

IN THE UNITED STATES DISTRICT COURT
FOR THE WESTERN DISTRICT OF TEXAS
MIDLAND/ODESSA DIVISION

DIALECT, LLC,

Plaintiff,

v.

META PLATFORMS, INC.,

Defendant.

Case No. 7:25-cv-00060-DC-DTG

**DECLARATION OF DR. MARK E. CROVELLA, PH.D. REGARDING
CLAIM CONSTRUCTION OF CERTAIN TERMS IN U.S. PATENT NO. 7,398,209**



Dr. Mark E. Crovella

Dated: November 21, 2025

I, Mark E. Crovella, Ph.D., hereby declare and state as follows:

1. I have been retained on behalf of Defendant Meta Platforms, Inc. (“Meta”), in connection with the above-captioned litigation, as an independent technical expert to offer my opinions regarding the scope and meaning that would have been given to certain terms in the claims of U.S. Patent Nos. 7,398,209 (“the ’209 patent”) by a person of ordinary skill in the art at the time of the alleged inventions.

2. I am being compensated for my time at a rate of \$650 per hour, which is my usual and customary charge for providing independent technical consulting and expert analysis and opinions. My compensation does not depend on the outcome of this litigation or on the opinions I offer, and I have no personal interest in the outcome of this litigation.

I. BACKGROUND & QUALIFICATIONS

3. I am a Professor in the Department of Computer Science at Boston University. From 2013 to 2018 I served as Chair of the Department. A brief summary of my background is as follows.

4. 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.”

5. 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.

6. 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 from 2013 to 2018.

7. I am the author of over two hundred research papers. According to Google Scholar, my work has been cited by over 30,000 other publications. Also 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 the field of networking: *Computer Communication Review*, *IEEE/ACM Transactions on Networking*, *Computer Networks*, and *IEEE Transactions on Computers*.

8. A more detailed overview of my education, employment, training, experience, and list of technical papers and books are contained in the attached Curriculum Vitae.

II. SCOPE OF ASSIGNMENT AND SUMMARY OF OPINIONS

9. I have been asked to consider and offer my opinion on whether certain terms appearing in claim 1 of the '209 patent inform, with reasonable certainty, those of ordinary skill in the art about the scope of the claimed invention.

10. Specifically, I have been asked to consider and offer my opinion on whether the element of claim 1 reciting “usable at each stage of processing the received user utterance” informs, with reasonable certainty, those of ordinary skill in the art about the scope of the claimed invention.

11. In my opinion, this claim element in the '209 patent fails to inform, with reasonable certainty, those of ordinary skill in the art about the scope of the claimed invention. The bases for my opinion are set forth below.

III. MATERIALS CONSIDERED

12. The analysis and opinions set forth herein are based on my knowledge, skill, experience, training, and education. I have also considered various materials for this assignment, which include the '209 patent itself, the prosecution history of the '209 patent, and the petition for *inter partes* review and supporting Declaration of Daniel Lopresti filed by Meta in proceeding IPR2025-01334, as well as other materials cited herein.

13. As noted below, I qualified as a person of ordinary skill in the art at the time of the alleged invention of the '209 patent. *See* ¶ 24, *infra*. Accordingly, in addition to viewing the issues discussed herein through the lens of a hypothetical person of ordinary skill in the art, I have also been able to apply my own knowledge and experience as a person of ordinary skill in the art as of the time of the alleged invention, based on my memory and recollection of the knowledge and experience I had at the time as a person of ordinary skill in the art.

IV. UNDERSTANDING OF RELEVANT LEGAL PRINCIPLES

14. The following paragraphs set forth my understanding of the relevant legal principles for construing terms in a U.S. patent, which I was informed of by counsel for Meta and asked to apply for purposes of my opinions.

15. It is my understanding that a purpose of claim construction is to determine the meaning and scope of patent claims.

16. It is my understanding that in district court litigation, claim terms in a U.S. patent are generally given the ordinary and customary meaning that the terms would have had, in light

of the claims, specification, and prosecution history, to a person of ordinary skill in the art in question as of the earliest claimed priority date of the patent.

17. It is my understanding that a person of ordinary skill in the art is deemed to read claim terms not only in the context of the particular claim in which the term appears, but in the context of the entire patent, including the specification. It is further my understanding that when construing the claims of a patent, the principal considerations are the plain language of the claim (including the surrounding claim language and context), the specification, and the prosecution history, which I understand are collectively referred to as the “intrinsic evidence.” It is also my understanding 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. It is also my understanding 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 or broader than it might otherwise be.

18. It is further my understanding that extrinsic evidence, such as dictionary definitions, may be considered in construing claim terms, although such evidence may be accorded less weight where it contradicts a meaning apparent from the intrinsic evidence.

19. It is my understanding that if a claim or claim term, read in light of the patent specification and the prosecution history, fails to inform, with reasonable certainty, those of ordinary skill in the art about the scope of the invention, then the claim is considered indefinite.

20. It is further my understanding that indefiniteness cannot be raised as a basis for challenging the validity of a patent in *inter partes* review proceedings before the U.S. Patent & Trademark Office.

V. '209 PATENT

A. Level of Ordinary Skill in the Art

21. According to the “Field of the Invention” section of the '209 patent specification, the alleged invention of the '209 patent relates to “the retrieval of online information or processing of commands through a speech interface.”¹ As further background, the “Background of the Related Art” section of the specification explains that “[c]ognitive research on human interaction shows that a person asking a question or giving a command typically relies heavily on context and the domain knowledge of the person answering,” whereas, “[o]n the other hand, machine-based queries (e.g., questions, commands, requests, and/or other types of communications) may be highly structured and are not inherently natural to the human user.”² Thus, according to the specification, “human questions and machine processing of queries may be fundamentally incompatible.”³ The specification asserts that, nonetheless, “the ability to allow a person to make natural language speech-based queries remains a desirable goal.”⁴ The specification further explains in its “Background of the Invention” section that “[m]uch work . . . has been done in the fields of natural language processing and speech recognition” and that “[s]peech recognition has steadily improved in accuracy and today is successfully used in a wide range of applications.”⁵ According to the specification, however, “current systems are generally unable to provide a complete environment for users to make natural language speech queries and receive natural sounding responses.”⁶ The specification asserts that, in one aspect of the

¹ '209 patent at 1:13–15.

² '209 patent at 1:17–34.

³ '209 patent at 1:34–35.

