An Improved Ec2s2 Framework for Secure Storage Virtualization

Gurvir Kaur
Research Scholar,
Punjabi University,
Patiala, Punjab, India

Gagandeep Kaur, PhD
Associate Professor,
Department of Computer Science,
Punjabi University, Patiala, India

ABSTRACT
The expanding interest for storage and computation has driven the development of vast information centers–the gigantic server cultivates that run huge numbers of today's Internet and business applications. This work proposes another upgraded design for storage virtualization. In a local customer server environment stores information, as well as oversees information appropriately. The principle center of this work is to accomplish four noteworthy goals. In the first place, security parameter where another chronicled data based approach is utilized which stores the MAC address so that there is one time security check along these lines lessening the time and expanding the property. Second metadata creation, in this duplication is uprooted by evacuating the copy words and utilizing the new indexing plan to diminish the span of metadata and hunt time where is enhanced the execution of the framework. At long last, the records are compacted with a misfortune less pressure technique and it is being scrambled and put away in the database therefore diminishing the storage room and securing the document by encoding the record.

Keywords
Storage virtualization, thin provisioning, indexing.

1. INTRODUCTION
In a turbulent economy, virtualization and distributed computing are turning out to be increasingly appealing for endeavors as a result of the comfort and adaptability over conventional figuring. Utilizing a virtual machine is advantageous contrasted with customary processing, for instance, when the client needs to have a machine with a specific arrangement of uncommon applications. Today it is regular to disseminate applications preinstalled on a virtual machine which the client can send without confounded establishments and arrangements.

Cloud Computing is a set of IT services that are provided over a network on a leased basis. As more and more services and information is placed on the cloud, storage and security issues come into force. Expense is the urgent variable in today's business world. IT framework is one of the critical areas where the associations are spending more to accomplish better execution and business congruity. The associations are constantly quick to lessen the expense without giving up the execution, which drives them to win the business contenders. Efficient utilization of resources reduces the cost associated with storage of data. In this regard virtualization technology plays an important role to achieve efficient consumption of the resources, by vitalizing the physical hardware.

In the present situation, associations are changing their customary datacenters to a virtualized datacenter to acquire the benefits from virtualization. IT framework virtualization incorporates server, stockpiling and system asset. Virtualization changes the customary datacenter into a more adaptable datacenter through server virtualization and combination in this way disentangling the provisioning of IT assets. Thus asset combination diminishes equipment cost. Consideration of element virtual machine relocation from servers to servers builds the adaptability altogether.

Distributed storage frameworks normally utilize uncommon equipment and programming alongside plate drives keeping in mind the end goal to give quick and dependable stockpiling for figuring and information preparing. The frameworks are very intricate, and might be considered as unique reason PC intended to furnish stockpiling limit alongside cutting edge information security highlights. Circle drives are a critical component inside a capacity framework, alongside equipment and unique reason inserted programming. They can give either piece got to capacity, or document got to capacity. Square get to is commonly conveyed over Fiber Channel, iSCSI, SAS, FICON or different conventions though document access is given utilizing NFS or CIFS conventions.

Storage Virtualization can occur at two different levels namely Block virtualization and file virtualization:

Block virtualization alludes to the deliberation of coherent stockpiling from physical stockpiling in a way that it can be gotten to without respect to physical stockpiling or heterogeneous structure. This partition permits the executives of the capacity framework more noteworthy adaptability away administration [1].

Record virtualization addresses the Network-Attached Storage (NAS) challenges by killing the conditions between the information got to at the document level and the area where the records are physically put away. This gives chances to advance stockpiling use and server solidification keeping in mind the end goal to perform non-problematic record movements. The strategy for moving the physical area is known as information movement. The genuine stockpiling is the capacity to exchange information while keeping up simultaneous I/O access.

2. SECURE STORAGE VIRTUALIZATION
With the advance of storage technologies to networked-attached storage, a recently emerging architecture that provides higher performance and availability than traditional direct-attached disks, new security concerns arise. In these environments, clients can no longer rely only on the storage servers to provide security guarantees, since these become now easier to compromise. In consequence, clients have to play a more proactive role in protocols designed for data protection and work together with the storage servers to ensure the confidentiality and integrity of the data.

