UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

MICROSOFT CORPORATION,
Petitioner,

v.

QOMPLX LLC,
Patent Owner.

IPR2026-00183
Patent: USP 11,539,663
Issued: December 27, 2022
Application No. 16/412,340
Filed: May 14, 2019

Title: SYSTEM AND METHOD FOR
MIDSERVER FACILITATION OF LONG-HAUL
TRANSPORT OF TELEMETRY FOR CLOUD-BASED SERVICES

DECLARATION OF MARK CROVELLA

MICROSOFT CORP.
EXHIBIT 1003

TABLE OF CONTENTS

Page

I. INTRODUCTION AND ENGAGEMENT 4

II. BACKGROUND AND QUALIFICATIONS 4

III. MATERIALS CONSIDERED AND INFORMATION RELIED UPON REGARDING '663 PATENT 7

IV. UNDERSTANDING OF PATENT LAW 9

V. SUMMARY OF OPINIONS 12

VI. LEVEL OF SKILL IN THE ART, AND STATE OF THE ART 13

A. Person Of Ordinary Skill In The Art 13

B. State Of The Art 14

1. Virtual Appliances Were Commonly Used For Simple And Rapid Deployment Of Resources 14

2. Open Virtual Appliances (OVAs) Were A Known And Standard Format For Automatically Installing And Configuring Applications In Virtual Machines 19

3. Midservers Were Well-Known And Facilitated Data Transfer Between Local Networks And Cloud Platforms 22

VII. THE '663 PATENT 27

A. The Claims' Earliest Effective Filing Date 27

B. The '663 Patent's Specification 27

VIII. CLAIM CONSTRUCTION 29

IX. GROUND 1: CLAIMS 1 AND 4 ARE OBVIOUS OVER MATURANA IN VIEW OF KASHYAP 29

A. Maturana (EX1004) 30

B. Kashyap (EX1005) 34

C. Combining Maturana And Kashyap.....38

D. Claim 1 47

1. Element [1.1].....47

2. Element [1.2].....49

3. Element [1.3].....52

4. Element [1.4].....59

5. Element [1.5].....61

6. Element [1.6].....64

7. Element [1.7].....68

E. Claim 4 69

1. Element [4.1].....69

2. Element [4.2].....70

3. Element [4.3].....71

4. Element [4.4].....71

5. Element [4.5].....71

6. Element [4.6].....71

7. Element [4.7].....72

I, Mark Crovella, do hereby declare as follows:

I. INTRODUCTION AND ENGAGEMENT

1. I have been retained as an independent expert on behalf of Microsoft Corporation in connection with the above-captioned Petition for *Inter Partes* Review ("IPR") to provide my analyses and opinions on certain technical issues related to U.S. Patent No. 11,539,663 (hereinafter "the '663 patent").

2. I am being compensated at my usual and customary rate for the time I spent in connection with this IPR. My compensation is not affected by the outcome of this IPR.

3. Specifically, I have been asked to provide my opinions regarding whether claims 1 and 4 (each a "Challenged Claim" and collectively the "Challenged Claims") of the '663 patent would have been obvious to a person having ordinary skill in the art ("POSITA") as of May 14, 2019. It is my opinion that each Challenged Claim would have been obvious to a POSITA after reviewing the prior art discussed herein.

II. BACKGROUND AND QUALIFICATIONS

4. I am an expert in the fields of computer science and information technology. In formulating my opinions, I have relied upon my training, knowledge, and experience in the relevant art. A copy of my curriculum vitae is appended to this declaration as Appendix A and provides a description of my professional experience,

including my academic and employment history, publications, conference participation, awards and honors, and more. The following is a brief summary of my relevant qualifications and professional experience.

5. I am a Professor in the Department of Computer Science at Boston University, and the Chair of Academic Affairs for the Faculty of Computing & Data Sciences. From 2013 to 2018 I also served as Chair of the Department of Computer Science. A brief summary of my background is as follows.

6. I earned an undergraduate degree in Biology from Cornell University, which I received in 1982. I received a master's degree in Computer Science from the University of Buffalo in 1989. I received a Ph.D. in Computer Science from the University of Rochester in 1994. The subject of my Ph.D. thesis was "Performance Prediction and Tuning of Parallel Programs."

7. From 1982 to 1984, I worked as a computer programmer for the State of Colorado. From 1984 to 1994, I was employed at Calspan Corporation, a research and development firm in Buffalo, NY, where I rose to the level of Senior Computer Scientist. My work at Calspan focused on development of experimental software and large-scale simulation software in support of contracts between Calspan and the U.S. Department of Defense.

8. In 1994, I joined the faculty of Boston University, initially as an Assistant Professor of Computer Science. I was promoted to the rank of Associate

Professor in 2000 and became a full Professor in 2006. I served as Chair of the Department of Computer Science from 2013 to 2018.

9. I have been pursuing research in the area of computer networking and data science since 1994. I have conducted research in a variety of areas related to the Internet and the World Wide Web, mainly through the application of data mining techniques and statistics. Among other areas, I have studied the efficient design of Web servers and content distribution systems. My research in network systems and data science includes research in the field of database and query systems, which are the backbone of services provided over the Internet and the World Wide Web. I have a published a number of research papers on topics related to distributed servers and datamining. My work in these areas is highly cited. From 2007 to 2009, I was the Chair of ACM SIGCOMM (the Special Interest Group in Computer Communication), the main professional organization for scientists in the field of computer networking.

10. I also spent time in industry working on topics related to network services and databases. I cofounded a startup (Commonwealth Network Technologies) which was acquired by Webmanage, Inc., where I had the title of Chief Scientist. At Webmanage, I designed content distribution systems. Webmanage was subsequently acquired by Network Appliance, Inc., as part of their Netcache division, where I became a Technical Director. At Network Appliance, I

helped design systems to distribute Web content and populate Web caches. I also served as Chief Scientist for the data science startup Guavus, developing big data solutions for the telecommunications industry.

11. I am co-author of Internet Measurement: Infrastructure, Traffic, and Applications (Wiley Press, 2006), and I am the author of over two hundred research papers. According to Google Scholar, my work has been cited by over 32,000 other publications. According to Google Scholar, as of May 2018, only 11 other researchers worldwide are more cited in the area of computer systems. I hold nine U.S. Patents derived from my Internet related research. I am a past editor of principal journals in related fields: Computer Communication Review, IEEE/ACM Transactions on Networking, Computer Networks, and IEEE Transactions on Computers.

III. MATERIALS CONSIDERED AND INFORMATION RELIED UPON REGARDING '663 PATENT

12. In preparing this declaration, I have reviewed the following materials bearing Exhibit Nos. that I understand are being referenced in the IPR to which my declaration accompanies:

LIST OF EXHIBITS

No.	Description
1001	U.S. Patent No. 11,539,663 ("the '663 patent")
1002	File History of U.S. Patent No. 11,539,663
1003	N/A

No.	Description
1004	U.S. Patent App. Pub. No. 2015/0281355A1 ("**Maturana**")
1005	U.S. Patent App. Pub. No. 2012/0246639A1 ("**Kashyap**")
1006	U.S. Patent App. Pub. No. 2015/0178113A1 ("**Dake**")
1007	B. Sosinsky, *Cloud Computing Bible* (2011) ("**Sosinsky**")
1008	U.S. Patent App. Pub. No. 2012/021618A1 ("**Arcese**")
1009	U.S. Patent App. Pub. No. 2013/0152080A1 ("**Sachindran**")
1010	X. Jin et al., *Reinventing Virtual Appliances*, IBM J. of Res. and Dev., 53(4), Paper 7 (2009) ("**Jin**")
1011	S. Mastrianni et al., *IT Autopilot: A Flexible IT Service Management and Delivery Platform for Small and Medium Business*, IBM Sys. J., 46(3), at 609-24 (2007) ("**Mastrianni**")
1012	U.S. Patent No. 10,298,670 to Ben-Shaul et al. ("**Ben-Shaul**")
1013	U.S. Patent App. Pub. No. 2016/0048408 A1 to Madhu et al. ("**Madhu**")
1014	Gabor Kecskemeti, et al., *Toward Efficient Appliance Delivery With Minimal Manageable Virtual Appliances*, IEEE Trans. On Services Computing, 7(2), at 279-92 (2014) ("**Kecskemeti**")
1015	U.S. Patent No. 10,579,403 to Antony et al. ("**Antony**")
1016	U.S. Patent App. Pub. No. 2007/0294676 A1 to Mellor et al. ("**Mellor**")
1017	U.S. Patent App. Pub. No. 2019/0026138 A1 to Cao et al. ("**Cao**")
1018	U.S. Patent App. Pub. No. 2019/0065213 A1 to Li et al. ("**Li**")
1019	DMTF, *Open Virtualization Format White Paper*, ver. 1.0.0 (2009) ("**OVF-2009**")
1020	DMTF, *Open Virtualization Format White Paper*, ver. 2.0.0 (2014) ("**OVF-2014**")
1021	DMTF, *Open Virtualization Format Specification*, ver, 2.1.1 (2015) ("**OVF-2015**")
1022	U.S. Patent App. Pub. No. 2013/0283364 A1 to Chang et al. ("**Chang**")
1023	N/A
1024	List of Challenged Claims

IV. UNDERSTANDING OF PATENT LAW

13. I am not an attorney. For the purposes of this declaration, I have been informed about certain aspects of the law that are relevant to my opinions. My understanding of the law was provided to me by the Petitioner's attorneys.

14. I understand that when considering the scope of the claims of a patent that the patent claim terms should generally be given the ordinary meaning that the terms would have to a person of ordinary skill in the art in question after reading the patent as of the earliest claimed priority date. As discussed below, it is my opinion that the challenged claims are not entitled to the earliest priority date asserted by the '663 patent. However, in my opinion, the meaning of the claim terms would not have materially differed to a person of ordinary skill in the art ("POSITA") whether the claims were evaluated at the time of the earliest claimed priority date (April 28, 2016) or at the time the application that led to the '663 patent was filed (May 14, 2019).

15. I understand that the person of ordinary skill in the art is deemed to read the claim term not only in the context of the particular claim in which the term appears, but in the context of the entire patent, including the other claims and the specification of the application as filed. I further understand that the principal considerations regarding the scope and meaning of the claims are the plain language of the claim (including the surrounding claim language and context), the patent

specification, and the prosecution history. I understand that while a claim is to be read in light of the specification, one must generally avoid importing limitations into the claim from the specification. I am also informed that the prosecution history can often inform the meaning of the claim by demonstrating how the inventor understood the invention and whether the inventor limited the invention in the course of prosecution, making the claim scope narrower than it would otherwise be. I applied these understandings when considering the scope and meaning of the claims of the ’663 patent.

16. I understand that a prior art reference anticipates an asserted claim, and thus renders the claim unpatentable, if all elements of the claim are disclosed in that prior art reference, either explicitly or inherently (i.e., necessarily present or implied).

17. I further understand that a claim is unpatentable if it would have been obvious. Obviousness of a claim requires that the claim would have been obvious from the perspective of a POSITA at the time the alleged invention was made. I understand that a claim could have been obvious from a single prior art reference or from a combination of two or more prior art references. I understand that an obviousness analysis requires an understanding of the scope and content of the prior art, any differences between the alleged invention and the prior art, and the level of ordinary skill in evaluating the pertinent art.

18. I further understand that a claim would have been obvious if it unites old elements with no change to their respective functions, or merely substitutes one element for another known in the field, and that combination yields predictable results. While it may be helpful to identify a reason for this combination, I understand that there is no strict requirement of finding an express teaching, suggestion, or motivation to combine within the references. When a product is available, design incentives and other market forces can prompt variations of it, either in the same field or different one. If a POSITA can implement a predictable variation, obviousness likely bars its patentability. For the same reason, if a technique has been used to improve one device and a POSITA would recognize that it would improve similar devices in the same way, using the technique would have been obvious. I understand that a claim would have been obvious if common sense directs one of ordinary skill, without hindsight bias caused by knowledge of the challenged claims, to combine multiple prior art references or add missing features to reproduce the alleged invention recited in the claims.

19. I further understand that certain factors may support or rebut the obviousness of a claim. I understand that such secondary considerations include, among other things, commercial success of the patented invention, skepticism of those having ordinary skill in the art at the time of invention, unexpected results of the invention, any long-felt but unsolved need in the art that was satisfied by the

alleged invention, the failure of others to make the alleged invention, praise of the alleged invention by those having ordinary skill or expertise in the art, and copying of the alleged invention by others in the field. I understand that there must be a nexus—a connection—between any such secondary considerations and the alleged invention. I also understand that contemporaneous and independent invention by others is a secondary consideration tending to show obviousness.

20. I am not aware of any allegations by the named inventors of the ’663 patent or any assignee of the ’663 patent that any secondary considerations tend to rebut the obviousness of any Challenged Claim of the ’663 patent.

21. I understand that in considering obviousness, it is important not to determine obviousness using the benefit of hindsight derived from the descriptions found in the patent being considered.

22. I understand that Petitioner has the burden of proving unpatentability by a preponderance of evidence, which means that the claims are more likely than not unpatentable.

23. The analysis in this declaration is in accordance with the above-stated legal principles.

V. <u>SUMMARY OF OPINIONS</u>

24. In forming the opinions set forth herein, I have considered:

a) The documents listed above;

b) My education, knowledge, skills, and experience in the field of computer science and engineering; and

c) The level of skill of a person having ordinary skill in the art at the time of the alleged invention of the ’663 patent, namely May 14, 2019.

25. As I explain in further detail below, it is my professional and expert opinion that claims 1 and 4 of the ’663 patent are obvious in view of the prior art.

VI. LEVEL OF SKILL IN THE ART, AND STATE OF THE ART

A. Person Of Ordinary Skill In The Art

26. I understand that in evaluating obviousness of the ’663 patent claims, the content of a patent or printed publication prior art should be interpreted the way a person of ordinary skill in the art would have interpreted the prior art as of the effective filing date of the challenged patent claims. I assume that the effective filing date is the date the application resulting in the ’663 patent was filed, namely May 14, 2019. My understanding of a person of ordinary skill in the art would not materially differ if the effective filing date were the earliest filing date of an alleged related application, namely April 28, 2016, or May 14, 2019.

27. I understand that factors that may be considered in determining the level of ordinary skill in the art at the time of the effective filing date of the challenged patent include: (1) the educational level of the inventor; (2) the types of problems

encountered in the art; (3) prior art solutions to those problems; (4) rapidity with which innovations are made; (5) sophistication of the technology; and (6) educational level of active workers in the field.

28. A person of ordinary skill in the art ("POSITA") in May 2019 would have had a bachelor's degree in computer science, computer engineering, electrical engineering, or an equivalent degree, with two to four years' experience designing, implementing, or otherwise working with cloud computing systems. More education, e.g., a Master's or Ph.D., would compensate for less work experience and vice versa.

29. A POSITA would have been familiar with common, then-existing technologies for cloud computing, including virtualization, as well as issues surrounding the design, implementation, and maintenance of such technologies. The state of the art below reflects certain technologies within the knowledge and skill of a POSITA.

B. State Of The Art

1. Virtual Appliances Were Commonly Used For Simple And Rapid Deployment Of Resources

30. Virtual appliances went hand in hand with virtual machines. A virtual appliance was known in the art as a virtual machine image containing a pre-installed operating system and one or more pre-installed software applications, such that the

software applications run when loaded in a virtual machine. As just one example, a 2012 patent publication explains this in its background section:

> A virtual appliance is a virtual machine image file that includes a pre-configured operating system and a single application. It is packaged as a pre-installed system image for a particular virtualized environment and hardware platform. A common file format for the virtual appliance is the open virtualization format (OVF). … Accordingly, the virtual appliance is a software image containing an application designed to run inside a virtual machine.

EX1005 (Kashyap), [0003]. As another example, a 2007 patent publication explained the benefits of building virtual appliances, such as failure containment, enhanced system security, and easier administration:

> Virtual machines (VMs) have attracted a great deal of attention lately for their potential to achieve better utilization of computer hardware while simultaneously isolating coarse-grained units of application software such as web servers, virus scanners, and email clients. *Isolating applications into individual VMs allows failure to be contained, enhances system security, and eases administration*. Virtual machines provide a means to build *virtual appliances*: complete applications, including customized Operating Systems (Oses), placed inside VM containers and administered as black boxes. Software

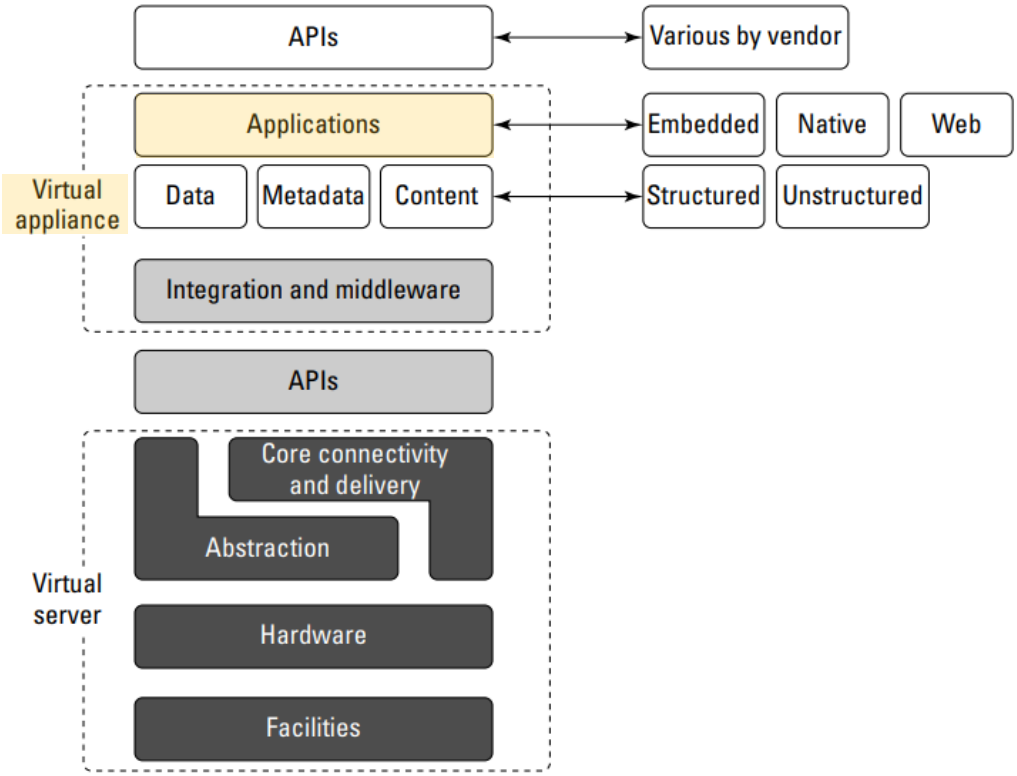
> developers may construct such VM-based appliances and ship completely customized software stacks to their customers.

EX1016 (Melor), [0002] (emphasis added). For other discussions reflecting how virtual appliances were understood in the art, see EX1009, [0003] and EX1008, [0003]-[0004].

31. Virtual appliances were already regarded as "a very important standard cloud computing deployment object" for cloud computing by 2011, as explained in Wiley Publishing's "Cloud Computing Bible" by Sosinsky. EX1007, 45; *id.*, a, e, xi. This is because virtual appliances provide the software applications comprising a cloud platform and thus "occupy the middle of the cloud computing stack" (*id*., 50-51), as shown in Sosinsky's Figure 3.2:

FIGURE 3.2

A virtual appliance is software that installs as middleware onto a virtual machine.



EX1007, 50-51 (highlighting added).

32. Sosinsky explained that "[o]ne of the major advantages of a virtual appliance is that you can use the appliances as the basis for assembling more complex services, the appliance being one of your standardized components." EX1007, 51. Virtual appliances thus enable a modular plug-and-play / "ready-to-use" approach for adding software functionality. EX1007, 52-55 (showing and discussing vendors providing virtual appliances).

33. POSITAs knew that "instantiating virtual appliances … that encapsulate the required applications or services (the *key functionality*) with their

support environment (e.g., an operating system and several software libraries)" was a standard technique for quickly and easily deploying software applications. EX1014 (Kecskemeti), 279 (emphasis original). Virtual appliances enable rapid provisioning because they "simplify and automate service deployment and eliminate the complexities in the service installation process." EX1010 (Jin), 5; *id*., 2 (explaining that virtual appliances simplify deployment and "eliminate the need for manual configuration and relieve the management burden"), 4-5 (discussing "complex and labor-intensive process" for manually deploying applications without virtual appliances); *see also* EX1007, 52 ("Virtual appliances are far easier to install and run than an application that you must set up yourself.").

34. It was well-known to use a bootstrap process, or boot process, to install and configure a virtual appliance automatically upon instantiation, so that a base virtual appliance can be deployed and then customized / configured for that specific instance. EX1006 (Dake), [0009]-[0010], [0018]-[0023]; *see also* EX1013 (Madhu), [0127]-[0129] (explaining that an initial "bootstrap process involves the automatic deployment, creation, and use of … virtual infrastructure," which begins when "a virtual machine [] is downloaded to an on-premise data center"); EX1016, [0019] (explaining how information provided by a virtual appliance is used to "instantiate[]" / "install" the virtual appliance on a virtual machine monitor /

hypervisor), [0027] (explaining that an OVA may be "responsible for its own configuration and modification").

2. Open Virtual Appliances (OVAs) Were A Known And Standard Format For Automatically Installing And Configuring Applications In Virtual Machines

35. Virtual appliances were such a standard tool for provisioning cloud resources that the Distributed Management Task Force (DMTF), an industry working group of over 160 companies and organizations, developed the Open Virtualization Format ("OVF") "for the packaging and distribution of virtual appliances composed of one or more virtual computer systems." EX1019, 3; EX1020, 8 ("OVF is a packaging format for virtual appliances."); EX1007, 243-44 (OVF is "a standard method for creating, packaging, and provisioning virtual appliances"). OVF was adopted as a national standard by ANSI in 2010.[1]

36. The resulting OVF package was referred to as an Open Virtual Appliance or "OVA." EX1021, 14 ("The file extension is .ova (open virtual appliance or application)"); *see also* EX1020, 14; EX1016 (Mellor), [0002]-[0003] (2007 patent describing an "Open Virtual Appliance (OVA)"); EX1017, [0001] (explaining that "OVA (open virtual appliance) template file is a single archive file

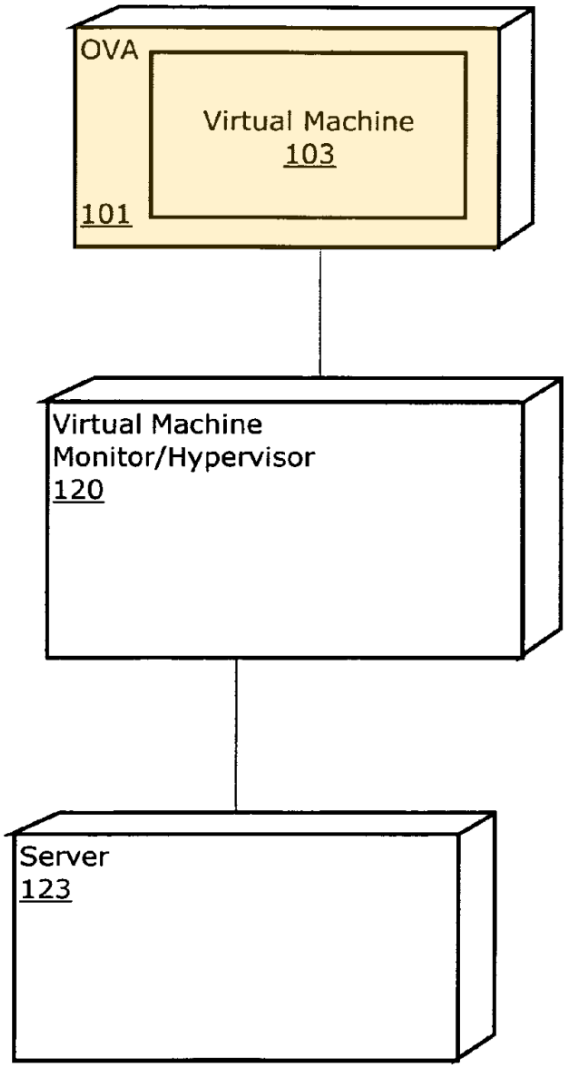
[1]See https://www.dmtf.org/news/pr/2010/8/dmtf%E2%80%99s-open-virtualization-format-achieves-ansi-adoption.

in TAR file format that contains files for portability and deployment of virtualization appliances," where the OVA template file includes "an OVF (open virtual format) descriptor file"). The '663 patent itself describes implementing the midserver with an "Open Virtual Appliance (OVA)," the file format for OVF, without any further detail or explanation. EX1001, 16:16-21.

37. OVAs are packaged with enough metadata "to automatically and securely install, configure, and run the virtual appliance on any virtualization platform." EX1019, 5; EX1020, 8. OVF thus provides "full support for *unattended installation* to ease the burden of deployment for users, and thereby enhance the user's experience." EX1019, 8 (emphasis added); EX1020, 10 (same).