⁴ '209 patent at 1:35–37.

⁵ '209 patent at 1:38–41.

⁶ '209 patent at 1:43–46.

invention, “systems are provided which may overcome the deficiencies of prior systems through the application of a complete speech-based information query, retrieval, presentation and command environment,” which “makes maximum use of context, prior information, domain knowledge, and user specific profile data to achieve a natural environment for one or more users making queries or commands in multiple domains.”⁷ As reflected by the foregoing, certain aspects of the alleged invention deal with automated speech recognition. For example, the specification explains that Figure 4A in the patent illustrates “a process **400** for receiving natural language speech-based queries and/or commands and generating a response using the system **90** according to another embodiment of the invention.”⁸ According to the specification, “[t]he process **400** may begin when a user’s utterance (i.e., user query and/or command) is captured through accurate speech recognition operating in a variety of real-world environments at **402**.”⁹

22. Based on my review of the ’209 patent and prosecution history, it is my opinion that a person of ordinary skill in the art (“POSITA”), at the time of the alleged invention and in the field of art of the ’209 patent, would have had at least a bachelor’s degree in computer science, electrical engineering, computer engineering, or a similar field, and at least two years of experience with automatic speech recognition or natural language processing, or equivalent education, experience, or knowledge. Because automatic speech recognition and natural language processing are (and have long been understood to be) interdisciplinary subjects,¹⁰

⁷ ’209 patent at 1:61–2:1.

⁸ ’209 patent at 22:34–37.

⁹ ’209 patent at 22:37–40.

¹⁰ See, e.g., L. Rabiner & B. Juang, *FUNDAMENTALS OF SPEECH RECOGNITION* at 2 (1993) (explaining that one of the challenges of research in automated speech recognition is “its interdisciplinary nature” and identifying, as relevant disciplines, “signal processing,” “physics (acoustics),” “pattern recognition,” “communication and information theory,” “linguistics,” “physiology,” “computer science,” and “psychology”); *id.* (“Successful speech-recognition systems require knowledge and expertise from a wide range of disciplines, a range far larger than

experience with one or more of the relevant disciplines would provide the necessary experience. For example, experience with technologies and techniques used in automatic speech recognition would provide the necessary experience. In my opinion, a person with less education but more relevant experience may also meet this standard, and vice versa.¹¹

23. The '209 patent issued from U.S. patent application Serial No. 10/452,147, which was filed on June 3, 2003, and claims priority to U.S. provisional application Serial No. 60/384,388 filed on June 3, 2002.¹² I have been informed and understand that Plaintiff in this case has asserted that "[t]he priority date of the '209 [p]atent is June 3, 2002"¹³ (the filing date of the above-referenced provisional application) and has not identified any earlier priority date for the alleged invention of the '209 patent, which I am informed that Plaintiff would have been required to disclose in their July 16, 2025 infringement contentions. Therefore, on the basis of the filing date of the provisional application and the absence of any contention by Plaintiff of an earlier date for the alleged invention, I have treated June 2002 as the relevant timeframe for applying the perspective of a POSITA for purposes of my opinion.¹⁴

any single person can possess. Therefore, it is especially important for a researcher to have a good understanding of the fundamentals of speech recognition (so that a range of techniques can be applied to a variety of problems), without necessarily having to be an expert in each aspect of the problem.”).

¹¹ I understand that in litigation involving U.S. Patent No. 8,015,006 (“the '006 patent”), which is in the same patent family as the '209 patent (more specifically, I understand the '006 patent is a continuation of the application that issued as the '209 patent and shares the same specification as the '209 patent), Plaintiff proposed that a POSITA would have had “at least a bachelor-level degree in computer science, computer engineering, electrical engineering, engineering, or a related field, and two years of experience with speech recognition, natural language processing, engineering, and/or equivalent education, research experience, or knowledge.” Plaintiff’s Opening Claim Construction Brief (ECF No. 159) at 8, *Dialect, LLC v. Amazon.com, Inc.*, Civil Action No. 1:23-cv-581 (E.D. Va. filed May 1, 2023). My opinions and analysis herein would remain the same under this slightly different and broader definition of a POSITA offered by Plaintiff for the related '006 patent.

¹² '209 patent at [21]–[22], [60].

¹³ Plaintiff’s Initial Infringement Contentions at 1 (July 16, 2025).

¹⁴ I note, however, that my opinions set forth herein would not change even if the claims of the

24. I note that I qualified as a POSITA at the time of the alleged invention of the '209 patent at least based on my level of education, knowledge, skill, training, and experience at that time. For example, as summarized above, I received a master's degree in Computer Science in 1989 and a Ph.D. in Computer Science in 1994. In addition, my Ph.D. work involved studying aspects of natural language processing. I also had several years of experience and familiarity with hidden Markov models (HMMs), which have been one of the primary and most successful techniques used in speech recognition.¹⁵ Just as one example of this, in early 2002 I co-authored a research paper that explored the application of HMMs to computer network transmissions.¹⁶ I reserve the right to provide further detail on my education, knowledge, skill, training, and experience that may be relevant.

B. “usable at each stage of processing the received user utterance” (claim 1)

25. In my opinion, the element of claim 1 reciting “usable at each stage of processing the received user utterance” fails to inform, with reasonable certainty, those of ordinary skill in the art about the scope of the claimed invention.

'209 patent were entitled to a priority date slightly earlier than June 2002.

¹⁵ See, e.g., H. Singer, *An Overview of Speech Recognition with Applications for Medical Professionals*, BME Vol. 12, No. 11, pp. 27–32 (1999), available at https://www.jstage.jst.go.jp/article/jsmbe1987/13/11/13_11_27/article (“The most successful and widely used acoustic models in recent years have been Hidden Markov Models (HMM). Practically all major speech recognition systems are implemented using some sort of HMM.”); L. Rabiner & B. Juang, *FUNDAMENTALS OF SPEECH RECOGNITION* at 322 (1993) (“In this chapter we will study one well-known and widely used statistical method of characterizing the spectral properties of the frames of a pattern, namely the hidden Markov model (HMM) approach. . . . The underlying assumption of the HMM (or any other type of statistical model) is that the speech signal can be well characterized as a parametric random process, and that the parameters of the stochastic process can be determined (estimated) in a precise, well-defined manner. We will show that the HMM method provides a natural and highly reliable way of recognizing speech for a wide range of applications and integrates well into systems incorporating both task syntax and semantics.”); *id.* at 439 (“As we have shown several times in this book, the most popular way in which speech is modeled is as a left-to-right hidden Markov model.”).

