunge traces of Estonia's subjugation to the Soviet Union. The disruptions from the DDoS, though temporary, were severe because Estonia by its own choice was one of the most wired countries in Europe. The shock of the attack was also felt elsewhere. NATO had to weigh a response to a cyber attack on one of its members; many governments, including the Bush administration, more sharply saw cyber vulnerability as a threat to national security.

Observers, as well as victims, of the Estonian DDoS and other high profile incidents, are likely to remember and to recount them as stories rather as data entries in a statistical table. There are several good reasons for this tendency, including the scale of the attacks, the prominence of their victims, the extent to which they were disclosed and the usefulness of the story form to organize complex data. Although there were precedents for the Estonia incident as an international conflict pursued in cyberspace, the scale and consequences of the attacks took matters to a new level. Attacks before Estonia, e.g., hactivist attacks on Israeli and Palestinian websites during the Second Intifada (2000-2001), mostly defaced the sites or redirected users to spoof sites; they rarely brought down the sites and halted their services to the public. Before the Chinese based exploits on Google, no premier American technology company publicly revealed announced a Chinese attempt to steal its intellectual property, although many experienced it.[1] Also, because targets of international cyber crime, industrial espionage and sabotage rarely report the type of security breaches or losses incurred, in a timely manner, if at all, the development of standardized formats and compilations for such incidents have also lagged. So when cyber incidents are well publicize there are few statistical resources for measuring their significance; they will stand out. Finally, the story form is a more intuitive and arguably a better means of bringing together the various dimensions and temporal sequence in an international cyber conflict than spatializing, categorical statistics.

Indeed, we argue *inter linea* in this paper against the notion that the account of a cyber incident is somehow "pre-scientific" Rather computational representations of such accounts, as specified below, especially when these representations are

---

[1] US-China Economic and Security Review Commission, Report to Congress, 2010. Washington: US Government, Nov., 2010, 236-240; US-China Economic and Security Review Commission, 2009 Report to Congress. Nov. 2009, 179-180; Northrop Grumman Corp., Capability of the People's Republic of China to Conduct Cyber Warfare and Computer Network Exploitation Prepared for The US-China Economic and Security Review Commission. McLean, VA: Oct 9, 2009, retrieved at http://www.uscc.gov/researchpapers/2009/NorthropGrumman_PRC_Cyber_Paper_FINAL_Approved%20Report_16Oct2009.pdf

enriched with commonsense and reflective knowledge, surpass the value added of statistics or feature vectors based upon such accounts, or more impoverished sources like surveys. While a feature vector might code for perpetrator, victim, action, damages, consequences and other relevant attributes of an incident, it suppresses the internal relations, e.g., causal logic, temporal unfolding, among these attributes that are captured in the language and structure of a story.

The problem with the accounts or similar incident reports is not their having less precision, but that their idiosyncrasies in language and narration might thwart comparisons and aggregations with formal, verifiable and replicable means. Put another way, identifying similar concepts and structures across texts remains a formidable computational challenge in artificial intelligence, regardless of our own natural abilities to understand analogies and recognize films in the same genre or having virtually identical plots. Of course, this challenge of so-called "unstructured data" is not specific to reports of cyber incidents. In established fields, like medicine and law, the challenge is reduced by having inputs created with controlled vocabularies. The field of international cyber incidents is apparently too young for the emergence of a standard vocabulary for reporting cyber incidents.[2] Organizations, like the United States military, that need to plan for and respond to cyber incidents, can require their personnel to use specific terms in discussing them,[3] but to limit data to only reports with such language could radically reduce the use of available sources and legacy reports.

A different solution would be a computational system that interprets the text of an incident report as a sequence of events – instances of various generic events (classes), which recur in incident reports, stories, accounts, etc. Such representation would

- Preserve relational information in the texts;
- Enable analysts to anticipate of the trajectory of a current incident through comparing the events already reported with those in past incidents.
- Enable analysts and decision makers to evaluate the potential effects of different interventions;

---

[2] R. Hurwitz & M. Seifter in unpublished research on international media descriptors for cyber conflict, 2005 – 2010, found little consistency in use and reference of terms like *attack, exploit, crime* with regard to cyber events. This was demonstrated by the low recall and accuracy when terms for hostile events that were learned in one year were used to retrieve media reports of such events in other years.

[3] "cyber's unique vocabulary doesn't discretely describe the nuances of its mission sets, lend itself to established legal interpretations of authorities and limitations, or reflect the standardized vernacular of the other domains… For these reasons, I have tasked the Joint Staff to develop the attached lexicon to align [Command Operations] vocabulary with standard joint terminology. This lexicon will be used as the starting point for normalizing terms in all cyber-related documents…" J. Cartwright, Vice Chairman of the Joint Chiefs of Staff, Cyberspace Operations Lexicon, 2010. Retrieved at http://www.nsci-va.org/CyberReferenceLib/2010-11-Joint%20Terminology%20for%20Cyberspace%20Operations.pdf

- Identify and support substantial differences in accounts and perceptions of the incidents. This could facilitate parties in the incident to better understand one another's motives and responses.

The next three sections of this paper discuss our work in developing such a system. First we describe the natural language processing system that builds a representation of a story's plot from surface text. Then we provide examples of inputs and outputs and follow with a discussion of features that can make its representations more relevant for international relations analysis. We conclude by noting other research, challenges and obstacles relevant to creating a library of incidents against which new ones can be evaluated.

I

The representations of the cyber incidents are produced from text inputs by the Genesis System for Natural Language Understanding and Visual Recognition, being developed by Winston and his students at MIT's Computer Science and Artificial Intelligence Laboratory. The development of the system is guided in part by the principle that "language enables description; description enables story telling and understanding; and story telling and understanding lie at the center of human education."[4] Stories integrate background information with the events described, relate characters' actions to their motives and have inherent organization of events and actions, in the sense that these create conditions for subsequent events, characters' motives and actions. An additional principle is "language enables imagination, and the deployment of visual and motor perceptions and actions on situations not directly experienced." Consequently, people can visualize and rehearse what they are told – including descriptions of incidents – even if they themselves have not experienced these incidents. Together these principles suggest the Genesis system's orientation toward *phronesis* or practical reasoning, in contrast to syllogistic reasoning. Such reasoning inheres in people's use of analogies and precedents to form expectations and select actions in situations.[5] The idea of "phronetic social science," which focuses on "context, practice, experience, common sense intuition and practical wisdom," has been proposed by contemporary social and political scientists, e.g., Habermas, Bourdieu, as a realizable, valuable alternative to a social science oriented toward theories, analysis and universals.[6] However, discussion of this point lies outside our present scope.