38. A patent application from 2007 described an "Open Virtual Appliance (OVA)" as a "packaging format for virtual machines 103 (VMs) or virtual appliances." EX1016, [0019]. As shown in FIG. 1 below, a "virtual appliance may include one or more virtual machines 103," and the VM 103 "may include an application and an operating system." EX1016, [0019].

Figure 1



EX1016, FIG. 1. The OVA is "instantiated on any virtual platform," such as the virtual machine monitor (VMM) / hypervisor 120. Further, "[t]he VMM 120 allocates the server's 123 resources that the VM requires for operation." EX1016, [0019]. An OVA is provided as "an executable instruction set (e.g., an xml file), including VM specific information (i.e. metadata), for directing installation in a

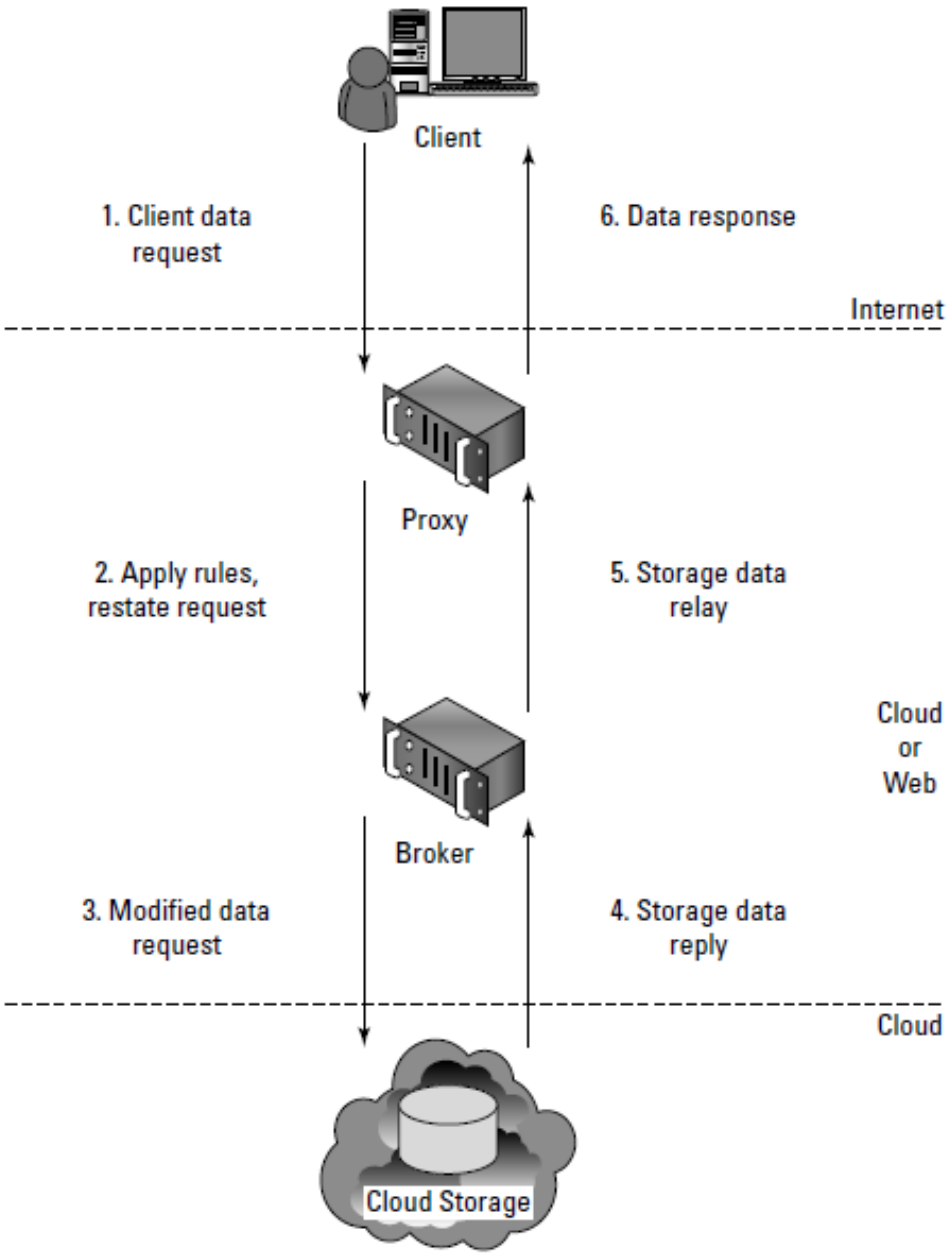
virtual environment of the VMs that the OVA encompasses." EX1016, [0021]. This early description is consistent with how the OVF standard describes packaging a virtual appliance. *See* EX1019, 9 ("The OVF descriptor is an XML document that describes meta-data about the software installed on the virtual disks."); *id.* ("The OVF environment is an XML document containing deployment time customization information for the guest software.").

39. The Open Virtualization Format (OVF) and Open Virtual Appliances (OVAs) continued to be used in the art for deploying / provisioning virtual appliances. For example, a 2017 patent application describes systems and methods for "deploying a virtual processing instance from an OVA template file." *E.g.*, EX1017, [0001]-[0005]. Similarly, a 2018 patent application describes "a virtual appliance that can be seamlessly provisioned" for different platforms and "distributed in a known format, such as the Open Virtualization Format (OVF)." EX1018, [0015].

3. Midservers Were Well-Known And Facilitated Data Transfer Between Local Networks And Cloud Platforms

40. The use of an intermediate system, whether it is termed a "midserver," "edge proxy," "cloud agent," or "cloud gateway," as an interface to facilitate data transfer from an enterprise network to a cloud system was a well-established practice. *E.g.*, EX1001, 11:28-32 (describing "midserver" as "a gateway to the

cloud-based service"); EX1004, [0048] (describing "on-premise cloud agent" as a "gateway that collects data items" and "move[s] the on-premise data to a cloud platform"); EX1007, 258 ("proxy"); EX1012, [0050] ("edge proxies"); EX1013, [0127] ("cloud gateways"); EX1015 (Antony), 6:15-19 (describing connection between "on-premise gateway" and "cloud-side gateway"); EX1022, [0016] ("cloud gateway"). As reflected in these references, POSITAs understood that using such an interface would simplify the connection between a local network and a cloud network (e.g., providing a single connection rather than individual connections to every device on the network), while ensuring security and other network services. For example, Sosinsky describes and shows a proxy, which "can be local or in the cloud," "in the direct data path between the client and data stored in the cloud" (EX1007, 257), as shown in its Figure 12.4:



EX1007, 258; *see also id.*, 257 (describing example where proxy referred to as a "Gatekeeper").

41. Such interfaces were commonly implemented, like other aspects of cloud systems, as a virtual appliance. For example, a 2017 published patent application described "two edge proxies connecting a remote 'cloud' network to a

local network that is typically on the premises of a particularly entity and termed an 'on-prem' network," where an "*on-prem edge proxy is a virtual appliance* that resides on premises (on-prem)," and "controls network traffic to and from a cloud edge proxy." EX1012 (Ben-Shaul), 3:61-67, 6:33-35 (emphasis added).

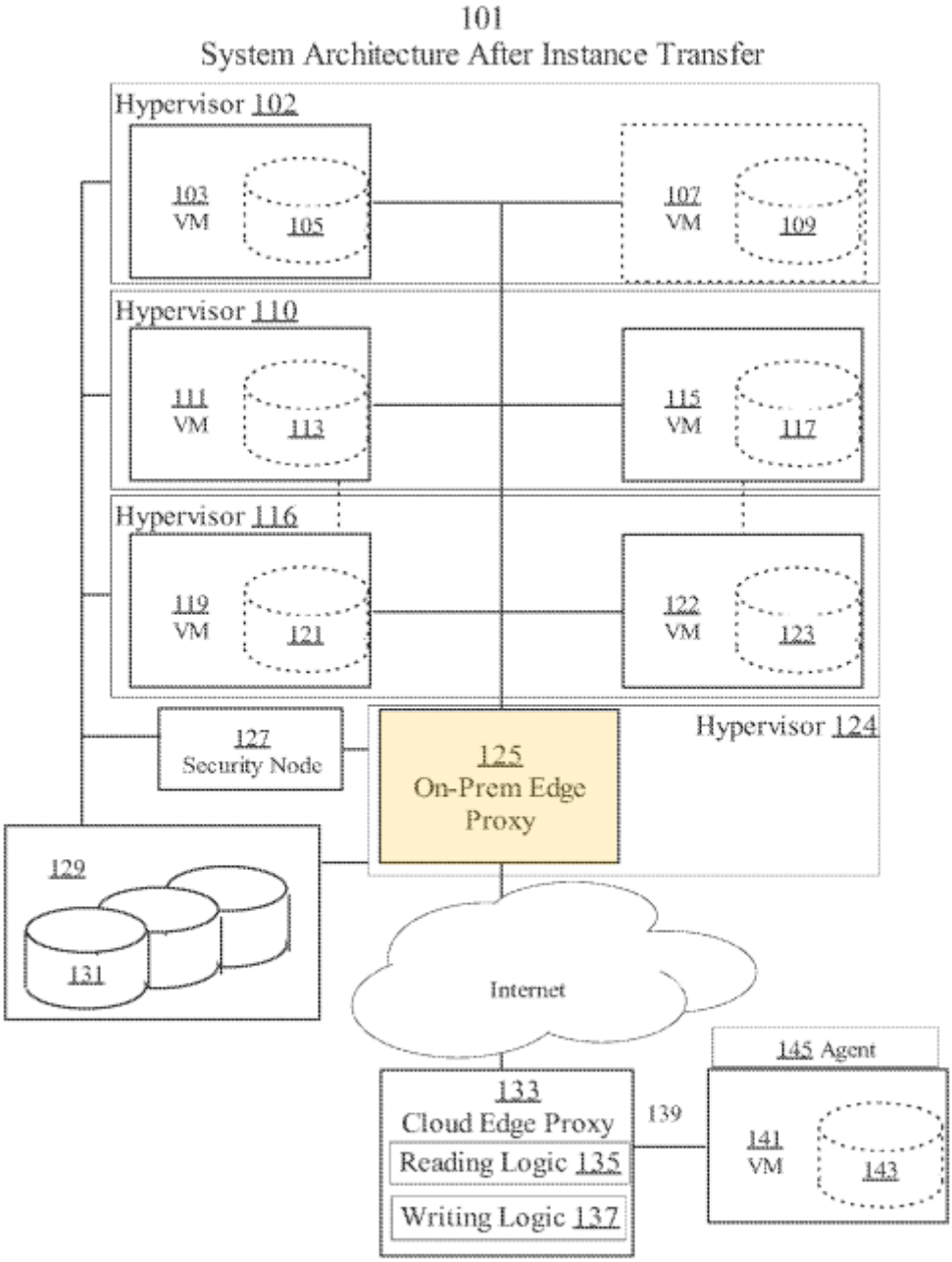


Fig. 1

EX1012, FIG. 1 (highlighting added).

42. Similarly, others described a "first cloud gateway in a private cloud" and a "second cloud gateway in a public cloud" would be connected with a secure

transport layer tunnel over a public network to enable secure network traffic between the private cloud and the public cloud (together forming a "hybrid cloud"). EX1022 (Chang), [0028]; *id.*, [0002] (describing hybrid clouds), [0019]-[0020] (same), [0033]; *see also* EX1015 (Antony), 4:8-20 (describing a gateway "executing as a virtual appliance" that provides components "with connectivity to an external network").

VII. THE '663 PATENT

A. The Claims' Earliest Effective Filing Date

43. For this declaration, I have been asked to assume that the claims' earliest possible effective filing date is May 14, 2019. That said, my opinions herein would not be materially different if the effective filing date was April 28, 2016.

B. The '663 Patent's Specification

44. The '663 patent describes using "midservers located between the business enterprise computer infrastructure and the cloud-based infrastructure to collect, aggregate, analyze, transform, and securely transmit data from a multitude of computing devices and peripherals at an external network to a cloud-based service." EX1001, Abstract.

45. The '663 patent describes "midserver" as "a server that functions as an interface between an external network and a cloud-based service," and performs standard operations such as "collecting, aggregating, analyzing, filtering,

transforming, and securely transmitting data." EX1001, 12:66-13:5. The "midserver deployment, automated onboarding of data, and endpoint transparent data transport" is disclosed only in terms of conventional applications of known technologies such as Open Virtual Appliance (OVA) and virtual private network (VPN), as well as known commercial and open-source software such as ZeroTier, Ansible, and Nginx. EX1001, 16:11-60; *id.*, 9:51-52 (describing using "standard VPN or similar encryption-based network transport technologies"). I note that ZeroTier software provides VPN connections between devices, Ansible software enables software provisioning and configuration management, and Nginx is open-source web server technology that can be used as a reverse proxy and load balancer, among other things. The Open Virtual Appliance (OVA) is just a virtual appliance using the standard Open Virtualization Format (OVF), packaged as a single file. *See above* Section VI.B.2.

46. The '663 patent describes the midserver as located "on-site" with the devices or endpoints from which it collects data (as shown in FIG. 1, reproduced below).

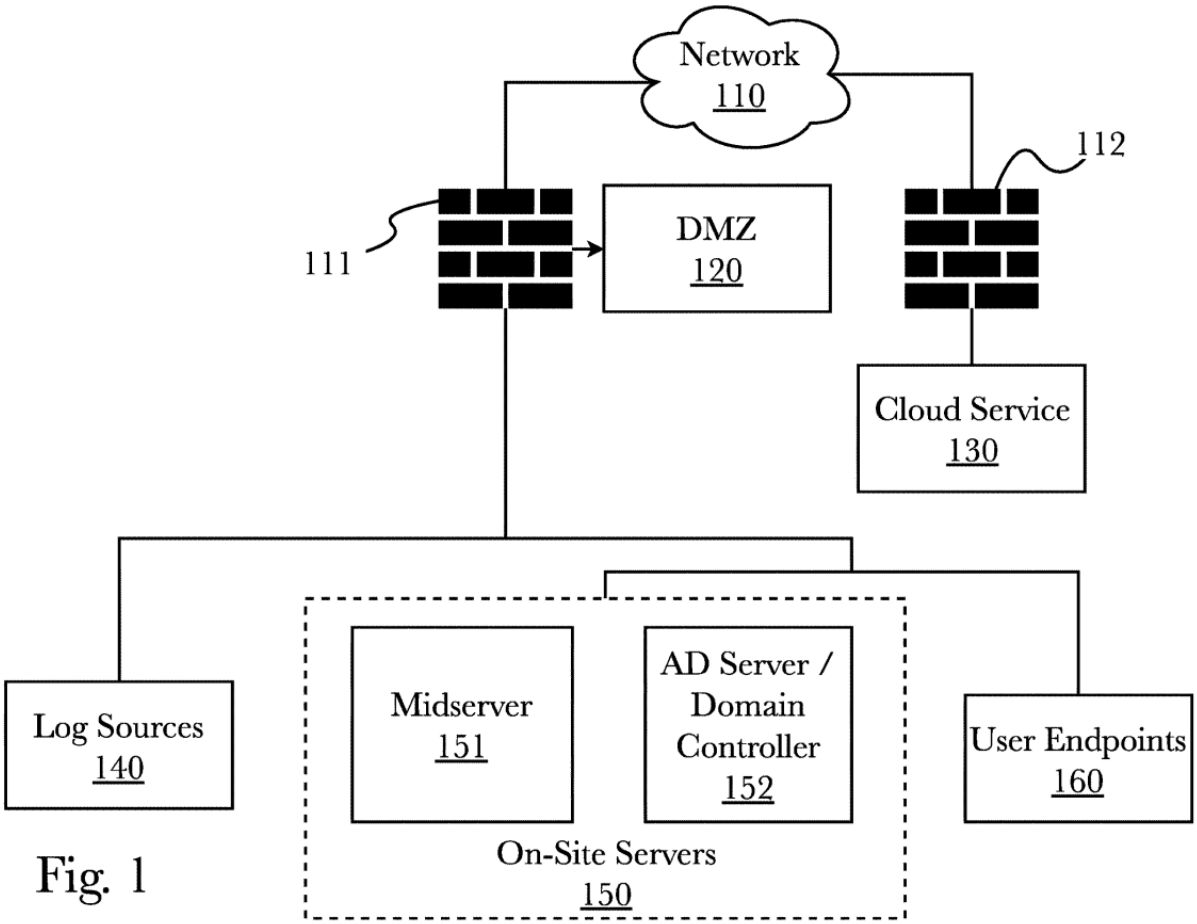


Fig. 1

EX1001, FIG. 1; *id.*, 13:7-43.

VIII. CLAIM CONSTRUCTION

47. For this declaration, I apply the plain and ordinary meaning of all claim terms, and do not rely on any specific constructions. I reserve the right to offer my opinions on any constructions or construction issues presented by Patent Owner.

IX. GROUND 1: CLAIMS 1 AND 4 ARE OBVIOUS OVER MATURANA IN VIEW OF KASHYAP

48. As explained below, claims 1 and 4 are obvious over Maturana in view of Kashyap.

A. Maturana (EX1004)

49. Maturana is a U.S. patent application publication that published October 1, 2015, from an application filed October 27, 2014. EX1004, Cover. Maturana is assigned to Rockwell Automation Technologies, Inc.

50. Maturana discloses "an on-premise cloud agent that facilitates collection and upload of industrial data to a cloud-based storage and processing infrastructure." EX1004, [0002]. As shown in FIG. 1 reproduced below, the "agent-based cloud infrastructure" of Maturana "enables remote monitoring and reporting of on-premise assets by implementing six general areas of functionality—data ingestion into the cloud, data prioritization, object modeling, data processing, data analytics, and reporting." EX1004, [0046].

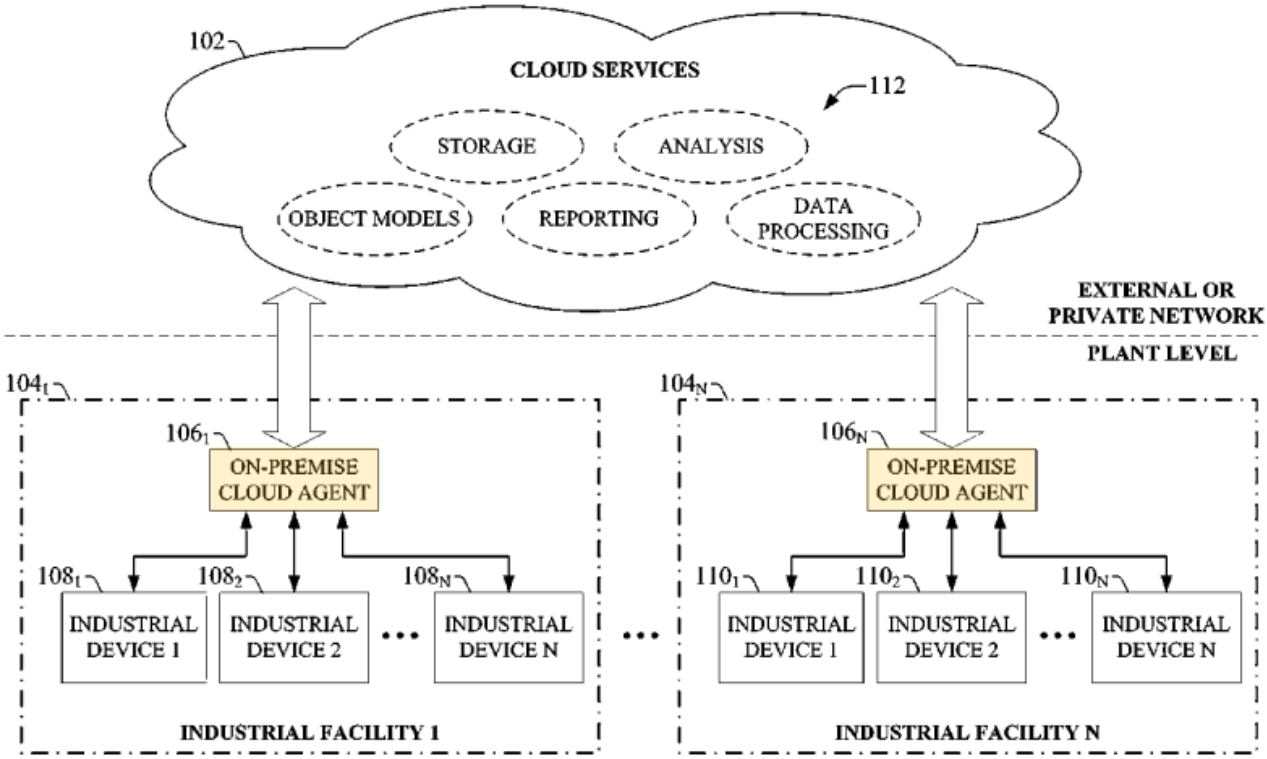


FIG. 1

EX1004, FIG. 1 (highlighting added). In this figure, cloud agents are installed and running at several locations, with each cloud agent collecting data from many different devices.

51. Maturana explains that "[i]ngestion of industrial device data in the cloud platform 102 through the use of cloud agents 106 can offer a number of advantages particular to industrial automation." EX1004, [0041]. These advantages include, as some of the examples provided by Maturana, (1) the easy scalability of cloud-based storage "to accommodate the large quantities of data generated daily by an industrial enterprise," (2) the ability for "multiple industrial facilities at different geographical locations [to] migrate their respective automation data to the cloud …

without the need to establish a private network between the facilities," and (3) the ability for cloud agents "to automatically detect and communicate with the cloud platform 102 upon installation at any facility, simplifying integration with existing cloud-based data storage, analysis, or reporting applications used by the enterprise." EX1004, [0041]. Maturana further explains that the cloud platform enables "software vendors to provide software as a service, removing the burden of software maintenance, upgrading, and backup from their customers." EX1004, [0041].

52. The cloud agent of Maturana, as shown in FIG. 2 reproduced below, includes various components such as a collection services component, a queue processing component, a subscription management component, a user interface component, one or more processors, and memory. EX1004, [0042]-[0045]. Maturana explains that the cloud agent can be implemented using "virtual machine(s)." EX1004, [0042].

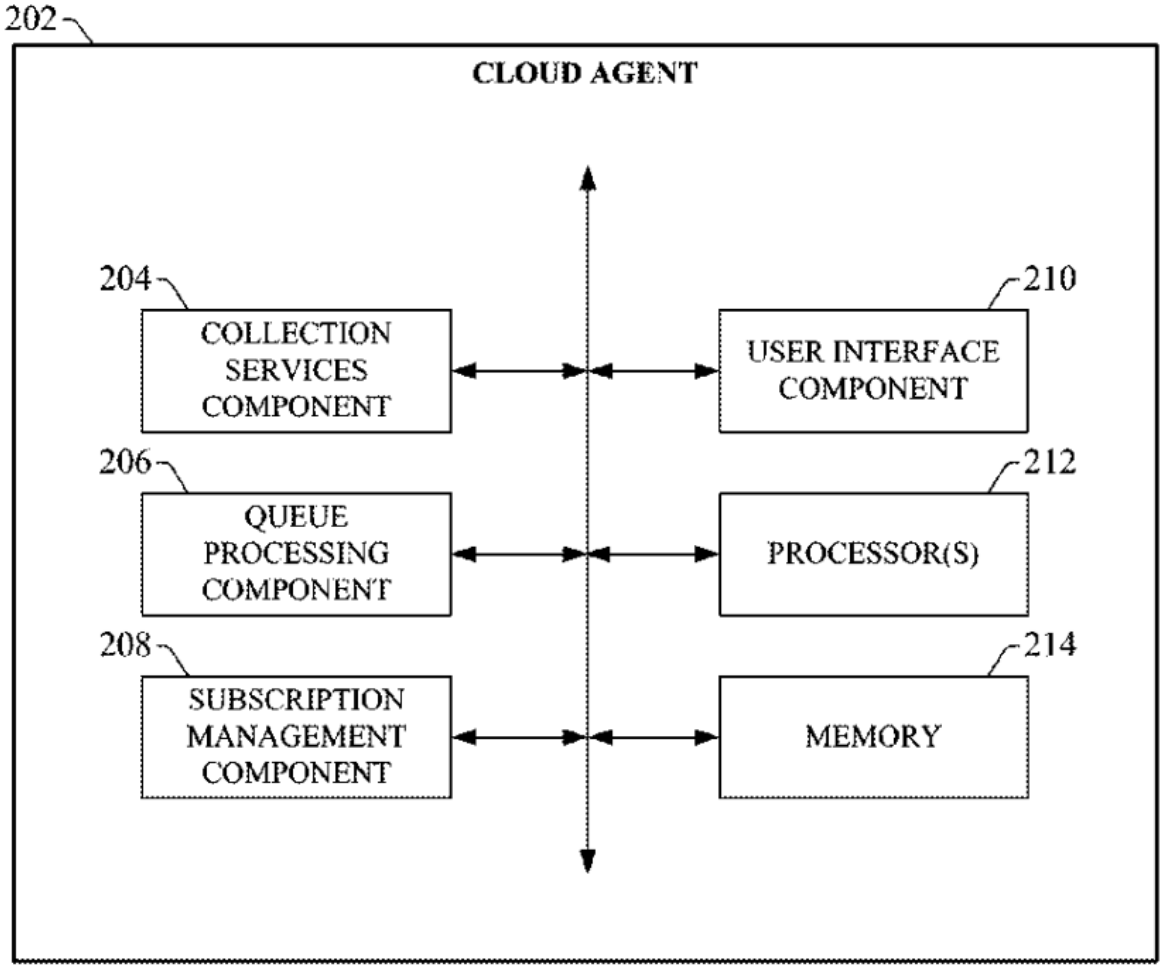


FIG. 2

EX1004, FIG. 2.