¹⁶ J. Liu, I. Matta, M. Crovella, *End-to-End Inference of Loss Nature in a Hybrid Wired/Wireless Environment*, <https://open.bu.edu/bitstream/handle/2144/1654/2002-008-loss-hmm.pdf?sequence=1>.

26. This element, as it appears in the context of claim 1, is shown below:

1. A method responsive to a user generated natural language speech utterance, comprising:

receiving the user generated natural language speech utterance, the received user utterance containing at least one request;

maintaining a dynamic set of prior probabilities or fuzzy possibilities *usable at each stage of processing the received user utterance*;

recognizing words and phrases contained in the received utterance using information in one or more dictionary and phrase tables;

parsing the recognized words and phrases to determine a meaning of the utterance, wherein determining the meaning includes determining a context for the at least one request contained in the utterance based on one or more keywords contained in the recognized words and phrases;

selecting at least one domain agent based on the determined meaning, the selected domain agent being an autonomous executable that receives, processes, and responds to requests associated with the determined context;

formulating the at least one request contained in the utterance in accordance with a grammar used by the selected domain agent to process requests associated with the determined context;

invoking the selected domain agent to process the formulated request; and

presenting results of the processed request to the user, the presented results generated as a result of the invoked domain agent processing the formulated request.

27. In my opinion, this claim element's failure to inform, with reasonable certainty, those of ordinary skill in the art about the scope of the claimed invention is due to the terms/phrases "usable" and "each stage of processing the received user utterance" as they are used in the context of claim 1. In my opinion, each of these terms/phrases independently creates ambiguity that prevents a POSITA from understanding the boundaries of claim 1 with reasonable certainty. I address each of these in turn below.

1. “usable”

28. I first address the term “usable.” As used in limitation 1(b),¹⁷ “usable” is an adjective that modifies the recited “dynamic set of prior probabilities or fuzzy possibilities” in that it specifies a criterion that a given “dynamic set of prior probabilities or fuzzy possibilities” must meet in order to satisfy the claim language—i.e., that the “dynamic set of prior probabilities or fuzzy possibilities” must be “usable at each stage of processing the received user utterance.” As explained above, I understand that if a claim or claim term, read in light of the patent specification and the prosecution history, fails to inform, with reasonable certainty, those of ordinary skill in the art about the scope of the invention, then the claim is considered indefinite. To put it differently, I understand that, in order to meet the definiteness requirement, a claim or claim term, read in light of the patent specification and the prosecution history, must inform, with reasonable certainty, those of ordinary skill in the art about the scope of the invention. Thus, the question a POSITA would face with respect to “usable” is how to decide, with reasonable certainty, whether a given “dynamic set of prior probabilities or fuzzy possibilities” are “usable” as provided by limitation 1(b), or not. In my opinion, a POSITA would not have been able to decide this question with reasonable certainty.

29. In my opinion, the lack of reasonable certainty that a POSITA would face stems principally from two independent but related aspects of how “usable” is recited in limitation 1(b). The first is that “usable” is recited in the claim language without specifying how, or for what purpose, the recited “dynamic set of prior probabilities or fuzzy possibilities” must be “usable.” To put it differently, to know with reasonable certainty the scope of the claims, a

¹⁷ I use “limitation 1(b)” as shorthand to refer to the second limitation in claim 1, i.e., “maintaining a dynamic set of prior probabilities or fuzzy possibilities usable at each stage of processing the received user utterance.”

POSITA would have asked, and would have needed a reasonably certain answer to: “‘usable’ to do what or for what purpose?” Because neither the claim language nor any intrinsic or extrinsic evidence would have provided a reasonably certain answer on this issue, a POSITA would have lacked reasonable certainty about the scope of claim 1. The second issue is that a POSITA would not have known with reasonable certainty what more is required, in order for a given “dynamic set of prior probabilities or fuzzy possibilities” to be “usable” as provided in limitation 1(b), beyond the “dynamic set of prior probabilities or fuzzy possibilities” merely existing (or being “maintain[ed]” as recited earlier in limitation 1(b)). In other words, a POSITA would not have known with reasonable certainty if it is enough that a given “dynamic set of prior probabilities or fuzzy possibilities” merely exists (or is being “maintain[ed]” as recited earlier in limitation 1(b)), in order for them to be considered “usable” as provided by limitation 1(b), even if they are never actually used, or if something more is required, and if so what that something more is. For example, if a given “dynamic set of prior probabilities or fuzzy possibilities” are “maintain[ed]” by the system in a database, but the system is otherwise not designed to access or utilize them, is that sufficient? Or must the system actually be constructed in such a way that, under appropriate circumstances, the “dynamic set of prior probabilities or fuzzy possibilities” will be accessed and utilized? A POSITA would not have known the answer to this question, one way or the other, with reasonable certainty, and neither the claim language nor any intrinsic or extrinsic evidence provide any guidance on this issue.

30. I first consider whether the intrinsic evidence (which I understand consists of the specification and the prosecution history) provides any guidance on these issues or would otherwise enable a POSITA to decide, with reasonable certainty, whether a given “dynamic set of prior probabilities or fuzzy possibilities” are “usable” as provided by limitation 1(b), or not.

31. In my opinion, the specification of the '209 patent would not have enabled a POSITA to understand how to decide, with reasonable certainty, whether a given “dynamic set of prior probabilities or fuzzy possibilities” are “usable” as provided by limitation 1(b), or not. For example, the specification does not define the term “usable,” nor does it provide any examples of when a given “dynamic set of prior probabilities or fuzzy possibilities” would be considered “usable” as required by limitation 1(b), versus when they would not be considered “usable” as required by limitation 1(b). One portion of the specification does discuss “probabilistic or fuzzy set decision and matching methods” being “applied” at “multiple steps”:

For each user utterance including a question or query or set of questions or queries, the system may perform multiple steps possibly including:

- capture of the user’s question or query through accurate speech recognition operating in a variety of real-world environments;
- parse and interpret the question or query;
- determine the domain of expertise required and context, invoking the proper resources, including agents;
- formulate one or more queries to one or more local and/or network data sources or sending appropriate commands to local or remote devices or the system itself;
- perform required formatting, variable substitutions and transformations to modify the queries to a form most likely to yield desired results from the available sources;
- execute the multiple queries or commands in an asynchronous manner and dealing gracefully with failures;
- extract or scrape the desired information from the one or more results, which may be returned in any one of a number of different formats;
- evaluate and interpret the results, including processing of errors, gathered and combine them into a single best result judged to be “best” even if the results are ambiguous, incomplete, or conflicting;
- perform required formatting, variable substitutions and transformations to modify the results to a form most easily understood by the user;
- and
- present the compound result, through a text to speech engine, to the user in a useful and expected manner.