In building representations of cyber incidents and other stories, the Genesis language system abstracts from passages in the surface text to intermediate

---

[4] P. Winston, "S3,Taking Machine Intelligence to the Next, Much Higher Level Version of January 31, 2011 (unpublished).

[5] Precedents would help evaluate how successful an action might be in the present case based on its prior use in similar ones, rather than to legitimate an action or decision, as occurs in the use of precedents in judicial processes.

[6] C. Geertz, Empowering Aristotle. *Science* 293:5527 (6 Jul 2001), p 53.

representations that derive from Wendy Lehnert's constructs of affective plot units.[7] "Plot units are conceptual structures that overlap with each other when a narrative is cohesive." The overlapping intersections of the plot units can be interpreted as arcs in a graph that encodes the plot(s) in the story, and the graph in turn can be analyzed to identify the story's main narrative thread, on one hand, and peripheral episodes or digressions, on the other. A plot unit itself configures a story character's affect regarding a situation/ event and a transition to another affective state or persistence of the same one. The transition can result from the character's thoughts about the situation or from changes in the situation produced by the character's or a related agent's action or by an external event. Lehnert distinguished three affect states: positive ("+" for events that please the character), negative ("—"marking events that hurt) and an affectively neutral mental state ("M" for thought or wish).

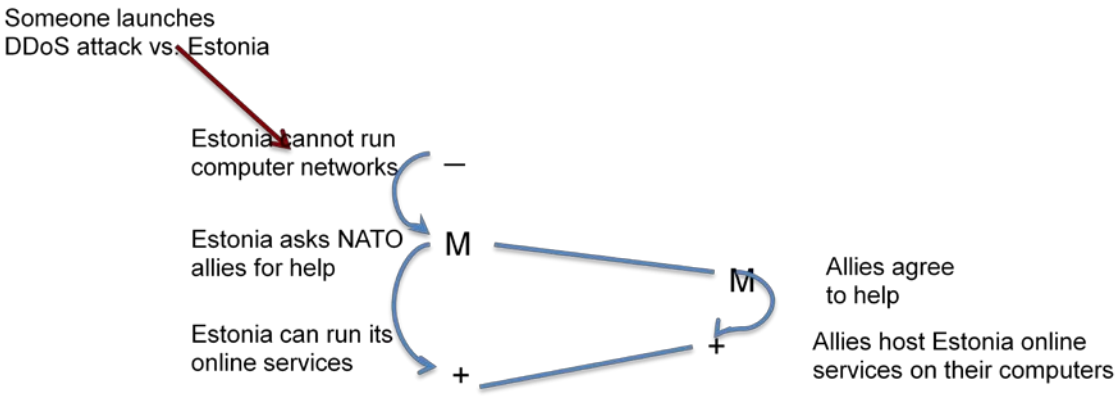
So the vignette:

> Estonia's computers are hacked. It wants to run online services. It gets hosting elsewhere and can now run services

can be represented, relative to the character Estonia:

> -- (computers hacked) → M(wants to run services) → + (gets hosting; services)

The affect states and action of other agents, e.g., Estonia's allies, can also be represented and tracked, thereby building more complex plot units, which can represent stereotypical interactions. The schema below




can be seen as a instance of interactions in which one agent helps another solve a problem.

On such basis, Lehnert extended a few simple monadic and dyadic plot units (see Appendix 1) to a set sufficiently rich for hand coding and characterizing

[7] W. Lehnert, Plot units and narrative summarization. *Cognitive Science* 5:4 (1981), 293-331.

episodes in complicated stories like O. Henry's "Gift of the Magi," and historian Arnold Toynbee's retelling the Jesus story.[8] Her research, in the field of Artificial Intelligence, was motivated by the question of how people remember and summarize stories. The plots units were proposed as identifying possible organized chunks of memory, with the sequence of chunks that comprised the main narrative thread being what people would likely remember.[9] Note, however, this work strongly resonates with the morphological or formalist tradition in literary analysis, that descends from Propp's studies of Russian folk tales.[10] That tradition understands stories as involving a small number of agents playing out stock roles, e.g., hero, villain, helper, through a sequence of episodes drawn from a limited number of types, e.g., departure, struggle, return.

The discovery of plot units within the (surface) textual account of a cyber incident is a challenge for the Genesis System, which Lehnert's research did not face. Despite the advances in computational linguistics in the nearly three decades since her original research, this challenge is still formidable for several reasons. First, computational linguistics has primarily dealt with syntactical analysis of sentences and statistically based inferences of tokens in texts, at the expense of in-depth semantic understanding of them. Second, because surface texts can articulate plot units (or other higher level representations) in countless ways, matching texts to a linguistically, but generically specified plot unit is not trivial. Third, comprehension of texts usually marshals the reader or listener's background knowledge and common sense reasoning. In the example above, one needs to know that "giving a snack" means giving John some food and this transfer implies that John will eat the food, thereby assuaging his hunger.

On the sentence level, the Genesis system sees the world through nearly two dozen frame-like representation, including representations for threads (classifying the nouns and verbs into entity and action hierarchies), trajectory, transition, transfer, location, time and cause, and coercion (see Appendix 2). Some representations are based on work by linguists and researchers in natural language processing; others derived from needs to capture meanings in stories input in the system, which the frames at hand did not represent. A particular sentence will satisfy (or match) only some, but not all, frames. On the contrary, many different frames are needed because there are many kinds of events that can be described. However sentences can instantiate these representations, whether they describe events in the physical world (the bird flew to a tree), a

---

[8] H. Alker, "Toynbee's Jesus: Computational hermeneutics and the continuing presence of classical Mediterranean civilization." In Alker, *Rediscoveries and Reformulations: Humanistic Methodologies for International Studies.* Cambridge: Cambridge U. Press, 1996, 104-143.