53. The collection services component is configured to "collect industrial device data, either from a data concentrator or directly from the devices themselves." EX1004, [0044]. The queue processing component is configured to "package a compressed data file prepared by the collection services component 204 into a data packet and push the data packet to the cloud platform." EX1004, [0044]. The subscription management component is "configured to maintain customer-specific

configuration and subscription information," which is used "to determine how the compressed data file should be packaged, and how to connect to the customer's cloud platform for transmission of the data packets." EX1004, [0044].

B. <u>Kashyap (EX1005)</u>

54. Kashyap is a U.S. patent application publication published September 27, 2012, from an application filed March 24, 2011. EX1005, Cover. Kashyap is assigned to International Business Machines Corporation.

55. The title of Kashyap is "CONFIGURATION OF VIRTUAL APPLIANCES," and Kashyap relates to "configuration of virtual appliances in a virtual machine." EX1005, [0001]; *see also id.*, [0004]. Kashyap describes a "packaging and distribution format" that supports installation and configuration of virtual appliances on virtual machines. *Id.*, [0005], Abstract; *see also id.*, [0021]. Kashyap explains that "[a] virtual appliance is a virtual machine image file that includes a pre-configured operating system and a single application." EX1005, [0003] ("[T]he virtual appliance is a software image containing an application designed to run inside a virtual machine."). Kashyap further teaches that virtual appliances allow users to carry out specific functions on a virtual machine: "A software virtual appliance is one or more virtual machines that collectively provide a specific function for a user through the services of a virtualization platform." EX1005, [0021].

56. Kashyap also describes the mechanics of installing a virtual appliance on a virtual machine: "Deploying a new virtual appliance requires copying the appliance image, including the operating system and application, to a pre-installed and pre-configured virtual machine." EX1005, [0003]. In this way, "the virtual appliance is a software image containing an application designed to run inside a virtual machine." *Id*.

57. Kashyap further teaches that "[a] common file format for the virtual appliance is the open virtualization format (OVF)." EX1005, [0003]. It also notes known benefits of OVF, including that it is "platform independent," allows users to deploy "on a virtualization platform of their choice," provides a standardized "description and format of virtual image appliances," and provides features that "allow for additional configuration parameters specific to the particular product." *Id.*, [0021].

58. Kashyap further teaches a virtual appliance packaging and distribution format that enables efficient "post-install configuration of virtual appliances in a virtual machine." EX1005, Abstract; *id.*, [0001], [0005]. This format involves installation of both an activation script, for "configuration of the virtual machine on boot" and an associated configuration script, for "configuration of an appliance application." *Id.*, [0020].

59. Kashyap's FIG. 3 (reproduced below) illustrates various aspects of a system that supports "post-install configuration of one or more virtual appliances with use of OVF." EX1005, [0032] (noting the use of "open virtualization format (OVF)" as a "packaging and distribution format"); *see also id.*, [0012].

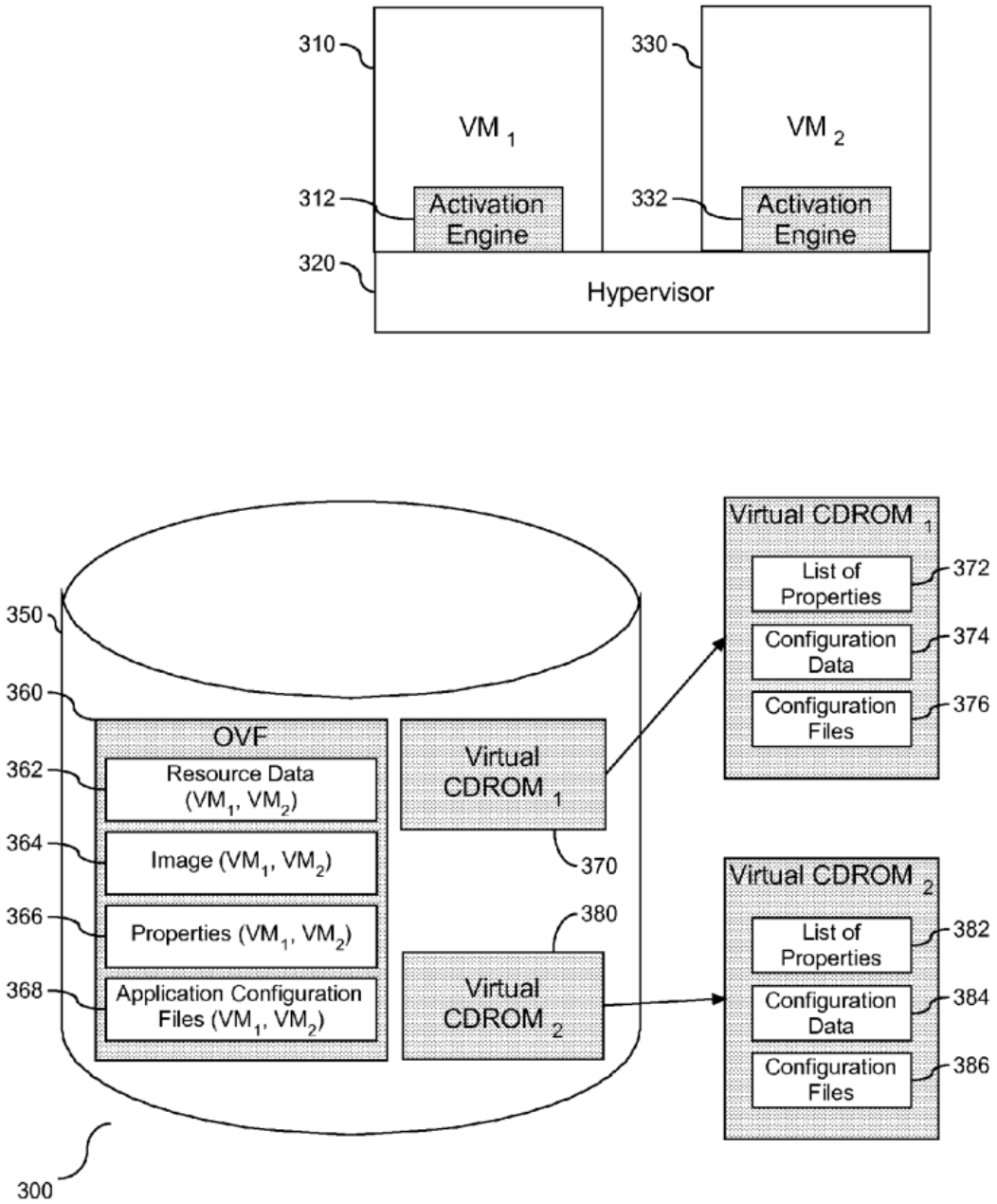


FIG. 3

EX1005, FIG. 3.

C. Combining Maturana And Kashyap

60. A POSITA implementing an agent-based cloud infrastructure as taught by Maturana would have naturally implemented the cloud agent as a virtual appliance with automatic post-installation configuration, as taught by Kashyap, to achieve a standardized, easily deployable service. The POSITA would have combined these teachings into an improved design (hereinafter "Maturana-Kashyap") that included benefits from each reference.

61. So implemented, the Maturana-Kashyap combination includes the following combined teachings:

- A server storing and executing a cloud agent, as taught by Maturana (EX1004, [0048]), configured to:
    - Automatically install the cloud agent as a virtual appliance software application, which automatically loads configuration data while booting, as taught by Kashyap (EX1005, [0003]-[0007], [0020]-[0022]);
    - Securely connect to a cloud platform via an external network, as taught by Maturana (EX1004, [0039], [0057], [0089], FIG. 1);
    - Collect data from a plurality of industrial devices via a local network, as taught by Maturana (EX1004, [0040], [0048]);

- Perform pre-processing and data refinement of the collected data, including aggregating data from multiple sources into a single cloud upload packet, as taught by Maturana (EX1004, [0054]-[0056]); and
- Send the data packet to the cloud platform over the secure connection, as taught by Maturana (EX1004, [0076]).

62. A POSITA would have been motivated, with a reasonable expectation of success, to combine the teachings of Maturana and Kashyap for at least the following reasons.

63. *First*, Maturana discloses cloud agents implemented with "virtual machine(s)" where execution of one or more functional components can "cause the machine(s) to perform the operations described." EX1004, [0042]. Maturana further describes implementing a cloud agent as "an application running on a server." EX1004, [0026]. Maturana expressly discloses deploying a persistent on-premise cloud agent that interfaces with multiple local data sources and communicates with a cloud platform, but Maturana provides little implementation detail on installation or configuration. *E.g.*, *id.*, [0041] (describing how the cloud agent operates and is configured "upon installation at any facility"); *id.*, [0043]. Kashyap addresses this exact problem by teaching how applications are packaged as virtual appliances and automatically installed and configured upon instantiation in a virtual machine

environment, thus providing detailed teachings on an efficient and reproducible approach for installation of software applications on virtual machines. EX1005, [0003] (describing a "virtual appliance"), [0020]-[0023] (describing implementing automated "application installation procedures" through activation and configuration scripts). As disclosed in Kashyap and known generally by POSITAs, a virtual appliance comprises "an application designed to run inside a virtual machine." EX1005, [0003] (describing how to deploy a virtual appliance); *supra* Sections VI.B.1 (explaining that virtual appliances were well known in the art for deploying applications), VI.B.2 (explaining that virtual appliances were known for providing gateway/interface functionality).

64. Kashyap teaches that OVF was "a packaging and distribution format for a virtual machine" and "standardizes the description and format of virtual image appliances." EX1005, [0021]. Within this standardized format, it teaches how to install and configure a virtual machine and associated software application using "activation and configuration scripts." *Id.*, [0020]. Using this flexible approach, "application installation procedures remain unchanged while allowing for changes and customization of the configuration script independent of the activation script." *Id.*; *id.*, [0005]-[0007], [0022], [0028]-[0029]. This approach allows a base virtual appliance to be deployed where needed and separately configured for specific instances.

65. POSITAs would have thus been motivated to implement Kashyap's detailed teachings on a flexible OVF-based framework for "application installation procedures" for the installation and configuration of cloud agents within the Maturana system. In view of the virtual-machine execution and "installation at any facility" taught by Maturana (EX1004, [0041]), a POSITA would have been motivated to implement the cloud agent / "application running on a server" (EX1004, [0026]) with a virtual appliance as taught by Kashyap to achieve the benefits of repeatable, site-by-site deployment. *E.g.*, EX1004, FIG. 1 (showing "cloud agent[s]" at each site), [0075] (describing the cloud agent being "moved to a different computing device," indicating portability of the cloud agent); *see* EX1005, [0003] (describing a virtual appliance as a packaged application ready for deployment); EX1020, 7-8 (describing the industry standard for packaging and distributing virtual appliances).

66. A POSITA would have understood that implementing Maturana's cloud agent as a virtual appliance, as taught by Kashyap, would predictably improve reliability, repeatability, and ease of installation across different environments. Such benefits were well-recognized in cloud and enterprise systems prior to the '663 patent's priority date, particularly for intermediary services that must be deployed on-premise but maintained and updated consistently across customer sites. *Supra* Sections VI.B.1 (describing the benefits of virtual appliances), VI.B.3 (describing

the known technique of on-premise intermediaries implemented as virtual appliances).

67. *Second*, the teachings of Maturana and Kashyap were complementary for the following reasons. The cloud agent of Maturana is "configured to automatically detect and communicate with the cloud platform 102 upon installation at any facility, simplifying integration with existing cloud-based data storage, analysis, or reporting applications used by the enterprise." EX1004, [0041]-[0042]. This requirement of Maturana is naturally complemented by Kashyap's teachings on an OVF-based virtual appliance that is easily installed on virtual machines, while allowing for easy configuration of a particular appliance. For example, Kashyap teaches how to complete installation and configuration of the application in a virtual appliance via activation and configuration scripts. EX1005, [0003], [0005]-[0006], [0020]-[0023]. These teachings of Kashyap naturally complement Maturana's teaching that a cloud agent should "automatically detect and communicate" with the cloud platform "upon installation at any facility," because they allow an application such as the cloud agent to be deployed to a server with automatic installation and configuration. EX1004, [0041]; *see also* EX1020, 10 (explaining that the standard format for a virtual appliance provides "full support for unattended installation to ease the burden of deployment for users").

68. Further, Maturana teaches "several cloud agents distributed throughout the customer's industrial enterprise" and recognizes that data volume "increases as additional plant facilities are added." EX1004, [0032], [0074]. Kashyap recognizes the need for "post-install configuration" of virtual appliances decoupled from the virtual appliance itself so that deployed virtual appliances can be customized without a proliferation of customized appliances, and provides a known solution where the configuration script and data are separate from the virtual appliance and loaded upon booting the virtual appliance. EX1005, [0001], [0005]-[0007], [0022], [0027]-[0028], [0031]. A POSITA implementing Maturana would naturally leverage the boot-time configuration taught by Kashyap to easily deploy the same base virtual appliance to different locations while "allowing for changes and customization independent of the" virtual appliance to be applied upon booting. EX1005, [0022].

69. Maturana further discloses "customer-specific configuration and subscription information" that controls packaging and transmission, and registry-stored configurations for secure connectivity. EX1004, [0044], [0052]-[0053], [0069]-[0070], [0074]-[0075]. Kashyap provides the mechanism to provide and apply such parameters automatically, where a "virtual transport mechanism external to the" virtual appliance supplies a plurality of properties, configuration data, and configuration files that the activation startup script and configuration script apply to the virtual appliance at boot. EX1005, [0031]; *id.*, [0020]-[0023], [0027], [0034],

FIG. 3. The technique for post-install configuration of the virtual appliance taught by Kashyap thus enables the specific configuration data taught by Maturana to be automatically provided to the cloud agent upon installation, thereby allowing the cloud agent to "automatically detect and communicate with the cloud platform 102 upon installation at any facility." EX1004, [0041]; *id.*, [0020], [0022] ("Applications included within virtual appliances may require post install configuration."). The teachings in Maturana and Kashyap therefore complemented each other, as the automatic installation and configuration techniques taught by Kashyap provide implementation details for achieving the express goals of automatic connection and simplified integration taught by Maturana. EX1004, [0041].

70. *Third*, Kashyap teaches using the "open virtualization format (OVF)" for packaging and deploying virtual appliances and describes how to implement its post-install configuration technique using OVF. EX1005, [0003], [0021]-[0022], [0026]-[0029], [0032]-[0034]. A POSITA would have been motivated to use the OVF-based virtual appliance technique taught by Kashyap to implement the cloud agent of Maturana because OVF was a widely adopted international standard for packaging and distributing "virtual appliances and the metadata required to *automatically and securely install, configure, and run them* on any virtualization platform." EX1020, 8 (emphasis added); *id.*, 7.

71. *Fourth*, the teachings of Maturana and Kashyap were compatible. The cloud agent of Maturana is expressly executable on "virtual machine(s)" (EX1004, [0042]) and operate as an "application running on a server" (EX1004, [0026]). Kashyap expressly describes deploying a virtual appliance onto a virtual machine and completing installation and configuration at boot. EX1005, [0003], [0005]-[0007]. The execution and deployment of a software application is consistent between both Maturana and Kashyap.

72. Maturana contemplates that a cloud agent may be "moved to a different computing device," with registry checks denying communication if the cloud agent is unregistered on the different device. EX1004, [0075]. Maturana thus teaches enforcing cloud agent identity per deployment to ensure secure communication. EX1004, [0074]-[0076]. Kashyap's teachings, where a virtual appliance is deployed and then specifically configured, are naturally compatible with the agent registry technique of Maturana because the same virtual appliance can be deployed across different locations, while per-site configuration data is loaded at boot as taught by Kashyap. EX1005, [0020]-[0023].

73. A POSITA implementing Maturana thus would have naturally turned to Kashyap for implementation details on how to deploy and customize the cloud agent for a specific site, with the additional advantage of not requiring storage and maintenance of a customized image / virtual appliance for every site in the cloud.

EX1005, [0020] (explaining that "application installation procedures remain unchanged while allowing for changes and customization of the configuration script"); *see, e.g.*, EX1006, [0009] (describing the impracticality of maintaining a plurality of customized images); EX1007, 52 (explaining that the file size of a virtual appliance may be at least ten times the size of an application alone). This motivation is further reinforced by industry practice recognizing virtual appliances as standard deployment vehicles for cloud-connected services, as they were understood to enable rapid provisioning, automated configuration, and scalable deployment across enterprise infrastructures. *Supra* Section VI.B.1 (citing and quoting EX1007, 45; EX1010, 2, 5; EX1014, 279).

74. *Finally*, a POSITA would have reasonably expected to succeed in implementing the Maturana-Kashyap combination because doing so would only entail implementing the respective complementary and compatible teachings of Maturana and Kashyap, as discussed in the preceding paragraphs. At most, a POSITA implementing Maturana-Kashyap would provide software code to implement the respective teachings of Maturana and Kashyap, which was well within the skill of a POSITA. Further, Kashyap provides specific teachings on using the "open virtualization format (OVF)" to deploy virtual appliances in accordance with its teachings (EX1005, [0006], [0021]-[0023]), and a POSITA would have reasonably expected to succeed implementing the cloud agent in this way because

OVF was a "standard for packaging such virtual appliances and the metadata required to automatically and securely install, configure, and run them on any virtualization platform." EX1020, 8; *id.*, 7. Further still, similar functionality has long been known in the art of cloud computing, as described above in the state of the art, and a POSITA would have been comfortable applying known and standard techniques. *Supra* Section VI.B.

75. So implemented, the Maturana-Kashyap combination includes a self-configuring virtual appliance-based cloud agent that collects data from a plurality of devices, processes the collected data to prepare it for transmission, and transmits the data over a secure connection to a cloud platform, and thus renders all challenged claims obvious.

D. Claim 1

1. Element [1.1]

[1.1] A system for ingestion of data into a cloud-based service from an external network, comprising:

76. To the extent the preamble is limiting, Maturana-Kashyap satisfies it. Maturana discloses "an on-premise agent architecture for collection of industrial data from plant-floor devices and transfer of the data to a cloud repository for processing and storage." EX1004, [0034]. Maturana expressly describes this collection and transfer of data as "*ingesting* industrial *data to a cloud platform* using

a cloud agent architecture." EX1004, [0019] (emphasis added). As shown in FIG. 1, Maturana teaches that the ingestion of data into the cloud platform ("cloud services") is done from a network external to the cloud services:

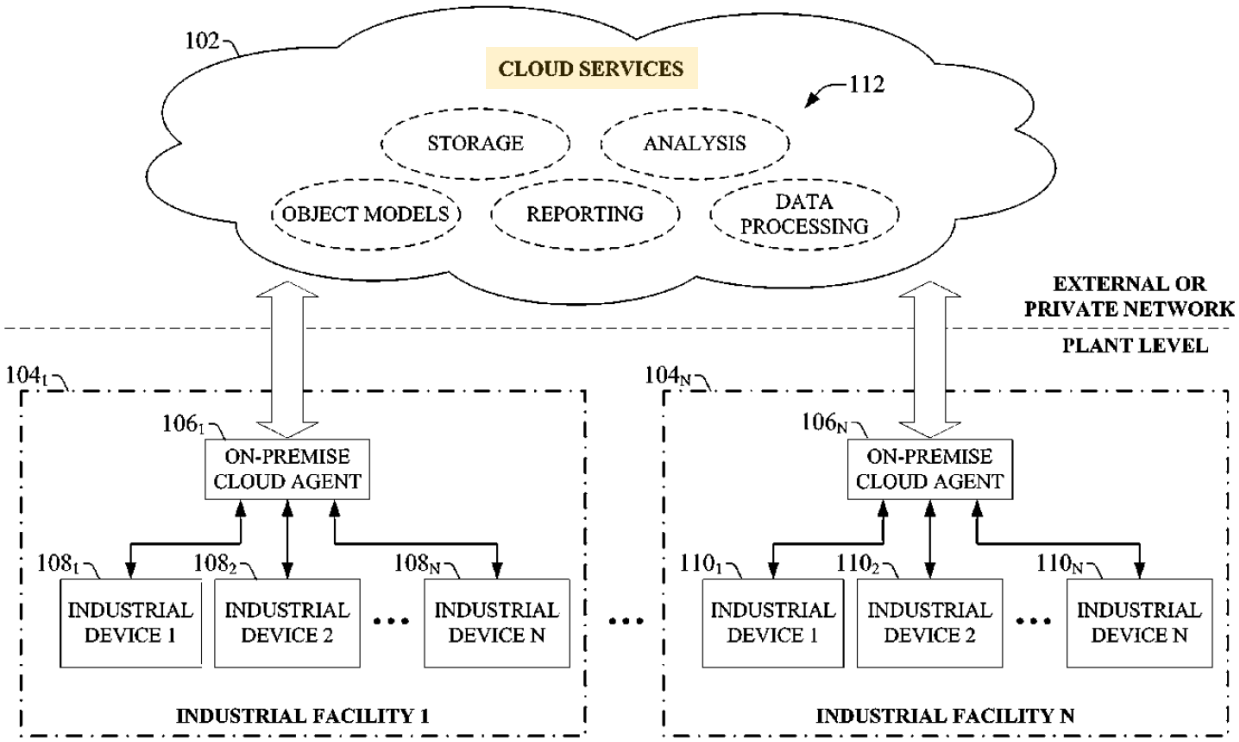


FIG. 1

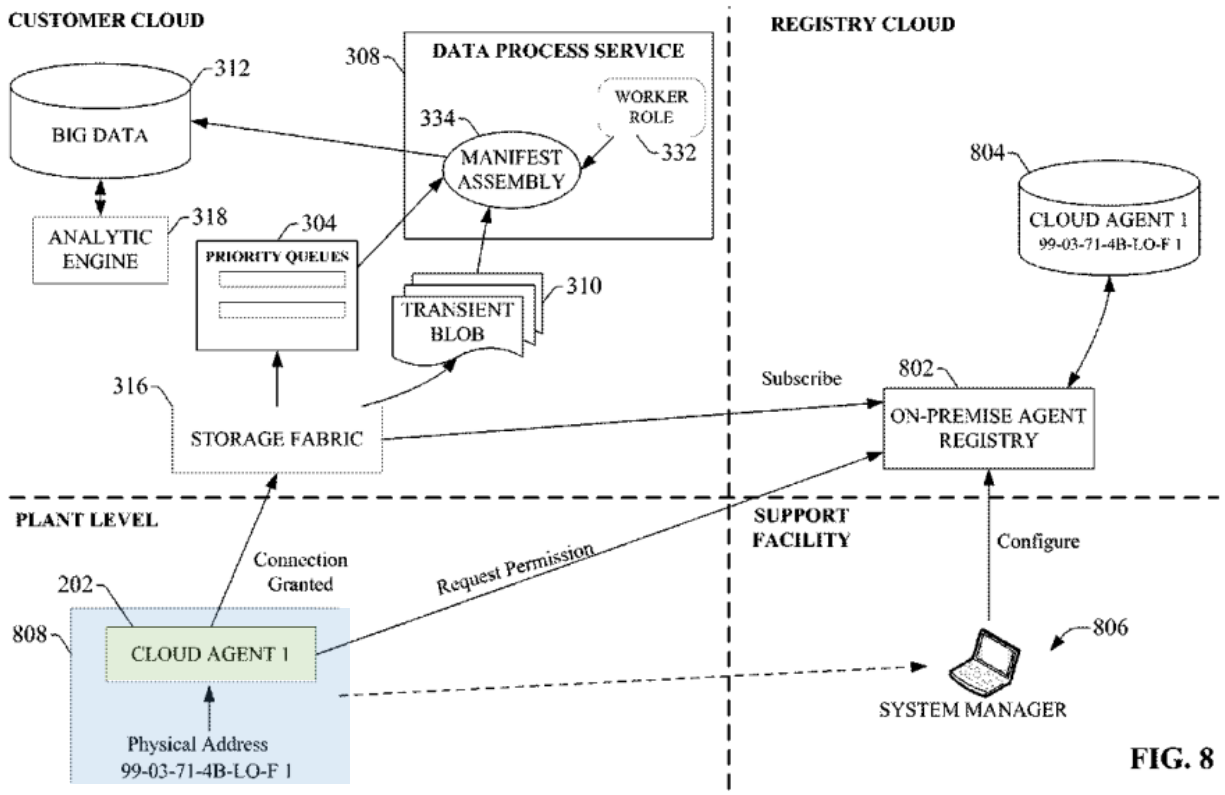
EX1004, FIG. 1 (highlighting added). In this figure, Maturana labels the cloud services as "external" to the network where the on-premise cloud agents reside and from which data originates (i.e., at the "plant level"). From the perspective of the cloud services, the cloud agents at the "plant level" are at an *external network*. *See* EX1004, [0037]; *id.*, [0033]-[0034] (discussing networks on which applications "collect and generate data" and how the cloud agents transfer "the data to a cloud repository for processing and storage"), [0039]-[0041].