The above steps preferably are performed while accounting for the domain of expertise required, the context for the question or command, domain specific information, the history of the user's interaction, user preferences, information sources or commands available, and responses obtained from the sources. At each stage *probabilistic or fuzzy*

set decision and matching methods can be applied to deal with inconsistent, ambiguous, conflicting and incomplete information or responses. In addition, the use of asynchronous queries that may result in rapid and graceful failure of some queries or commands may allow the system to robustly return results quickly and in a manner that seems natural to the user.¹⁸

32. In my opinion, however, this portion of the specification would not have enabled a POSITA to understand how to decide, with reasonable certainty, whether a given “dynamic set of prior probabilities or fuzzy possibilities” are “usable” as provided by limitation 1(b), or not, because it discusses and discloses that probabilistic or fuzzy set decision and matching methods can be “applied,” not whether or when probabilistic or fuzzy set decision and matching methods are (or are not) “usable.” Thus, for example, this portion of the specification would not contribute anything to a POSITA’s understanding of whether or when a given “dynamic set of prior probabilities or fuzzy possibilities” would be “usable” at each stage of processing the received user utterance even when they are not actually “applied” at each stage of processing the received user utterance. In addition, while this portion of the specification refers to probabilistic or fuzzy set decision and matching methods being applied “to deal with inconsistent, ambiguous, conflicting and incomplete information or responses,” that language (“to deal with inconsistent, ambiguous, conflicting and incomplete information or responses”) does not appear in the claim, nor does the specification confine the application of probabilistic or fuzzy set decision and matching methods exclusively to “deal[ing] with inconsistent, ambiguous, conflicting and incomplete information or responses.” (Indeed, as I note below, the specification refers to applying “probabilistic or fuzzy set matching methods” to other tasks.)

33. There are other portions of the ’209 patent specification that similarly refer to “probabilistic or fuzzy set matching methods,” but in each case the specification does not discuss

¹⁸ ’209 patent at 3:46–4:19 (emphasis added).

the concept of “usable” or convey a reasonably certain meaning for this term. For example, one portion of the specification explains:

According to another aspect of the invention, special procedures to present information that is in the form of long text strings, tables, lists or other long response sets may be employed. Simply presenting a long set of information in an ordered manner is not natural nor what most users have in mind. The system, ***using for example, probabilistic or fuzzy set matching methods***, may extract the most relevant information to the user and presents [*sic*] these subsets first.¹⁹

34. In my opinion, this portion of the specification would also not have enabled a POSITA to understand how to decide, with reasonable certainty, whether a given “dynamic set of prior probabilities or fuzzy possibilities” are “usable” as provided by limitation 1(b), or not. The reasons for this are similar to those previously explained for the portion of the specification discussed above (*see* ¶ 32, *supra*). Specifically, this portion of the specification discusses and discloses “using” probabilistic or fuzzy set matching methods, not whether or when probabilistic or fuzzy set matching methods are (or are not) “usable.” Thus, for example, this portion of the specification would not contribute anything to a POSITA’s understanding of whether or when a given “dynamic set of prior probabilities or fuzzy possibilities” would be “usable” at each stage of processing the received user utterance when the system is not actually “using” them. To put it differently, the specification’s discussion of “using” probabilistic or fuzzy set matching methods in relation to a particular activity implies that such methods are “usable” for that activity, but it would not convey anything to a POSITA about when such methods remain “usable” even when not “using” them. For similar reasons, the other portions of the specification that discuss probabilistic or fuzzy set decision methods would not have enabled a POSITA to understand how to decide this issue with reasonable certainty either.²⁰

¹⁹ ’209 patent at 5:36–43 (emphasis added).

²⁰ *See, e.g.*, ’209 patent at 5:22–35 (“In order to make the responses to user’s questions and commands seem more natural, the system may employ one or more dynamically invokeable

35. Based on my review, the prosecution history of the '209 patent likewise would not have enabled a POSITA to understand how to decide, with reasonable certainty, whether a given “dynamic set of prior probabilities or fuzzy possibilities” are “usable” as provided by limitation 1(b), or not. For example, the claims as originally submitted in the application for the '209 patent did not recite “usable” at all.²¹ Following the Examiner’s May 30, 2007 non-final rejection of all pending claims, the applicant amended the claims by (among other things) adding application claim 44 (which eventually issued as claim 1 of the '209 patent).²² As added, application claim 44 recited what is now limitation 1(b) in claim 1 of the '209 patent, “maintaining a dynamic set of prior probabilities or fuzzy possibilities usable at each stage of processing the received user utterance.”²³ However, the applicant’s remarks accompanying the amendment did not discuss limitation 1(b), nor did the Examiner do so at any subsequent point during prosecution. The only discussion of newly added application claim 44 in the applicant’s remarks was that “Claims 41-59 are newly added.”²⁴ Following the addition of application claim 44, in an April 18, 2008 Office Action, the Examiner allowed this claim (among others), while

personalities. Personalities have specific characteristics, which simulate the behavioral characteristics of real humans. Examples of these characteristics include, sympathy, irritation, and helpfulness. The personality also randomizes aspects of responses, just as a real human would do. This behavior includes randomization of terms used and the order of presentation of information. Characteristics of the personality are invoked **using** probabilistic or fuzzy set decision and matching methods” (emphasis added)); *id.* at 6:58–61 (“When multiple users are engaged in interleaved sessions, the system gracefully resolves conflicts **using** a probabilistic or fuzzy set decision method.” (emphasis added)); *id.* at 20:57–60 (“The domain agent 156, may select the presentation template, determine order of presentation for tokens and determine variable substitutions and transformations **using** probabilistic or fuzzy set decision methods.” (emphasis added)).

²¹ '209 patent prosecution history, June 3, 2003 Claims.

²² '209 patent prosecution history, Nov. 30, 2007 Response to Non-Final Office Action at 18.

²³ '209 patent prosecution history, Nov. 30, 2007 Response to Non-Final Office Action at 23.

