Contextual Binding and Deception Detection

Jim Q. Chen, Ph.D.

National Defense University, U.S.A.

Abstract

Deception is frequently used in cyber attacks. Detecting deception is always a challenge, as witnessed in attacks in social media and other online environments. Contexts can help to identify deception. Unfortunately, there is not much literature available in this aspect. This paper explores the unique properties of contextual binding. It examines roles that it plays. It also proposes a novel approach in detecting deception utilizing contextual binding in the cyber domain.

Introduction

A context is defined in Webster dictionary as "the interrelated conditions in which something exists or occurs" or as "the parts of a discourse that surround a word or passage and can throw light on its meaning". (http://www.merriam-webster.com/dictionary/context) It is also defined in Brézillon (1999) and Brézillon (2002) as "a collection of relevant conditions and surrounding influences that make a situation unique and comprehensible". In this sense, a context helps to disambiguate meaning and find out the actual referent. Hence, it is essential in data mining and big data analytics. There are various approaches in context analysis. Brézillon (2003) uses contextual graphs to address the dynamic of context. Gaifman (2008) uses syntactically represented context operators in the analysis of contextuality. Grossi, Dignum, and Meyer (2006) propose a notion of contextual terminology to "reason within contexts (intra-contextual reasoning)" and to "reason also about contexts and their interplay (inter-contextual reasoning)". Rebuschi and Lihoreau (2009) address "the connections between knowledge and context" with the contextual epistemic logic. McCarthy (1993) discusses "formalizing contexts as first class objects". From Linguistic perspective, Key (1989) mentions contextual operators, which, in his term, "are lexical items or grammatical constructions whose semantic value consists, at least in part, of instructions to find in, or impute to, the context a certain kind of information structure and to locate the information presented by the sentence within that information structure in a specified way". All these researches scrutinize contexts from varied perspectives. They all show the significance of contexts in nailing down meaning or interpretation. In the same spirit, this paper proposes the Contextual Binding Conditions and the Detection Condition utilizing contextual operators, on the basis of linguistic samples. These conditions are then applied to both language disambiguation and deception detection in the cyber domain. The success of the application confirms the validity of these conditions.

Challenges

As discussed previously, contexts are used in linguistic analysis, in building intelligent agents, and in other artificial intelligence fields. However, no formal methods that employ contexts or contextual operators have ever been used in detecting deception in the cyber domain. This paper intends to explore this possibility.

Before moving on, let us be aware of two types of deception.

Caddell (2004) defines two types of deception. They are fabrication and manipulation. He states, "If false information is created and presented as true, this is fabrication." "Manipulation, on the other hand, is the use of information which is technically true, but is being presented out of context in order to create a false implication." Almeshekah and Spafford (2014) also state: "Deception always involves two basic steps, hiding the real and showing the false".

How can deception, specifically fabrication and manipulation, be detected in the cyber domain? This is one of the challenges we are facing. Even though there is not much literature available in this aspect, researches in the similar fields may be looked at. Burgoon, Blair, Qin, and Nunamaker (2003) use decision trees in detecting deception within online chat messages. They utilize "16 linguistic features that can be automated to return assessments of the likely truthful or deceptiveness of a piece of text".

They find out that “deceivers do utilize language differently than truth tellers”. However, it has to be pointed out the 16 linguistic features are pre-defined so that deception employing features other than these 16 features may not be detected.

Zhou, Shi, and Zhang (2008) develop the Statistical Language Models (SLMs), which consider “all of the words in a text as potential features without relying on the extraction of a predefined set of cues to deception”. Word dependencies are learned to “capture semantic relationships and dependency relationships among words so as to approximate the meaning of sentences, which can benefit deception detection”. This method is better than the method used in Burgoon, Blair, Qin, and Nunamaker (2003). However, it is relatively time-consuming, as “SLMs consider all possible n-grams as features and implicitly represent the importance of those features according to their contribution to the quality of language modelling”. In addition, this approach does not address the detection of fabrication or manipulation as it is not designed for that purpose.

To address all these challenges and to figure out a holistic and dynamic solution, Chen and Duvall (2014) propose the Operational-Level Cybersecurity Strategy Formation Framework, which consists of a Contextual Analysis Component among other components in making strategic decisions. This paper further explores the inner-workings of the Contextual Analysis Component.