2. Element [1.2]

[1.2] a midserver comprising at least a processor, a memory, and a plurality of programming instructions stored in the memory and operating on the processor, wherein the plurality of programming instructions, when operating on the processor, cause the processor to:

77. Maturana-Kashyap satisfies this element. The server / hardware platform executing the cloud agent taught by Maturana is a *midserver* because it "facilitates collection and upload of industrial data to a cloud-based storage and processing infrastructure" (EX1004, [0002]), just as the midserver of the '663 patent is used to "collect, aggregate, analyze, transform, and securely transmit data from a multitude of computing devices and peripherals at an external network to a cloud-based service" (EX1001, 7:60-66). *See* EX1004, [0048] (explaining that "[c]loud agent 202 can execute on any suitable hardware platform (e.g., a *server*, a LINUX box, etc.)"). The '663 patent describes a midserver as "an interface between an external network and a cloud-based service" (EX1001, 8:5-7, 8:17-19), and the server hosting the cloud agent of Maturana meets this description, as it "acts as a generic gateway that collects data items from the various industrial assets on the plant network and packages the collected data according to a generic, uniform data packaging schema used to move the on-premise data to a cloud platform 302." EX1004, [0048].

78. Maturana depicts the cloud agent 202 as executing on a hardware server 808 in FIG. 8:



EX1004, FIG. 8 (highlighting added); *id.*, [0074].

79. Within this hardware host, as depicted in FIG. 2 of Maturana below, the cloud agent 202 includes "one or more processors 212, and memory 214" (*at least a processor, a memory*), and various functional components that "comprise software instructions stored on memory 214 and executed by processor(s) 212" (*a plurality of programming instructions stored in the memory and operating on the processor*). EX1004, [0043] (color added).

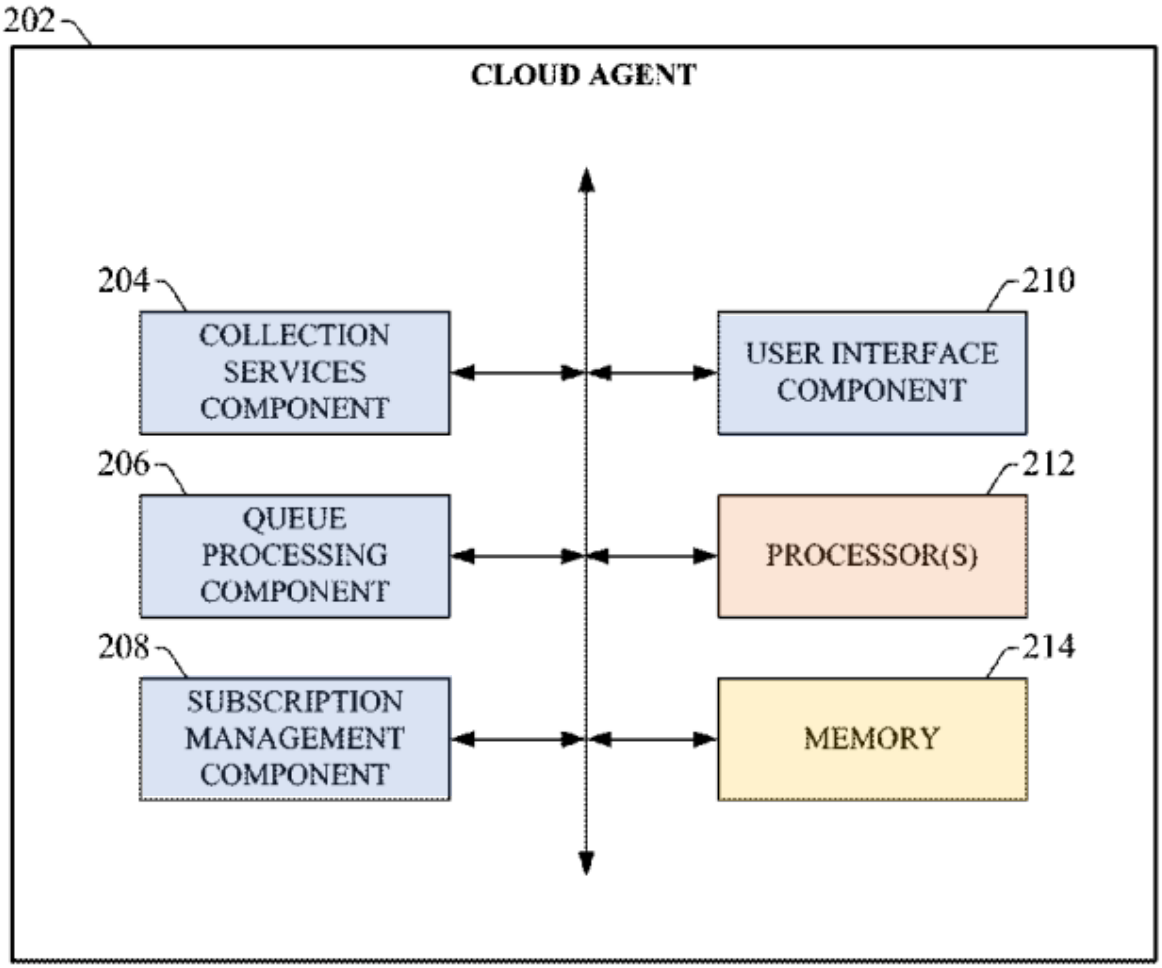


FIG. 2

EX1004, FIG. 2 (highlighting added).

80. A POSITA would have understood the underlying gateway computer (e.g., server) with at least one processor and memory as a *midserver*, and the cloud-agent components stored as a virtual appliance in that memory and executed by the processor to perform the gateway / midserver functions (collection, transformation, and transmission of data to the cloud platform) as *a plurality of programming instructions*. *Supra* Section IX.C (explaining that a POSITA would have been

motivated to implement the cloud agent with a virtual appliance). As explained in the following sections, in the Maturana-Kashyap system these instructions cause the midserver to perform the recited functions.

3. Element [1.3]

[1.3] automatically install a virtual appliance software application, the virtual appliance software application configured to automatically load a plurality of stored configurations on the midserver;

81. Maturana-Kashyap satisfies this element. As discussed *supra* Section IX.C, a POSITA implementing Maturana's teachings of a cloud agent comprising "an application running on a server" and executed with "virtual machine(s)" (EX1004, [0026], [0042]) would have been motivated to use Kashyap's teachings regarding a standardized way to install and configure "an application designed to run inside a virtual machine," namely via a virtual appliance (EX1005, [0003]). EX1005, [0020]-[0021]. The virtual appliance is deployed to the server of Maturana-Kashyap via OVF, as taught by Kashyap, which involves "copying the appliance image, including the operating system and *the application*, to a pre-installed and pre-configured virtual machine," such as the virtual machine on the server disclosed by Maturana. EX1005, [0003] (emphasis added); *id.*, [0021]; EX1004, [0042]. The cloud agent of Maturana-Kashyap is therefore a *virtual appliance software application* because it is the software application provided within the virtual

appliance, as taught by Kashyap, and executed with a virtual machine, as taught by Maturana. *Supra* Section IX.C.

82. **Regarding "automatically install a virtual appliance software application,"** the server (*midserver*) automatically installs the cloud agent (*virtual appliance software application*) of Maturana-Kashyap when the virtual appliance is deployed to and executed by the server. EX1020, 8 (explaining that a virtual appliance distributed via OVF is "automatically and securely install[ed]").

83. Kashyap discloses "application installation procedures" that *automatically install the virtual appliance software application* when the virtual appliance boots. EX1005, [0020]. Specifically, the "application included within a virtual appliance requires installation of activation and configuration scripts" to complete installation of the application. EX1005, [0020]; *id.* [0025]. The activation script, provided with the virtual appliance and "instantiated during boot operation" of the virtual appliance, configures the virtual machine when the virtual appliance boots, while the configuration script, provided separately from the virtual appliance and called by the activation script during boot, configures the application itself. EX1005, [0005]; *id.*, [0006]-[0007], [0020]-[0034]. A POSITA would have understood that these scripts are executed *automatically* because they are called "during boot operation," which is the automatic process that initially occurs when the server executes the virtual appliance to set up the virtual appliance. EX1005,

[0005]-[0007]; *id.*, [0020] ("The installation of the activation script pertains to configuration of the virtual machine on boot."), [0023]-[0025]. This approach taught by Kashyap is consistent with how POSITAs have long understood the "activation" of virtual appliances, which "uses the metadata and scripts in the virtual appliance package created to customize and activate a VM appliance for a new environment." EX1011, 620.

84. This process of activating the virtual appliance is depicted in FIG. 1 of Kashyap:

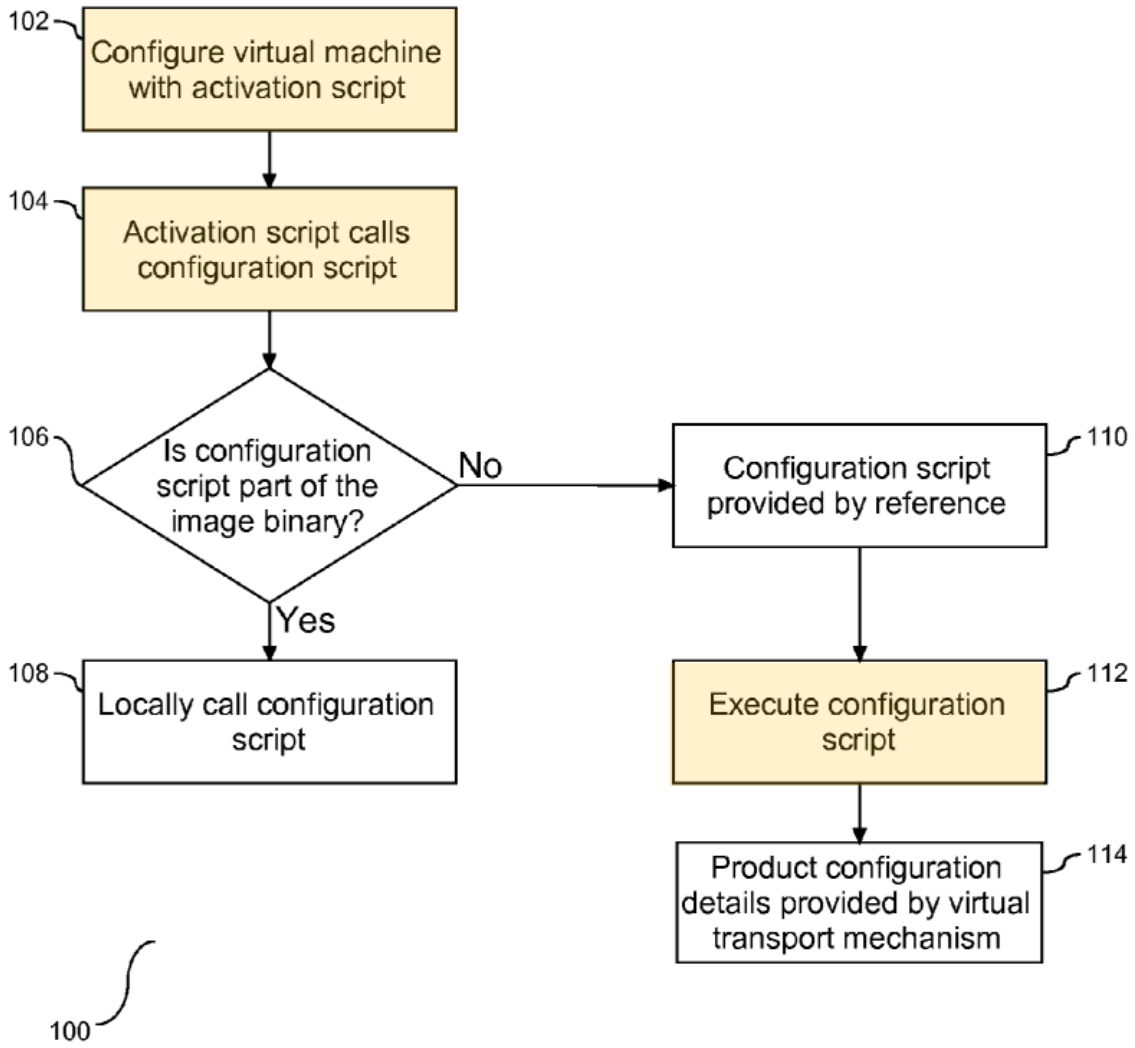


FIG. 1

EX1005, FIG. 1 (highlighting added). Together, these automatically executed scripts *automatically install the virtual appliance software application* / cloud agent.

85. **Regarding "the virtual appliance software application configured to automatically load a plurality of stored configurations on the midserver,"** Kashyap discloses automatically loading a plurality of stored configurations. Kashyap explains that "[a]pplications included within virtual appliances may require

post install configuration," and that "[t]o enable the customization, the configuration framework must not be embedded or hardcoded within the virtual machine binary as this would not support post-install configuration." EX1005, [0022]. To address this, Kashyap teaches using a virtual transport mechanism specified in an OVF specification to load configuration information into the booting virtual appliance (*the virtual appliance software application configured to automatically load a plurality of stored configurations on the midserver*). EX1005, [0022], [0027].

86. Specifically, during installation of the virtual appliance software application, the activation script calls the configuration script. EX1005, [0020], [0022]-[0023]. The configuration script loads configuration data via a virtual transport mechanism. EX1005, [0026]-[0028], [0031], [0034]. This configuration data is loaded in one embodiment as a virtual CD-ROM, and used to configure the virtual appliance software application automatically during the boot operation. EX1005, [0005]-[0007], [0020], [0022] ("An OVF specification has a mechanism to include the configuration parameters and a transport mechanism, such as a virtual CD-ROM, to transfer these parameters to the booting virtual machine."). During this process, where the cloud agent is configured using the configuration data provided via the virtual transport mechanism, the cloud agent (*virtual appliance software application*) thus *automatically loads a plurality of stored configurations on the midserver*. FIG. 1 of Kashyap depicts the process of loading the configuration data:

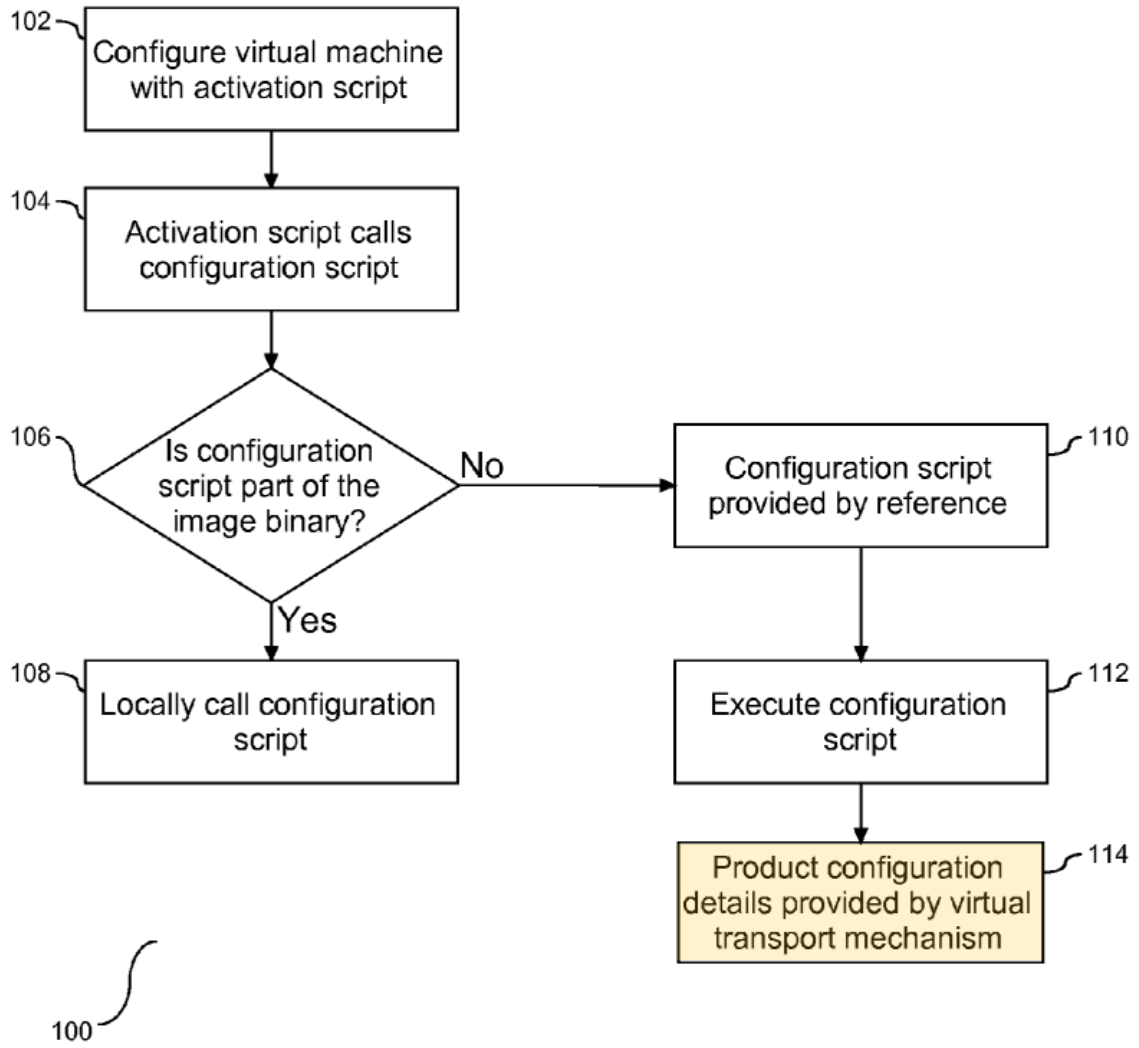


FIG. 1

EX1005, FIG. 1 (highlighting added).

87. FIG. 3 of Kashyap depicts a list of properties, configuration data, and configuration files (*a plurality of stored configurations*) in a virtual CD-ROM that are used to configure the application:

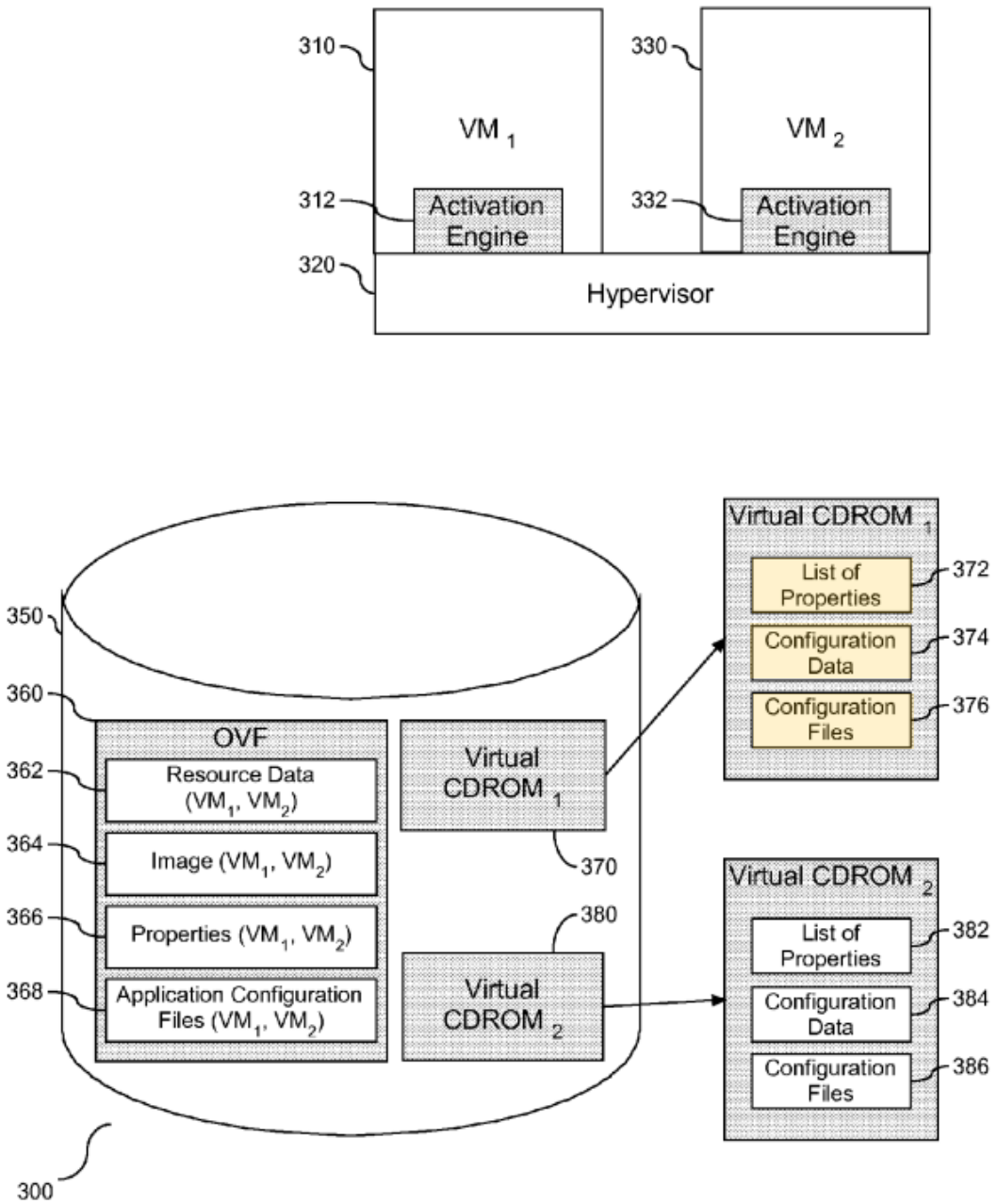


FIG. 3

EX1005, FIG. 3 (highlighting added). In combination, the list of properties, configuration data, and configuration files taught by Kashyap further include

"customer-specific configuration and subscription information" taught by Maturana. EX1004, [0044], [0053].

88. So implemented, the cloud agent of Maturana-Kashyap is provided to an on-premise server as a virtual appliance software application that is activated in the virtual machine environment of the server (*automatically install a virtual appliance software application*). The cloud agent (*virtual appliance software application*) then automatically loads configuration data and configuration files, including customer-specific configuration information, as part of its boot process (*the virtual appliance software application configured to automatically load a plurality of stored configurations on the midserver*), where this configuration data enables the cloud agent to "automatically detect and communicate with the cloud platform." EX1004, [0041].

4. Element [1.4]

[1.4] establish a secure network connection to an external network;

89. Maturana-Kashyap satisfies this element. Maturana discloses establishing a secure network connection to an external network, referred to by Maturana as a "cloud platform" or "the cloud," through a "methodology for securely connecting a cloud agent to a cloud platform by handshaking with an agent registry" (*establish a secure network connection to an external network*). EX1004, [0021]. Specifically, "[t]o ensure *secure outbound traffic to the cloud*, one or more

embodiments of cloud agent 202 can support HTTPS/SSL, certificate authority enabled transmission, and/or unique identity using MAC addresses." EX1004, [0057] (emphasis added).