[9] Lehnert's work built on her mentor Roger Schank's work in natural language (R. Schank & R. Abelson, *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*, 1977) which largely treats sentences as describing purposive actions producing changes in states of affairs. Stories or accounts of activities as agents' acting in ways, required by a setting or situation; the agents are following scripts. The problem of language understanding is then to discover (or abstract) the actions or scripts that are expressed in the surface text.

[10] V. Propp, *Morphology of the Folktale* (1928). Austin, TX: U. of Texas Press, 1968.

world of artifacts (the computer ran quickly) or an abstract symbolic world (the country moved toward democracy). Such capability is, of course, quite important in dealing with reports on cyber incidents.

The path from sentences to instantiated representations goes through the Start Parser, developed over a 25-year period by Boris Katz and his students.[11] In comparison to other, statistically trained parsers, Start blunders less, and, instead of producing a parse tree, it outputs a semantic net, which greatly facilitates instantiating frame-like representations for the sentences. WordNet is a also used in this stage as a source of classification information, a move that is a temporary shortcut in lieu of writing classifications in English ("a province is part of a country") or inferring them from texts already acquired.

In developing Genesis, Winston and his associates have used simple plot summaries of Shakespeare plays to benchmark system capabilities and to identify areas for additional work. The plots have the virtue of being both familiar, so researchers can readily recognize misrepresentations, and rich in universally important factors such as power, emotions, consequences, and ties between people. Because many plots involve high politics, the same bodies of knowledge that are needed to understand these plots appear appropriate, if not sufficient, for understanding international conflicts. That reduced the stretch in using the system to represent and analyze accounts of recent high profile cyber incidents, such as the alleged 2007 Russian cyber attack on Estonia's network infrastructure and the cyber attacks on Georgian websites during the 2008 war between Georgia and Russia over South Ossetia.

Like the Shakespeare plot summaries, news stories and almost all but the most exacting legal or technical texts, do not include some information that people use to make inferences about changes in the situation or agents' affects, as part of understanding what they are reading or hearing. The system already had acquired some of this information or *background knowledge* because the developers had supplied it in English sentences as the system needed it to understand the Shakespeare plots. Other bits needed to be added, e.g.,

> If XX, an entity, harms BB, an entity, and BB belongs to YY, XX has harmed YY
> Defacement of a website is a kind of harm to the website.

Equipped with these rule-like, reflexive statements, Genesis can produce *an elaboration* graph of predictions and explanations, in which cascades of inferences augment the explicit elements provided in the input, going well beyond the information given. The elaboration graph can then be analyzed for the presence of Lehnert-like plot units, respecified in English sentences, to yield reflective knowledge of the meaning (purport) of events in an incident or of an entire incident itself. Specifically the system was able to characterize the attack on Estonia as a case of revenge.

[11] http://start.csail.mit.edu/start-system.html

II

The inputs for the cyber incidents were redacted from short accounts of selected incidents based on open sources. These incidents received wide coverage in the Western press and impacted government officials and professionals responsible for cyber security. As a set, they begin to represent the diversity of threats to states in cyber space and responses to them, per a common [Charney's, 2010] four-way typology of cyber threats: cybercrime, industrial espionage, political espionage, cyber war/ sabotage. To these, we add cyber activism at the national and international levels against established governments. In addition to the previously mentioned Estonia and China-Google incidents, they include the attack on Georgian networks during the war over South Ossetia (2008), cyber activism in the Iranian (failed) Green Revolution (2009), Stuxnet (2010), and the TJMaxx/ Heartland Payment cases (2006, 2007). Obvious candidates for extending this set are Ghostnet/ Shadow in the Cloud, the Wikileaks publication of US diplomatic reports (2010) and cyber fueled revolutions in Tunisia and Egypt (2010-2011).

The Estonia and Georgia cases were processed first and reported here, because they are similar both with regard to apparent motivations for the cyber attacks and the suspected involvements of the Russian government, hereafter Russia, or Russian hactivists. We believed that the system's discovering a similarity at the plot level would be a promising achievement. We were also interested in learning how much background knowledge would be needed to successfully represent the first case and how much additional background knowledge the second case required. The language is admittedly very elemental to assure that Start would parse the sentences and the relations between events and affective states would be transparent. The input for Estonia is shown below, the considerably longer input for the Georgia case is in Appendix 3.

> Estonia and Russia are countries. Estonia built many computer networks. Estonia insulted Russia because Estonia relocated a war memorial. Estonia relocated the war memorial because Estonia did not respect Russia. Someone damaged Estonia's valuable computer networks after Estonia harmed Russia. The computer networks did not work because of the damage. The damage harmed Estonia. Estonia created a center to study computer security. Estonia believed other states would support the center.

Outputs of Start and Genesis processing are shown in the screen shots below. These include (part of) the input in white boxes, the inferences on the basis of reflexive (rule-like) knowledge and the boxed information that is used in instantiating a sequence of actions and affect states that constitute a large, complex plot unit or narrative thread. The black lines connect the explicit and inferred information in both logical and temporal sequences. The red lines connect the yellow boxes that comprise the instantiated script, which is named at the bottom of the screen. As seen, the two incidents (Fig. 1 & 2) shared the

same plot, “getting even,” if the analyst is a friend of Russia, or “revenge,” if the analyst identifies with the cyber attacks’ targets

