International Journal for Modern Trends in Science and Technology, 6(8S): 178-184, 2020

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTSTCIET35
Available online at: http://www.ijmtst.com/ncracse2020.html

An Efficient Management and Querying of Encrypted Architecture Text Platform by Encrypted Communication using ELK

A Ravi Kishore | B Prashanth Babu | K Ashok Babu

Computer Science and Engineering, Bapatla Engineering College, Bapatla, AP, India

To Cite this Article
A Ravi Kishore, B Prashanth Babu and K Ashok Babu, "An Efficient Management and Querying of Encrypted Architecture Text Platform by Encrypted Communication using ELK", *International Journal for Modern Trends in Science and Technology*, 6(8S): 178-184, 2020.

Article Info
Received on 16-July-2020, Revised on 15-August-2020, Accepted on 25-August-2020, Published on 30-August-2020.

ABSTRACT

Due to its rapid evolution and rapid penetration in the industry, Big Data has become an important and essential instrument for data analysis and decision making. Providing data protection has therefore been a major problem for Large Data platforms. This research proposes a stable communication network by applying a particular scenario consisting of a series of software applications relevant to collecting, converting, and processing vast volumes of knowledge, often known as stack ELK; More specifically, in our plan, the data is sent encrypted from its source in workstations to be processed in Elasticsearch in encryption mode, thereby guaranteeing its secrecy. The findings indicate that the encryption method does not impact the productivity in the method of producing and distributing data packets. The method to protect the message information needs less than 2 milliseconds per data packet which meets the real-time monitoring requirements.

KEYWORDS: *Big Data, Security Information, Vulnerabilities, Threats, Security and Privacy, Management and Querying of Encrypted Data, Key Management, Computer Systems Organization, Availability.*

INTRODUCTION

The details created and collected in the numerous technical scenarios are on the verge of reaching the imaginable limits of a digital environment that is rapidly challenged by the exponential growth in knowledge currently produced[13].Big data enters the scene as a consequence of this uncontrollable development, as a series of technology-based resources for processing structured, unstructured, and semi-structured data produced from numerous knowledge sources[3].Today, knowledge channels reach outside the conventional organised repositories, including resources including email , social networks, server logs, sensor generated data, etc. Unstructured knowledge, missing a specific format [5]. The protection challenges that companies are subjected to are increasingly rising as they face both internal and external attacks, which is why the knowledge deemed a strategic asset in a company needs to be safeguarded. Their disruption or failure could disrupt activities, paralyse infrastructure and trigger significant economic harm[2].

Given all the benefits of utilising Big Data due to increased efficiency and profitability, security concerns are a major problem for this technology, particularly as an enterprise collects and retains vast volumes of information. Such details will become the object of attackers and offenders. Nowadays, one of the most powerful methods for handling Big Data is Elasticsearch, which along

with Logstash and Kibana make up the ELK framework, an open-source network that enables data extraction practically in real time of knowledge obtained from a single source[8].
Elasticsearch was developed for the key goal of processing vast quantities of knowledge in a distributed computing system. Since it is an open-source tool with free access to the public, security was not a concern for its creators, which translates into security gaps in the information it stores. This document examines the operation of the ELK stack, focusing in particular on the transmission of information from workstations to Elasticsearch, and analyses the level of security provided by this platform.in the management of big data.

RELATED WORK

Generally, data is put away in databases to process and deal with its relations; a few data are named a profoundly significant data that should be exceptionally made sure about or on a degree of security, the most ideal approach to make sure about such data is to encode it. Numerous encryption calculations were examined and numerous structures of databases have arranged to put the contemplations of encryption and security of the databases.In [1] the significant difficulties and plan contemplations about database encryption were portrayed. The article first presents an assault model and the principle significant difficulties of data security, encryption overhead, key management, and joining impression. Next, the article audits related scholarly work on elective encryption designs; ordering encrypted data; and key management. At long last, the article finishes up with a benchmark utilizing the accompanying structure rules: encryption setup, encryption granularity, and keys stockpiling. First light Xiaodong Song [2] proposes another encryption technique that permits looking encrypted data without decoding. Nonetheless, the strategy isn't adjusted for database encryption. HankenHacijumus [3] proposes a way that has a shortcoming; it will yield bogus joining records, which prompts the enormously expanded expense of decoding records and debased execution of the inquiry. They propose an outline of executing SQL over encrypted data in the database-specialist co-op model. At that point in [4], the journalists proposed another question technique, wherein the inquiry is finished on the server-side and the customer side together, they have proposed pail file, which bolsters the range question for the numeric data. At that point they include a strategy that bolsters math calculation [5].