90. Maturana depicts the cloud agent connecting to cloud services on an "external network" in FIG. 1:

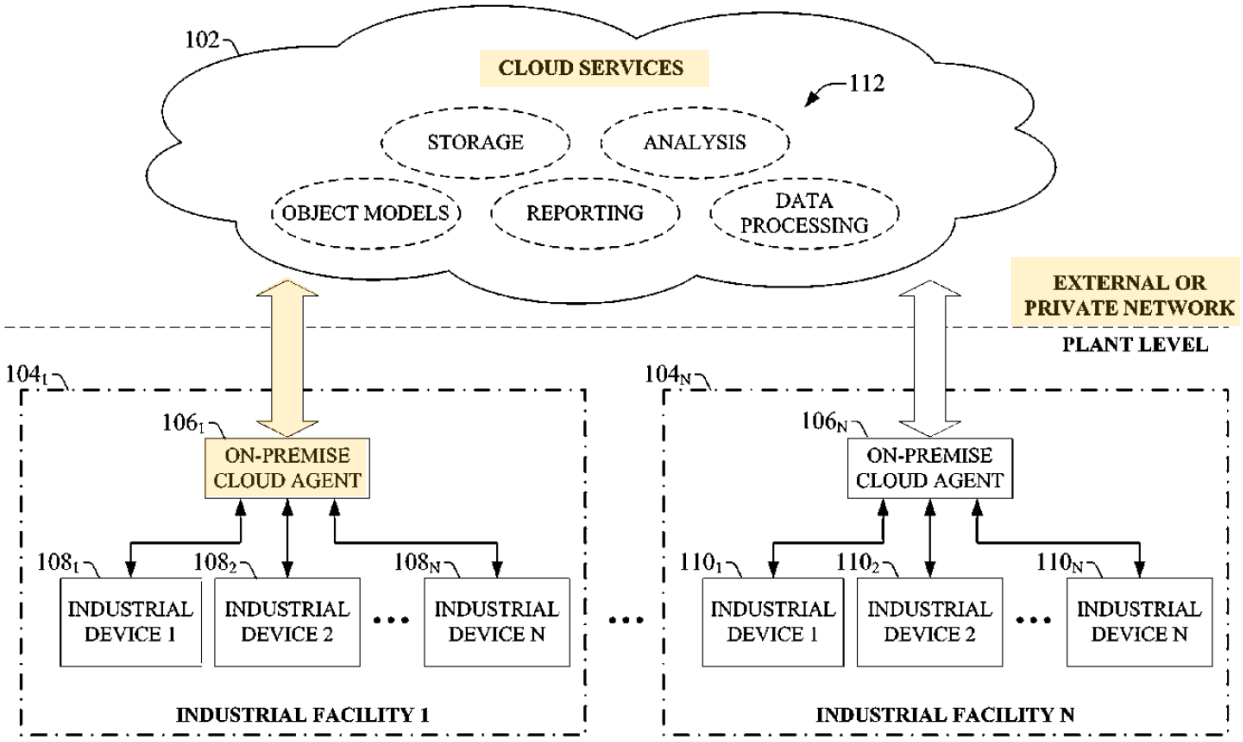


FIG. 1

EX1004, FIG. 1 (highlighting added);

91. Maturana-Kashyap thus discloses and teaches establishing a secure network connection to an external network, at least because Maturana expressly discloses doing so.

5. Element [1.5]

[1.5] receive data over a local network from a plurality of computing devices;

92. Maturana-Kashyap satisfies this element. Maturana discloses the cloud agent receiving data from a plurality of computing devices over a local network. Specifically, the cloud agent "collect[s] data from industrial devices" (*receive data from a plurality of computing devices*). EX1004, [0039]. Figure 1 of Maturana depicts a plurality of industrial devices providing data to a cloud agent over a local network:

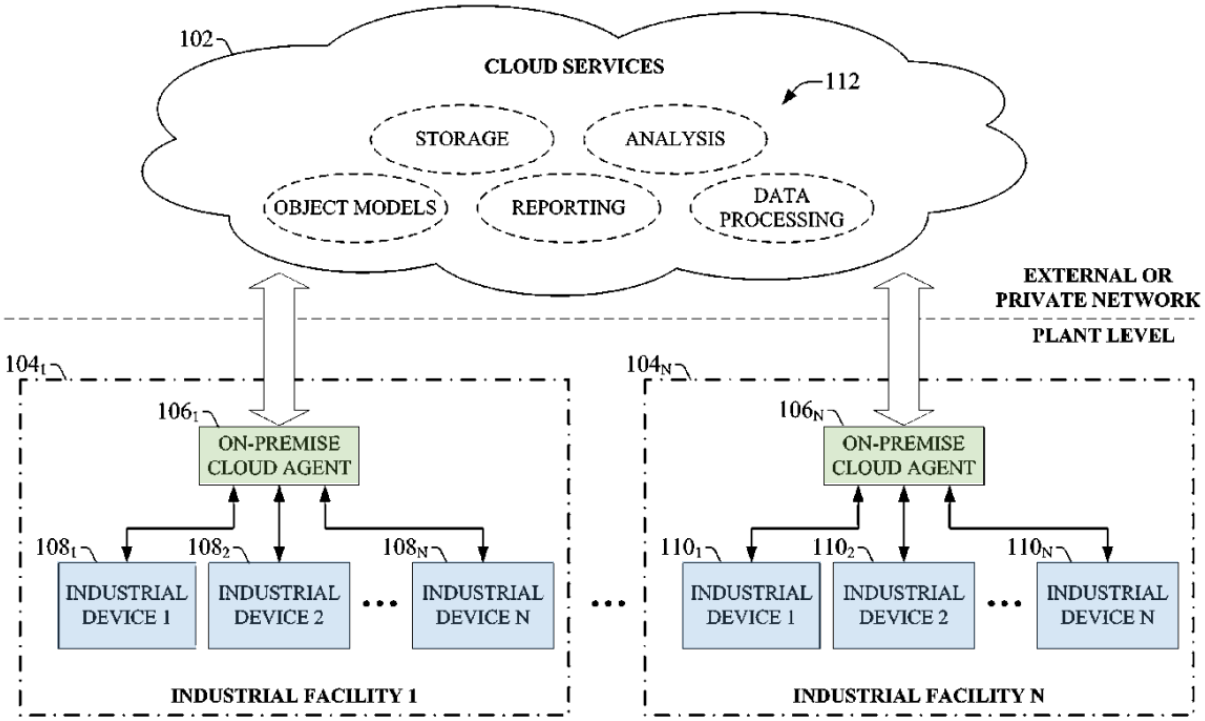


FIG. 1

EX1004, FIG. 1 (highlighting added).

93. The "number of industrial devices" (e.g., industrial device 1, industrial device 2, through industrial device N) providing data to the cloud agent (e.g., on-premise cloud agent) include "industrial controllers that facilitate monitoring and control of their respective processes" (*a plurality of computing devices*). EX1004, [0037]; *id.*, [0036]. The industrial controllers are *computing devices* because the term "controller" refers to "a computer-related entity," including "a computer" (EX1004, [0026]), and such "[i]ndustrial controllers store and execute user-defined control programs to effect decision-making in connection with the controlled process" (EX1004, [0031]). FIG. 12 of Maturana, reproduced below, expressly discloses receiving collected industrial data at a cloud agent from one or more on-premise data sources:

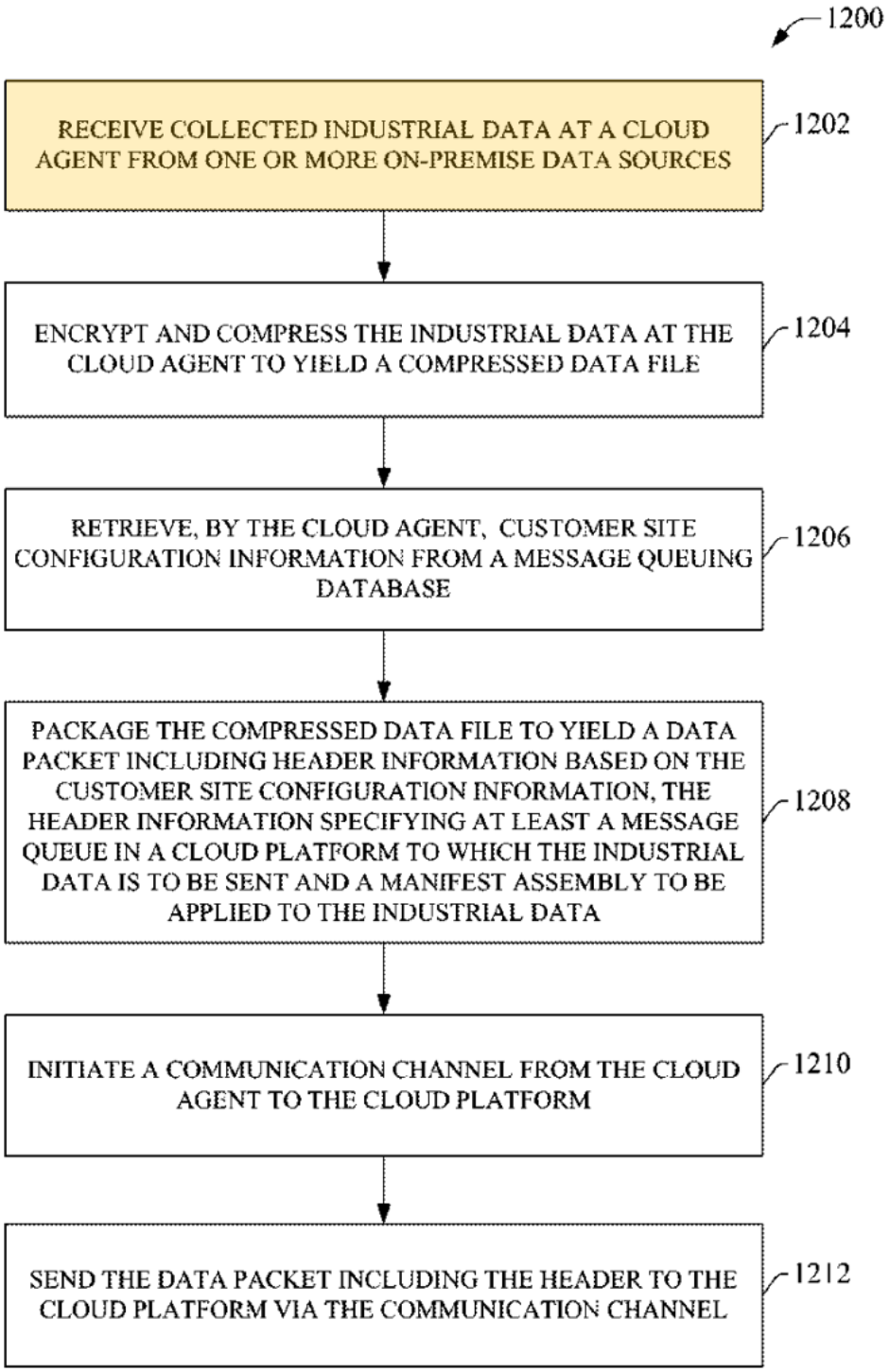


FIG. 12

EX1004, FIG. 12 (highlighting added).

94. Regarding a *local network*, Maturana discloses collecting the data on the "plant network." EX1004, [0048]. Specifically, the cloud agent "collects data items from the various industrial assets on the plant network and packages the collected data according to a generic, uniform data packaging schema used to move the on-premise data to a cloud platform 302." EX1004, [0048]. The plant network is a *local network* because the on-premise cloud agent and the industrial controllers are connected "through a physical or wireless *local area network* or radio link." EX1004, [0040] (emphasis added); *id.*, [0037] (describing plant network).

95. Maturana-Kashyap thus discloses and teaches receiving data over a local network from a plurality of computing devices, at least because Maturana expressly discloses doing so.

6. <u>Element [1.6]</u>

[1.6] apply a plurality of transformations to at least a portion of the received data; and

96. Maturana-Kashyap satisfies this element. Maturana discloses applying a plurality of transformations to at least a portion of the received data. Maturana discloses that the "cloud agent 202 can also perform local analytics on the data prior to moving the data to the cloud platform," where "[t]his can comprise substantially any type of pre-processing or data refinement that may facilitate efficient transfer of the data to the cloud, prepare the data for enhanced analysis in the cloud, reduce the

amount of cloud storage required to store the data, or other such benefits" (*apply a plurality of transformations to at least a portion of the received data*). EX1004, [0054]; *see also id.*, [0049] (cloud agent "can intelligently sort and organize the data"), [0052] (the cloud agent "packages the compressed data file 412 into a data packet and pushes the data packet to the cloud platform").

97. Maturana expressly teaches *a plurality of transformations* that the cloud agent applies to at least a portion of the received industrial data:

- "[C]loud agent 202 may also *transform a specified subset of the industrial data* from a first format to a second format" in order to "reduce the amount of processing load on the cloud side." EX1004, [0055] (emphasis added).
- "[C]loud agent 202 may be configured to compress the collected data using any suitable data compression algorithm prior to migrating the data to the cloud platform," where "[t]his can include detection and deletion of redundant data bits, truncation of precision bits, or other suitable compression operations." EX1004, [0054].
- "[C]loud agent 202 may be configured to aggregate data by combining related data from multiple sources." EX1004, [0054].
- "Cloud agent 202 may also encrypt sensitive data prior to upload to the cloud." EX1004, [0054].

- "[C]loud agent 202 may filter the data according to any specified filtering criterion." EX1004, [0055].
- "Cloud agent 202 may also associate metadata with selected subsets of the data prior to migration to the cloud, thereby contextualizing the data within the industrial environment." EX1004, [0056]. This "layered processing of the collected data to generate meta-level knowledge … can subsequently be leveraged by cloud-based analysis tools to facilitate enhanced analysis of the data in view of a larger plant context." EX1004, [0056].

Each of these example processing steps in Maturana is a transformation.

98. Maturana discloses applying more than one such transformation to at least a subset of data by using "also" to describe the different transformations (e.g., "cloud agent 202 may *also* transform" in addition to "compress," "aggregate," "encrypt," and so on). EX1004, [0054]-[0056]. Maturana also expressly discloses applying a plurality of transformations to data in FIG. 12, including encrypting, compressing, and packaging:

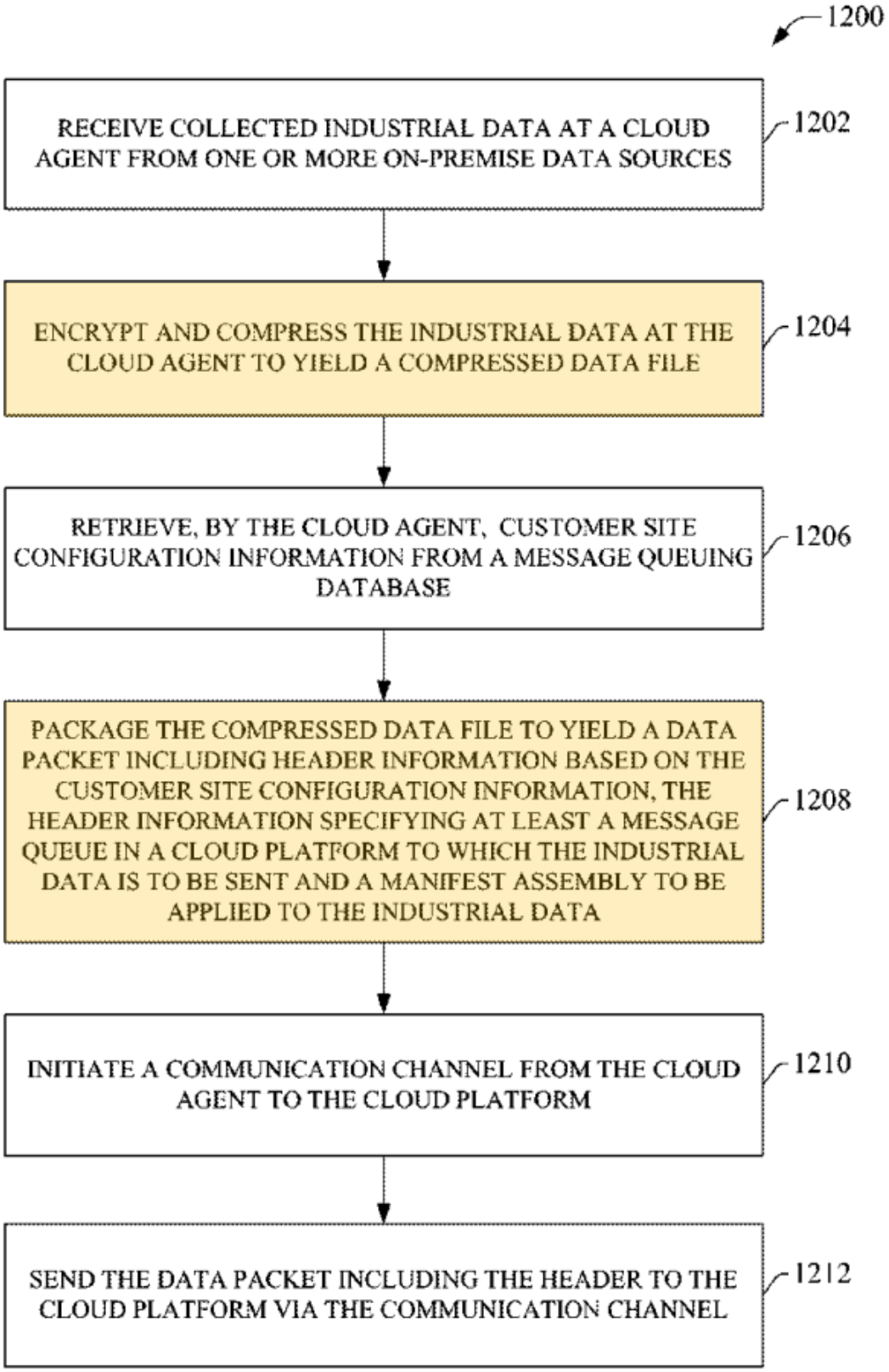

1200
RECEIVE COLLECTED INDUSTRIAL DATA AT A CLOUD AGENT FROM ONE OR MORE ON-PREMISE DATA SOURCES
1202
ENCRYPT AND COMPRESS THE INDUSTRIAL DATA AT THE CLOUD AGENT TO YIELD A COMPRESSED DATA FILE
1204
RETRIEVE, BY THE CLOUD AGENT, CUSTOMER SITE CONFIGURATION INFORMATION FROM A MESSAGE QUEUING DATABASE
1206
PACKAGE THE COMPRESSED DATA FILE TO YIELD A DATA PACKET INCLUDING HEADER INFORMATION BASED ON THE CUSTOMER SITE CONFIGURATION INFORMATION, THE HEADER INFORMATION SPECIFYING AT LEAST A MESSAGE QUEUE IN A CLOUD PLATFORM TO WHICH THE INDUSTRIAL DATA IS TO BE SENT AND A MANIFEST ASSEMBLY TO BE APPLIED TO THE INDUSTRIAL DATA
1208
INITIATE A COMMUNICATION CHANNEL FROM THE CLOUD AGENT TO THE CLOUD PLATFORM
1210
SEND THE DATA PACKET INCLUDING THE HEADER TO THE CLOUD PLATFORM VIA THE COMMUNICATION CHANNEL
1212

FIG. 12

EX1004, FIG. 12 (highlighting added); *id.*, [0085]. This express disclosure of the cloud agent performing compression, encryption, and data packetization with structured header information in Maturana is express disclosure of *applying a plurality of transformations* because each alters the data in a different and meaningful way: (1) Compression is a transformation because it changes the data's representation and size while preserving content; (2) encryption is a transformation that converts the original data into a different form that cannot be interpreted without a corresponding decryption key, thus altering the data's structure and accessibility; and (3) packaging the compressed and encrypted data into a data packet further transforms the data by embedding it within a structured container that adds metadata governing routing, prioritization, and cloud-side processing. EX1004, [0054], [0069]-[0070], [0085]-[0086]. Maturana specifically teaches applying these transformations prior to transmitting the data to the cloud platform. *E.g.*, EX1004, [0086].

99. Maturana-Kashyap thus discloses and teaches applying a plurality of transformations to at least a portion of the received data, at least because Maturana expressly discloses doing so.

7. <u>Element [1.7]</u>

[1.7] retransmit the received data over the secure connection as a single data stream.

100. Maturana-Kashyap satisfies this element. Maturana discloses retransmitting the received data over the secure connection as a single data stream. Maturana discloses that a queue processing component of the cloud agent "packages the compressed data file 412 into a data packet and pushes the data packet to the cloud platform" (*retransmit the received data over the secure connection as a single data stream*). EX1004, [0052]. The data is provided as a single data stream because "data from multiple sensors measuring related aspects of an automation system can be identified and *aggregated into a single cloud upload packet* by cloud agent 202." EX1004, [0054] (emphasis added).

101. Thus, Maturana-Kashyap discloses and teaches retransmitting the received data over the secure connection as a single data stream, at least because Maturana expressly discloses doing so.

E. <u>Claim 4</u>

1. <u>Element [4.1]</u>

[4.1] A method for ingestion of data into a cloud-based service from an external network, comprising the steps of:

102. Maturana-Kashyap discloses and teaches this element, and thus renders this claim obvious. *See Supra* Section IX.D.1 (element 1.1). Maturana-Kashyap discloses, as taught by Maturana, a "methodology for ingesting industrial data to a cloud platform using a cloud agent architecture." EX1004, [0019], FIG. 12; *id.*,

[0005] (describing "a method for sending industrial data to a cloud platform"), Claim 12 (claim a "method for sending industrial data to a cloud platform").

2. Element [4.2]

[4.2] installing a midserver comprising at least a processor, a memory, and a plurality of programming instructions stored in the memory and operating on the processor;

103. Maturana-Kashyap discloses and teaches this element, and thus renders this claim obvious. *See Supra* Section IX.D.2 (element 1.2). A POSITA would have understood that the computing device / server executing the cloud agent, as taught by Maturana, is installed at a facility where data is to be collected, and thus the method of ingesting data to a cloud platform taught by Maturana-Kashyap includes *installing* the server at the facility. Maturana discloses the "installation at any facility" (EX1004, [0041]) of a cloud agent executing on a "hardware platform" such as "a server" (*id.*, [0048]). *See also* EX1004, [0026]. Maturana further discusses the physical devices / servers executing the "cloud agents distributed throughout the customer's industrial enterprise." EX1004, [0074]. Each server "has a physical address (e.g., a MAC address or other physical address) that uniquely identifies the computing device," and a given cloud agent is associated with that physical address so that the cloud agent "and the physical device on which the agent executes are logically linked." *Id.* These disclosures show that, in Maturana-Kashyap, a server is *installed* at a particular facility to execute a cloud agent.

3. Element [4.3]

[4.3] automatically installing, using the midserver, a virtual appliance software application, the virtual appliance software application configured to automatically load a plurality of stored configurations on the midserver;

104. Maturana-Kashyap discloses and teaches this element, and thus renders this claim obvious. *See Supra* Section IX.D.3 (element 1.3).

4. Element [4.4]

[4.4] establishing a secure network connection from the midserver to an external network;

105. Maturana-Kashyap discloses and teaches this element, and thus renders this claim obvious. *See Supra* Section IX.D.4 (element 1.4). Maturana discloses establishing the secure network connection from the server with the cloud agent (*from the midserver*) to the cloud platform / external network. EX1004, [0057] ("[t]o ensure secure outbound traffic to the cloud"), FIG. 1.

5. Element [4.5]

[4.5] receiving data over a local network from a plurality of computing devices;

106. Maturana-Kashyap discloses and teaches this element, and thus renders this claim obvious. *See Supra* Section IX.D.5 (element 1.5).

6. Element [4.6]

[4.6] applying a plurality of transformations to at least a portion of the received data; and

107. Maturana-Kashyap discloses and teaches this element, and thus renders this claim obvious. *See Supra* Section IX.D.6 (element 1.6).

7. Element [4.7]

[4.7] retransmitting the received data over the secure connection as a single data stream.

108. Maturana-Kashyap discloses and teaches this element, and thus renders this claim obvious. *See Supra* Section IX.D.7 (element 1.7).

I, Mark Crovella, do hereby declare and state, that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Dated: January 13, 2026

Signature

APPENDIX A

MARK E. CROVELLA

Duan Family Fellow
Professor, Computer Science
Professor, Faculty of Computing and Data Sciences

Boston University
665 Commonwealth Ave.
Boston, MA 02215

Current as of January 13, 2026

EDUCATION

Ph.D. Computer Science, University of Rochester, October 1994.

M.S. University of Rochester, May 1991.

M.S. State University of New York at Buffalo, May 1990.

B.S. Cornell University, May 1982.

PROFESSIONAL EXPERIENCE

Professor Department of Computer Science, Boston University, 2006 to present. **Department Chair** 2013 – 2018; **Associate Professor** 2000 – 2006; **Assistant Professor** 1994 – 2000. On leave 2000 – 2001.

Professor Faculty of Computing & Data Sciences, Boston University, 2022 to present. **Duan Family Fellow** 2024 – present; **Chair of Academic Affairs** 2022 – present; **Founding Faculty Member**, 2020 – 2022; **Data Science Faculty Fellow** 2019 – 2022.

Affiliated Faculty Department of Electrical and Computer Engineering, Graduate Program in Bioinformatics, Center for Information Systems and Engineering, and Division of Systems Engineering, Boston University.

Chief Scientist Guavus, Inc., February 2012 to August 2014.

Visiting Faculty Laboratoire d'Informatique de Paris 6 (LIP6); 2003 – 2004 and 2018 – 2019; INRIA Paris 2018 – 2019; Laboratory of Information, Networking and Communication Sciences (LINCS) Paris 2018 – 2019.

Technical Director Network Appliance, Tewksbury, MA. November 2000 to April 2002.

Chief Architect Webmanage Technologies, Chelmsford, MA. May 2000 to acquisition by Network Appliance in November 2000.

Lead Research Scientist epicRealm, Inc., Richardson, TX. January to May 2001.

Senior Computer Scientist Calspan Corporation, Buffalo, NY. August 1984 to August 1994.

HONORS

Duan Family Faculty Fellow, 2024, Boston University.

Distinguished Alumni Award, 2017, University at Buffalo, Department of Computer Science and Engineering.

Innovator of the Year, 2014, Boston University.

IETF/IRTF Applied Networking Research Prize for "Routing State Distance: A Path-based Metric for Network Analysis," ACM Internet Measurement Conference (IMC), 2012.