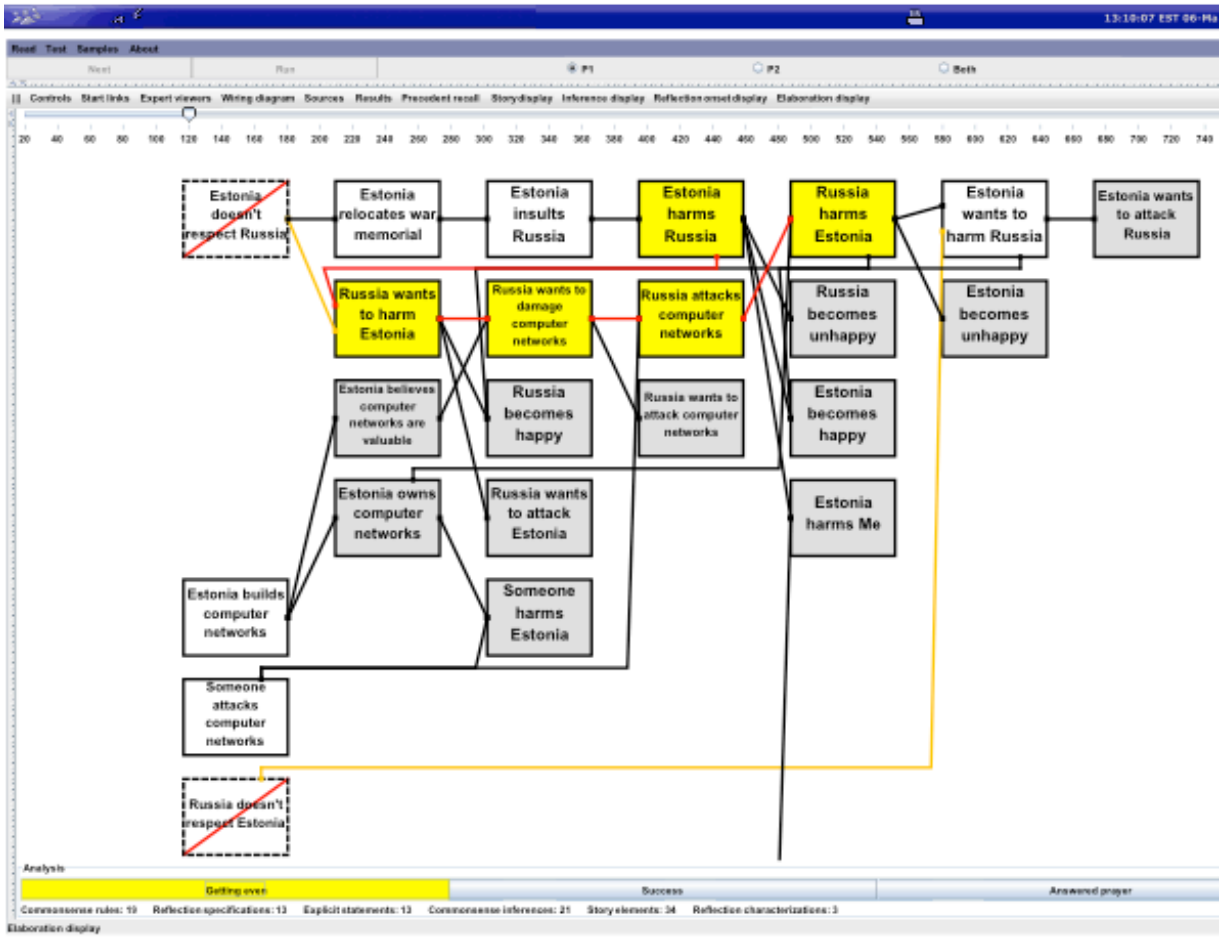



Fig 1: The chain of events and motivations leading to the cyber attack on Estonia is composed of the yellow boxes, connected by the red line, starting at “Estonia harms Russia.” This plot is identified at the bottom of the screen as “getting even.” Note in this interpretation of the text, the cyber attack is a means to Russia’s end of harming Estonia.

Fig 2: Genesis finds a similar chain of events and motivations leading to the cyber attack on Georgia

Several features of the output are worth noting. First, only those input statements and inferences that are related to the instantiation of the plot appear (with their boxes' initial color having been white or grey, respectively). Second, Russia is inferred to have cyber attacked both Estonia and Georgia. The inference is based on a common sense rule that if an agent has a stated or inferred desire to harm the target and no other actors are known to have such desire, that agent is responsible. The inference is both debatable and reveals how the system is input bound. Had the inputs indicated that the Russian people were angry about Estonia's and Georgia's actions and the background knowledge included that Russian "hactivists" were part of the Russian people and could launch cyber attacks, the conclusion would have been different. Third, the discovery and naming of a plot depends on a point of view, which tracks one of the actors. A friend of Estonia finds the incident satisfies the definition of (senseless) revenge as defined for the module for plot instantiation,[12] where XX, in this case, is Estonia.

> Start description of "revenge". XX and YY are entities. XX's harming YY led to YY's wanting to harm XX. YY's wanting to harm XX lend to YY's harming XX. The end.

A friend of YY, or Russia in this case, would, however, define the incident as one of "getting even." Furthermore, the plot could change relative to one of the actors if that actor figured in subsequent events. Had NATO responded to a purported Estonian call for assistance, by imposing some meaningful sanction on Russia,

[12] Developed by Winston's student David Nackoul.

the affect of Russia and its friends would have changed from happy to sad. The would then have been a Pyrrhic victory for them, with their "getting even" recognized as a stage in a longer plot.

By understanding input information as actions, affects or mental states, e.g., desires, and their configurations in plot units, the system can also answer specific questions about an incident. So "who attacked Georgia's computer networks?" and "Why?" can be answered by reading off the conditions that precede the cyber attack in the plot unit (See Appendix 4).

III

Genesis has proved it can understand cyber incidents. However because of its resemblance to expert systems, whose brittleness is notorious and whose effectiveness is best in limited, well specified domains, we need question the extent of this success.

1. Was the bar set too low by hand crafting the inputs?
2. Might newly reported incidents require so much additional background information, as to thwart the possibilities of understanding new incidents in terms of previous ones?
3. What payoffs for understanding international cyber conflicts (and cooperation) derive from working with this system?
4. What extensions to this system could make it more useful for the study of international cyber conflicts?

1. Some editing of accounts was needed, because Start, despite its relative robustness, does not parse all sentences. It fails on certain syntax, like elliptical clauses: "the people in France are French and not English" is not parsed as "the people in France are French; the people in France are not English." If we choose to omit such editing, so Genesis works on only sentences that Start parses, we risk losing information that contributes to the realization of a plot.