²⁴ '209 patent prosecution history, Nov. 30, 2007 Response to Non-Final Office Action at 30.

rejecting other claims.²⁵ On May 7, 2008, the applicant cancelled all pending claims besides the claims that the Examiner had allowed in the April 18, 2008 Office Action, and the '209 patent issued thereafter, with application claim 44 renumbered as issued claim 1.²⁶ Based on my review, and as reflected in the foregoing summary, neither the applicant nor the Examiner discussed limitation 1(b) or the term “usable” during prosecution of the '209 patent. Thus, in my opinion, nothing in the prosecution history would have enabled a POSITA to understand how to decide, with reasonable certainty, whether a given “dynamic set of prior probabilities or fuzzy possibilities” are “usable” as provided by limitation 1(b), or not.

36. In addition, at the time of the alleged invention, there was no generally accepted understanding available to a POSITA of what “usable” meant in relation to “prior probabilities or fuzzy possibilities” (or a “dynamic set of prior probabilities or fuzzy possibilities”), or even more generally within the context of a “method responsive to a user generated natural language speech utterance,” which the preamble sets forth as the method that the alleged invention is directed to. In other words, a POSITA would not have understood “usable” to have any particular ordinary and customary meaning within the art to which the '209 patent is directed. For example, Fundamentals of Speech Recognition (1993) by Rabiner and Juang, which a POSITA would have considered to be one of the canonical texts on automated speech recognition, discusses “a priori probabilities” in the context of speech recognition, but does not provide any understanding of what “usable” might mean in relation to such probabilities.²⁷ Other texts that a POSITA

²⁵ '209 patent prosecution history, April 18, 2008 Final Rejection.

²⁶ '209 patent prosecution history, May 7, 2008 Response After Final Action at 2–7.

²⁷ See L. Rabiner & B. Juang, FUNDAMENTALS OF SPEECH RECOGNITION at 363 (1993) (“Consider recognition of a vocabulary of V words, each of which is represented by an HMM [Hidden Markov Model], with parameter set λ_v , $v = 1, 2, \dots, V$. We assume $P(v)$ to be the a priori probability for word v , $v = 1, 2, \dots, V$. The set of HMMs $\Lambda = \{\lambda_v\}$ together with the a priori probabilities thus defines a probability measure for an arbitrary observation sequence $\mathbf{O}$ ”).

would have looked to as informative if not authoritative likewise fail to provide any understanding of what “usable” might mean in relation to “prior probabilities or fuzzy possibilities” (or a “dynamic set of prior probabilities or fuzzy possibilities”), or in the context of speech recognition or natural language processing more generally.²⁸

37. Looking even more generally to the field of computer technology, a POSITA would still have not found or been aware of any particular ordinary and customary meaning for “usable.” For example, I have found no contemporaneous technical dictionaries relating to computing technology that would have conveyed to a POSITA any particular ordinary and customary meaning for this term.²⁹

38. Given the absence of any particular ordinary and customary meaning for “usable” in the art, and the absence of any intrinsic evidence in the specification or prosecution history that would have enabled a POSITA to understand how to decide, with reasonable certainty, whether a given “dynamic set of prior probabilities or fuzzy possibilities” are “usable” as provided by limitation 1(b), or not, a POSITA would have been left with trying to apply a lay understanding of “usable.” Such a lay understanding, at the time of the time of the alleged invention, would have been along the lines of “able to be used” or “capable of being used.”³⁰ This lay understanding, however, suffers from the same two issues I described above (*see* ¶ 29,

²⁸ *See generally, e.g.*, J. Holmes & W. Holmes, SPEECH SYNTHESIS AND RECOGNITION (2d ed. 2001); D. Jurafsky & J. Martin, SPEECH AND LANGUAGE PROCESSING: AN INTRODUCTION TO NATURAL LANGUAGE PROCESSING, COMPUTATIONAL LINGUISTICS, AND SPEECH RECOGNITION (2000); C. Baber & J. Noyes, INTERACTIVE SPEECH TECHNOLOGY: HUMAN FACTORS ISSUES IN THE APPLICATION OF SPEECH INPUT/OUTPUT TO COMPUTERS (1993).

²⁹ *E.g.*, D. Pountain, THE NEW PENGUIN DICTIONARY OF COMPUTING: AN A-Z OF COMPUTING JARGON AND CONCEPTS (2001) (no definition for “usable”); IEEE STANDARD DICTIONARY OF ELECTRICAL AND ELECTRONICS TERMS (6th ed. 1996) (same).

³⁰ *See, e.g.*, PAPERBACK OXFORD ENGLISH DICTIONARY at 928 (2002) (defining “usable” as “able to be used”); MERRIAM-WEBSTER’S COLLEGIATE DICTIONARY at 1301 (10th ed. 1996) (defining “usable” as “capable of being used”).

supra). With respect to the first issue, this lay understanding begs the question of how, or for what purpose, the recited “dynamic set of prior probabilities or fuzzy possibilities” must be “usable,” but neither the claim language nor any intrinsic or extrinsic evidence provide a reasonably certain answer to a POSITA. Just as a simple analogy to illustrate the point, a pair of sunglasses may or may not be considered “usable” while mowing a lawn, depending on how, or for what purpose, the sunglasses must be “usable.” For example, the requirement could be that the sunglasses must be “usable” to cut grass (which would likely not be met by most typical sunglasses). The requirement could alternatively be that the sunglasses must be “usable” for any purpose whatsoever, in which case the answer could be yes—a pair of sunglasses are “usable” in the ordinary, everyday sense that one could wear them while mowing a lawn in order to reduce glare, provide a physical barrier from dust and debris (e.g., grass clippings), and provide protection from the sun’s UV rays. Without knowing how, or for what purpose the sunglasses must be “usable,” it is not possible to say with reasonable certainty whether they are “usable” or not. An analogous problem would face a POSITA trying to discern, with reasonable certainty, whether or not a given “dynamic set of prior probabilities or fuzzy possibilities” are “usable” as provided by limitation 1(b): without knowing with reasonable certainty how, or for what purpose, the “dynamic set of prior probabilities or fuzzy possibilities” must be “usable,” a POSITA would be unable to decide (let alone decide with reasonable certainty) whether the “dynamic set of prior probabilities or fuzzy possibilities” are “usable.”