Best Student Paper Award for "Intrusion as (Anti)social Communication: Characterization and Detection," ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2012.

IEEE Fellow, 2011, "For contributions to the measurement and analysis of networks and distributed systems."

ACM Fellow, 2010, "For contributions to the measurement and analysis of networks and distributed systems."

ACM SIGMETRICS Test of Time Award for "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes," (with Azer Bestavros). One of the three inaugural Test of Time Awards given by SIGMETRICS in 2010. The award is given for a paper published between 1973 and 1999 that has had the greatest impact over the subsequent period.

CNRS Fellowship, Centre National de la Recherche Scientifique (Department Sciences et Technologies de l'Information et de la Communication), for 2003-2004 academic year as Visiting Associate Professor at Laboratoire d'Informatique de Paris 6 (LIP6).

NSF CAREER Award, 1995.

ARPA Fellowship in High Performance Computing, 1993-1994.

DARPA Parallel Processing Fellowship, 1992-1993.

Editorial Boards

Steering Committee, IEEE Transactions on Network Science and Engineering, 2016–2019.

Editor, Computer Communication Review, 2005 – 2008.

Editor, Computer Networks, 2004 – 2005.

Editor, IEEE/ACM Transactions on Networking, 2000 – 2006.

Editor, IEEE Transactions on Computers, 2000 – 2004.

Funding

"Mechanistic Interpretability of Large Language Models," Open Philanthropy, 2025 – 2026, $52,646.

"Large Scale Analysis of Configurations and Management Practices in the Domain Name System," NSF CISE/CNS NeTS Medium, 10/1/23 – 9/30/26, $525,000.

"Methods for Active Measurement of the Domain Name System," NSF CISE/CNS IMR, 10/1/23 - 9/30/26, $256,463.

Google Research Award, 10/1/20, $30,000.

"Traveling Back to the Future: Using Network Science to Unravel Evolutionary Conserved RNA Binding Proteins and Complexes Associated with Disease," BU Hariri Institute Research Incubation Award, 2/9/18, $35,280.

"Analytic Tools for Evolving Path-based Networks," NSF CISE/CNS NeTS Small, 8/15/16 – 7/31/19, $499,951.

"Structural Matrix Completion for Data Mining Applications," NSF CISE/III Small, 9/1/14 – 8/31/17, $499,702 with 1 other PI.

"Understanding Communication Strategies for Ad Hoc Networks," NSF CISE/CNS NeTS Small, 8/1/11 – 7/31/14, $390,853.

"Information and Software Assurance: A Coordinated Approach to Cyber-Situation Awareness Based on Traffic Anomaly Detection,", DoD ARO, 9/1/2011 - 8/31/2015, $800,000 with 3 other PIs.

"Securing the Open Softphone," NSF CISE/CNS TC Large, 7/1/10 – 6/30/15, $2,992,896, with 8 other PIs. REU Supplement, 4/13/11 – $73,125.

"New Directions in Network Dimensionality Reduction for Routing and Beyond," NSF CISE/CNS NeTS Small, 7/1/10 – 6/30/13, $450,000.

"Wide-Aperture Traffic Analysis for Internet Security," NSF CISE/CNS TC Medium, 9/1/09 – 8/31/13, $723,053, with 2 other PIs.

Sponsoring Advisor for Postdoctoral Fellow under the *Computing Innovations Fellowship Program,* NSF CIF, 9/1/09 – 8/31/11, $267,500.

"Passive Methods for Internet Topology Discovery," National Geospatial Intelligence Agency, 8/1/09 – 7/31/12, $360,000.

"Instrumentation and Measurement for GENI," NSF/BBN, 8/1/08 – 8/31/11, $80,932

"Whole-Network Anomaly Diagnosis," Intel, 1/1/05 – 12/31/07, $158,000.

"Coordinate Systems for the Internet," NSF CISE/ANI, 8/15/03 – 7/31/06, $222,035.

"Collaborative Research: Modular Strategies for Global internetwork Monitoring," with 6 other PIs, NSF CISE/CCR ITR, 9/1/03 – 8/31/08, $1,994,513.

"Research Support for IP Network Characterization," Sprint Advanced Technology Laboratories, 9/1/03 – 8/31/04, $29,592.

"Research Support for IP Network Characterization," with 3 other PIs, Sprint Advanced Technology Laboratories, 9/1/02 – 8/31/03, $80,000.

"A Control Theoretic Approach to the Design of Internet Traffic Managers," with 3 other PIs, NSF CISE/ANIR, 9/1/01 – 8/31/05, $719,994.

"Diagnosis and Control of Network Variability by Massively Accessed Servers," with 2 other Principal Investigators, NSF CISE/ANI, 9/1/00 – 8/31/05, $1,210,768.

"Wide Area Web Measurement," GTE Laboratories, 9/1/00 – 8/31/01, $20,000.

"Wide Area Web Measurement," Xerox PARC, 9/1/99 – 8/31/00, $15,000.

"COMMONWEALTH: Architecture and Protocols for Scalable WWW Service," with 2 other Principal Investigators, NSF CISE/CCR Experimental Software Systems, 7/15/97 – 6/30/00, $590,087.

"Research Support for Network Performance Characterization, Analysis, and Evaluation," Principal Investigator, Hewlett-Packard Company, 9/1/96 – 9/1/97: $69,404. Renewed for 9/1/97 – 9/1/98: $88,154. Renewed for 9/1/98 – 9/1/99: $29,873.

"Practical Performance Prediction for Parallel Programmers," Principal Investigator, NSF CISE/CCR CAREER, 9/1/95 – 5/31/98, $91,066

"Research Infrastructure for Parallel and Distributed Systems: Real-Time, Multimedia, and High-Performance," with 4 other Principal Investigators, NSF CISE Institional Infrastructure, $874,019

"Research Instrumentation for Real-Time, Multimedia and High-Performance Computing in Distributed Systems," with 4 other Principal Investigators, NSF CISE IGR, $142,744

"Graduate Assistantships in High Performance Computing," with 5 other Principal Investigators, US Dept. of Education, 9/1/95 – 8/31/98, $354,645

Professional Affiliations

Association for Computing Machinery (Fellow)
IEEE (Fellow)
ACM SIGCOMM and SIGMETRICS
IEEE Computer Society and Communication Society

Invited Conference Talks

Invited Speaker, *TMA Experts Summit,* Paris, June 18, 2019.

Invited Panelist, *The 3rd Future Network Development Conference,* Nanjing, China, May 22, 2019.

Keynote Speaker, *IEEE Workshop on Data Science for Networking (DS4N),* Boston, MA, December 14, 2017.

Invited Panelist, *CSE 50th Anniversary Celebration,* University at Buffalo, Department of Computer Science and Engineering, September 30, 2017.

Keynote Speaker, *DATANET 16: Workshop on Data-Driven Networking,* Arizona State University, May 13, 2016.

Invited Speaker, *Workshop on Mathematics of Data Analysis in Cybersecurity,* ICERM, Brown University, October 24, 2014.

Invited Speaker, *51st Annual Allerton Conference on Communication, Control, and Computing,* University of Illinois at Urbana-Champaign, October 3, 2013.

Invited Speaker, *Network Links: Connecting Social, Communication and Biological Network Analysis,* Institute for Mathematics and its Applications (IMA), Minneapolis, MN, February 29, 2012.

Invited Speaker, *Statistics of Networks Workshop,* Isaac Newton Institute for Mathematical Sciences, Cambridge, UK, June 24-25, 2010.

Keynote Speaker, *NetEval 2009 Workshop, The First Workshop on Performance Evaluation of Next-Generation Networks,* Boston, MA, April 26, 2009.

Keynote Speaker, *2006 Euro-NGI Workshop on Visions of Future Generation Networks (EuroView2006),* Würzburg, Germany, August 1, 2006.

Invited Speaker, *Maryland Hybrid Networks Center,* Advanced Networks Colloquium Series, University of Maryland, December 16, 2005.

Keynote Speaker, *Workshop on End-to-End Monitoring Techniques and Services (E2EMON),* Nice, France, May 15, 2005.

Invited Speaker, *Workshop on Internet Measurement, Modeling, and Analysis,* Statistical Research Center for Complex Systems (SRCCS), Seoul National University, Seoul, South Korea, January 11, 2005.

Invited Speaker, *Workshop on Measurement, Modeling and Analysis of the Internet,* Institute for Mathematics and its Applications (IMA), Minneapolis, MN, January 12, 2004.

Keynote Speaker, *First Workshop on Performance of Computational Systems and Communication,* part of the XXII Congress of the Brazilian Computer Society, Florianopolis, Brazil, July 18, 2002.

Invited Speaker, Institute for Pure and Applied Mathematics (IPAM) Program on Large Scale Communication Networks, Los Angeles, CA, March 18-22, 2002.

Invited Speaker, *DIMACS Workshop on Internet and WWW Measurement, Mapping and Modeling,* New Brunswick, NJ, February 13-15, 2002.

Invited Speaker, *Workshop on Job Scheduling Strategies for Parallel Processing,* Cambridge, MA, June 16, 2001.

Invited Speaker, *COST 257 Final Conference,* Würzburg, Germany.

Invited Plenary Speaker, *Performance Tools 2000:* The 11th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation, Schaumburg, Illinois, March 27 - 31, 2000.

Invited Speaker, *Institute for Mathematics and Its Applications Workshop: "Scaling Phenomena in Communication Networks,"* Minneapolis, MN, October 22-24, 1999.

Invited Speaker, Boston University Minority Engineers' Society, 2nd Annual Technical Exposition, Boston, MA, April 10, 1999.

Invited Speaker, Workshop on Intelligent Internet Infrastructure, IBM T.J. Watson Laboratory, Yorktown Heights, NY, March 22, 1999.

Invited Speaker, Workshop on Architectures, Performance, and Applications in Globally Distributed Web Services, Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil, March 11 - 12, 1998.

Invited Speaker, Long-Range Dependence and Self-Similar Network Traffic Session, Winter Simulation Conference, Atlanta, GA, Dec. 7-10, 1997

Invited Speaker, DIMACS/NSF Workshop on "End-to-End Network Modeling and Simulations," Princeton University, Princeton, NJ, October 23-25, 1997.

Invited Panelist, Hot Topic Session On Traffic Characterization, ITC-15, Washington DC, June 22-27, 1997

Invited Speaker, IFIP Working Group 7.3 Workshop, June 19-20, Orcas Island, WA.

Invited Panelist, "Engineering Implications of Long-Range Dependence in Teletraffic Data," MASCOTS '97, January 1997, Haifa, Israel

Invited Speaker, Applied Probability Session of Informs, the Operations Research Society, May 1997, San Diego, CA.

Invited Speaker, Joint Research Conference on Statistics in Quality, Industry and Technology, May 29-31 1996, NIST, Gaithersburg, MD. Sponsored by the *American Statistical Association*, the *Institute of Mathematical Statistics*, E. I. du Pont, and the *National Institute of Standards and Technology.*

Professional Activities

Chair, Scientific Advisory Board of the Max Plank Institute for Software Systems, 2024.

Scientific Advisory Board of the Max Plank Institute for Software Systems, 2016 – 2025.

Member, IETF/IRTF Applied Networking Research Prize Selection Committee, 2014 – 2016.

Executive Steering Committee of the Hariri Institute for Computing and Computational Science & Engineering, 2011 – 2014, 2024 – present.

Member of the External Review Visiting Committee for the Department of Computer Science at Amherst College, February 2013.

Organized Workshop: "GPU Programming with CUDA" at Boston University, July 24, 2012.

Member, SIGCOMM Rising Star Award Selection Committee, 2011.

Chair, SIGCOMM Test of Time Award Selection Committee, 2011.

Team Mentor for Conrad Awards, Cyber Security Challenge 2011.

External Judge for FCC's Open Internet Apps Challenge, 2011.

General Co-Chair, CoNEXT 2011.

Internet2 perfSONAR Workshop, Executive Committee, 2010.

Chair, ACM SIGCOMM 2007–2009.

Program Co-Chair, IFIP Networking 2010.

Internet2 Network Research Strategic Planning Team, 2009-present.

Internet2 Network Research Review Committee, `http://www.internet2.edu/networkresearch/nrrc.html`, 2009-present.

Advisory Board of CRAWDAD (Community Resource for Archiving Wireless Data At Dartmouth).

Steering Committee, Passive and Active Measurement Workshop (PAM), 2004-2006.

General Chair, Passive and Active Measurement Workshop (PAM) 2005.

Steering Committee, Internet Measurement Conference, 2003-2007.

ACM SIGCOMM Executive Committee, Award Committee Chair, 2003-2006. b

Program Comittee Chair, Internet Measurement Conference 2003.

Co-organizer (with C. Lindemann and M. Reiser), Workshop on Internet Performance Modeling, Dagstuhl Castle, Germany, September 26-29, 1999.

Co-organizer (with five others), BU/NSF Workshop on Internet Measurement, Instrumentation, and Characterization; Boston University, August 30, 1999.

Tutorials Chair, SIGMETRICS '99, Atlanta GA.

Program Committees:

- SIGCOMM: 2002, 2006, 2008, 2012, 2017
- SIGMETRICS: 1999, 2000, 2002, 2007, 2009, 2010, 2011, 2012, 2014, 2015, 2016
- CoNEXT: 2005, 2007, 2010, 2011

- Internet Measurement Conference: 2002, 2004, 2007, 2010, 2012, 2015, 2017, 2018, 2020, 2022
- ICDM: 2018
- The Web Conference: 2000, 2001, 2002, 2019, 2022 (Senior PC Member)
- KDD: 2017, 2023
- Intelligent Systems for Molecular Biology (ISMB): 2015, 2017
- First IMC Workshop on Internet Visualization: 2012
- Performance: 2011
- Passive and Active Measurement Workshop (PAM): 2009, 2010
- Infocom: 2009 (TPC Area Chair), 2010 (TPC Area Chair)
- IFIP Networking: 2009
- NetEval: 2009
- HotMetrics: 2008
- E2EMON: 2007
- First Workshop on Experimental Computer Science: 2007
- Workshop on Mining Network Data (MineNet-06): 2006
- Workshop on Large Scale Network Inference: 2005
- 6th Global Internet Symposium: 2001
- 6th International Workshop on Web Caching and Content Distribution: 2001
- Workshop on Caching, Coherence and Consistency: 2001
- Workshop on Network-Related Data Management: 2001
- ITC Specialist Seminar on IP Traffic Management and Modeling: 2000
- SPIE Performance and Control of Network Systems Conference: 1998, 1999
- 18th International Conference on Distributed Computing Systems (ICDCS-98): 1998
- Third IEEE Workshop on Services in Distributed and Networked Environments (SDNE '96): 1996

Course Texts and Materials

"Linear Algebra, Geometry, and Computation," Electronic Text, one-semester course on Linear Algebra from a Computer Science standpoint. 2015-2025. Available at `http://mcrovella.github.io/CS132-Geometric-Algorithms`. Includes contributions from Wayne Snyder and students.

"Computational Tools for Data Science," Electronic Text, one-semester course on applied methods in machine learning using Python. 2021-2022. Available at `http://mcrovella.github.io/CS506-Computational-Tools-for-Data-Science`. Includes contributions from Evimaria Terzi and George Kollios.

"Foundations of Data Science III," Electronic Text, one-semester course on probability and statistics for data science, 2023. Available at `https://mcrovella.github.io/DS122-Foundations-of-Data-Science-III/`. Co-developed with Pawel Przytycki and with contributions by Lisa Wobbes.

Tutorials and Short Courses

"Data Mining for Measurement Data," presented at the 5th PhD School on Traffic Monitoring and Analysis, associated with the 7th International Workshop on Traffic Monitoring and Analysis, Barcelona, April 21-22, 2015. Available at `https://github.com/mcrovella/mining-low-dim-network-data`.

"Traffic Modeling: Methods and Results for Single Links and Whole Networks," presented at SIGCOMM 2004, Portland OR.

"Traffic Modeling," short course presented at Aalborg University, February 2, 2004.

"Performance Characteristics of World Wide Web Information Systems," presented at SIGMETRICS '97, Seattle WA, June 1997; SIGMETRICS '98, Madison WI, June 1998; and Performance Tools '98, Palma de Mallorca, Spain, September 1998.

CONSULTING

Member of the Technical Advisory Board, Guavus, Inc. 2006 to 2017.

Consultant, Fujitsu Laboratories of America, Sunnyvale, CA. 2002 to 2004.

Consultant and Expert Witness, in litigation cases involving software patents, software trade secrets, and intellectual property. Ongoing.

Member of the Technical Advisory Board, epicRealm, Inc., Richardson, TX. March 2000 to May 2001.

Consultant and Member of the Technical Advisory Board, WebManage Technologies, Inc., Chelmsford MA. July 1999 to November 2000.

Founder and Principal Technical Officer, Commonwealth Network Technologies, Inc., Boston MA. Incorporated April 27, 1998; acquired by WebManage Technologies, October 12, 1999.

Consultant and Member of the Technical Advisory Board, VitalSigns Software Inc., Santa Clara, CA; acquired by International Network Services, Inc.; now part of Lucent. November 1996 to October 1999.

PATENTS

Methods for Storing and Reading Digital Data on a Set of DNA Strands, US Patent # 10,027,347, issued July 17, 2018.

Method and Apparatus for Whole-Network Anomaly Diagnosis and Method to Detect and Classify Network Anomalies Using Traffic Feature Distributions, US Patent # 8,869,276, issued October 21, 2014.

Compressed Set Representation for Sets as Measures in OLAP Cubes, US Patent # 8,533,167, issued September 10, 2013.

Method and Apparatus for Dynamic Resource Discovery and Information Distribution in a Data Network, US Patent # 7,747,741, issued June 29, 2010.

System for Creating and Distributing Prioritized List of Computer Nodes Selected as Participants in a Distribution Job, US Patent # 7,346,682, issued March 18, 2008.

Method and Apparatus for Election of Group Leaders in a Distributed Network, US Patent # 6,993,587, issued January 31, 2006.

Method and Apparatus for Scalable Distribution of Information in a Distributed Network, US Patent # 6,748,447, issued June 8, 2004.

Method and Apparatus for Reliable and Scalable Distribution of Data Files in Distributed Networks, US Patent # 6,718,361, issued April 6, 2004. European Patent # EP1305924 issued March 31, 2010.

Distributed Routing, US Patent # 6,370,584, issued April 9, 2002.

Method and Apparatus for Assigning Tasks in a Distributed Server System, US Patent # 6,223,205, issued April 24, 2001.

Doctoral Students

Ahmed Youssef, Ph.D. June 2023. "Computational Methods to Uncover Cell State Proteomes and Profile Protein Interaction Dynamics." Co-advised with Andrew Emili.

Bashir Rastegarpanah, Ph.D. August 2021. "Tools for Responsible Decision-Marking in Machine Learning."

Larissa Spinelli, Ph.D. June 2019. "Empirical Studies of Factors Affecting Opinion Dynamics."

Giovanni Comarela, Ph.D. July 2017. "On the Dynamics of Interdomain Routing in the Internet."

Natali Ruchansky, Ph.D. August 2016. "Matrix Completion with Structure." Co-advised with Evimaria Terzi.

Andrej Cvetkovski, Ph.D. May 2013. "Graph Embeddings for Low-Stretch Greedy Routing."

Gonca Gürsun, Ph.D. March 2013. "Inferring Hidden Features in the Internet."

Vijay Erramilli, Ph.D. December 2008. "Forwarding in Mobile Opportunistic Networks."

Nahur Fonseca, Ph.D. December 2008. "Stochastic Modeling Applied to Detection Problems in Network Protocols and Traffic."

Anukool Lakhina, Ph.D. August 2006. "Network-Wide Traffic Analysis: Methods and Applications."

Jun Liu, Ph.D. August 2002. "Characterizing Network Elements and Paths Using Packet Loss Behavior."

Paul Barford, Ph.D. June 2000. "Modeling, Measurement and Performance of World Wide Web Transactions."

Robert L. Carter, Ph.D. January 1998. "Performance Measurement and Prediction in Packet-Switched Networks: Techniques and Applications."

Other Doctoral Students Advised

Mengfei Cao, Ph.D, August 2016, "New Graph Metrics Improve Network-based Protein Function Prediction" Tufts University, Member of the Committee.

Bimal Viswanath, Ph.D, April 2016, "Towards Trustworthy Social Computing Systems." Universität des Saarlandes, Member of the Committee.

Yonjun Liao, Ph.D., January 2013, "Learning to Predict End-to-End Network Performance." Université de Liège, Belgium, Member of the Jury.

Diana Zeaiter Joumblatt, Ph.D., December 2012, "Prediction of User Dissatisfaction with Internet Application Performance at End-Hosts." Université Pierre et Marie Curie, Rapporteur.

Chrisil Arackaparambil, Ph.D., July 2011, "Anomaly Detection in Network Streams Through a Distributional Lens." Dartmouth College, Member of the Examining Committee.

Italo Cunha, Ph.D., July 2011, "Tracking Internet Routes and Path Reachability for Network Tomography." Université Pierre et Marie Curie (U. Paris VI), Rapporteur.

Hung Nguyen, Ph.D., February 2008, "User-level Internet Tomography and Overlay Routing." Ecole Polytechnique Fédérale de Lausanne (EPFL), Rapporteur.

Benoit Donnet, Ph.D., September 2006, "Algorithms for Large-Scale Topology Discovery." Université Pierre et Marie Curie (U. Paris VI), Président du Jury.

David Chua, Ph.D. (Mathematics), May 2006, "Statistical Analysis for Whole Networks." Second Reader.

Stilian Stoev, Ph.D. (Mathematics) May 2005.

Abhishek Kumar, Ph.D., October 2005, "Network Data Streaming: Algorithms for Network Monitoring and Coordination." Georgia Tech, Committee Member.

Marjan Bozinovski, Ph.D., May 2004, "Fault-Tolerant Platforms for IP-Based Session Control Systems." Aalborg University, Committee Member.

Artur Ziviani, Ph.D. December 2003, "Qualité de Service et Localisation d'Hotes." University of Paris VI, Examinateur.

Liang Guo, Ph.D. April 2003, "Size-Aware Scheduling of TCP Flows." Second Reader.

Alberto Medina, Ph.D. March 2003, "Practical Estimation of Internet Traffic Demands." Third Reader.

Vladas Pipiras, Ph.D. (Mathematics) April 2002, "Symmetric Stable Self-Similar Processes." Third Reader.

Khaled A. Harfoush, Ph.D. November 2001, "A Framework for End-to-End Characterization of Metric-Induced Network Topologies." Committee Member.

Santiago M. Pericas-Geertsen, Ph.D. November 2001, "XML-Fluent Mobile Agents." Committee Member.

Alia K. Atlas, Ph.D. August 1998, "Statistical Rate Monotonic Scheduling." Second Reader.

Ghulam Bhatti, Ph.D. January 1998, "Parallel and Distributed Discrete Event Simulation: An Event-Reservation Approach." Second Reader.

Sulaiman Mirdad, Ph.D. June 1997, "WebWave: Globally Load-Balanced, Fully-Distributed Caching". Second Reader.

Gi Tae Kim, Ph.D. June 1997, "Framework for Adaptive Forward Erasure Recovery (AFER) for Fault-Tolerant Real-Time Communications". Second Reader.

Carlos Cunha, Ph.D. June 1997, "Trace Analysis and its Application to Performance Enhancements of Distributed Information Systems". Second Reader.

Susan Nagy, Ph.D. April 1997, "Admission Control and Scheduling Strategies for Real-Time Database Systems". Second Reader.

Vadim Teverovksy, Ph.D. February 1997 (Mathematics), "Detection and Estimation of Long-Range Dependence." Third Reader.

Euthimios Panagos, Ph.D. May 1995, "Client-Based Logging: A New Paradigm for Distributed Transaction Management". Third Reader.

Masters Students

Megan Hacker, M.S. May 2012, "Methods for Pairing Mobile Devices Through Shared Sensing"

Lauren Walters, MSc. May 2004, "A Web Browsing Workload Model for Simulation," University of Cape Town, South Africa, external examiner.

Alan Charbonneau, M.A. May 2003, "Simulation of an SJF Proxy"

Iana Vitkova, M.A. May 2003, "Simulation of an SJF Proxy"

Robert Frangioso, M.A. Sept 2002, "Connection Scheduling in Web Servers"

Xiaohui Zhang, M.S. May 1999, "Cachability of Web Objects"

Scott Harrison, M.S. May 1998, "DPR Monitor: A Web-based DPR Network Monitoring Tool"

Stanley Gambarin, M.S. May 1997, "Implementing a Single-Process Non-Multithreaded World Wide Web Server"

Kostas Magoutis, M.S. May 1996, "Economies of Scale In Partitioned Disk Caches with Self-Similar References"

Fred Bierhaus, M.S. May 1996, "Improving Interactive Traffic Through the EQL Driver"

Books

M. Crovella and B. Krishnamurthy, "Internet Measurement: Infrastructure, Traffic, and Applications," John Wiley and Sons, 2006, 512 pp.