A more substantial problem is the validity of adding information to the inputs to facilitate the system's instantiation of plot units or narrative thread. That is, some sentences need explication in order to trigger inferences or be attached to a plot unit. For example, a sentence was inserted in the Georgia case to say that Georgia considered its computer networks important, because the system infers that an agent is harmed by damage to an artifact it owns only if the agent values that artifact.[13] We believe this *ad hoc* move is justified, because such information could be generalized in the (reflexive) knowledge base for cyber incidents and be available for processing of other cyber incident reports, viz.,

[13]The relatively minor dependence of Georgia government and military on its computer networks argues against the cyber attack as part of Russia's military operations and for the likelihood of the attackers being Russian hactivists, who wanted to avenge Georgia's harm of people under Russian protection.

> States value infrastructure. Infrastructure includes computer networks, telecommunications, power grids…

Ironically this step would replicate the "real world" learning from the Estonia and Georgia incidents that components of a state's infrastructure are vulnerable, because of their dependence on computer communications and control.

An even stronger criticism is that Genesis might work in the laboratory, but is likely to fail under "field conditions," that is, in handling a news stream regarding a cyber incident, in which causes of events and motivations of agents are under-articulated. However, given a more robust parser and broader reflexive knowledge base, the system could, in principle, be outstanding in "field conditions." It would be able to generate inferences about the various agents and events and determine which of these could be coerced into a coherent, connected sequence a plot units, spanning the incident from its beginning to present stage, regardless of the noise in the information. Indeed the processing of the Georgia input demonstrated this capability in instantiating a plot that ignored almost all information in the input not directly related to the progress of the main events.

2. The need for domain knowledge has for decades plagued efforts to automate understanding and analysis of messages, reports, news accounts, etc. In regard to international cyber incidents, needed domain knowledge includes knowing the names of states, that states are entities, that they have governments, armies, land, provinces, infrastructures, economies; that other agents include NGOs and criminal organizations; that there are different types of cyber attacks and exploits and that these may have different objectives, etc. This is just a beginning, but much of this information is low hanging fruit. Some can be acquired from machine parsed data accumulated by international events data research,[14] and more will become available with the development of ontologies and taxonomies for cyberspace. A relevant concern, however, is whether a broad but shallow knowledge base will suffice for processing reports on a large number of incidents or will most new, high profile incidents require additional information? The success of the events data projects in covering kinetic-world events and our own work gives some reason for optimism. To process the Georgia incident, only one political fact needed to be added to knowledge previously acquired for processing the Estonia incident, viz., that South Ossetia was a part of Georgia, or, expressed more generally, regions are parts of states. Nevertheless, cyberspace is a rapidly changing domain. As evidenced by Stuxnet, cyber threats are becoming increasingly sophisticated and potent, and as evidenced by Wikileaks, even a very small group of individuals can use the technology to precipitate international incidents. New incidents are therefore likely to establish

[14] Events data research projects, directed by Philip Schrodt, first at University of Kansas and now at Penn State, http://eventdata.psu.edu/ have compiled extensive lexicons of states, government agencies, non-state actors at the international and sub-national levels and categories of conflict & cooperative actions in kinetic, political and economic spaces, but yet in cyberspace.

new facts. Fortunately, Genesis can easily accommodate these. One needs only type them in English to the knowledge base.

Domain knowledge for the cyber incidents also includes the reflective knowledge that the system uses to recognize plot units or a narrative thread in a sequence of events, affective states and motivations among agents. To the extent that such knowledge is a theory of the domain, it is clearly inadequate. On a reductive, realist, state centric, unitary actor view of international relations, much of international relations might resemble Shakespeare's more violent plays, but even on this view, many international incidents, including some cyber conflicts, do not. For example, the accounts of cyber crimes and of international cooperation in the takedown of the criminals are not tales of revenge, but rather police procedurals, wherein bureaucratic administration replaces fate and passion as a driving force. Similarly Google's announcement it would stop complying with Chinese censorship, as a response to the China based attempt to steal its intellectual property, appears less a futile stab at "getting even" than an attention getting complaint, that may avoid a drastic exit, per Hirschman's model.[15] Also the accounts of activists using cyber technologies to evade censorship, organize democratic revolutions and elicit international support are not easily coerced into traditional stories, though they do bear some resemblance to populist novels and films.

The system therefore needs constraints against premature recognition and naming of plots. One such constraint would be a rule at the level of reflective knowledge that identified the main plot as the longest connected sequence of plot units which included the most recent observed event. There is also need to develop support for different theoretical perspectives, which is distinct from taking the previously discussed system capability of taking viewpoints of different agents. For a start, theoretical perspectives could be differentiated by the types of motivations inferred for agents, e.g., a problem solving interest vs. the desire to strike back, and by means available for achieving goals, e.g., mutual assistance vs. duel-like confrontations. More generally and vaguely, system development needs to acknowledge the possibility that new types of stories are unfolding in cyberspace.

3. Our use of Genesis to understand and analyze reports of cyber incidents has led to more attentive readings of these reports, particularly with respect to indications that the affordances of cyberspace can change the character of international relations. Take, for example, the representation of the Georgia incident, for which the system identified Russia as the cyber attacker, based on knowledge that treats a state as a unitary actor that can suffer harm and be in a lousy mood. This inference corresponds to assumptions in both realist and idealist theories of international relations that people can act at the international level only through organizations of their states. But suppose the system had also read a report which noted the negative affect toward Georgia in online

[15] A. Hirschman, *Exit, Voice and Loyalty*. Cambridge, MA: Harvard, 1970.

conversations among Russian "hactivists,"[16] and the knowledge base included rules that groups of individuals could launch international cyber attacks. The resulting representation then might have identified Russians as the cyber attackers, acting to help the Russian military or the people under attack in South Ossetia. The story of the cyber incident might then have been one of "piling on," in a Georgian perspective, and "rallying for our brothers," in a Russian perspective. The capability of influencing international relations that a loosely organized group now has via their actions in cyberspace may become a recurrent feature in cyber incidents: e.g., the hacker group Anonymous's denial of service attacks on financial and government institutions in Tunisia was prompted by the revolution and may have contributed to its victory.[17]