2. “each stage of processing the received user utterance”

39. I next address “each stage of processing the received user utterance.” As used in limitation 1(b), this phrase purports to define when the claimed “dynamic set of prior probabilities or fuzzy possibilities” must be “usable.” (As I explain above, however, the claim does not address how, or for what purpose, the claimed “dynamic set of prior probabilities or

fuzzy possibilities” must be “usable,” which is another reason that, in my opinion, a POSITA would lack reasonable certainty about the scope of the claim. *See* ¶ 29 *et seq.*, *supra.*) To determine whether a given “dynamic set of prior probabilities or fuzzy possibilities” meets the requirement of being “usable at each stage of processing the received user utterance,” an important threshold question that a POSITA would face is what are the “stage[s] of processing the received user utterance.” In my opinion, a POSITA would not have been able to decide this question with reasonable certainty, even reading the claim language in light of the patent specification and the prosecution history. In other words, a POSITA would not have known, with reasonable certainty, what counts as a “stage of processing the received user utterance,” and what doesn’t. This, in turn, would prevent a POSITA from being able to decide, with reasonable certainty, whether any given “dynamic set of prior probabilities or fuzzy possibilities” is “usable” at each of those “stage[s].” Just to illustrate this point, consider a scenario where a given “dynamic set of prior probabilities or fuzzy possibilities” are “usable” only at stages X, Y, and Z. If X, Y, and Z are the only “stage[s] of processing the received user utterance,” then the requirement of “maintaining a dynamic set of prior probabilities or fuzzy possibilities usable at each stage of processing the received user utterance” as recited in limitation 1(b) could be satisfied. However, if A, B, and C are also “stage[s] of processing the received user utterance” (in addition to X, Y, and Z) then the requirement of “maintaining a dynamic set of prior probabilities or fuzzy possibilities usable at each stage of processing the received user utterance” as recited in limitation 1(b) would not be satisfied, because the “dynamic set of prior probabilities or fuzzy possibilities” are “usable” only at stages X, Y, and Z. Without knowing whether the “stage[s] of processing the received user utterance” consist of A, B, C, X, Y, and Z, or just X, Y, and Z, a POSITA simply cannot determine whether a given “dynamic set of prior

probabilities or fuzzy possibilities” are “usable” at “each stage of processing the received user utterance,” as recited in limitation 1(b). In this way, knowing with reasonable certainty what are the “stage[s] of processing the received user utterance” would be essential for a POSITA to understand, with reasonable certainty, the scope of the claim.

40. To begin with, I note that nothing in the claim language itself identifies “each stage of processing the received user utterance” or even a single such “stage.” For example, there is no previous limitation of claim 1 that expressly introduces this claim element and identifies what the “stage[s]” are (which I understand is referred to as an antecedent basis), nor is there any such limitation later in the claim. In fact, the word “stage” does not appear anywhere else in claim 1 or in any other claim of the ’209 patent.

41. There is also no language in the claim that a POSITA would view as implicitly supplying an antecedent basis. While the body of claim 1 recites a number of steps, none of the limitations recite a step of “processing,” much less a step of “processing the received user utterance.” (I note that the same is true of all other claims in the ’209 patent, all of which depend either directly or indirectly from claim 1.) For example, the body of claim 1 recites steps of “receiving,” “maintaining,” “recognizing,” “parsing,” “selecting,” “formulating,” “invoking,” and “presenting.” While other limitations in claim 1 use the root word “process” (or variants such as “processes” or “processed”), a POSITA would not view any of these as implicitly supplying an antecedent basis for “each stage of processing the received user utterance” as recited in limitation 1(b). For example, limitation 1(e)³¹ recites “the selected domain agent being an autonomous executable that receives, processes, and responds to requests associated with the

³¹ I use “limitation 1(e)” as shorthand to refer to the fifth limitation in claim 1, i.e., “selecting at least one domain agent based on the determined meaning, the selected domain agent being an autonomous executable that receives, processes, and responds to requests associated with the determined context.”

determined context.” This use of “processes” in limitation 1(e) is referring to a feature of “the selected domain agent”—i.e., that it must be “an autonomous executable that receives, processes, and responds to requests associated with the determined context.” Accordingly, a POSITA would not view “processes” as it appears in limitation 1(e) as supplying an antecedent basis for “each stage of processing the received user utterance.” The same is true of other uses of “process[es/ed/ing]” in claim 1.³²

42. In my opinion, the specification of the ’209 patent would not have enabled a POSITA to determine, with reasonable certainty, what are the “stage[s] of processing the received user utterance” as recited in limitation 1(b). In other words, the specification would not have enabled a POSITA to know, with reasonable certainty, what counts as a “stage of processing the received user utterance,” and what doesn’t.

43. To begin with, the specification never explicitly discusses “processing the received user utterance,” nor does it ever identify the “stage[s]” thereof. There are a handful of instances where the specification refers to a “stage” of something, but as I explain below, none of these would have enabled a POSITA to determine, with reasonable certainty, what are the “stage[s] of processing the received user utterance” as recited in limitation 1(b).

44. In the first instance, the specification explains in its “Summary of the Invention” section that “several stages” of the “process” use probabilistic and fuzzy reasoning to achieve “robustness to partial failure”:

³² For example, limitation 1(h) recites “presenting results of the processed request to the user, the presented results generated as a result of the invoked domain agent processing the formulated request.” A POSITA would not have viewed the term “processed” as it appears in limitation 1(h) (in “processed request”) as supplying an antecedent basis for “each stage of processing the received user utterance.” Nor would a POSITA have viewed the term “processing” as it appears in limitation 1(h) (in “processing the formulated request”) as supplying an antecedent basis for “each stage of processing the received user utterance.”

In another aspect of the invention, systems are provided which may overcome the deficiencies of prior systems through the application of a complete speech-based information query, retrieval, presentation and command environment. This environment makes maximum use of context, prior information, domain knowledge, and user specific profile data to achieve a natural environment for one or more users making queries or commands in multiple domains. Through this integrated approach, a complete speech-based natural language query and response environment may be created. Further, at each step in the process, accommodation may be made for full or partial failure and graceful recovery. The robustness to partial failure is achieved through the use of probabilistic and fuzzy reasoning at *several stages of the process*. This robustness to partial failure promotes the feeling of a natural response to questions and commands.³³

45. In my opinion, this paragraph of the specification would not have enabled a POSITA to understand, with reasonable certainty, what are the “stage[s] of processing the received user utterance” as recited in limitation 1(b). To begin with, this paragraph does not specifically discuss “processing the received user utterance,” nor does it identify “each stage” or the “stage[s]” thereof. Instead, this paragraph refers to “the process” vaguely, without specifying what “process” is being referred to or even discussing any particular “process.” A POSITA would have no way to know if “the process” in this paragraph is referring to some particular process, or simply the general idea of “a complete speech-based information query, retrieval, presentation and command environment” mentioned earlier in the paragraph. In addition, this paragraph asserts only that probabilistic and fuzzy reasoning are used at “several stages” (of the unspecified “process”) to achieve “robustness to partial failure,” but does not specify what those “several stages” are. As a result, this paragraph of the specification would not have enabled a POSITA to understand, with reasonable certainty, what are the “stage[s] of processing the received user utterance” as recited in limitation 1(b).