BOOKS EDITED

M. Crovella and L. M. Feeney and D. Rubenstein (editors), "Proceedings of the 9th International IFIP TC 6 Networking Conference," Springer-Verlag, 410pp, 2010.

JOURNAL ARTICLES

Hui Peng, Sergei Kotelnikov, Megan E. Egbert, Shany Ofaim, Grant C. Stevens, Sadhna Phanse, Tatiana Saccon, Mikhail Ignatov, Shubham Dutta, Zoe Istace, Mohamed Taha Moutaoufik, Hiroyuki Aoki, Neal Kewalramani, Jianxian Sun, Yufeng Gong, Dzmitry Padhorny, Gennady Poda, Andrey Alekseenko, Kathryn A. Porter, George Jones, Irina Rodionova, Hongbo Guo, Oxana Pogoutse, Suprama Datta, Milton Saier, Mark Crovella, Sandor Vajda, Gabriel Moreno-Hagelsieb, John Parkinson, Daniel Segre, Mohan Babu, Dima Kozakov, and Andrew Emili, "Ligand interaction landscape of transcription factors and essential enzymes in *E. coli*," *Cell,* 2025.

A. Youssef, I. Paul, M. Crovella, and A. Emili, "DESP: Demixing Cell State Profiles from Dynamic Bulk Molecular Measurements," *Cell Reports Methods,* 2024.

L. Salamatian, S. Anderson, J. Matthews, P. Barford, W. Willinger, and M. Crovella, "A Manifold View of Connectivity in the Private Backbone Networks of Hyperscalers," *Communications of the ACM*, August 2023.

N. Kewalramani, A. Emili, M. Crovella, "State-of-the-Art Computational Methods to Predict Protein-Protein Interactions with High Accuracy and Coverage," *Proteomics,* 2023.

A. Youssef, F. Bian, P. Havugimana, M. Crovella, and A. Emili, "Dynamic Remodeling of *Escherichia coli* Interactome in Response to Environmental Perturbations," *Proteomics,* 2023.

L. Salamatian, S. Anderson, J. Mathews, P. Barford, W. Willinger, and M. Crovella, "A Manifold View of Connectivity in the Private Backbone Networks of Hyperscalers," *Communications of the ACM*, Research Highlights, 6(7), July 2023.

J. Jung, M. Guo, M. Crovella, J.G. McDaniel, and K.M. Warkentin. "Frog Embryos Use Multiple Levels of Temporal Pattern in Risk Assessment for Vibration-Cued Hatching," *Animal Cognition,* 2022.

J. N. Law, K. Akers, N. Tasnina, C. M. Della-Santina, S. Deutsch, M. Kshirsagar, J. Klein-Seetharaman, M. Crovella, P. Rajagopalan, S. Kasif, and T. M. Murali, "Interpretable Network Propagation with Application to Expanding the Repertoire of Human Proteins that Interact with SARS-CoV-2," *GigaScience,* 10(12), December 2021.

D. Lancour, J. Dupuis, R. Mayeux, J. Haines, M. Pericak-Vance, G. Schellenberg, M. Crovella, L. Farrer, and S. Kasif, "Analysis of Brain Region-Specific Co-Expression Networks Reveals Clustering of Established and Novel Genes Associated with Alzheimer Disease," *Alzheimer's Research & Therapy,* 2020.

J. Fan, X. C. Li, M. Crovella, and M. D. M. Leiserson, "Matrix (Factorization) Reloaded: Flexible Methods for Imputing Genetic Interactions with Cross-Species and Side Information," *OUP Bioinformatics,* 2020.

A. Ramirez, S. Dankel, B. Rastegarpanah, W. Cai, R. Xue, M. Crovella, Y.-H. Tseng, R. Kahn, and S. Kasif, "Single-Cell Transcriptional Networks in Differentiating Preadipocytes Suggest Drivers Associated with Tissue Heterogeneity," *Nature Communications,* 2020.

J. Fan, A. Cannistra, I. Fried, T. Lim, T. Schaffner, M. Crovella, B. Hescott, and M.D.M. Leiserson, "Functional Protein Representations from Biological Networks Enable Diverse Cross-Species Inference," *Nucleic Acids Research,* 2019.

D. Lancour, A. Naj, R. Mayeux, J. L. Haines, M. A. Pericak-Vance, G. C. Schellenberg, M. Crovella, L. A. Farrer, and S. Kasif, "One for All and All for One: Improving Replication of Genetic Studies through Network Diffusion," *PLoS Genetics*, Vol 14, April 2018.

A. Cvetkovski and M. Crovella, "Low-stress Data Embedding in the Hyperbolic Plane Using Multidimensional Scaling," *Applied Mathematics and Information Sciences,* Vol 11, Number 1, January 2017.

A. Cvetkovski and M. Crovella, "Multidimensional Scaling in the Poincaré Disk," *Applied Mathematics and Information Sciences,* Vol 10, Number 1, January 2016.

M. Cao, H. Zhang, J. Park, N. Daniels, M. Crovella, L. Cowen, and B. Hescott, "Going the Distance for Protein Function Prediction: A New Distance Metric for Protein Interaction Networks," *PLoS One,* Vol 8, Number 10, October 2013.

A. Cvetkovski and M. Crovella, "On the Choice of a Spanning Tree for Greedy Embedding of Network Graphs," *Networking Science,* February 2013.

B. Donnet, P. Raoult, T. Friedman, and M. Crovella, "Deployment of an Algorithm for Large-Scale Topology Discovery," *IEEE Journal on Selected Areas in Communications, Special Issue on Sampling the Internet*, Vol 24, Number 12, December 2006.

D. Chua, E. Kolaczyk, and M. Crovella, "Network Kriging," *IEEE Journal on Selected Areas in Communications, Special Issue on Sampling the Internet*, 2006.

B. Gueye, A. Ziviani, M. Crovella, and S. Fdida, "Constraint-Based Geolocation of Internet Hosts," *IEEE/ACM Transactions on Networking,* Vol 14, Number 6, December 2006.

R. Fonseca, V. Almeida, and M. E. Crovella, "Locality in a Web of Streams," *Communications of the ACM*, 48(1), pp. 82–88, January 2005.

M. Fayed, P. Krapivsky, J. Byers, M. Crovella, D. Finkel, and S. Redner, "On the Emergence of Highly Variable Distributions in the Autonomous System Topology," *Computer Communcation Review*, 33(2), pp. 41–49, July 2003.

A. Lakhina, J. W. Byers, M. E. Crovella, and I. Matta, "On the Geographic Location of Internet Resources," *IEEE Journal on Selected Areas in Communications, Special Issue on Internet and WWW Measurement, Mapping, and Modeling*, 2003.

P. Barford and M. E. Crovella, "Critical Path Analysis of TCP Transactions," *IEEE/ACM Transactions on Networking*, 9 (3), pp. 238–248.

M. E. Crovella, C. Lindemann, and M. Reiser, "Internet Performance Modeling: the state of the art at the turn of the century," *Performance Evaluation,* 42 (2–3), pp. 91-108, September 2000.

R. L. Carter and M. E. Crovella, "On the Network Impact of Dynamic Server Selection," *Computer Networks,* 31 (23–24), pp. 2529–2558.

M. Harchol-Balter, M. E. Crovella and C. D. Murta, "On Choosing a Task Assignment Policy for a Distributed Server System," *Journal of Parallel and Distributed Computing,* Special Issue on Software Support for Distributed Computing, September 1999.

P. Barford and M. E. Crovella, "Measuring Web Performance in the Wide Area," *Performance Evaluation Review,* Special Issue on Network Traffic Measurement and Workload Characterization, August 1999.

P. Barford, A. Bestavros, A. Bradley, and M. E. Crovella, "Changes in Web Client Access Patterns: Characteristics and Caching Implications," *World Wide Web,* Special Issue on World Wide Web Characterization and Performance Evaluation, 1999.

M. E. Crovella and M. S. Taqqu, "Estimating the Heavy Tail Index from Scaling Properties," *Methodology and Computing In Applied Probability,* 1(1), 1999.

M. E. Crovella and A. Bestavros, "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes," *IEEE/ACM Transactions on Networking,* 5(6):835–846, December 1997.

R. L. Carter and M. E. Crovella, "Measuring Bottleneck Link Speed in Packet-Switched Networks," *Performance Evaluation,* 27&28:297–318, October 1996.

L. A. Crowl, M. E. Crovella, T. J. LeBlanc, and M. L. Scott, "The Advantages of Multiple Parallelizations in Combinatorial Search," *Journal of Parallel and Distributed Computing,* special issue on Data-Parallel Algorithms and Programming, 21(1):110–123, April 1994.

Book Chapters

M. E. Crovella, M. S. Taqqu, and A. Bestavros, "Heavy-Tailed Probability Distributions in the World Wide Web," Chapter 1 of *A Practical Guide To Heavy Tails,* pages 3–25, Chapman & Hall, New York, 1998.

M. E. Crovella and L. Lipsky, "Simulations with Heavy-Tailed Workloads," Chapter 3 of *Self-Similar Network Traffic and Performance Evaluation,* Wiley/Wiley-Interscience, 1999.

K. Park, G. Kim and M. Crovella, "The Protocol Stack and its Modulation Effect on Self-Similar Traffic," Chapter 16 of *Self-Similar Network Traffic and Performance Evaluation,* Wiley/Wiley-Interscience, 1999.

M. E. Crovella, "Performance Characteristics of the World Wide Web," in *White Book on Performance Evaluation,* Günter Haring, Christoph Lindemann, and Martin Reiser, editors, Springer-Verlag, 1999.

Invited Papers

M. E. Crovella, "Performance Evaluation with Heavy Tailed Distributions," in *Lecture Notes in Computer Science 1786*, Proceedings of the 11th International Conference on Modelling Tools and Techniques for Computer and Communication System Performance Evaluation (TOOLS '2000), pp. 1-9, March 2000. Revised version appeared in *Job Scheduling Strategies for Parallel Processing : 7th International Workshop,* pp. 1-10, June 16, 2001.

Refereed Conference Papers

N. Tasnina, M. Crovella, and T. M. Murali, "Provenance Tracing in Network Diffusion Algorithms," *Proceedings of the Pacific Symposium on Biocomputing (PSB),* Hawaii, January, 2026.

G. Franco and M. Crovella, "Pinpointing Attention-Causal Communication in Language Models," *Proceedings of NeurIPS*, San Diego, CA, December 2025

S. Jasina, L. Salamatian, P. Barford, M. Crovella, and W. Willinger, "A Breath of Fresh Air: Visualizing How Networks 'Breathe'", *Proceedings of the 24th ACM Workshop on Hot Topics in Networking (HotNets 2025)*, College Park, MD, November, 2025.

S. Buxbaum, L. Tassis, L. Boschelli, D. Christenson. G. Comarela, M. Crovella, and M. Varia, "Privacy-Preserving Machine Learning on Web Browsing for Public Opinion," *Proceedings of The International Symposium on Cyber Security, Cryptology and Machine Learning (CSCML),* December, 2025.

P. Calais, G. Franco, Z. Tang, T. Nikas, W. Meira Jr., E. Terzi, and M. Crovella, "Disentangling Text and Math in Word Problems: Evidence for the Bidimensional Structure of Large Language Models' Reasoning," *Findings of the Association for Computational Linguistics,* June, 2025.

C. Kranig, E. Pauley, W.-S. Wang, P. Barford, M. Crovella, and J. Sommers, "Toward a Representative DNS Data Corpus: A Longitudinal Comparison of Collection Methods, *Proceedings of the Network Traffic Measurement and Analysis Conference,* Copenhagen, Denmark, June 2025. Recipient of the TMA Community Contribution Award.

W.-S. Wung, C. Kranig, E. Pauley, P. Barford, M. Crovella, and J. Sommers, "Squatspotting: Towards the Systematic Measurement of Typosquatting Techniques," *Proceedings of IFIP Networking,* Limassol, Cyprus, May 2025.

A. Anderson, A. S. Mondal, P. Barford, M. Crovella, and J. Sommers, "An Elemental Decomposition of DNS Name-to-IP Graphs," *Proceedings of INFOCOM*, Vancouver, Canada, May 2024.

G. Franco, M. Crovella, and G. Comarela, "Dependence and Model Selection in LLP: The Problem of Variants," *Proceedings of KDD*, Los Angeles, CA, August 2023.

K. Quinn, E. Terzi, and M. Crovella, "Characterizing Covid Waves via Spatio-Temporal Decomposition," *Proceedings of KDD*, Washington DC, August 2022.

L. Salamatian, S. Anderson, J. Matthews, P. Barford, W. Willinger, and M. Crovella, "Curvature-based Analysis of Network Connectivity in Private Backbone Infrastructures," *Proceedings of ACM SIGMETRICS / IFIP Performance,* Mumbai, India, June 2022.

B. Rastegarpanah, K. P. Gummadi, and M. Crovella, "Auditing Black-Box Prediction Models for Data Minimization Compliance," *Proceedings of NeurIPS,* Online, December 2021. **Selected as a Spotlight Presentation (3% acceptance rate).**

L. Amichi, A. Carneiro Viana, M. Crovella, and A. A. F. Loureiro, "From Motion Purpose to Perceptive Spatial Mobility Prediction," *Proceedings of ACM SIGSPATIAL*, Beijing, China, November 2021.

G. Grassi, C. Barakat, M. Crovella, and R. Teixeira, "Leveraging Website Popularity Differences to Identify Performance Anomalies," *Proceedings of Infocom,* Online, May 2021.

L. Amichi, A. Carneiro Viana, M. Crovella, and A. A. F. Loureiro, "Understanding Individuals' Proclivity for Novelty Seeking," *Proceedings of ACM SIGSPATIAL,* Online, November 2020.

J. Fan, X. C. Li, M. Crovella, and M. D. M. Leiserson, "Matrix (Factorization) Reloaded: Flexible Methods for Imputing Genetic Interactions with Cross-Species and Side Information," *Proceedings of the 19th European Conference on Computational Biology,* Online, August 2020.

B. Rastegarpanah, M. Crovella, and K. P. Gummadi, "Fair Inputs and Fair Outputs: the Incompatibility of Fairness in Privacy and Accuracy," *Proceedings of the Workshop on Fairness in User Modeling, Adaptatation, and Personalization (FairUMAP),* Online, July 2020.

L. Spinelli and M. Crovella, "How YouTube Leads Privacy-Seeking Users Away from Reliable Information," *Proceedings of the Workshop on Fairness in User Modeling, Adaptatation, and Personalization (FairUMAP),* Online, July 2020.

J. Law, M. Kshirsagar, M. Crovella, P. Rajagopalan, J. Klein-Seetharaman, S. Kasif and T. M. Murali, "A Network-Based Label Propagation Framework to Reposition Drugs for COVID-19 (Poster)," Proceedings of RECOMB, Online, June 2020.

L. Amichi, A. Viana, M. Crovella, and A. A. F. Loureiro, "Mobility profiling: Identifying Scouters in the Crowd (Poster)," *Proceedings of the CoNEXT Student Workshop,* Orlando, FL, December 2019.

B. Rastegarpanah, K. P. Gummadi, and M. Crovella, "Fighting Fire with Fire: Using Antidote Data to Improve Polarization and Fairness of Recommender Systems," *Proceedings of WSDM 2019,* Melbourne, Australia, February 2019.

G. Comarela, R. Durairajan, P. Barford, D. Christenson, and M. Crovella, "Assessing Candidate Preference through Web Browsing History," *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD),* London, UK, August 2018. Press coverage: BU Today: `http://www.bu.edu/today/2018/better-election-predictions`.

L. Spinelli and M. Crovella, "Unravelling the Dynamics of Online Ratings," Proceedings of the 4th International Symposium on Social Media Mining and Analysis (SMMA 2018), Exeter, UK, June 2018.

M.D.M. Leiserson, J. Fan, A. Cannistra, I. Fried, T. Lim, T. Schaffner, M. Crovella, and B. Hescott, "A Multi-Species Functional Embedding Integrating Sequence and Network Structure," Proceedings of RECOMB, Paris, France, April 2018.

B. Rastegarpanah, K. Gummadi, and M. Crovella, "Exploring Explanations for Matrix Factorization Recommender Systems" (Position Paper), FATREC Workshop on Responsible Recommendation, Como Italy, August 2017.

L. Spinelli and M. Crovella, "Closed-Loop Opinion Formation", *Proceedings of the 9th International ACM Web Science Conference (WebSci),* Troy, NY, June 2017.

N. Ruchansky, E. Terzi, and M. Crovella, "Targeted Matrix Completion," *Proceedings of the SIAM International Conference on Data Mining (SDM),* Houston, TX, April 2017.

C. Mavroforakis, D. Erdös, M. Crovella, and E. Terzi, "Active Positive-Definite Matrix Completion," *Proceedings of the SIAM International Conference on Data Mining (SDM),* Houston, TX, April 2017.

G. Comarela, E. Terzi, and M. Crovella, "Detecting Unusually-Routed ASes: Methods and Applications," *Proceedings of the ACM Internet Measurement Conference (IMC),* Santa Monica, CA, November 2016.

N. Ruchansky, M. Crovella, and E. Terzi, "Matrix Completion with Queries," *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD),* Sydney, Australia, August 2015.

G. Comarela and M. Crovella, "Identifying and Analyzing High Impact Routing Events with PathMiner," *Proceedings of the ACM Internet Measurement Conference (IMC),* Vancouver, BC, November 2014.

B. Viswanath, M. A. Bashir, M. Crovella, S. Guha, K. Gummadi, B. Krishnamurthy, and A. Mislove, "Towards Detecting Anomalous User Behavior in Online Social Networks," *Proceedings of the 23rd USENIX Security Symposium*, San Diego, CA, August 2014.

Y. Zhang, T. Lappas, M. Crovella, and E. Kolaczyk, "Online ratings: convergence towards a positive perspective?," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP),* Florence, Italy, May 2014.

G. Comarela, G. Gürsun, and M. Crovella, "Studying Interdomain Routing over Long Timescales," *Proceedings of the ACM Internet Measurement Conference (IMC),* Barcelona, Spain, October 2013.

L. Spinelli, M. Crovella and B. Eriksson, "AliasCluster: A Lightweight Approach to Interface Disambiguation," *Proceedings of the Global Internet Symposium,* Turin, Italy, April 2013.

B. Eriksson and M. Crovella, "Understanding Geolocation Accuracy using Network Geometry," in *Proceedings of the Infocom 2013 Miniconference,* Turin, Italy, April 2013.

J. M. Agosta, J. Chandrasekar, M. Crovella, N. Taft, and D. Ting, "Mixture Models of Endhost Network Traffic," in *Proceedings of the Infocom 2013 Miniconference,* Turin, Italy, April 2013.

G. Gürsun, N. Ruchansky, E. Terzi, and M. Crovella, "Routing State Distance: A Path-based Metric for Network Analysis," in *Proceedings of the ACM Internet Measurement Conference (IMC),* Boston, MA, November 2012. **Winner of a 2013 IETF/IRTF Applied Networking Research Prize.**

G. Gürsun and M. Crovella, "On Traffic Matrix Completion in the Internet," in *Proceedings of the ACM Internet Measurement Conference (IMC),* Boston, MA, November 2012.

T. Lappas, M. Crovella, and E. Terzi, "Selecting a Characteristic Set of Reviews," in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2012,* Beijing, China, August 2012.

Q. Ding, N. Katenka, P. Barford, E. Kolaczyk, and M. Crovella, "Intrusion as (Anti)social Communication: Characterization and Detection," in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2012,* Beijing, China, August 2012. **Best Student Paper Award.**

G. Gürsun, N. Ruchansky, E. Terzi, and M. Crovella, "Inferring Visibility: Who's (not) Talking to Whom?," in *Proceedings of SIGCOMM 2012,* Helsinki, Finland, August 2012.

B. Eriksson and M. Crovella, "Estimating Intrinsic Dimension Via Clustering," in *IEEE Statistical Signal Processing Workshop (SSP),* Ann Arbor, MI, August 2012.

C. Thomas, J. Sommers, P. Barford, D. Kim, A. Das, R. Segebre, M. Crovella, "A Passive Measurement System for Network Testbeds," in *Tridentcom 2012,* Thessaloniki, Greece, June 2012.

G. Comarela, M. Crovella, V. Almeida and F. Benevenuto, "Understanding Factors that Affect Response Rates in Twitter," in *ACM Hypertext 2012,* Milwaukee, WI, June 2012.

A. Cvetkovski and M. Crovella, "Low-Stretch Greedy Embedding Heuristics," in *Fourth International Workshop on Network Science for Communication Networks (NetSciCom 2012),* Orlando, FL, March 2012.

V. Manfredi, M. Crovella, and J. Kurose, "Understanding Stateful vs Stateless Communication Strategies for Ad hoc Networks," in *Proceedings of Mobicom 2011,* Las Vegas, NV, September 2011.

G. Gürsun, M. Crovella, and I. Matta, "Describing and Forecasting Video Access Patterns," in *Proceedings of the Infocom 2011 Miniconference,* Shanghai, China, April 2011.

V. Bharti, P. Kankar, L. Setia, G. Gürsun, and M. Crovella, "Inferring Invisible Traffic," in *Proceedings of CoNEXT 2010,* Philadelphia, PA, December 2010.

J. Sommers, P. Barford and M. Crovella, "Router Primitives for Programmable Active Measurement," in *Workshop on Programmable Routers for Extensible Services of Tomorrow (PRESTO '09)*, Barcelona, Spain, August, 2009.

A. Cvetkovski and M. Crovella, "Hyperbolic Embedding and Routing for Dynamic Graphs," in *Proceedings of Infocom 2009*, Rio de Janeiro, Brazil, April 2009.

V. Erramilli and M. Crovella, "Forwarding in Opportunistic Networks with Resource Constraints," in *Proceedings of the Fourth ACM Workshop on Challenged Networks (CHANTS 08)*, San Francisco, CA, September, 2008.

N. Fonseca, M. Crovella, and K. Salamatian, "Long Range Mutual Information," in *Proceedings of the First Workshop on Hot Topics in Measurement and Modeling of Computer Systems (Hotmetrics '08),* Annapolis, MD, June, 2008.

V. Erramilli, A. Chaintreau, M. Crovella, and C. Diot, "Delegation Forwarding," in *Proceedings of the 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing (ACM MobiHoc),* Hong Kong SAR, China, May 2008.

P. Chhabra, C. Scott, E. D. Kolaczyk, and M. Crovella, "Distributed Spatial Anomaly Detection," in *Proceedings of Infocom 2008*, Phoenix, AZ, April 2008.

V. Erramilli, A. Chaintreau, M. Crovella, and C. Diot, "Diversity of Forwarding Paths in Pocket Switched Networks," in *Proceedings of the ACM SIGCOMM Internet Measurement Conference*, San Diego, October, 2007.

B. Eriksson, P. Barford, R. Nowak, and M. Crovella, "Learning Network Structure from Passive Measurements," in *Proceedings of the ACM SIGCOMM Internet Measurement Conference*, San Diego, October, 2007.

V. Erramilli, M. Crovella, and N. Taft, "An Independent-Connection Model for Traffic Matrices," in *Proceedings of the ACM SIGCOMM Internet Measurement Conference*, Rio de Janeiro, October, 2006.

X. Li, F. Bian, M. Crovella, C. Diot, R. Govindan, G. Iannaccone, and A. Lakhina, "Detection and Identification of Network Anomalies Using Sketch Subspaces," in *Proceedings of the ACM SIGCOMM Internet Measurement Conference*, Rio de Janeiro, October, 2006.

A. Lakhina, M. Crovella, and C. Diot, "Detecting Distributed Attacks using Network-Wide Flow Traffic," in *Proceedings of FloCon 2005 Analysis Workshop*, New Orleans, LA, September, 2005.

A. Lakhina, M. Crovella, and C. Diot, "Mining Anomalies Using Traffic Feature Distributions," in *Proceedings of ACM SIGCOMM*, Philadelphia, PA, August 2005.

D. B. Chua, E. D. Kolaczyk, and M. Crovella, "A Statistical Framework for Efficient Monitoring of End-to-End Network Properties," (Poster) in *Proceedings of ACM SIGMETRICS*, Banff, Canada, June 2005.

B. Donnet, P. Raoult, T. Friedman, and M. Crovella, "Efficient Algorithms for Large-Scale Topology Discovery," in *Proceedings of ACM SIGMETRICS*, Banff, Canada, June 2005.

A. Soule, A. Lakhina, N. Taft, K. Papagiannaki, K. Salamatian, A. Nucci, M. Crovella, and C. Diot, "Traffic Matrices: Balancing Measurements, Inference and Modeling," in *Proceedings of ACM SIGMETRICS*, Banff, Canada, June 2005.

B. Donnet, T. Friedman, and M. Crovella, "Improved Algorithms for Network Topology Discovery," in *Lecture Notes in Computer Science, Proceedings of the 2005 Passive and Active Measurement Workshop*, Boston, MA, March, 2005.

D. B. Chua, E. D. Kolaczyk, and M. Crovella, "Efficient Monitoring of End-to-End Network Properties," in *Proceedings of Infocom 2005*, Miami, FL, March 2005.

