Defining A New Type of Global Information Architecture for Contextual Information Processing

Gregory Vert
Center for Secure Cyberspace
Computer Science
Louisiana State University
(206) 409-1434
gvert12@csc.lsu.edu

S.S Iyengar
Center for Secure Cyberspace
Computer Science
Louisiana State University
(225) 578-1252
iyengar@csc.lsu.edu

Vir Phoha
Center for Secure Cyberspace
Computer Science
Louisiana Tech. University
(318) 257-2298
phoha@coes.latech.edu

ABSTRACT

In this paper, we introduce a new model for global integration of disparate data about thematic events into an information context. Such a model has a context that controls the processing and derivation of knowledge from a context. The context of the information also can control the dissemination the information and the knowledge derived from it. A semantic based processing grammar is then developed that defines how processing of contextual may be contextually derived. This architectures processing model is generalized such that the specific processing actions for a given system can be mapped onto the grammar by an entity using this model as its core processing paradigm. Finally, because contextual driven processing is meant to operate at a global information sharing level, we examine and propose a novel method for determination of the level of security a context may require as is disseminated across the internet. Again this model is open architected such that the level of security is suggested but the specifics of application are tailored to the needs of an entitiy.

Keywords

global contextual processing, contextual processing security, security brane.

1. INTRODUCTION

As the second millennium has been dawning there has been a remarkable shift in the computing paradigm away from the concepts of hardware processing data in a structured monotonic fashion. This evolution has become increasingly spurned on by some of the most spectacular natural and manmade disasters our civilization has ever seen. Among such disasters there is 911, tsunamis in Asia, volcanic eruptions and catastrophic nuclear accidents. In all of the these events there has been a structure of information distribution, usage and processing that has not kept up with the needs for information content and a context to determine how it will be processed. Some instances of this are information not being shared among countries and entities, uncorrelated inferences of meaning and criticalities of information processing in a fashion that truly serves various perspectives needs. Context driven processing is driven by the environment and semantics of meaning describing an event. Often this type of processing requires a context which may contain meta data about the events data. Such meta data usually has a spatial and temporal component to it but is actually much more complicated. The key is that contextual meta data describes the environment that the event occurred in and thus can drive how information is stored, processed and disseminated.

The concept of context has existed in computer science for many years especially in the area of artificial intelligence. The goal of research in this area has been to link the environment a machine exists in to how the machine may process information. An example typically given is that a cell phone will sense that its owner is in a meeting and send incoming calls to voicemail as a result. Application of this idea has been applied to robotics and to business process management [1].

Some preliminary work has been done in the mid 90's. Schilit was one of the first researchers to coin the term context-awareness [2,3]. Dey extended the notion of a context with that of the idea that information could be used to characterize a situation and thus could be responded to [4]. In the recent past more powerful models of contextual processing have been developed in which users are more involved [5]. Most current and previous research has still largely been focused on development of models for sensing devices [6] and not contexts for information processing.

Little work has been done on the application of contexts to that of how information is processed. The model that we have developed is that of creating a model for describing information events, storage of meta data and processing rules, thus giving them a context. This context then can be used to control the processing and dissemination of such information in a hyper distributed global fashion. The next section will provide a general overview of the newly developed model and how contexts are defined. Section three will present a semantic grammar that can be utilized to process contexts. Section four will present an open architected model for semantic processing and section five will propose a novel method for determination of the level of security a context needs as it is disseminated.

2. CONTEXTUAL PROCESSING

2.1 Overview

To understand the issues connected with contexts we introduce some details about the newly developing model for contextual processing.

The initial development of a context was to examine the natural disasters of the Indian Ocean tsunami, three mile island nuclear plant and 9/11 to determine what elements could be used to categorize these events. After analysis it was realized that all of them had the following categories, which refer to as the dimensions of a context. They are:

time – the span of time and characterization of time for an event

space – the spatial dimension

impact – the relative degree of the effect of the event on surrounding events

similarity – the amount by which events could be classified as being related or not related.

Each one of the dimensions can be attributed which can be used to derive the semantic processing rules presented later. These dimensions were discovered to be critical in the derivation of knowledge about an event because they affected the process of reasoning about an event. For instance, the time space dimensions can be utilized to reason that a tsunami in the middle of a large ocean may not have the *impact* or *similarity* to that of one just off the coast of Thailand and therefore the processing and dissemination of that information will be different. The reasoning is based in this case on the context defined by the dimensions.