³³ '209 patent at 1:61–2:9 (emphasis added).

46. In the second instance where the specification refers to a “stage” of something, which I have excerpted above (*see* ¶ 31, *supra*), the specification’s “Summary of the Invention” section explains that “[a]t each stage probabilistic or fuzzy set decision and matching methods can be applied to deal with inconsistent, ambiguous, conflicting and incomplete information or responses”:

For each user utterance including a question or query or set of questions or queries, the system may perform multiple steps possibly including:

- capture of the user’s question or query through accurate speech recognition operating in a variety of real-world environments;
- parse and interpret the question or query;
- determine the domain of expertise required and context, invoking the proper resources, including agents;
- formulate one or more queries to one or more local and/or network data sources or sending appropriate commands to local or remote devices or the system itself;
- perform required formatting, variable substitutions and transformations to modify the queries to a form most likely to yield desired results from the available sources;
- execute the multiple queries or commands in an asynchronous manner and dealing gracefully with failures;
- extract or scrape the desired information from the one or more results, which may be returned in any one of a number of different formats;
- evaluate and interpret the results, including processing of errors, gathered and combine them into a single best result judged to be “best” even if the results are ambiguous, incomplete, or conflicting;
- perform required formatting, variable substitutions and transformations to modify the results to a form most easily understood by the user;
- and
- present the compound result, through a text to speech engine, to the user in a useful and expected manner.

The above steps preferably are performed while accounting for the domain of expertise required, the context for the question or command, domain specific information, the history of the user's interaction, user preferences, information sources or commands available, and responses obtained from the sources. ***At each stage*** probabilistic or fuzzy set decision and matching methods can be applied to deal with inconsistent, ambiguous, conflicting and incomplete information or responses. In addition, the use of asynchronous queries that may result in rapid and graceful failure of some queries or commands may

allow the system to robustly return results quickly and in a manner that seems natural to the user.³⁴

47. In my opinion, these paragraphs of the specification would likewise not have enabled a POSITA to understand, with reasonable certainty, what are the “stage[s] of processing the received user utterance” as recited in limitation 1(b). To begin with, these paragraphs do not specifically discuss “processing the received user utterance,” nor do they identify “each stage” or the “stage[s]” thereof. While these paragraphs identify a variety of steps that the system “may perform” when a “user utterance includ[es] a question or query or set of questions or queries,” it does not equate those steps with “stage[s] of processing the received user utterance” as recited in limitation 1(b). Moreover, the specification indicates that the list of steps in these paragraphs is non-exhaustive, such that even if the listed steps constituted “stage[s] of processing the received user utterance,” a POSITA would not know with reasonable certainty whether—and if so what—other steps not found in the list might also constitute “stage[s] of processing the received user utterance.” Specifically, this portion of the specification begins by explaining that “[f]or each user utterance including a question or query or set of questions or queries, the system may perform multiple steps *possibly including . . .*”³⁵ These paragraphs would thus convey to a POSITA that “the system” may perform “multiple steps,” such as those appearing in the immediately following list, but that the listed steps are not exhaustive of all the steps that “the system” may perform. In addition, these paragraphs are not clear whether “probabilistic or fuzzy set decision and matching methods” can be applied at all of the listed “steps,” or only those “steps” in which there is a need to “deal with inconsistent, ambiguous, conflicting and

³⁴ '209 patent at 3:46–4:19 (emphasis added).

³⁵ '209 patent at 3:46–48 (emphasis added).

incomplete information or responses.”³⁶ The specification states at the end of these paragraphs that “[a]t each stage probabilistic or fuzzy set decision and matching methods can be applied *to deal with inconsistent, ambiguous, conflicting and incomplete information or responses.*”³⁷ A POSITA would not have been reasonably certain whether the specification is suggesting that “probabilistic or fuzzy set decision and matching methods can be applied” at all of the listed steps, or only the subset of listed steps where there is a need to “to deal with inconsistent, ambiguous, conflicting and incomplete information or responses.”³⁸ As a result, these paragraphs of the specification would not have enabled a POSITA to understand, with reasonable certainty, what are the “stage[s] of processing the received user utterance” as recited in limitation 1(b).

48. In the third instance where the specification refers to a “stage” of something, the specification refers to “each stage of the dialog,” and in particular discusses “dynamically updat[ing]” prior probabilities or fuzzy set possibilities “for the words in the dictionary and phrase tables,” in order to “maximize the probability of correct recognition at each stage of the dialog”:

A user may start a dialog with the system 90 when they first address it. This can be done by speaking a generic word (“computer”) or addressing a specific name (“Fred”), which may be generally tied to a system personality 108. Once the user starts the dialog, it may be recognized by the speech recognition engine 120, using unique characteristics of the user's speech. At the end of a dialog or to interrupt a dialog, the user may utter a dismissal word (“good bye”).

³⁶ '209 patent at 4:14–17.

³⁷ '209 patent at 4:14–17 (emphasis added).

³⁸ In these paragraphs, the specification expressly identifies only one of the listed steps as needing to deal with ambiguous, incomplete, or conflicting information—specifically, the “evaluat[ing] and interpret[ing]” step, which is described as: “evaluate and interpret the results, including processing of errors, gathered and combine them into a single best result judged to be ‘best’ even if the results are ambiguous, incomplete, or conflicting.” '209 patent at 3:66–4:2.