2.1. MAC Address Filter

The basic principle of MAC address filter is recording all valid MAC addresses into a list. The clients on the list are authenticated and can be connected to the target AP. Other clients whose MAC addresses are not on the list are not allowed to connect to the AP. An authentication mechanism is responsible for creating a credential, which is used to discern whether a client is who it claims to be. This is the first line of defines conducted before the connection between clients and AP is established. After the clients pass the authentication mechanism, either legal or illegal, they will confront the second security defines measure, the encryption mechanism.

2.2. Encryption of Data using SSL

This strategy gives secure correspondence connects and permits encryption of information utilizing SSL. The utilization of SSL gives extra cryptographic security. The steps are given below:

- Initially, virtual machines request for file.
- Then (Token + Path) is generated by file server and is forwarded to the client
- A session key is established with the help of symmetric key cryptography by Client component and storage area network (SAN) component. When session is established, the component server continues secure communication by using symmetric key encryption. The session key is valid for single session only.
- Token used during the session establishing phase is encrypted by the client component.
- To storage area network component E (Token + Path) is passed by virtual machine.
- SAN component authenticates and validates this token.
- Storage Network releases the files after performing successful token validation; otherwise operation will is denied.
- VM file will be available once storage network is allowed.

The whole process of key generation and cross validation ensures that the identified user is authentic to be approving the access. Here the application of cryptographic technique identifies the valid users, then authenticate for the present session and finally allows the users to access the data. In the other words, this technique ensures the session based access and cross certification of the uniqueness of the access request. This helps in the securing of data.

2.3. Storage Virtualization Challenges

Capacity systems administration and highlight rich savvy stockpiling clusters have tended to and given particular answers for business issues. As an enabling means, virtualization needs to build the estimation of the present course of action, yet bringing virtualization into a domain includes new challenges. The capacity virtualization arrangement must be fit for tending to the issues, for example, versatility, usefulness, sensibility and backing.

2.3.1. Scalability:

Consider the adaptability of a situation with no virtualization. Its environment may have a few stockpiling clusters that give stockpiling autonomously of each other. Every cluster is overseen autonomously and meets application necessities as far as IOPS and ability [7]. Once in the past virtualization, a capacity exhibit won't be seen as an individual substance. The surroundings as a full ought to as of now be examined. Therefore, the framework that is upheld each at a physical level and from a virtualization viewpoint ought to be set up to viably handle the administration.

2.3.2. Functionality:

Usefulness is another test away virtualization. Capacity cluster gives an extensive fluctuates of advanced usefulness crucial for meeting an application's administration levels. This incorporates neighborhood replication, developed separation remote replication. In exceedingly virtualized surroundings, the virtual gadget ought to offer consistent or propelled execution than what is in a matter of seconds open on the capacity cluster, and it ought to in any case influence existing usefulness on the exhibits.

2.3.3. Manageability

The administration of the capacity base in a virtualized domain is a fundamental thought for capacity chairmen. Introducing a virtualization gadget breaks the end-to-end view into three unique regions: the server to the virtualization machine, the virtualization gadget to the physical stockpiling, and the virtualization machine itself. The virtualized stockpiling area must be expert for meeting these issues and should arrange with existing organization mechanical assemblies to engage organization of a conclusion to-end virtualized environment.

2.3.4. Support

Virtualization is not a stand-alone innovation anyway one thing that must work at interims partner degree existing setting that is entangled and occasionally needs numerous administration apparatuses that presents capacity issues. Without a virtualization arrangement, a few partnerships endeavor to combine items from a solitary merchandiser to facilitate these difficulties [9]. Presenting a virtualization arrangement lessens the imperative to institutionalize on individual merchandiser. Be that as it may, supportability issues in an abundantly virtualized heterogeneous environment set up difficulties in coordination and similarity of items and arrangements from totally diverse producers and merchants.

2.3.5. Backup

The most important difficult part of any storage services is that it should be essential obtainable and recoverable. With the consolidation of data will classically effect in a huge quantity of records existing in a particular location. According to the data size, backing up and replicating this data could be a major time upload. Other key issues are like bandwidth ease of use and storage appliance backplane bottlenecks.