The time and space dimension context driven processing will have the factors of geospatial and temporal elements to them. The geospatial domain can mean that information is collected and stored at a distance from where it may be processed and used in decision support as well as a description of the region that a context may pertain to. This means that context based information processing (CBIP) processing must have a comprehensive model to route information based on semantic content to the appropriate processing location and dissemination channels. CBIP processing can and often does have a temporal component. It can be collected over periods at regular or irregular intervals (the attribution of the dimension) and the time that the information is collected also may determine where the information is sent and the context of how the information is processed. For instance information that is collected as simply monitoring information may in the case of the Tsunami flow to research institutions around the world for storage and analysis at some point in the future. Whereas, noticing earthquakes on the ocean floor may route collected information to countries surrounding an ocean for immediate high speed analysis, critical real time decision making and rapid dissemination. Some factors that should be considered in CBIP processing are referred to as information criticality factors (ICF). These factors are further developed in ongoing research but are primarily used to drive processing decision making. They may include such attribution among other attributes as:

- time period of information collection
- criticality of importance,
- impact e.g. financial data and cost to humans
- ancillary damage
- spatial extent
- spatial proximity to population centers

These factors and many others in the model could be used to evaluate threat, damage, and criticality of operational analysis. Other factors affecting CBI processing might be based on the *quality of the data* such as:

- currency, how recently was the data collected, is the data stale and smells bad
- ambiguity, when things are not clear cut – e.g. does a degree rise in water temperature really mean global warming
- contradiction, what does it really mean when conflicting information comes in different sources
- truth, how do we know this is really the truth and not an aberration
- confidence that we have the truth

In order to analyze the above factor and their effect on CBIP, it was useful to examine three different natural and manmade disasters most people are familiar with in which a concept shift on how information is processed could have remediated the situation if not all together avoided it. We initially considered the 9/11 incident where information about the attackers and their operations and activities were stored everywhere from Germany, to Afghanistan to Florida. If the information could have been orchestrated into a contextual collection of data, the context and relationships of the data would have given a very different interpretation or knowledge about what was really going on. Of course the goal of our model does not examine how that information would be located and integrated, that can be the subject of future work. The model only proposes a paradigm for data organization, processing semantics and security. For the initial analysis of 9/11 we came up with the following descriptive factors which eventually lead to the derivation of the context of contextual dimension presented earlier. These were:

temporality – defined to be the time period that the event unfolded over from initiation to conclusion

damage – the relative damage of the event both in terms of casualties, and monetary loss

spatial impact – defined to be the spatial extent, regionally that the event occurs over.

policy impact – directly driving the development of IA (security) policy both within a country and among countries. This directly led to the evolution of security policy driving implementation because of the event.

2.2 Defining a Context

Contextual processing is based on the idea that information can be collected about natural or abstract events and that meta information about the event can then be used to control how the information is processed and disseminated. In its simplest form, a context is composed of a feature vector

$$F_n\langle a_1,..a_n\rangle$$

where the attributes of the vector can be of any data type describing the event. This means that the vector can be composed of images, audio, alpha-numeric etc. Feature vectors can be aggregated via similarity analysis methods into super contexts $S_{c.}$ The methods that might be applied for similarity reasoning can be statistical, probabilistic (e.g. Baysian), possibilistic (e.g fuzzy sets) or machine learning and data mining based (e.g. decision trees). Aggregation into super sets is done to mitigate collection of missing or imperfect information and to minimize computational overhead when processing contexts.

definition: A context is a collection of attributes aggregated into a feature vector describing a natural or abstract event.

A super context can be described as a triple denoted by:

$$S_n = (C_n, R_n, S_n)$$

where C is the context data of multiple feature vectors, R is the meta-data processing rules derived from the event and contexts data and S is controls security processing. S is defined to be a feature vector in this model that holds information about security levels elements or including overall security level requirements.

definition: A super context is a collection of contextual data with a feature vector describing the processing of the super context and a security vector that contains security level and other types of security information.