In [6] Hore enhanced the can record technique on the best way to parcel the can to get the exchange between the security and question execution. The strategies dependent on the list is upheld by DBMS (DataBase Management System) and concentrated on the inquiry execution at the expense of extra room. There are additionally a few investigates on the fluffy inquiry of the character string. Zhengfei Wang proposed a capacity to help a fluffy inquiry over the encrypted character data [7] [8]. Their technique named blending coding strategy encodes each adjoining two characters in arrangement and changed over the first string legitimately to another trademark string by a hash work. This strategy can't manage a few characters and could perform gravely for a big character string.

Paper [9] had proposed attributes lattice to communicate string and the network will likewise be packed into a paired string as list. Each character string needs a network size of 259x256, it is enormous and will prompt a lot of calculation; likewise, the length of the record has come to in excess of a hundred bits, which isn't reasonable for capacity in the database.In [10] the paper chips away at a gathering of clients that needs to get to make sure about data on a server. The common delicate information requires greater security and privacy insurance, In that paper, two plans were proposed which can look through the encrypted archives without re-encoding all reports in a server regardless of whether gathering keys must be refreshed. The plans can bolster general database standardization for the encrypted database. Their investigations show that their plans are significantly more productive than the comparables ones.

Paper [11] just scrambles the touchy field and it is likewise utilizing pail list to improve inquiry execution. The request on numeric data is valuable. In any case, on the character data, it has little impact. So the technique in [11] isn't good for the character data. [12] Creates a B+ tree list for the data before encoding them. While querying the encrypted data, right off the bat, it finds the encrypted records identified with the querying predicate dependent on the B+ tree list; furthermore, it unscrambles the encrypted records to achieve the outcomes. Likewise, it must encode the B+ tree itself to shield it from releasing secret

information. As per the structure of the B+ tree, it encodes every hub of the B+ tree independently. The consequences of tests in [12] show that the inquiry execution over the encrypted data diminishes around 20 percent contrasted and the plaintext question execution. The conventional method to look encrypted data is to unscramble all the data to plain content at that point discover the objective records. Along these lines is costs very time and has a terrible execution particularly with an enormous number of records. We are proposing another strategy to question encrypted data with numerous data types (string, character, numeric, and date). Our strategy will have a decent practically identical reaction time with the conventional way. We likewise will utilize a file over the data, the ordering information ought to be connected with the data well International Journal of Computer Applications (0975 – 8887) Volume 41–No.4, March 2012 47 enough to give a viable inquiry execution system; on the opposite side, the connection among records and data ought not make the way for connecting that can include the insurance. The assailants shouldn't figure the first info esteem from the yield esteem if utilizing a similar capacity for encryption/unscrambling. We have a good test, we don't have a clue how the current DBMSs work and we can't add changes to its centers, that need an open-source DBMS. To take care of this difficult we need to ensure that our new technique can adjust effectively with the DBMS.Our proposed route executed on a standard database from a general benchmark, a few tests will be done to demonstrate the hypothetical thought behind our work and this will follows by examination with the customary way.

PROPOSAL WORK

The Big Data framework consists of four modules: data collection, transmission, storage and simulation as shown in Figure 1, which will enable us to conduct a security analysis of information flowing through the different stages of the network.

3.1 Data Collection Module

Using an operator introduced in the work stations of the clients, the extraction of data produced from the various wellsprings of information that lives in these computers will be performed. This application is answerable for sending the data to the vehicle module who will be liable for data assortment.

3.2 Data Transmission Module

This module is liable for the transmission of data produced by the specialists, gets the information sent by the data assortment module, forms the information, changing it into a conceivable organization, and then sending it to the data stockpiling layer. For this reason, we use Logstash as fundamental information preparing software.