4. One useful enhancement of the system would be its measuring the various actions in an incident along a scale for their degrees of international conflict or cooperation. Such scales have been developed in events data research for measuring kinetic, political and rhetorical actions at the international level,[18]but none has been developed for actions in or about cyber space. The scale would need to rank order the actions in according to their inherent hostility or amity. It would also need to indicate the perceived equivalents of the ranks in the other spaces, so, among other things, patterns of escalation and de-escalation could be identified over the sequence of actions in the incident. This enhancement would benefit analysts and decision makers alike, especially at the onset of incidents, when responses and interventions are being weighed. However its development will likely prove difficult. Because conflict is cyberspace is so recent, few people have strong intuitions regarding the relative severities of various cyber actions. A Stuxnet weapon with the capability of breaking critical infrastructure facilities seems particularly ominous, but would a narrowly targeted attack with such a weapon be more severe than a denial of service attack that deprives doctors of the information they need to save their patients' lives? Even if the IR theorists, practitioners and security experts whom we intend to poll can agree on marking levels of cyber enmity and friendship, the results might be not be directly applicable to the texts upon which representations of cyber incidents will be built. In other research we found that the language describing cyber events over the past five years has lacked specificity and stability. For example, the references of the generic "cyber attack" can range in some years from DDoS to malware and phishing, but are more constricted in others. Only recently has the general press begun distinguishing between cyber exploits and attacks. Consequently, to the extent that there is reason to differentiate among cyber actions, the system will need in most cases to learn what the action was from the description of its effects.

---

[16] Such as J. Carr, *Inside Cyber Warfare: Mapping the Cyber Underworld.* Sebastopol, CA: O'Reilly, 2009.

[17] BBC News, Anonymous activists target Tunisian government sites, 4 Jan. 2011. Retrieved at http://www.bbc.co.uk/news/technology-12110892, 3/1/2011.

[18] Most notably Goldstein's revision of the WEIS, see J. Goldstein, A conflict-cooperation scale for WEIS events data. *J. of Conflict Resolution*, 36:2 (1992), 369-385.

## Conclusions

The Genesis system can compute the plot units in textual accounts of cyber incidents and on their basis instantiate the type of plot or narrative thread for an incident. The plot discloses the motivations of agents and the logic that drives the events. This achievement grounds some guarded optimism for developing a library of cyber incident representations (or templates) against which reports of an unfolding current incident might be matched to produce expectations about its future course. The project of producing formal representations of narrative accounts comes from our recognition that much current knowledge about cyber conflict and cooperation derives from anecdotes and stories rather than statistical data and theories. We also expect that retelling such stories will be an important part of building the cultures within organizations that will address the challenges of conflict and cooperation in cyberspace, so knowledge of the social practices of agents in these stories and the discursive practices of their tellers can provide insight into these organizations.

Such stories, particularly as justifications of actions or outcomes, can be one-sided and present irrational or inappropriate actions for emulation.[19] Genesis can ameliorate this tendency, for the benefit of the self-reflective analyst, with its present capability of representing the same incident from different characters' points of view. These representations might differ in their plot units as well as the evaluation of affects before and after actions.

However we need to distinguish between understanding a particular perception of cyber incidents and understanding cyber incidents in terms of their causes, consequences and textures. We have already seen that the traditional narratives are not adequate for the new types of international interactions that cyberspace affords: the rapid creation of oppositions whose mobilization transcends political differences; the intimate involvement of an international public and the sometimes effective participation of outside hackers. These suggest the need for new plots. Will that be enough? The late political sociologist Charles Tilly noted that stories are reductive by virtue of having a few characters and attributing actions and outcomes to these characters intentions.[20] They are thus inadequate and incompatible explanations of social processes and causality. Put another way, a plot is not a theory.

## Contributions

This paper has

---

[19] G.Akerlof & R. Shiller, *Animal Spirits: How Human Psychology Drives the Economy and Why It Matters for Global Capitalism.* Princeton, NJ: Princeton, 2009.

[20] C. Tilly, *Stories, Identities and Political Change.* Lanham, MD: Rowman & Littlefield Publishers, Inc, 2002.

- Argued for the value of story related data in the study of cyber conflicts and incidents;
- Argued that computational representations of incident stories and reports can be vital for understanding how such incidents unfold and the cultures of organizations that relate the stories and reports;
- Noted that the use of the Genesis in story understanding has led to more attentive readings of cyber stories and reports by human investigators;
- Identified concerns and limitations with the story understanding system on technical and conceptual levels;
- Suggested reasons for optimism and next steps in developing system to understand accounts of cyber incidents.

## Appendix 1

W. Lehnert, Plot units and narrative summarization. *Cognitive Science* 5:4 (1981), 298-299.



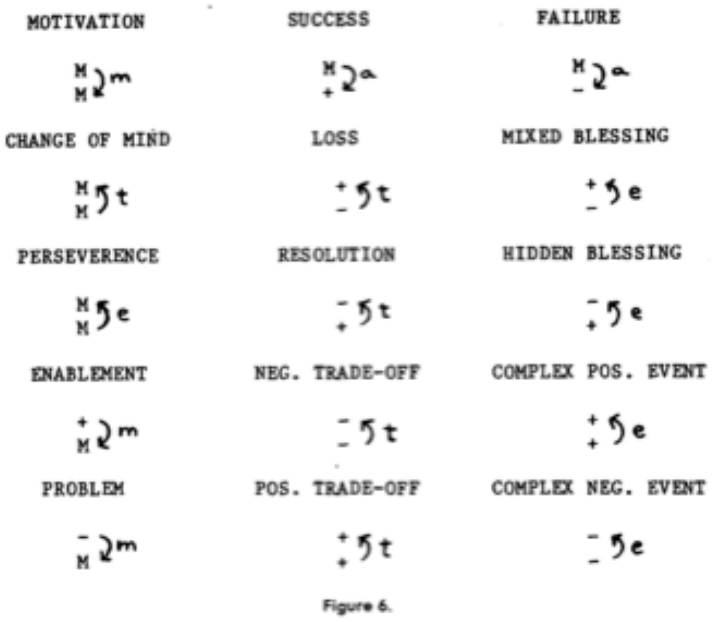



Figure 6.