The cardinality of F with C is:

m:1

which when substituted into S creates a (C, R, S) cardinality of:

m:1:1

for the proposed model. However, we have not examined the impact, constraints of implications of having an

m:n:o

type of cardinality.

All of the above are a *type* of feature vectors where the elements of the vector can contain any type of information including the derived contextual processing rules and security methods for the given super context.

3. SUPER CONTEXTS AND TIME

3.1 Overview

A super context is composed of context data from many sensing event objects, Eo_i, as shown in figure 1. As such contextual information collection works in a similar fashion to sensor networks and can borrow from theory in the field. Figure 1 shows the nature of collection of event object data over time. One can visualize a region of interest, e.g. the Indian Ocean tsunami for which event object data is collected which is centered over a thematic event object, e.g. the origin of the tsunami.

definition: A thematic event object (Teo) is the topic of interest for which event objects are collecting data. An example of a Teo would be the center of a tsunami.

As time passes (moving to the right) in figure 1, event object data collection can be visualized as extruding the region of interest to the right and that event objects operate sporadically in collection of context information. This is the core idea for construction of super contexts.

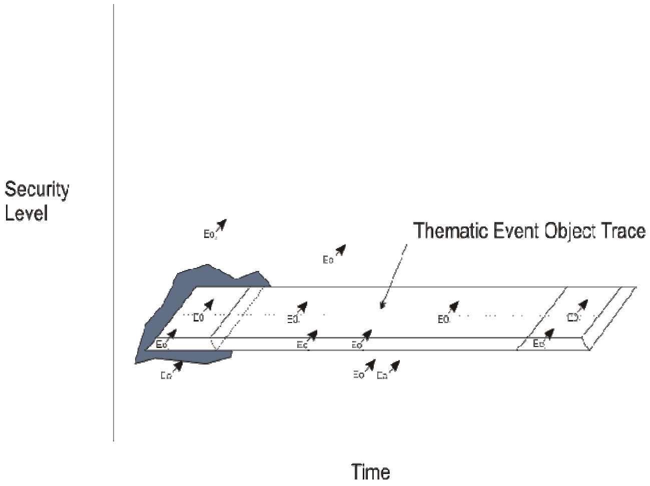

Figure 1. Visualization of the sporadic nature of streaming collection of context data, from event objects as a function of time (moving to the right) for a region of interest.

Figure 1 visualizes several important concepts behind contexts. First, event objects may collect and send data at sporadic or regular intervals. Secondly objects may have relationships with other objects e.g. falling under a mathematical surface, that may be used to determine security levels for information being collected. In figure 1, the security level axis is shown to start introducing the concept of security by use of branes discussed in section 5. The raised pyramid feature in the figure is a type of brane that can be useful in security (discussed later and shown in figure 2). Of note is that some event objects for reasons including technical limitations may stream intermittently or fail to stream. This is visualized by the event objects on the same horizontal lines bursting data at various points in time. This a core concept that directly maps contexts onto the defining dimension of time and space. As such the contextual based model has to address in its development the ideas visualized in figure 1. The security level access is discussed in the final section of the paper.

4. SEMANTIC AND SYNTACTICAL MODELS FOR CONTEXTUAL PROCESSING

4.1 General Operation and Concept

In addition to the complexities of data types that can comprise a context, the classifications and categorizations, data can often have a geospatial and temporal aspect to it. This is due to the fact that data often represents complex events. Events can have multiple *meta-characterizations* that can drive the derivations of a semantic model that can be utilized for contextual processing. some of the these so far defined are:

- Singular – an event that happens a point in time, at a singular location
- Regional
- Multipoint Regional
- Multipoint Singular – events that occur at a single point in time but with multiple geographic locations
- Episodic – events the occurs in bursts for given fixed or unfixed lengths of time
- Regular – as suggested these events occur at regular intervals
- Irregular – the time period on these type of events is never the same as previous t
- Slow Duration - a series of event(s) that occupy a long duration, for example the eruption of a volcanoes
- Short Duration – example an earthquake
- Undetermined
- Fixed Length
- Unfixed Length
- Bounded
- Unbounded
- Repetitive - these types time events generate streams of data – graph of attributes change in value over time

These meta-characterization can be applied to data for the previously discussed original 9/11 analysis (temporality, damage, spatial impact, policy impact) and thus to the final set of dimension that were derived for characterization of contextual processing.