Logstash [9] is an instrument that permits the exchange of information by gathering, preparing, and sending data to a particular goal. This amazing asset underpins a wide assortment of sources of info and data handling, for example, separating occasions, having as yield the fundamental goal like Elasticsearch.

This module is the gathering purpose of all the information sent by the data assortment module, so it is important to consider shortcomings in the transmission brought about by a potential drop-in administration. For which an extra Logstash server is added to this module giving burden adjusting and high availability abilities.

3.3 Data Storage Module

This module is liable for putting away the data created by the specialists situated in the work stations.

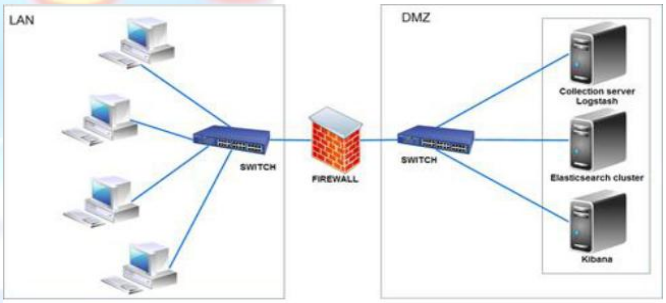


Figure 2.ELK deployment diagram.

For this reason, we utilized Elasticsearch [14], a Lucene based open source web crawler intended to work in groups by imitating data to different hubs offering continuous support capacity [10]. Otherwise called the core of the ELK stack, it is a hunt stage nearly continuously and works with the idea of modified list permitting a quick addition and recuperation of data, likewise, it utilizes a strategy for numerous duplicates to ensure the availability and dependability of put away data.

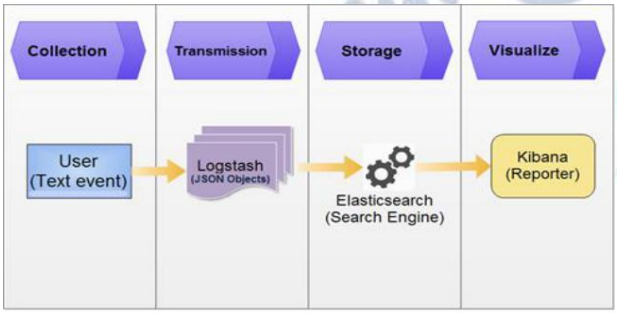
3.4 Display Module

This module plays out the representation of the data put away in Elasticsearch, giving a graphical

interface to the examination and search of information. Kibana [11] was utilized in the usage of this module, an extremely flexible and instinctive opensource apparatus that communicates with Elasticsearch to play out the investigation and representation of data.

MANAGEMENT AND QUERYING OF ENCRYPTED ARCHITECTURE

4.1 Platform Deployment



The stage is actualized in a corporate system gave in its framework of security gear and a virtualized situation in which the arrangement has been conveyed as can be found in Figure 2. The workstations send the data created by the various wellsprings of information to the server Logstash who thusly will transmit this data to Elasticsearch for its stockpiling and its later perception in Kibana.

4.2 Data Source and Collection

The information produced in the workstations is sent to the Logstash server using an application made to recognize any movement that a client performs on his computer, that is, this program will tune in to everything the client types and sends it to the data assortment module.

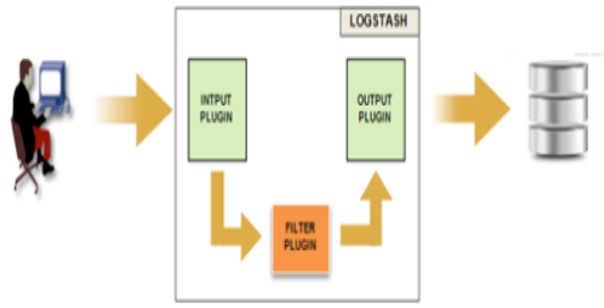


Figure 3. Data collection

4.3 Data Transmission

This module gathers the data sent by the operators introduced in the workstations, giving burden adjusting and ensuring transmission dependability since it is arranged in high availability. Logstash is comprised of three parts: information, yield, and channel, the last configurations the data as indicated by specific particulars. The information segment expends the information originating from an information source and the yield segment sends the data to a goal [4]. The data is commonly not organized and often contains mistaken information that isn't pertinent for legitimate use, so the channel plays out the examination of the info fields and permits disposing of superfluous information, as can be found in Figure 3.