Proposal

A novel approach is proposed here in this section, where the contextual binding relationship is explored and then used in detecting abnormal behavior. Within a contextual binding relationship, a contextual binding operator plays a crucial role. It helps to set up the baseline in a context, in which the Detection Condition can be applied.

The Contextual Binding Conditions

A contextual binding operator is deterministic in disambiguation. Let us take a look at a linguistic example demonstrated in the English sentence in (1) below.

(1) John likes to buy a book$_i$ and read it$_i$ within three days.

Any English speaker knows that the following interpretation is acceptable: “John likes to buy a book and read the book that he buys within three days.” Any English speaker also knows that the interpretation below is not acceptable: “John likes to buy a book and read the map within three days.”

The pronoun “it” in this sentence refers back to the noun phrase “the book”, which precedes the pronoun in the same sentence. To a certain extent, pieces of information provided previously can serve as contextual components for entities in the same sentence or following sentences. In this particular case, the agent of the action is the noun phrase “John” in the subject position; the patient of the action is the noun phrase “a book” in the object position; and the predicate, i.e. the action, is “purchasing and reading x, which is a book in this case”. Now, the noun phrase “the book” plays the role of a contextual operator, which becomes available for the interpretation of the pronoun “it”, as it satisfies the condition of being a singular non-human entity, just like the pronoun “it”, in this particular context. Based on the observation of the contextual relationship, it may be claimed that the pronoun “it” is contextually bound by the contextual operator of “what”, namely, “a book” in this particular case. In this contextual binding case, if the pronoun “it” refers to another entity, such as “a map”, rather than the entity “a book”, the interpretation is immediately recognized as being abnormal or regarded as being unacceptable. This clearly reveals a contextual binding relationship, which can be defined as follows:

The Basic Contextual Binding Condition:

Assume X is an entity, and CO is a contextual operator.

(i)If X is directly related to CO in such a setting:

$CO_i \{\ldots\ldots X_i \ldots\ldots\}$

then X_i is *contextually bound* by CO_i.

The entity X_i is *directly related* to CO_i iff CO_i provides a context that the interpretation of X_i solely depends on.

Now, the contextual relationship in (1) can be captured in the following schematic configuration:

$CO_{book} \{\ldots\ldots X_{book} \ldots\ldots\}$

Applying the Basic Contextual Binding Condition, the pronoun “it” is contextually bound by the contextual operator “book”. This means that the pronoun “it” has to be interpreted as “the book” if the Basic Contextual Binding Condition is obeyed.

Given COs = {agent, patient, activity, time, location, environment, background, precedence, etc.}, the Restrictive Contextual Binding Condition can also be defined.

Before we define this condition, let us see how McCarthy (1993) handles the time component. In discussing the relations among contexts, McCarthy (1993) examines the specialize-time (t, c), which he refers to as “a context related to c in which the time is specialized to have the value t”. The axiom that he comes up with is as follows:

C0: specialize-time (t, c1, c2) $\wedge$ ist (p, c1) $\supset$ ist (c2, at-time (t, p))

This axiom refers to two assertions. The first one is in Context1, i.e. c1, the proposition p is true in the context of c1. The second one is in Context2, i.e. c2, which is a subset of the set c1, and which has the specialize-time t, the proposition is true at time t. The second assertion is a subset of the first assertion.

From this perspective, the time component further narrows down the interpretation of the entity with the time aspect.

Let us take a look at another linguistic example demonstrated in the English sentence in (2) below.

(2) Yesterday John bought a $book_i$ at the bookstore. He enjoyed reading it_i.

Here, the pronoun "it" in the second sentence refers back to the noun phrase "a book", sitting in the object position of the first sentence and serving as the patient of the action. The temporal adverbial phrase "yesterday" refers to the time component of the context. The locality adverbial phrase "at the bookstore" refers to the locality component of the context. Comparing the noun phrase "a book" in (1) with the noun phrase "a book" in (2), one may notice that the former refers to a general book while the latter refers to a specific book, i.e. the book bought yesterday at the bookstore. In this sense, the latter in (2) may be considered as a subset of the former in (1).