3. RESOURCE MANAGEMENT IN VIRTUAL STORAGE

The dynamic workloads seen by large portions of the applications running inside server farm imply that a little while later, the underlying position and asset offers given to a virtual machine may get to be inadequate for its developing interest. Hotspots structure inside a server farm when the asset necessities of one or more VMs on a physical host surpass the host's ability [10]. The extensive scale nature and the rate with which workloads can change imply that the server farms require computerized asset administration methods to keep these hotspots.

3.1. Challenges

Organizations that have adopted virtualization technologies are facing a number of management issues. One of management issue in organization is server consolidation in which the number of (physical) server scan be significantly reduced to smaller number of (physical) servers that are running multiple virtual servers. While the hardware can significantly be reduced, an organization still has to manage the large number of virtual servers. Furthermore, server virtualization allows organizations to easily deploy new virtual servers with the click of a mouse button. Although server consolidation reduces number of physical server but due to a large number of virtual servers it become hard to manage. Server virtualization introduced a new issue, called VM sprawl.

Traditionally, if someone wants to deploy more servers there is a whole process of ordering hardware and procedures, but now IT administrators can deploy virtual servers with a click on the mouse [11]. There is often the deception that virtual machines are free. However, the deployment of new virtual machines can lead to the requirement or purchase of additional hardware resources. Deploying a couple new virtual servers might not require additional resources, but in case of VM sprawl, the virtual environment is characterized by many redundant virtual machines that inefficiently use resources. This could lead to the purchase of more hardware, consisting of servers and storage, and possibly additional licensing cost. Also, while the virtual environment can be centrally managed more IT staff might be needed to address and solve the VM sprawl problem.

Another problem is the sharing of IT resources. As resources are shared, the applications and data of different departments of an organization can reside anywhere in the virtual environment. Different departments of an organization are often also reluctant of the idea on sharing resources with other departments and losing control of their own resources. However, new development in management tools has addressed some of these problems.

Different difficulties, like provisioning and scaling of administrations inside framework mists alongside cost-touchy asset provisioning for versatile information gushing applications in virtualized situations. Resource provisioning can include three dimensions as per hardware resources, the software obtainable on those resources, inclusion the instance throughout which those resources must be assured to be accessible. Resource provisioning model should permit resource clients to identify supplies across these three dimensions as well as the resource contributor to resourcefully assure those necessities.

4. PROPOSED METHODOLOGY

Technique is the methodical investigation of strategies that are, can be, or have been connected inside an order. To accomplish the set down destinations, an orderly method is to be taken after. The examination approach is the trial study in which a structure is intended to store information safely on the servers. The accompanying strides are to be taken after:

4.1. Set-up of Client - Server Environment:

An adjacent client server environment can be made using .net framework. For making this environment, Microsoft Visual Studio 2010 and SQL Server Management Studio 2008 have been used. Microsoft Visual Studio is a joined progression environment from Microsoft. It is used to make PC programs for Microsoft Windows, locales, web applications and web organizations. It supports differing programming vernaculars and grants the code administrator and debugger to reinforce about any programming tongue. Visual Studio 2010 goes with .NET Framework 4 and support IBM DB2 and Oracle Databases, despite Microsoft SQL Server. It fuses instruments for investigating parallel applications. In this client server outline; server has a limit pool which contains resources. Virtualization is a basic segment for limit organization; thusly virtual machines have been created by of advantages means number of virtual machines of any benefit depends on upon their capacity and temperature. In this Framework, 30 Resources has been added to the capacity pool where diverse number of Virtual machines has been created for every asset.

4.2. Secure Storage Virtualization based on Historical information with SSL Authentication:

Existing component permits secure access control to the capacity virtualization approach. This method gives secure correspondence connects and permits encryption of information utilizing SSL. Proposed framework works as follows:

- Initially, virtual machines request for file.
- Then (Token + Path) is generated by file server and is forwarded to the client.
- A session key is established with the help of symmetric key cryptography by Client component and storage area network (SAN) component. When session is established, the component server continues secure communication by using symmetric key encryption. The session key is valid for single session only.
- Token used during the session establishing phase is encrypted by the client component.
- To storage area network component E (Token + Path) is passed by virtual machine.
- SAN component authenticates and validates this token.
- Storage Network releases the files after performing successful token validation; otherwise operation will is denied.
- VM file will be available once storage network is allowed.