Logstash offers assistance for events and records made by various framework shows, correspondence between methods, talk, and email. It supports UDP, Web sockets, HTTP, and more. It has a module that grants messages to be examined as events on the framework through UDP and the principle course of action field required for this module is the port, for our circumstance we structure it with 5965 by which Logstash checks out events. It similarly uses a plain kind codec that works with direct substance without delimitation and the default character encoding configuration is Unicode Transformation Format UTF-8, as can be found in Figure 4.

```
1 input {
2  udp {
3   port => 5965
4   codec => plain { charset => "UTF-8" }
5   type => "TextEvent"
6  }
7 }
```

Figure 4. Logstash input.

When the information has been gotten and changed into level items, Logstash utilizing the grok channel investigates the unstructured data and changes over it into organized utilizing standard articulations as can be found in Figure 5, these examples are consolidated in Logstash and permit to channel words, numbers, and dates.

Put away in a "logstash-test-text-2019.01.13" file, as can be found in Table 1. Likewise, it very well may be distinguished that the information put away is totally intelligible in clear content "Hi", which speaks to a genuine security hazard, so as a measure to ensure the privacy of the information in the fields considered basic it is important to encode this data. Elasticsearch in its business mode has an apparatus that permits encoding the put away information known as X-Pack [6], installment

supplement that gives security to the data, however which thus is prohibitive for clients who can't get to this asset at their significant expense. This work gives an option of free access to take care of this issue, permitting it to give security by scrambling the data facilitated in Elasticsearch.

Field	Data
_index	logstash-test-text-2019.01.13
_type	TextEvent
_sourceTime stamp	2019-01-12 17:55:02,082
_ typeword	Hello

Table 1. Information stored in the clear text index

In the output, we send Elasticsearch the information that will be stored in an index with the format "logstash-test-text-% + YYYY.MM.dd" as you can see in Figure 5. This will be the document that will contain the data collected by Logstash.

```
output {
 if [type] == "TextEvent" {
  elasticsearch {
   index => "logstash-test-text-%{+YYYY.MM.dd}"
   document_type => "TextEvent"
   hosts => "localhost"
  }
 }
```

Figure5. Logstash output

aper [11] just scrambles the touchy field and it is likewise utilizing pail list to improve inquiry execution. The request on numeric data is valuable. In any case, on the character data, it has little impact. So the technique in [11] isn't good for the character data. [12] Creates a B+ tree list for the data before encoding them. While querying the encrypted data, right off the bat, it finds theencrypted records identified with the querying predicate dependent on the B+ tree list; furthermore, it unscrambles the encrypted records to achieve the outcomes. Likewise, it must encode the B+ tree itself to shield it from releasing secret information. As per the structure of the B+ tree, it encodes every hub of the B+ tree independently. The consequences of tests in [12] show that the inquiry execution over the encrypted data diminishes around 20 percent contrasted and the plaintext question execution. The conventional method to look encrypted data is to unscramble all the data to plain content at that point discover the objective records. Along these lines is costs very time and has a terrible execution particularly with an enormous number of records. We are proposing another strategy to question encrypted data with numerous data types (string, character, numeric, and date). inoersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity in an equation.

The SI unit for magnetic field strength H is A/m. However, if you wish to use units of T, either refer to magnetic flux density B or magnetic field strength symbolized as $\mu_0 H$. Use the center dot to separate compound units, e.g., "A·m^2."

PROTECTING DATA

Big Data stages ought to have the option to have information in their hubs dependably and safely, in any case, this data is presented to threats that could harm its uprightness. This exploration proposes a technique that permits putting away the information in an encrypted structure utilizing an encryption calculation that permits us to give security to the data produced in the assortment module for later capacity in Elasticsearch.

The methods used to ensure security in the Big Data stage depend on the Asymmetric Encryption Standard Advanced Encryption (AES) calculation otherwise called Rijndael, which can process 128-piece squares utilizing encryption keys of 128, 192 and 256 bits [1], which permits scrambling the information from its inception in the workstations until its stockpiling in the group characterized in Elasticsearch.