Sometimes, a primitive plot unit will appear without other interceding affect states. This occurs most commonly with the units "problem," "enablement," and "motivation." Other primitive plot units tend to be broken up by interceding affect states. For example, it may take months (with lots of interceding emotional reactions) to find out that a job promotion is now leading to an ulcer. This would be an example of a mixed blessing, or a good thing turned sour.

EXAMPLES OF PRIMITIVE PLOT UNITS

| | |
|---|---|
| PROBLEM: | You get fired and need a job.<br>You bounce a check and need to deposit funds.<br>Your dog dies and you long for companionship. |
| SUCCESS: | You ask for a raise and you get it.<br>You fix a flat tire.<br>You need a car so you steal one. |
| FAILURE: | Your proposal of marriage is declined.<br>You can't find your wallet.<br>You can't get a bank loan. |

| | |
|---|---|
| RESOLUTION: | Your broken radio starts working again.<br>They catch the thief who has your wallet.<br>You fix a flat tire after a blow out. |
| LOSS: | Your big income tax refund is a mistake.<br>The woman you love leaves you.<br>The car you just bought is totaled. |
| POS. TRADE-OFF: | You buy a new Toyota and then inherit a Porsche.<br>You take a day off and then realize it's a holiday.<br>You get a raise and then win the Irish Sweepstakes. |
| NEG. TRADE-OFF: | You get fired so you don't have to take a lousy job assignment.<br>Your car blows up so you don't have to make the next insurance payment.<br>You lose the election so you don't have to placate demanding voters. |
| PERSEVERENCE: | You want to get married (again).<br>You reapply to Yale after being rejected.<br>You want to ski again after a bad skiing accident. |
| HIDDEN BLESSING: | You get audited and they owe you.<br>You sprain an ankle and win damages.<br>Your mother dies and you inherit a million. |
| MIXED BLESSING: | You buy a car and it turns out to be a lemon.<br>You fall in love and become insanely jealous.<br>Your book is reviewed but they hate it. |
| CHANGE OF MIND: | You apply to Harvard and then to Yale.<br>You want to buy a car but decide against it.<br>You want to see a movie until a friend pans it. |
| MOTIVATION: | You need advice so you decide to ask a friend.<br>You want to buy a car so you apply for a loan.<br>You want to reach a client so you call him. |
| ENABLEMENT: | You decide to celebrate after a raise.<br>You receive a book and decide to read it.<br>You get a loan and have to pay it back. |
| COMPLEX POS: | A gift is indicative of close friendship.<br>Your raise signifies recognition.<br>You win respect by getting a rolls royce. |
| COMPLEX NEG: | You lose $100 when your wallet is stolen.<br>You break an arm in a car accident.<br>Your house burns down and you aren't covered. |

Lehnert, 315: plot units in “Gift of the Magi”

When a narrative embodies total symmetry, we detect this immediately and remember it as a salient feature of the story. For example, consider *The Gift of the Magi* by O. Henry. This is a story about a young couple who want to buy each other Christmas presents. They are both very poor. Della has long beautiful hair, and Jim has a prized pocket watch. To get money for the presents, Della sells her hair and Jim sells his pocket watch. Then, she buys him a gold chain for his watch, and he buys her an expensive ornament for her hair. When they find out what they've done, the are consoled by the love behind each other's sacrifices.

The story exhibits an extreme symmetry:

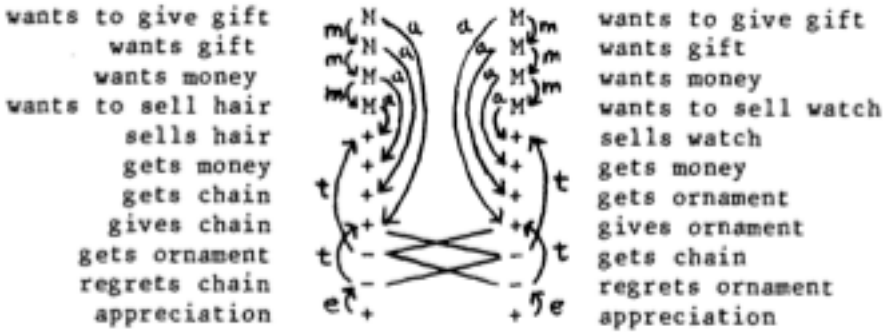
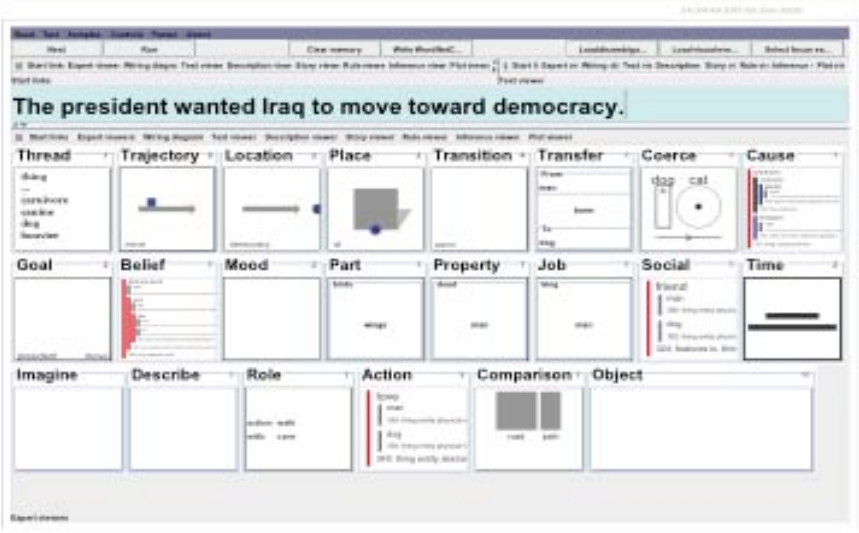



Figure 34.