Following McCarthy (1993), the assertion within a specialized time, location, environment, and/or background is treated as a subset of a general assertion.

This can be captured schematically as follows in defining the Restrictive Contextual Binding Condition.

The Restrictive Contextual Binding Condition:

Assume X is an entity, and CO is a contextual operator.

(ii) In a specialized time, location, environment, background, if X[Y,Z] is directly related to CO[Time, Locality] in such a setting:

$CO_i[Time_{j,}\ Locality_k]$ {……$X_i[Y_{j,}\ Z_k]$……}

then $X_i[Y_{j,}\ Z_k]$ is *contextually bound* by $CO_i[Time_{j,}\ Locality_k]$.

The entity $X_i[Y_{j,}\ Z_k]$ is *directly related* to $CO_i[Time_{j,}\ Locality_k]$ iff (if and only if) $CO_i[Time_{j,}\ Locality_k]$ provides a context that the interpretation of $X_i[Y_{j,}\ Z_k]$ solely depends on.

Obviously, $CO_i[Time_{j,}\ Locality_k]$ is more restrictive than CO_i. In this sense, the Restrictive Contextual Binding Condition is a subset of the Basic Contextual Binding Condition.

Now, the contextual relationship in (2) can be captured in the following schematic configuration:

$CO_{book}[Time_{yesterday,}\ Locality_{atbookstore}]$
{……$X_{book}[Y_{yesterday,}\ Z_{atbookstore}]$……}

Applying the Restrictive Contextual Binding Condition, the pronoun "it" is contextually bound by the contextual operator "$CO_{book}[Time_{yesterday,}\ Locality_{atbookstore}]$". This means that the pronoun "it" has to be interpreted as "the book bought yesterday at the bookstore" if the Restrictive Contextual Binding Condition is obeyed.

The Detection Condition

The Detection Condition can be derived from the above two conditions:

(iii) If X is in such a contextual binding configuration:

CO_i {……X_m……}

where X_m is supposed to be contextually bound by CO_i but not so, then an abnormal circumstance is detected.

Likewise,

(iv) If X is in such a contextual binding configuration:

$CO_i[Time_{j,}\ Locality_k]$ {……$X_m[Y_{j,}\ Z_k]$……}

where $X_m[Y_{j,}\ Z_k]$ is supposed to be contextually bound by $CO_i[Time_{j,}\ Locality_k]$ but not so, then an abnormal circumstance is detected.

As shown above, a contextual operator helps to form the contextual binding relationship and to resolve ambiguity. A deception can be detected if the entity is supposed to be contextually bound by a contextual operator but it is not so in a configuration.

Let us apply these conditions to the case in (1).

In (1), there is such a configuration:

CO_{what} {……X_{what}……}

This can be rewritten as follows:

CO_{book} {……X_{book}……}

Here, the pronoun "it" possesses the property "X_{book}", which is *contextually bound* by CO_{book}. Hence, this interpretation is valid and acceptable.

However, if the pronoun "it" in (1) refers to another entity, say "the map", rather the entity "the book" that is mentioned in the first sentence, this contextual binding relationship immediately ceases to exist. Below is the configuration:

CO_{book} {……X_{map}……}

Here, the pronoun "it", which possesses the property "X_{map}", is supposed to be contextually bound by the contextual operator "CO_{map}" but not be contextually bound by the contextual operator "CO_{book}". However, the contextual operator "CO_{map}" is not available. Hence, such a configuration triggers the Detection Condition. The interpretation is thus regarded as being invalid and unacceptable.

Let us look at another linguistic example demonstrated in the English sentence in (3) below.

(3) * John likes to buy a $book_i$ and read $them_i$ within three days.

Any speaker of English knows that this sentence is awkward in the context where the pronoun "them" refers back to the noun phrase "a book", as there is a mismatch between the third-person singular form and the third-person plural form.

The contextual relationship can be captured in the following schematic configuration:

CO_{book} {……X_{books}……}

Here, X_{books} is supposed to be contextually bound by CO_{book} but not so. Hence, an abnormal circumstance is detected.

Let us have a look at still another linguistic example demonstrated in the English sentences in (4) below.

(4) * John likes to buy a cookbook$_i$ and cook it$_i$ following the instruction.