4.1.1 Semantic Processing Syntax

The above meta-characterization of context data in our model was developed into a semantic syntax that determines how the processing rules associated with a super context are applied to determine processing and security. These can drive the context processing engine during the process of the transformation of thematic data to dissemination and knowledge. The semantic model contains the following operational elements:

Event Class < abstract, natural>

Event Type < spatial, temporal>

Periodicity < regular, irregular>

Period < slow, short, medium, long, undeterminable, infinite, zero >

Affection<regional, point, global, poly nucleated, n point>

Activity < irregular, repetitive, episodic, continuous, cyclic, acyclic>

Immediacy < catastrophic, minimal, urgent, undetermined >

Spatiality < point, bounded, unbounded >

Dimensionality <1, 2, 3, n>

Bounding < Fixed Interval, Bounded, Unbounded, Backward Limited, Forward Limited, continuous>

Directionality < linear, point, polygonal >

Figure 2: Modeling the semantic categories of context based meta-characterizations of data in a context

The semantic categories were then developed into a syntax of production rules that can populate the R vector in a super context and thus be used to control the contextual processing of a super context. The syntax takes the form of the following:

R1: <event class>, <event type>, <R2>

R2: (<periodicity> <period>) <R3>

R3:(<affection><activity>) <spatiality> <directionality> <bounding> <R4>

R4: <dimensionality> <immediacy>

Figure 3: Syntax for application data meta-characterizations to derive super context processing rules R in S(C, R, S).

The above semantic grammar and syntax can be developed into sentences that affect the processing of contextual information. For instance, a tsunami in the Indian ocean might generate a processing context sentence such as:

"R1 = natural, spatial-temporal, irregular-slow, regional episodic catastrophic unbounded 3D linear"

The application of this model and its production rules can then be mapped to resource actions with such things as priorities on use, computation, notification, etc. It is envisioned that the entities using such a model (e.g. government) would determine how to map sentence to actions specific to their mission. For instance the above production R1 may produce the following mapped response rules affecting the processing and collection of contextual information:

natural spatial temporal => {notify and activate associated computational resources}

irregular-slow – tsunami => {notify hotels, activate alert system sensors, satellite tracking data computers activated, high priority collection mode}

catastrophic => {context processing elevated to kernel mode}

unbounded => {distributed notification and transfer of context processing to surrounding countries }

There can be are many other of meta-characterization classifiers of event that could be developed for the contextual processing model that could control mapping of contextual data into knowledge and actions.. The goal of this initial work is to provide a model that is open architected enough to accommodate further growth and refinement. New classifiers should be developed to fit in a member of the semantic model proposed above. Even more importantly the selection of the transformation method to knowledge will be predicated on the characteristics of the events context..

5. SECURITY ON CONTEXTS

5.1 Overview

The final component of the context model is that of security. This is the S in the (C,R,S) that defines a super context. It was realized early that definition of security for information flowing globally around the planet on the internet would be a tough challenge. It was also realized that a lot of good security techniques already existed that could be applied to contextual data e.g. encryption. However, because contextual information is streaming as shown in Figure 1, encryption can be computationally intractable. What was needed was a method to determine that security measures were good enough, which we call *pretty good security*. This method had to be architected so that it suggested to users the level of security required and allowed the users to define how they accomplished that level e.g. what security methods they wanted to apply. The concept of using a mathematically defined surface, a brane was developed for the model.

A brane is a term borrowed from Cosmology. It can have multiple mathematical dimensions (1,2,3, …n) and can be thought of a mathematically described boundary between n number of spaces.

definition: A brane can be is a three dimensional surface that is overlaid above a two dimensional object. Finding the intersection of the projection of event objects on the 2D surface with the brane can provide a value that can be utilized to calculate security level for the context of a given event object.

Branes have been applied to modeling universes. In considering the application of branes to contexts it was realized that a brane of order three (three dimensional) could be utilized to determine the levels of security required for a given contexts data. This could be done by superimposing a brane over a thematic region of interest and then projecting event objects onto its surface. Calculation of the intersection point then becomes a security level that reflects the geo-spatial relationship of any given event object to the thematic object over which the brane is geo-referenced.