This configuration involves (1) nested subgoals and (2) achievement (in getting and giving the gifts), (3) loss (in no longer having the things they sold), (4) another loss (in no longer having pleasure from the act of giving) (5) regrettable mistakes (the bad gifts), and (6) hidden blessings (in realizing what the gifts signify). Not only is there complete symmetry across both characters, but there are ironic causalities across the plot units. For example, the sense of loss does not occur until the top-level goals are achieved (when the gifts are exchanged). At the same time, this loss is also the basis for a hidden blessing at the end of the story, when they realize how the gifts signify their unselfish love for each other.

## Appendix 2




The Genesis system features a large suite of representations: threads, trajectories and transitions appear frequently in written and spoken natural language.

## Appendix 3

Georgia and Russia are countries. Georgia owns Georgia's army units. Georgia owns Georgia's computer networks. South Ossetia is a region. Georgia believes its computer networks are important. Telecommunication websites are
artifacts and computer networks are artifacts. Georgia's telecommunication websites are part of Georgia's computer networks.

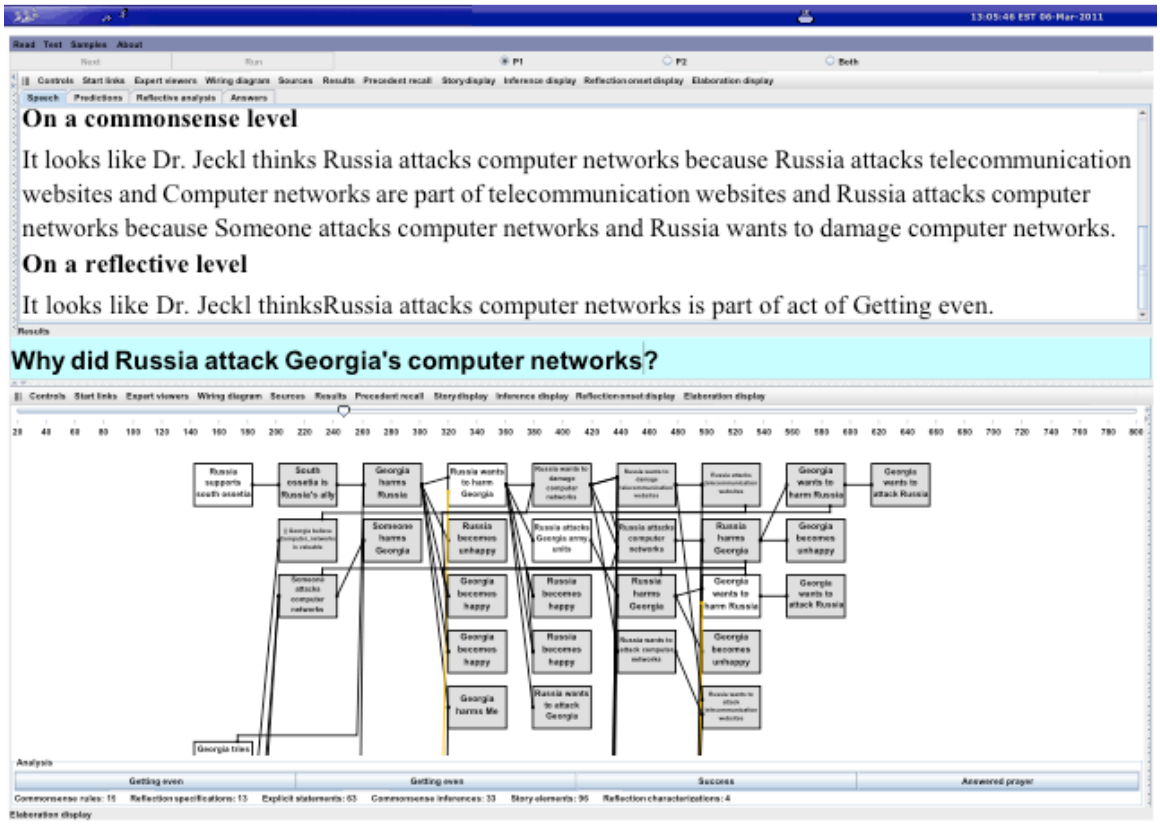
Georgia became independent when the Soviet Union broke into separate states in 1989. In the 1990s, the government of Georgia lost control of South Ossetia. Most of the people in South Ossetia were not ethnic Georgians, but members of other ethnic groups, including Russians. They set up their own government.
South Ossetia was supported by Russia. Russia gave many of the people in South Ossetia Russian passports. Russia was committing itself to protect these people.

Georgia tried to control South Ossetia. This strategy led to occasional fights between army units of Georgia and army units of South Ossetia. In June and July army units of Georgia and South Ossetia attacked each other several time with artillery fire. In late July, there were several denial of service attacks on Georgian government web sites. In August, 2008, Georgian army units entered South Ossetia, after the Georgian government said the army of South Ossetia had attacked villages in South Ossetia where ethnic Georgians lived. Russian army units entered South Ossetia. Russia attacked Georgia's army units because Russia wanted to harm Georgia.

Then, someone attacked Georgia's telecommunication websites. There were also attacks on Georgian government websites. These attacks included defacement, denial of service and corruption of databases. These attacks kept the Georgian government

from using the Internet to communicate with the rest of the world or with the small number of people in Georgia with Internet access. These attacks continued for five days, until a cease fire between Russia and Georgia was declared. The people who organized the cyber attacks were probably Russian activists and not Russian army units. The attacks did not damage Georgia very much because Georgia was less dependent on the Internet than are technologically advanced countries.

## Appendix 4




Genesis answers questions about specific events, motivations and moods either at the common sense level by backtracking through the assertions and inferences on the story line and at the reflective level by putting the event/ affect in the context of the plot . Since Dr. Jeckl* has identified himself as a friend of Russia, the plot is “getting even.”

* The traditional spelling of Dr. Jekyll, who is a recurrent interlocutor in Genesis work, was altered to facilitate speech synthesis of the name from its appearance in text.