Any speaker of English knows that it is awkward to have the pronoun “it” in this context to refer back to the noun phrase “a cookbook”, because the noun phrase “a cookbook” possesses the features: [+Object, -edible] while the pronoun “it” possesses the features: [+Object, +edible] in the sub-context of “cooking”. This mismatch in features indicates that the noun phrase “a cookbook” and the pronoun “it” refer to different entities. In other words, the pronoun “it” is not contextually bound by the contextual operator “a cookbook” in this particular case.

The contextual relationship can be captured in the following schematic configuration:

$CO_{cookbook}$ {……$X_{ediblething}$……}

As the pronoun “it” is not contextually bound by $CO_{cookbook}$ in its contextual domain, another abnormal circumstance is detected.

Assuming whatever is within the contextual operator is normal, the variable is normal if and only if it is contextual bound by its corresponding contextual operator. As shown above, in order to be properly bound in its contextual domain, the variable has to possess the same features or properties as those of the contextual operator. Any deviation triggers the Detection Condition.

Deception Detection

In this section, the Contextual Binding Conditions and the Detection Condition are applied in the detection of deception. Assume that what an application or an executable is expected to do on the basis of its functional requirement is included in the feature set of the contextual operator. As a result, this sets up the baseline for the application or the executable. The actual execution of the application or the executable is a variable, which should be bound by the contextual operator. If the actual execution involves more features than or different features from what is contained in the contextual operator, the deviation from being normal is identified, the Detection Condition is triggered, and a deception is detected.

Let us examine fabrication first. A piece of malware is a good example of fabrication. For instance, appended to the executable “notepad.exe” is a piece of code that makes possible for the executable to perform file transfer in addition to its original functionality of text file editing. This abnormal behavior can be easily detected with the help of the Basic Contextual Binding Condition and the Detection Condition.

Assume what is expected for the original functionality of the executable is contained inside a contextual operator as a feature set. Assume the actual functionality of the executable is contained within a variable as current features. The variable, by definition, should be contextually bound by the contextual operator. Schematically, this relationship is represented below:

$CO_{TextEditing}$ {……$X_{TextEditing}$……}

This represents a normal situation, in which an executable is doing what it is expected to do.

When an extra functionality is added into this executable, the contextual relationship gets changed, as shown below:

$CO_{TextEditing}$ {……$X_{TextEditing+FileTransfer}$……}

Here, one of the actual functionalities of the executable, i.e. “FileTransfer”, is not contextually bound by the contextual operator. Thus, an unacceptable behavior is immediately detected at the application level, even before it is executed and at the time when a request for extra resource utilization is made.

This also applies to other pieces of malware, which always make requests for additional resource utilization. If this contextual analysis component is implemented within the kernel of an operating system, anytime when a request for resource utilization is received, if it is not contextually bound by a contextual operator, the request is denied immediately, an investigation is launched, and this activity is logged.

Let us check manipulation now. Stegonography is a good example of manipulation. For instance, one may hide a text file inside a graphic file. After this operation, the file size of the modified graphic file may remain the same as the file size of the original graphic file. At the first glance, nothing seems to have happened. However, using a digital forensic tool, one would see the systematic change of hexadecimal code even though the change for each byte is minor, say the change from “0x52” to “0x51” in one byte and the change from “0x73” to “0x72” in another byte that is 3 bytes after the previously changed byte. This becomes obvious when one compares the code for the original graphic file with the code for the modified graphic file. In addition, the original file timestamp pattern, consisting of the date created time, the date accessed time, the date modified time, and the date last saved time, is changed. Evidently, the Restrictive Contextual Binding Condition is violated. Hence, the abnormal behavior in this type of cases can also be detected.

Assume both the expected code pattern and the expected timestamp pattern are contained within the feature set of the contextual operator. Assume the actual code pattern and the actual timestamp pattern are contained as current features in the variable. Also assume that the timestamps are used to further restrict the actual code pattern, as illustrated in the Restrictive Contextual Binding Condition. By definition, the variable should be contextually bound by the contextual operator. Schematically, this relationship is represented below:

$$CO_{Pattern1}[Time_{Pattern2}]\{\ldots\ldots X_{Pattern1}[Y_{Pattern2}]\ldots\ldots\}$$

This represents a normal situation, in which an expected pattern is obtained.