This technique is highly secure but **this authentication process takes so much time which were cause of delay**. So, to control this historical based access system is used with SSL. Where server stores the historical access information about user and on this basis user gets authentication. Basically historical access information is the information about the user i.e. their personal information or any login information so that the user can be authenticated once only rather that again and again. This information also includes MAC address information and security key. MAC address is defined as a physical address of a machine which is unique for every machine. **This scheme is secure as well as fast.**

4.3. Thin Provisioning using Indexing

Thin provisioning is an architecture which uses virtualization technology to show the more physical resources than actually available resources. A system is not thin provisioned if it contains enough resource to support all of the virtualized resources. Thin provisioning is applied to secured

architecture. The technology of thin provisioning is based on thin-store [5]. Storage provisioning is the technique of allocating storage space to virtual machines, servers or any other computing device. It ensures the capability of the system to utilize the resource intelligently in case of huge availability. Thin store component of the proposed method comprises of four parts.

4.3.1. Metadata Manager:

For the administration of metadata the metadata director assumes a significant part which is key for virtualization and controls legitimate volume and mapping table.

4.3.2. Address Mapper:

Utilize for load balancing and forms mapping demand from legitimate volume. It powerfully dispenses physical location from the capacity.

4.3.3. Storage Reclaimer:

It basically deals with the free space. This diminishes provisioning a productive way to deal with use stockpiling in the better way.

4.3.4. Resource Monitor:

It investigates the condition of capacity gadget and deals with the storage rooms when its aggregate limit is going to wrap up.

Metadata is information about information. It is abnormal state data that incorporates when something was done, where it was done, the document sort and configuration of the information, the first source, and so on[12]. The thought of metadata can be extended to incorporate data about how substance is being utilized, who is utilizing the substance, and when various bits of substance are being utilized can important and profitable affiliations be watched. In this secure system while user upload data, metadata has been generated. Metadata is defines as a data about data in which all the keywords of the stored files has been indexed. It means all the keywords of the files has been saved and maps with its location. Sometimes different files have same keywords and they stored different location. It means same keywords are on the different location which results as a **duplication of data and heavy storage problem**. So, to remove this heavy load in this work duplication has been removed. This removal of duplication is achieved by new indexing scheme as shown in table1.

Table 1: Indexing Schemes

Metadata Indexing			
Original		**New Enhanced**	
Cloud	1	cloud	1,2,3
Energy	1	Energy	1,2
Storage	2	virtualization	3
Cloud	2		
Energy	2		
Virtualization	3		
Cloud	3		

Due to removal of duplicate words from the metadata, the size of the metadata decreases. On implementing the system the metadata size was observed to be 264 KB and when same number of files was uploaded in the proposed system the size of metadata was reduced to 136 KB. This will show how much storage requirement is decreased as compared to previous work as shown in table 2.

Table 2: Storage Reduction for Metadata

Metadata Size		
Original Size (KB)	**Enhanced Size (KB)**	**Enhancement %**
264	136	48.48

4.4. Resource Management using Data Compression

Information pressure includes encoding data utilizing fewer bits than the first representation. Pressure can be either lossy or lossless. Lossless pressure decreases bits by distinguishing and killing factual repetition. No data is lost in lossless pressure. As such information can pack and uncompressed without loss of data. Lossy pressure decreases bits by distinguishing superfluous data and evacuating it. The procedure of lessening the span of an information record is alluded to as information pressure. In this work, lossless information pressure procedure is utilized to pack information which likewise lessens capacity prerequisite.

5. RESULTS AND DISCUSSION

In performance analysis, write time, read time and response time of Enhanced Cloud Control and Security System (EC2S2) Framework and Advanced Framework has been analyzed.

5.1. Write Time Analysis

Write time can be measured for putting away every record on the server. Here Time can be characterized as a period taken to store information on the server where information is some place composed on a plate at the server side. Write time can be calculated as:

$$Time(W_{Time}) = size\ of\ data * Data\ Transfer\ rate$$

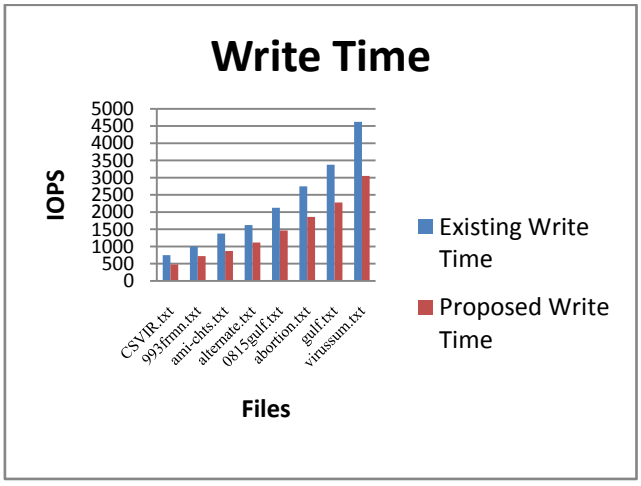

Figure 1: Write Time Analysis per File (IOPS)

5.2. Read Time Analysis:

Read time can be measured for removing document from the server. Here Time can be characterized as a period taken to concentrate information from the server where information is some place composed on a plate at server side. Read time can be calculated as:

$$Time(R_{Time}) = (size\ of\ data * data\ Transfer\ rate) + response\ time$$

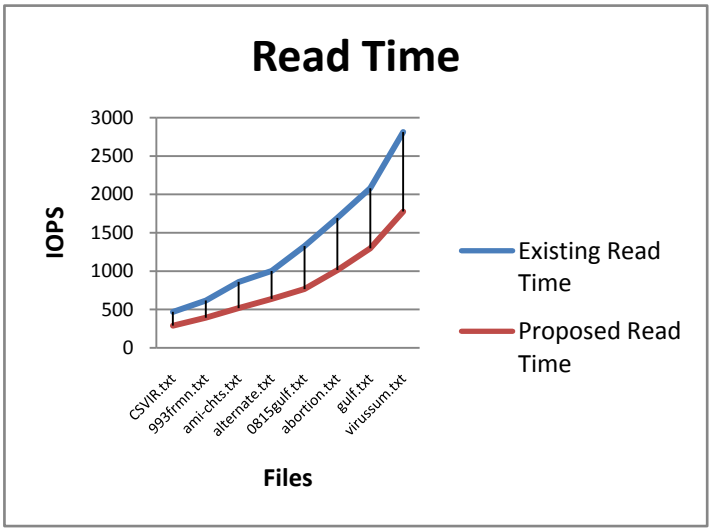

Figure 2: Read Time Analysis per File (IOPS)

This predicts shows that the read time is decreased as compared to previous work. Table 3 shows the percentage improvement or enhancement in write time & read time for different files.

Table 3: Percentage Improvement

Sr. No.	File name	Existing			Enhanced			% Improvement (Write Time)	% Improvement (Read Time)
		File Size (KB)	Write Time	Read Time	Compressed File size (KB)	Write Time	Read Time		
1	CSVIR.txt	101	750	468	71	480	288	36.00	38.46
2	993frmn.txt	141	1000	616	99	720	392	28.00	36.36
3	ami-chts.txt	178	1375	858	125	868	517	36.87	39.74
4	alternate.txt	220	1625	1001	154	1116	637	31.32	36.36
5	0815gulf.txt	284	2125	1326	199	1464	765	31.11	42.31
6	abortion.txt	359	2750	1694	252	1860	1012	32.36	40.26
7	gulf.txt	439	3375	2079	308	2280	1296	32.44	37.66
8	virussum.txt	592	4625	2812	415	3050	1776	34.05	36.84

The average improvement in read and write time for 16 KB block is as shown in table 4.

Table 4: Average Improvement

Write & Read Time for 16 KB Block			
	Existing (IOPS)	**Proposed (IOPS)**	**Percentage Improvement**
Write Time	125	122	2.4
Read Time	77	69	10.4

5.3. Performance Analysis

Response time can be measured while user searches their data from server. Here Time can be defined as a time taken to search data on the server where data is somewhere written on a disk at the server side. Response time can be calculated as:

$$Time(R_{Time}) = time\ to\ search\ string\ from\ meadata + read\ data\ from\ disk$$

When the performance of the proposed system is analyzed with the existing system there is an improvement of 17.75% as compared to the existing system.