This strategy will secure against assaults that are proposed to be utilized in the data assortment and capacity process in Elasticsearch to take touchy and secret information. This proposition depends on scrambling the interchanges that are produced from the endpoints by the operator giving an encryption calculation to all the information created by the end-clients.

The specialist introduced in the workstations was created in C#, a programming language planned by Microsoft for its .NET stage [7], which is tuning in to any action produced by a client

gathering and transmitting through a channel unreliable all the information entered, for this specific a program was structured utilizing the Rijndel class situated in the library System Security Cryptography, which permits us to perform symmetric encryption or mystery key encryption.

Once encrypted, the information is sent to Logstash, which thusly transmits it to Elasticsearch and as can be found in Table 2, the put away information is encrypted, which ensures the privacy of the information.

Field	Data
_index	logstash-test-text-2019.01.13
_type	TextEvent
_sourceTimestamp	2019-01-12 17:55:02,082
_ typeword	37SrtQPZoVLQYgqCoJXexQ==

Table 2. Information put away in the encrypted text list

The data channel is tuned in to by catching traffic utilizing the Wireshark [12] traffic investigation apparatus, which permits catching the live data on the system interface of the beginning hardware, it is prove by Figure 6 that the information transmitted is encrypted, ensuring along these lines that the data originating from the work stations is transmitted through the system safely, ensuring its privacy.

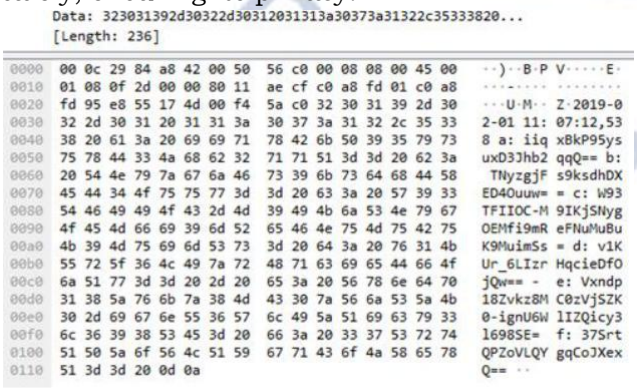


Figure 6. Wireshark encrypted text.

RESULT ANALYSIS

When the proposed arrangement is executed, it is important to approve that its activity doesn't influence the presentation of the ELK stage, explicitly the conveyance time of information from the assortment module to the capacity module. Since we are doing the encryption of data this procedure could cause delays, so it is important to break down the time it takes the information without scrambling structure some portion of the work stations until you show up at Elasticsearch, versus the time it takes encrypted information and recognizes if there are over the top deferrals.

The tests did comprised of estimating the reaction time by sending 4 gatherings of data parcels in which each gathering contains 10 bundles of a similar size, the length of the parcel is corresponding to the substance message which will increment upwards starting with one gathering then onto the next. In the principal gathering will be sent bundles with the message "Hi", the second "Hi World", the third "Hi World Here" and the fourth "Hi World Here Now", from the workstations to the Elasticsearch server. This first situation of experimentation was managed without applying an encryption calculation to then recurrent the analysis with the encrypted messages.

Figure 8 presents the normal reaction time for every data gathering. Noticing that there is a distinction of 2 milliseconds between the gathering of bundles with messages in clear content and encrypted. Demonstrating that the extra time required to encode the messages is indistinct (2 milliseconds) doesn't bargain the exhibition of the stage, getting a satisfactory reaction and guaranteeing consistence with the prerequisites of checking progressively.

When the proposed arrangement is executed, it is important to approve that its activity doesn't influence the exhibition of the ELK stage, explicitly the conveyance time of information from the assortment module to the capacity module. Since we are doing the encryption of data this procedure could cause delays, so it is important to investigate the time it takes the information without encoding structure some portion of the work stations until you show up at Elasticsearch, versus the time it takes encrypted information and distinguishes if there are extreme deferrals.