When a graphic file is altered to accommodate a hidden text file, the actual code pattern gets changed. Now, the contextual relationship also gets changed, as shown below:

$$CO_{Pattern1}[Time_{Pattern2}]\{\ldots\ldots X_{Pattern6}[Y_{Pattern7}]\ldots\ldots\}$$

Here, the actual representation of the file is not contextually bound by the contextual operator, because the actual code pattern represented by "$X_{Pattern6}$" is different from the expected pattern "$X_{Pattern1}$" contained in the contextual operator and the actual timestamp pattern represented by "$X_{Pattern7}$" is different from the expected pattern "$X_{Pattern2}$" contained in the contextual operator. Hence, the unacceptable behavior is detected at the code level.

As shown above, the Contextual Binding Conditions and the Detection Condition can successfully detect deception such as fabrication and manipulation.

Conclusion

Detecting deception in cyberspace is a challenge. Based on the analysis of the unique property of contextual operators in a natural language, this paper proposes a contextual binding mechanism that can be used to disambiguate interpretation and identify invalid and unacceptable interpretation in a natural language. The same mechanism can also be used to detect deception in the cyber domain, specifically fabrication and manipulation. This mechanism can not only aid the decision-making in cyber conflicts or cyber competitions but also lay the foundation for employing contextual operators in an artificial intelligence system.

References

Almeshekah, M. and Spafford, E. 2014. Planning and Integrating Deception into Security Defenses. The New Security Paradigm Workshop (NSPW 2014), Retrieved from http://www.meshekah.com/wp-content/uploads/2014/10/planning_and_integrating_deception_into_computer_security_defenses_Almeshekah_Spafford.pdf.

Brézillon, P. 1999. Context in problem solving: A survey. *The Knowledge Engineering Review*, 14 (1), pp.1-34.

Brézillon, P. 2002. Modeling and using context: Past, present, and future. Rapport de Recherche du LIP6 2002/010, Université Paris 6, France.

Brézillon, P. 2003. Context Dynamic and Explanation in Contextual Graphs. *Context 2003*. *Lecture Notes in Artificial Intelligence (LNAI) 2680*. Blackburn, P. et al. (Eds). pp.94-106. Berlin: Springer-Verlag.

Burgoon, J., Blair, J., Qin, T., and Nunamaker, J. 2003. Detecting Deception through Linguistic Analysis. *Proceedings of First NSF/NIJ Symposium on Intelligence and Security Informatics*. pp.91-101, Berlin: Springer-Verlag.

Caddel, J. 2004. *Deception 101 - Primer on Deception*. Strategic Studies Institute, U.S. Army War College.

Chen, J. & Duvall, G. 2014. On Operational-Level Cybersecurity Strategy Formation. *Journal of Information Warfare*, 13 (3), pp.79-87.

Gaifman, H. 2008. Contextual Logic with Modalities for Time and Space. *Review of Symbolic Logic*, 1 (4), pp.433-458.

Grossi, D., Dignum, F., and Meyer, J. 2006. Contextual Terminologies. *Computer Logic in Multi-Agent Systems (CLIMA) VI*, *Lecture Notes in Artificial Intelligence (LNAI) 3900*. Toni, F & Torroni, P. (Eds). pp.284-302. Berlin: Springer-Verlag.

Kay, P. 1989. Contextual Operators: Respective, Respectively, and Vice Versa. *Proceedings of the Fifteenth Annual Meeting of the Berkeley Linguistics Society*. pp.181-192.

McCarthy, J. 1993. Notes on Formalizing Context. *Proceedings of the 13th International Joint Conference on Artificial Intelligence – Volume 1*. pp.555-560. San Francisco, CA: Morgan Kaufmann Publishers Inc.

Rebuschi, M. & Lihoreau, F. 2009. Contextual Epistemic Logic. Retrieved from http://www.academia.edu/8052225/Contextual_Epistemic_Logic.

Zhou, L., Shi, Y., and Zhang, D. 2008. A Statistical Language Modeling Approach to Online Deception Detection. *IEEE Transactions on Knowledge and Data Engineering*, 20 (8). pp.1077-1081.