Table 5: Performance Analysis

Performance Analysis(Average Response Time)			
	Existing(ms)	**Proposed(ms)**	**Percentage Improvement**
Response Time	7.38	6.07	17.75

6. CONCLUSION

This local client-server environment not only stores data, but also manages data properly. This work is majorly focused to achieve four main objectives. First, security parameter where a new historical information based approach is used which stores the MAC address so that there is one time security check thus reducing the time and increasing the property. Second metadata creation, in this duplication is removed by removing the duplicate words and using the new indexing scheme to reduce the size of metadata and search time where is improved the performance of the system. Finally, the files are compressed with a loss-less compression method and it is being encrypted and stored in the database thus reducing the storage space and securing the file by encrypting the file. Further the proposed setup can be experimented with different forms of data like image, video and pdf files etc.

7. REFERENCES

[1] Chuanyi LIU, Dapeng JU, Yu GU, Youhui ZHANG, Dongsheng WANG and David H.C.Du, (2008), Semantic Data de-Duplication for Archival Storage Systems, 13th IEEE Asia-Pacific Computer Systems Architecture Conference, pp. 1–9.

[2] Deepa Raj and Seema Gupta, Framework of Adaptive Lossless Compression with Block Wise Data, 2015 International Conference on Advances in Computer Engineering and Applications (ICACEA), pp. 103–107, March 2015.

[3] Felipe Díaz, Elias A. Doumith and Maurice Gagnaire, (2011), Impact of Resource Over-Reservation (ROR) and Dropping Policies on Cloud Resource Allocation, 2011 3rd IEEE International Conference on Cloud Computing Technology and Science, pp. 470–476.

[4] Jun Wei Ge, Yong Long Deng, and Yi Qiu Fang, (2010), Research on Storage Virtualization Structure in Cloud Storage Environment, 2010 International Conference on Multimedia Technology, ICMT 2010, pp. 1-4.

[5] Kai Quian, Letian Yi and Jiwu Shu, ThinStore: Out-of-Band Virtualization with Thin Provisioning, 6th IEEE International Conference on Networking, Architecture, and Storage, pp. 1–10, July 2011.

[6] Lei Liu, (2008), An Autonomous Search Method for Storage Virtualization, International Conference on High-Performance Clustered Computing, pp.1-10.

[7] Lin Qiao, Balakrishna R. Iyer, Divyakant Agrawal and Amr El Abbadi, SVL: Storage Virtualization Engine Leveraging DBMS Technology. 21st International Conference on Data Engineering (ICDE), pp. 1048–1059, April 2005.

[8] R.Aishwarya and Dr.S.Malliga, (2014), Intrusion detection system- An efficient way to thwart against Dos/DDos attack in the cloud environment, 2014 International Conference on Recent Trends in Information Technology,

[9] Sandeep Kumar and Syam Kumar P, Secure and Efficient Design and Implementation of out-of-band Storage Virtualization, 2015 IEEE International Advance Computing Conference (IACC), pp. 1021–1025, June 2015.

[10] Sho Kawahara and Kenichi Kourai, (2014). The Continuity of Out-of-band Remote Management Across Virtual Machine Migration in Clouds, IEEE/ACM 7th International Conference on Utility and Cloud Computing, pp. 176–185.

[11] Tzu-Fang Sheu, Nen-Fu Huang and Hsiao-Ping Lee, 2010. In-Depth Packet Inspection Using a Hierarchical Pattern Matching Algorithm, IEEE Transactions on Dependable and Secure Computing, Vol. 7, No. 2, June 2010, pp. 175-188.

[12] Xiaolong Xu and Qun Tu, (2015). Data Deduplication Mechanism for Cloud Storage Systems, 2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, pp. 286-294.

[13] Yingping LU, Chuanyi LIU, Chunhui Shi, Guanlin LU, David H.C.Du and Dongsheng WANG, (2008), ADMAD: Application-Driven Metadata Aware de-Duplication Archival Storage System, 5th IEEE International Workshop on Storage Network Architecture and Parallel I/Os, pp. 29–35.

[14] Zhen Qin, Xian-ping Tao, Yu Huang, Jian Lu and Tao Wu, (2010) Policy Generation for Privacy Protection Based on Granular Computing, 7th Web Information Systems and Applications Conference, WISA 2010, pp. 53–58.

[15] Songchun Gong, Songyin Fu, and Zheng Chen, Research on the Key Technologies of Storage Virtualization, 2012 International Conference on Cybernetics and Informatics, Springer, Volume 163, 2014, pp. 1273-1279.