For instance, in figure 2, the event object Eo_i inside the brane (on the x axis) projects onto the brane producing a security level of about .5. The thematic object, T_{eo} projects onto the brane with security level 1.

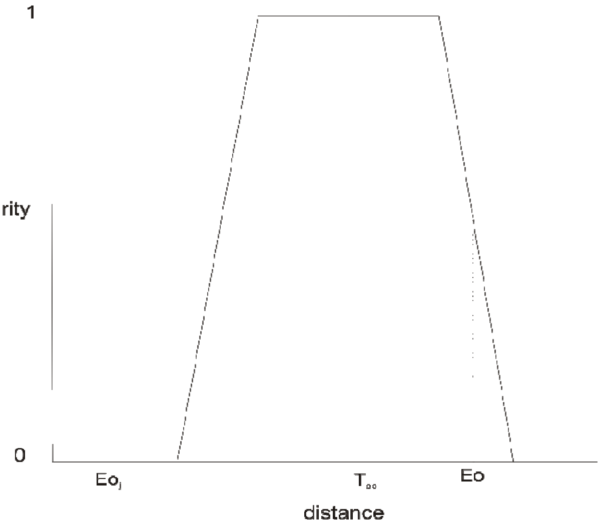

Figure 2. Projection of event objects onto a branes surface to calculate its security level.

Objects outside of the brane receive a security level of 0 because the project of the object does not intersect the branes surface. Such an object in this simple model might be of a different thematic type or too far away from the T_{eo} to be of interest thus not requiring a security level. A security level of 1 suggests to the consumer of a super context that it will need maximum security methods applied as defined by the consumer. For instance the T_{eo} need full security e.g. full encryption or other measures, whereas event objects with security level 0 can be ignored or sent in clear text.

6. CONCLUSIONS

6.1 Future Work

The modeling of contextual processing and is a broad new area of computer science and can be the beginning of many new research threads. This paper provides a simple introduction to the subject and an overview of architecturally how the model for contextual processing is thought to operate. Any particular component of the model presented in this paper can lead to many unresolved research questions that need further definition and empirical validation. As an example, research could be done on i) how security levels from branes correlate with application of the R processing rules in $S_n = (C_n, R_n, S_n)$, ii) development of “spot security” based on security level to limit computational overhead and iii) the integration of contextual similarity into the brane models, what the semantics might mean and how they can be related to streaming contexts with high computational overhead for security processing. Because the concept of contextual processing is so broad it offers the possibility to draw many disciplines of computer science more tightly together.

7. REFERENCES

1. Rosemann, M., & Recker, J. (2006). "Context-aware process design: Exploring the extrinsic drivers for process flexibility". T. Latour & M. Petit *18th international conference on advanced information systems* engineering. proceedings of workshops and doctoral consortium*: 149-158, Luxembourg: Namur University Press.*
2. Schilit, B.N. Adams, and R. Want. (1994). "Context-aware computing applications" (PDF). *IEEE Workshop on Mobile Computing Systems and Applications (WMCSA'94), Santa Cruz, CA, US*: 89-101.
3. Schilit, B.N. and Theimer, M.M. (1994). "Disseminating Active Map Information to Mobile Hosts". *IEEE Network* **8** (5): 22–32. doi:10.1109/65.313011.
4. Dey,Anind K. (2001). "Understanding and Using Context".*Personal Ubiquitous Computing* **5** (1): 4–7. doi:10.1007/s007790170019.
5. Cristiana Bolchini and Carlo A. Curino and Elisa Quintarelli and Fabio A. Schreiber and Letizia Tanca (2007). "A data-oriented survey of context models" (PDF). *SIGMOD Rec.* (ACM) **36** (4): 19--26. doi:10.1145/1361348.1361353. ISSN 0163-5808. http://carlo.curino.us/documents/curino-context2007-survey.pdf.
6. Schmidt, A.; Aidoo, K.A.; Takaluoma, A.; Tuomela, U.; Van Laerhoven, K; Van de Velde W. (1999). "Advanced Interaction in Context" (PDF). *1th International Symposium on Handheld and Ubiquitous Computing (HUC99), Springer LNCS, Vol. 1707*: 89-101.