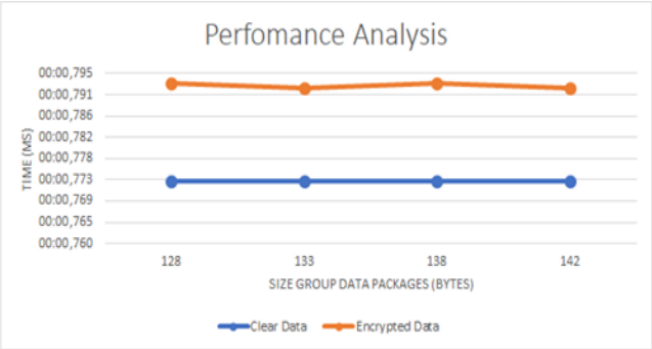


Figure 8. Performance analysis

The tests completed comprised of estimating the reaction time by sending 4 gatherings of data bundles in which each gathering contains 10 bundles of a similar size, the length of the parcel is relative to the substance message which will increment upwards starting with one gathering then onto the next. In the principal gathering will be sent bundles with the message "Hi", the second "Hi World", the third "Hi World Here" and the fourth "Hi World Here Now", from the workstations to the Elasticsearch server. This first situation of experimentation was managed without applying an encryption calculation to then recurrent the investigation with the encrypted messages.

Figure 8 presents the normal reaction time for every data gathering. Taking note of that there is a distinction of 2 milliseconds between the gathering of bundles with messages in clear content and encrypted. Demonstrating that the extra time required to scramble the messages is intangible (2 milliseconds) doesn't bargain the presentation of the stage, acquiring a sufficient reaction and guaranteeing consistence with the prerequisites of observing continuously.

CONCLUSION

The handling of data from end-clients is touchy, because of the significance that this information speaks to individual keys, financial balances, classified information, and so forth. Thus, it is essential to ensure clients' privacy through the execution of security systems. Elasticsearch is an extremely useful asset that gives numerous advantages to handling a lot of information, yet it just ensures the secrecy of information under the utilization of extra installment enhancements to which not all clients approach. Along these lines, in this paper an answer is planned and actualized to scramble the information at source (work stations) which courses through the system in a secured way and is put away at the goal (Elasticserch) encrypted, accordingly ensuring the classification of the information. The tests performed show that the way toward encoding the information doesn't bargain the exhibition of the ELK stage, exhibiting that the extra time required to scramble the messages is subtle of under 2 milliseconds for every data parcel, acquiring a satisfactory reaction and guaranteeing consistence with the observing prerequisites continuously.

REFERENCES

[1] G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.

[2] W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

[3] H. Poor, *An Introduction to Signal Detection and Estimation.* New York: Springer-Verlag, 1985, ch. 4.

[4] B. Smith, "An approach to graphs of linear forms (Unpublished work style)," unpublished.

[5] E. H. Miller, "A note on reflector arrays (Periodical style—Accepted for publication)," *IEEE Trans. Antennas Propagat.*, to be published.

[6] J. Wang, "Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication)," *IEEE J. Quantum Electron.*, submitted for publication.

[7] C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.

[8] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interfaces(Translation Journals style)," *IEEE Transl. J. Magn.Jpn.*, vol. 2, Aug. 1987, pp. 740–741 [*Dig. 9th Annu. Conf. Magnetics* Japan, 1982, p. 301].

[9] K. K. Kumar, S. G. B. Kumar, S. G. R. Rao and S. S. J. Sydulu, "Safe and high secured ranked keyword searchover an outsourced cloud data," 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, 2017, pp. 20-25, doi: 10.1109/ICICI.2017.8365348.

[10] M. Young, *TheTechincal Writers Handbook.* Mill Valley, CA: University Science, 1989.

[11] (Basic Book/Monograph Online Sources) J. K. Author. (year, month, day). *Title* (edition) [Type of medium]. Volume(issue). Available: http://www.(URL)

[12] J. Jones. (1991, May 10). Networks (2nd ed.) [Online]. Available: http://www.atm.com

[13] (Journal Online Sources style) K. Author. (year, month). Title. *Journal* [Type of medium]. Volume(issue), paging if given. Available: http://www.(URL)

[14] R. J. Vidmar. (1992, August). On the use of atmospheric plasmas as electromagnetic reflectors. *IEEE Trans. Plasma Sci.* [Online]. *21(3).* pp. 876—880. Available: http://www.halcyon.com/pub/journals/21ps03-vidmar