N. Fonseca and M. Crovella, "Bayesian Packet Loss Detection for TCP," in *Proceedings of Infocom 2005*, Miami, FL, March 2005.

B. Gueye, A. Ziviani, M. Crovella, and S. Fdida, "Constraint-Based Geolocation of Internet Hosts," in *Proceedings of the ACM SIGCOMM Internet Measurement Conference*, Taormina, Italy, October 2004.

A. Lakhina, M. Crovella, and C. Diot, "Characterization of Network-Wide Anomalies in Traffic Flows," in *Proceedings of the ACM SIGCOMM Internet Measurement Conference*, Taormina, Italy, October 2004.

A. Lakhina, M. Crovella, and C. Diot, "Diagnosing Network-Wide Traffic Anomalies," in *Proceedings of ACM SIGCOMM*, Portland, OR, August 2004.

A. Lakhina, M. Crovella, and C. Diot, "Exploring the Subspace Method for Network-Wide Anomaly Diagnosis," (Poster) in *Network Troubleshooting Workshop: Research, Theory and Operations*, Portland, OR, August 2004.

A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E. D. Kolaczyk, and N. Taft, "Structural Analysis of Network Traffic Flows," in *Proceedings of ACM SIGMETRICS / Performance 2004*, New York, NY, June 2004.

L. Tang and M. E. Crovella, "A Geometric Exploration of the Landmark Selection Problem," in *Lecture Notes in Computer Science 3015, Proceedings of the 2004 Passive and Active Measurement Workshop (PAM2004)*, pp. 63–72, Juan-les-Pins, France, April 2004.

L. Tang and M. E. Crovella, "Virtual Landmarks for the Internet," in *Proceedings of the 2003 Internet Measurement Conference*, Miami, Florida, October 2003.

R. Fonseca, V. Almeida, M. E. Crovella, and B. Abrahão, "On the Intrinsic Locality Properties of Web Reference Streams," in *Proceedings of IEEE Infocom 2003,* San Francisco, California, April 2003.

A. Lakhina, J. Byers, M. E. Crovella, and P. Xie, "Sampling Biases in IP Topology Measurements," in *Proceedings of IEEE Infocom 2003,* San Francisco, California, April 2003.

M. E. Crovella and E. Kolaczyk, "Graph Wavelets for Spatial Traffic Analysis," in *Proceedings of IEEE Infocom 2003,* San Francisco, California, April 2003.

J. Liu, I. Matta, and M. E. Crovella, "End-to-End Inference of Loss Nature in a Hybrid Wired/Wireless Environment," in *Proceedings of Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt'03)*, Sophia-Antipolis, France, March 2003.

A. Lakhina, J. W. Byers, M. E. Crovella, and Ibrahim Matta, "On the Geographic Location of Internet Resources," (Extended Abstract) in *Proceedings of the ACM SIGCOMM Internet Measurement Workshop 2002,* Marseilles, France, November 2002.

J. Liu and M. Crovella, "Using Loss Pairs to Discover Network Properties," in *Proceedings of the ACM SIGCOMM Internet Measurement Workshop 2001,* pp. 127–138, San Francisco, CA, November 2001.

P. Barford, A. Bestavros, J. Byers, and M. Crovella, "On the Marginal Utility of Network Topology Measurements," in *Proceedings of the ACM SIGCOMM Internet Measurement Workshop 2001,* pp. 5–17, San Francisco, CA, November 2001.

L. Guo, I. Matta, and M. E. Crovella, "How Does TCP Generate Pseudo-Self-Similarity?," in *Proceedings of MASCOTS 2001*, Cincinnati, Ohio, August 2001.

P. Barford and M. E. Crovella, "Critical Path Analysis of TCP Transactions," in *Proceedings of the 2000 ACM SIGCOMM Conference*, Stockholm, Sweden, pp. 127–138, September 2000. Also appeared by invitation as "Análisis de secuencia crítica de transacciones TCP" in *Proceedings of SIGCOMM Latin America,* San Antonio de Belén, Costa Rica, April 2001.

M. E. Crovella, R. Frangioso, and M. Harchol-Balter "Connection Scheduling in Web Servers," in *1999 USENIX Symposium on Internet Technologies and Systems (USITS '99).*

P. Barford and M. E. Crovella, "A Performance Evaluation of Hyper Text Transfer Protocols," in *Proceedings of ACM SIGMETRICS '99,* pp. 188–197, Atlanta, Georgia, May 1999.

A. Bestavros, M. E. Crovella, J. Liu, and D. Martin, "Distributed Packet Rewriting and its Application to Scalable Server Architectures," in *Proceedings of the Sixth International Conference on Network Protocols (ICNP '98),* pp. 290–297, October, 1998.

M. Harchol-Balter, M. E. Crovella, and C. D. Murta, "On Choosing a Task Assignment Policy for a Distributed Server System," in *Proceedings of Performance Tools '98, Lecture Notes in Computer Science,* No. 1469, pp. 231–242, September 1998.

P. Barford and M. E. Crovella, "Generating Representative Web Workloads for Network and Server Performance Evaluation,"in *Proceedings of ACM SIGMETRICS/Performance '98,* pp. 151–160, Madison, Wisconsin, July 1998.

M. E. Crovella, M. Harchol-Balter, and C. D. Murta, "Task Assignment in a Distributed System: Improving Performance by Unbalancing Load," in *Proceedings of SIGMETRICS '98 (Poster paper),* Madison, Wisconsin, July 1998.

M. E. Crovella and P. Barford, "The Network Effects of Prefetching," in *Proceedings of IEEE Infocom '98, the Seventeenth Annual Joint Conference of the IEEE Computer and Communication Societies,* San Francisco, California, pp. 1232–1240, April 1998.

M. E. Crovella and L. Lipsky, "Long-Lasting Transient Conditions in Simulations with Heavy-Tailed Workloads," in *Proceedings of the 1997 Winter Simulation Conference*, December 1997.

K. Park, G. Kim, and M. E. Crovella, "On the Effect of Traffic Self-Similarity on Network Performance," in *Proceedings of the SPIE International Conference on Performance and Control of Network Systems*, November, 1997.

P. Fiorini, L. Lipsky, and M. E. Crovella, "Consequences of Ignoring Self-Similar Data Traffic In Communications Modeling," In *Tenth International Conference on Parallel and Distributed Computing Systems (PDCS-97),* New Orleans, LA, pp. 322-327, October 1997.

R. L. Carter and M. E. Crovella, "Server Selection using Dynamic Path Characterization in Wide-Area Networks," in *Proceedings of IEEE Infocom '97, the Sixteenth Annual Joint Conference of the IEEE Computer and Communication Societies,* Kobe, Japan, April 1997.

V. Almeida, A. Bestavros, M. Crovella, and A. de Oliveira, "Characterizing Reference Locality in the WWW," in *Proceedings of the Fourth International Conference on Parallel and Distributed Information Systems (PDIS '96)*, December 1996.

K. Park, G. Kim, and M. E. Crovella, "On the Relationship Between File Sizes, Transport Protocols, and Self-Similar Network Traffic," in *Proceedings of the Fourth International Conference on Network Protocols (ICNP '96),* pp. 171–180, October, 1996.

R. L. Carter and M. E. Crovella, "Measuring Bottleneck Link Speed in Packet-Switched Networks," in *PERFORMANCE '96, the International Conference on Performance Theory, Measurement and Evaluation of Computer and Communication Systems,* October 1996.

M. E. Crovella and A. Bestavros, "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes," in *Proceedings of the 1996 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems,* May 1996. **Winner of a 2010 ACM SIGMETRICS Test of Time Award.**

M. E. Crovella and R. L. Carter, "Dynamic Server Selection in the Internet," In *Proceedings of the Third IEEE Workshop on the Architecture and Implementation of High Performance Communication Subsystems (HPCS '95),* August 1995.

A. Bestavros, R. L. Carter, M. E. Crovella, C. R. Cunha, A. Heddaya, and S. A. Mirdad, "Application-Level Document Caching in the Internet," In *Proceeedings of the Second Workshop on Services in Distributed and Networked Environments (SDNE '95),* June 1995.

M. E. Crovella and T. J. LeBlanc, "Parallel Performance Prediction Using Lost Cycles Analysis," In *Proceedings of Supercomputing '94,* November, 1994.

R. Bianchini, M. E. Crovella, L. Kontothanassis, and T. J. LeBlanc, "Software Interleaving," In *Proceedings of the 6th IEEE Symposium on Parallel and Distributed Processing*, October, 1994.

M. E. Crovella and T. J. LeBlanc, "Performance Debugging using Parallel Performance Predicates," In *3rd ACM/ONR Workshop on Parallel and Distributed Debugging*, May, 1993, pp. 140-150.

M. Crovella, R. Bianchini, T. LeBlanc, E. Markatos, and R. Wisniewski, "Using Communication-to-Computation Ratio in Parallel Performance Prediction and Program Design," *Proceeedings of the Fourth IEEE Symposium on Parallel and Distributed Processing*, December 1992, pp. 238-245.

M. Crovella, P. Das, C. Dubnicki, T. LeBlanc, and E. Markatos, "Multiprogramming on Multiprocessors," *Proceeedings of the Third IEEE Symposium on Parallel and Distributed Processing*, December 1991, pp. 590–597, Also published as TR 385, Department of Computer Science, University of Rochester, February 1991 (revised May 1991).

E. Markatos, M. Crovella, P. Das, C. Dubnicki, and T. LeBlanc, "The Effects of Multiprogramming on Barrier Synchronization," *Proceeedings of the Third IEEE Symposium on Parallel and Distributed Processing*, December 1991, pp. 662-669. Also TR 380, University of Rochester Computer Science Department, May 1991.

N. Bashir, M. Crovella, G. DeTitta, F. Han, H. Hauptman, J. Horvath, H. King, D. Langs, R. Miller, T. Sabin, P. Thuman and D. Velmurugan, "Parallel Solutions to the Phase Problem in X-Ray Crystallography," *Proceedings of The Fifth Distributed Memory Computing Conference*, IEEE Computer Society Press, 1990, pp. 513–521.

M. Crovella, M. S. Summers, E. M. Camardo, and S. S. Weissman, "Exploiting a Restricted Application Domain in SDE Design," *Advance Working Papers of the Second International Workshop on Computer Aided Software Engineering,* IEEE Computer Society Press, July 1988, pp. 2-3–2-4.

J. G. Neal and M. E. Crovella, "An Expert System for Tactical Support Jammer Configuration," *Proceedings of the 1987 IEEE National Aerospace and Electronics Conference (NAECON '87),* May 1987, pp. 1308–1314.

J. G. Neal, M. E. Crovella, and C. Y. Thielman, "A Knowledge-Based Approach to an EW Mission Planning Task," *Proceedings of the Knowledge Based Information Systems Symposium (KBIS-87),* SHAPE Technical Centre, The Hague, Netherlands, 1987.

J. G. Neal, D. F. Funke, and M. E. Crovella, "JAMEX: An Expert Advisor on Tactical Support Jammer Configuration," *Proceedings of the Second Annual Conference on Expert Systems in Government,* IEEE Computer Society Press, November 1986, pp. 383–394.

Refereed Presentations

I. Fried, A. Cannistra, C. Casey, A. Piel, T. Lim, M. Crovella, B. Hescott, "CANDL: Coarsely Aligning Networks with Diffusion and Landmarks," Theme Presentation, ISMB/ECCB, Dublin, July 2015.

UNREFEREED PUBLICATIONS

R. Canetti, A. Cohen, C. Conley, M. Crovella, S. L. Dogan, M. Gaboardi, W. Hartzog, R. Van Loo, C. T. Robertson, K. Baird Silbaugh, "National Telecommunications and Information Administration: Comments from Researchers at Boston University and the University of Chicago," Boston Univ. School of Law Research Paper No. 23-24, June 2023.

M. Crovella, "The Skillful Interrogation of the Internet," in *Computer Communication Review,* vol. 49, no. 3, October 2019.

kc Claffy, M. Crovella, T. Friedman, C. Shannon and N. Spring, "Community-oriented network measurement infrastructure (CONMI) workshop report," in *Computer Communication Review,* vol. 36, no. 2, pp. 41–48, April, 2006.

M. E. Crovella, "How Big is the Average Web Page?" in *Inclusions, the Arts of Mathematics,* Journal of the Mathematical Association of America Student Chapter at Boston University, May 1999.

T. J. LeBlanc, M. L. Scott, B. Marsh, E. Markatos, C. Dubnicki, M. Crovella, and T. Becker, "The Psyche Parallel Operating System," *IEEE TC on Operating Systems Newsletter,* vol. 6, no. 1, 1992, pp. 11-13.

TECHNICAL REPORTS

G. Grassi, C. Barakat, M. Crovella, and R. Teixeira, "Leveraging Website Popularity Differences to Identify Performance Anomalies," (Extended Version), Available at `https://hal.inria.fr/hal-03109717`, January 2021.

J. N. Law, N. Tasnina, M. Kshirsagar, J. Klein-Seetharaman, M. Crovella, P. Rajagopalan, S. Kasif, and T. M. Murali, "Identifying Human Interactors of SARS-CoV-2 Proteins and Drug Targets for COVID-19 using Network-Based Label Propagation," arXiv:2006.01968, June 2, 2020.

B. Rastegarpanah, M. Crovella, and K. P. Gummadi, "Fair Inputs and Fair Outputs: the Incompatibility of Fairness in Privacy and Accuracy," arXiv:2005.09209, May 19, 2020.

M.D.M. Leiserson, J. Fan, A. Cannistra, I. Fried, T. Lim, T. Schaffner, M. Crovella, and B. Hescott, "A Multi-Species Functional Embedding Integrating Sequence and Network Structure," bioRxiv, December 4, 2017.

J.-M. Agosta, J. Chandrasekar, M. Crovella, N. Taft, and D. Ting, "Mixture Models of Endhost Network Traffic," arxiv:1212.2744, December 13, 2012.

M. Hacker, M. Crovella, and L. Reyzin, "Secure Pairing of Mobile Devices," BUCS-TR-2012-011, May 16, 2012.

A. Cvetkovski and M. Crovella, "Multidimensional Scaling in the Poincare Disk," arxiv:1105.5332, May 26, 2011.

A. Lakhina, M. Crovella, and C. Diot, "Diagnosing Network-Wide Traffic Anomalies," BUCS-TR-2004-008, February 24, 2004. Also published as a Sprint Research Report.

A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E. D. Kolaczyk, and N. Taft, "Structural Analysis of Network Traffic Flows," BUCS-TR-2003-021, November 10, 2003. Also published as Sprint Research Report RR03-ATL-110708.

M. Fayed, P. Krapivsky, J. Byers, M. Crovella, and D. Finkel, "On the Size Distribution of Autonomous Systems," BUCS-TR-2003-001, January 17, 2003.

R. Fonseca, V. Almeida, M. E. Crovella, and B. Abrahão, "On the Intrinsic Locality Properties of Web Reference Streams," BUCS-TR-2002-022, July 15, 2002.

A. Lakhina, J. Byers, M. E. Crovella, and P. Xie, "Sampling Biases in IP Topology Measurements," BUCS-TR-2002-021, July 15, 2002.

M. E. Crovella and E. Kolaczyk, "Graph Wavelets for Spatial Traffic Analysis," BUCS-TR-2002-020, July 15, 2002.

A. Lakhina, J. W. Byers, M. E. Crovella, and Ibrahim Matta, "On the Geographic Location of Internet Resources," BUCS-TR-2002-015, 2002.

J. Liu, I. Matta, and M. E. Crovella, "End-to-End Inference of Loss Nature in a Hybrid Wired/Wireless Environment," BUCS-TR-2002-008, March 14, 2002.

M. E. Crovella, R. Frangioso, and M. Harchol-Balter "Connection Scheduling in Web Servers," BUCS-TR-1999-003, March 31, 1999.

M. Harchol-Balter, M. E. Crovella, and C. D. Murta, "On Choosing a Task Assignment Policy for a Distributed Server System," MIT Laboratory for Computer Science, MIT-LCS-TR-757, 1998.

P. Barford and M. E. Crovella, "Generating Representative Web Workloads for Network and Server Performance Evaluation," BUCS-TR-1997-006, November 4, 1997.

M. E. Crovella, M. Harchol-Balter, and C. D. Murta, "Task Assignment in a Distributed System: Improving Performance by Unbalancing Load," BUCS-TR-1997-018, October 31, 1997.

M. Harchol-Balter, M. E. Crovella, and C. D. Murta, "To Queue or Not to Queue: When Queueing is Better Than Timesharing in a Distributed System," BUCS-TR-1997-017, October 31, 1997.

K. Park, G. Kim, and M. E. Crovella, "On the Effect of Traffic Self-Similarity on Network Performance," CSD-TR 97-024, Dept. of Computer Sciences, Purdue University, July 1997.

M. Crovella and P. Barford, "The Network Effects of Prefetching," BUCS-TR-97-002, February 7, 1997.

K. Park, G. Kim, and M. E. Crovella, "On the Relationship Between File Sizes, Transport Protocols, and Self-Similar Network Traffic," BUCS-TR-96-016, August 7, 1996.

V. Almeida, A. Bestavros, M. Crovella, and A. deOliveira, "Characterizing Reference Locality in the WWW," BUCS-TR-96-011, June 21, 1996.

A. Bestavros, M. Chen, M. Crovella, A. Heddaya, S. Sclaroff, and J. Cowie, "Responsive Web Computing: Resource Management, Protocol Techniques, and Applications (A research statement)," BUCS-TR-96-008, March 21, 1996.

R. L. Carter and M. E. Crovella, "Dynamic Server Selection Using Bandwidth Probing in Wide-Area Networks," BUCS-TR-96-007, Boston University Computer Science Department, March 18, 1996.

R. L. Carter and M. E. Crovella, "Measuring Bottleneck Link Speed in Packet-Switched Networks," Technical Report BUCS-TR-96-006, Boston University Computer Science Department, March 15, 1996.

M. E. Crovella and A. Bestavros, "Explaining World Wide Web Traffic Self-Similarity," Technical Report BUCS-TR-95-015 (Revised), Boston University Department of Computer Science, October 1995.

M. E. Crovella, T. J. LeBlanc, and W. Meira Jr., "Performance Measurement and Modeling with the Lost Cycles Toolkit," Technical Report 580, University of Rochester Computer Science Department, June 1995.

C. Cunha, A. Bestavros, and M. Crovella, "Characteristics of WWW Client-based Traces," Technical Report BUCS-TR-95-010, Boston University Department of Computer Science, April 1995.

A. Bestavros, R. L. Carter, M. E. Crovella, C. R. Cunha, A. Heddaya, and S. A. Mirdad, "Application-Level Document Caching in the Internet," Technical Report BUCS-TR-95-002, Boston University Department of Computer Science, February 1995.

M. E. Crovella, "Performance Prediction and Tuning of Parallel Programs," Technical Report 573 and Ph.D. Thesis, Computer Science Dept., U. Rochester, February 1995.

M. E. Crovella and T. J. LeBlanc, "The Search for Lost Cycles: a New Approach to Parallel Program Performance Evaluation," Technical Report 479, University of Rochester Computer Science Department, December 1993.

L. A. Crowl, M. E. Crovella, T. J. LeBlanc, and M. L. Scott, "Beyond Data Parallelism: The Advantages of Multiple Parallelizations in Combinatorial Search," Technical Report 92-80-06, December 1992, Department of Computer Science, Oregon State University. Revised version appears as Technical Report 451, U. Rochester Department of Computer Science, April 1993.

R. Bianchini, M. E. Crovella, L. Kontothanassis, and T. J. LeBlanc, "Alleviating Memory Contention in Matrix Computations on Large-Scale Shared-Memory Multiprocessors," Technical Report 449, U. Rochester Computer Science Department, April 1993.

R. Bianchini, M. E. Crovella, L. Kontothanassis, and T. J. LeBlanc, "Memory Contention in Scalable Cache-Coherent Multiprocessors," Technical Report 448, U. Rochester Computer Science Department, April 1993.

D. Bergmark and M. Crovella, "Case Study in KSR Programming: Finding Outliers by the Minimum Volume Ellipsoid Method," Technical Report CTC92TR107, October 1992, Center for Theory and Simulation in Science and Engineering, Cornell University, Ithaca, NY 14853.

Recent Talks

"Beyond accuracy: understanding the performance of LLMs on exams designed for humans,"

- UFMG, Department of Computer Science, March 8, 2024.

"Estimating Traffic Flows that Can't Be Measured"

- Google Telemetry Seminar Series, May 17, 2021.

"Assessing Candidate Preference Through Web Browsing History"

- Sorbonne Université (UPMC) LIP6, April 15, 2019.
- INRIA Sophia-Antipolis, March 20, 2019.
- INRIA Saclay, December 13, 2018.
- INRIA Paris, November 8, 2018.
- LINCS Laboratory, Paris, October 10, 2018.

"Human Behavior is Low Dimensional"

- TMA Experts Summit 2019, Paris, June 18, 2019.
- Chinese Academy of Sciences, Institute of Computing Technology, Beijing, May 20, 2019.
- Huawei Network AI Workshop, Paris, April 9, 2019.
- LINCS Laboratory, Paris, February 13, 2019.
- Wesleyan University Department of Mathematics and Computer Science, April 13, 2017.
- IEEE Workshop on Data Science for Networking (DS4N), December 14, 2017.

"The Statistics of Human Activity and Implications for Security"

- BU Statistics Club, Professor Speaker Series, November 10, 2016.

"Analyzing the Evolution of Internet Routing through Path Analysis"

- Cisco FAST Seminar Series, July 17, 2015.
- Raytheon BBN Technologies, SDP Seminar, September 28, 2015.

"Reversing the Curse of Dimensionality for Security Analysis"

- NYU Poly Department of Computer Science, December 5, 2014.

"Anomaly Detection and Dimensionality Reduction"

- Workshop on Mathematics of Data Analysis in Cybersecurity, ICERM, Brown University, October 24, 2014.

"Targeted Matrix Completion"

- 51st Annual Allerton Conference on Communication, Control, and Computing, University of Illinois at Urbana-Champaign, October 3, 2013.

"Better Routing with Hyperbolic Geometry"

- Laboratory of Information, Networking and Communication Sciences (LINCS) Paris, December 19, 2012.

"A Fine Grained Distance Metric for Small Worlds"

- Université de Liège, Liège Belgium, January 11, 2013.
- Laboratory of Information, Networking and Communication Sciences (LINCS) Paris, Mar 20, 2012.

"Inferring Invisible Traffic"

- Williams College, Dept. of Computer Science, Williamstown, MA, November 22, 2013.
- Tufts University, Dept. of Computer Science, Medford, MA, November 7, 2013.
- University of Miami, Dept. of Computer Science, Miami, FL, March 2, 2011.
- International Computer Science Institute (ICSI), Berkeley, CA, July 20, 2010.
- Statistics of Networks Workshop, Isaac Newton Institute for Mathematical Sciences, Cambridge, UK, June 24, 2010.

"Network Monitoring: Challenges and Opportunities"

- Indian Institute of Technology (IIT), Delhi, India, June 9, 2008.

"Exploiting Feature Distributions in Anomaly Diagnosis"

- Dartmouth, November 28, 2005.
- MIT CSAIL-LIDS Networking and Systems Seminar, Cambridge MA, April 11, 2005.

"Network Anomaly Diagnosis"

- Boston Area Networking and Networkers Annual Summit (BANANAS-2005), BBN Technologies, Cambridge MA, January 20, 2005.

"Applying the Subspace Method to Network Traffic Analysis"

- Statistical Research Center for Complex Systems (SRCCS), Workshop on "Internet Measurement, Modeling, and Analysis," Seoul National University, Seoul, South Korea, January 11, 2005.
- Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea, January 13, 2005.

"Diagnosing Network-Wide Traffic Anomalies"

- Columbia University, Nov. 17, 2004.
- Workshop on Internet Signal Processing (WISP), CAIDA, Nov. 11, 2004.
- Ecole Normale Supérieure, Paris, Apr. 13, 2004.
- France Telecom R&D, Paris, Mar. 18, 2004.

"Structural Analysis of Network Traffic Flows"

- Project Metropolis Meeting, École Nationale Supérieure des Télécommunications de Bretagne (ENST), Rennes, Feb. 10, 2004.

- Institute for Mathematics and its Applications, Workshop on "Measurement, Modeling and Analysis of the Internet," Minneapolis, MN, January 12, 2004.

"Virtual Landmarks for the Internet"

- Intel Research, Cambridge UK, October 1, 2003.
- Ecole Polytechnique Fédérale de Lausanne (EPFL), November 10, 2003.
- Laboratoire d'Informatique de Paris VI (LIP6), November 14, 2003.

"Sampling Biases in IP Topology Measurements"

- Séminaire Graphes, Réseaux et Modélisation, Laboratoire de Recherche en Informatique, Université d'Orsay, January 8, 2004.
- IBM T.J. Watson Labs, April 11, 2003.
- BU Statistics Colloquium, April 10, 2003.
- National Institute of Statistical Sciences, March 28, 2003.