According to another embodiment of the invention, the system 90 may employ a speech recognition engine 120 1.24 seeding for improved word recognition accuracy, using data from the dictionary and phrase tables 112, user profiles 110, and the agents 106. At the same time, *the fuzzy set possibilities or prior probabilities for the words in the dictionary and phrase tables may be dynamically updated to maximize the probability of correct recognition at each stage of the dialog.*³⁹

49. In my opinion, these paragraphs of the specification would likewise not have enabled a POSITA to understand, with reasonable certainty, what are the “stage[s] of processing the received user utterance” as recited in limitation 1(b). To begin with, these paragraphs do not specifically discuss “processing the received user utterance,” nor do they identify “each stage” or the “stage[s]” thereof. Rather, in these paragraphs, the specification refers to “each stage of the dialog,”⁴⁰ which the prior paragraph makes clear is referring to a user’s “dialog with the system 90.”⁴¹ At most, the discussion in these paragraphs of recognizing words in the user’s dialog with the system using “fuzzy set possibilities or prior probabilities for the words in the dictionary and phrase tables” could suggest to a POSITA that limitation 1(c)⁴² might be considered a “stage of processing the received user utterance” as recited in limitation 1(b), but in my opinion a POSITA could not be reasonably certain of that conclusion, because these paragraphs are only discussing “an[] embodiment of the invention”⁴³ and do not identify recognizing words in the user’s dialog with the system as a “stage of processing the received user utterance.”⁴⁴ In addition, a POSITA

³⁹ ’209 patent at 17:1–16 (emphasis added).

⁴⁰ ’209 patent at 17:16.

⁴¹ ’209 patent at 17:1.

⁴² I use “limitation 1(c)” as shorthand to refer to the third limitation in claim 1, i.e., “recognizing words and phrases contained in the received utterance using information in one or more dictionary and phrase tables.”

⁴³ ’209 patent at 17:9.

⁴⁴ I note that in the Declaration of Daniel Lopresti filed in support of Meta’s petition for *inter partes* review in proceeding IPR2025-01334, Dr. Lopresti similarly opined that “[w]hile the ’209 Patent is not clear which of its claimed steps are considered a ‘stage of processing the received user utterance,’ the specification discusses the use of ‘prior probabilities’ and ‘fuzzy

still would not know with reasonable certainty whether—and if so what—other activities might constitute “stage[s] of processing the received user utterance” as recited in limitation 1(b). As a result, these paragraphs of the specification would not have enabled a POSITA to understand, with reasonable certainty, what are the “stage[s] of processing the received user utterance” as recited in limitation 1(b).

50. In the fourth instance where the specification refers to a “stage” of something, the specification essentially repeats the statement I discussed earlier from column 4, with reference to Figure 4A (*see* ¶¶ 46–47, *supra*):

FIG. 4A is a process 400 for receiving natural language speech-based queries and/or commands and generating a response using the system 90 according to another embodiment of the invention. . . . At each ***stage of the process*** 400, probabilistic or fuzzy set decision and matching methods may be applied to deal with inconsistent, ambiguous, conflicting and incomplete information or responses.⁴⁵

51. In my opinion, this paragraph of the specification would likewise not have enabled a POSITA to understand, with reasonable certainty, what are the “stage[s] of processing the received user utterance” as recited in limitation 1(b). To begin with, neither this paragraph nor Figure 4A specifically discusses or describes “processing the received user utterance,” nor do they identify “each stage” or the “stage[s]” thereof. Moreover, the flow diagram (400) depicted in Figure 4A is described by the specification as the process of “receiving natural language speech-based queries and/or commands and generating a response using the system 90,”⁴⁶ not “processing the received user utterance” as recited in limitation 1(b). In addition, this paragraph

possibilities’ as it relates to ‘recognizing step’ 1[c],” citing the same portion of the specification at column 17 that I discuss above. Decl. of Daniel Lopresti ¶ 138, IPR2025-01334 (Aug. 18, 2025).

⁴⁵ ’209 patent at 22:34–23:33.

⁴⁶ ’209 patent at 22:34–36.

is not clear whether “probabilistic or fuzzy set decision and matching methods” may be applied at all of the steps of the process 400 depicted in Figure 4A, or only those steps in which there is a need to “deal with inconsistent, ambiguous, conflicting and incomplete information or responses.”⁴⁷ This paragraph of the specification states that “[a]t each stage of the process 400, probabilistic or fuzzy set decision and matching methods may be applied *to deal with inconsistent, ambiguous, conflicting and incomplete information or responses.*”⁴⁸ A POSITA would not have been reasonably certain whether the specification is suggesting that “probabilistic or fuzzy set decision and matching methods may be applied” at all of the steps depicted in Figure 4A, or only at an unidentified subset of steps where there is a need to “to deal with inconsistent, ambiguous, conflicting and incomplete information or responses.”⁴⁹ As a result, this paragraph of the specification would not have enabled a POSITA to understand, with reasonable certainty, what are the “stage[s] of processing the received user utterance” as recited in limitation 1(b).

52. In another part of column 17, the specification describes a parser “dynamically receiv[ing] keyword and associated prior probability or fuzzy possibility updates” and “determin[ing] a context for an utterance by applying prior probabilities or fuzzy possibilities to keyword matching, user profile 110, and dialog history”:

Once the words and phrases have been recognized by the speech recognition engine 120, the tokens and user identification is passed to the parser 118. The parser 118 examines the tokens for the questions or commands, context

⁴⁷ '209 patent at 23:30–33.

⁴⁸ '209 patent at 23:30–33.

⁴⁹ In this paragraph, the specification expressly identifies only one of the steps depicted in Figure 4A as needing to deal with ambiguous, incomplete, or conflicting information—specifically, the “EVALUATE AND INTERPRET THE RESULTS” step (box 416), which the specification describes as follows: “Next, the results may be evaluated and interpreted including processing of errors, and gathering and combining them into a single best result judged to be ‘best’ even if the results are ambiguous, incomplete, or conflicting at 416.” '209 patent at 23:15–19.

and criteria. *The parser 118 determines a context for an utterance by applying prior probabilities or fuzzy possibilities to keyword matching, user profile 110, and dialog history.* The context of a question or command determines the domain and thereby, the domain agent 156, if any, to be evoked. For example, a question with the keywords “temperature” implies a context value of weather for the question. *The parser dynamically receives keyword and associated prior probability or fuzzy possibility updates* from the system agent 150 or an already active domain agent 156. Based on these probabilities or possibilities the possible contexts are scored and the top one or few are used for further processing.⁵⁰

53. In my opinion, this paragraph of the specification would likewise not have enabled a POSITA to understand, with reasonable certainty, what are the “stage[s] of processing the received user utterance” as recited in limitation 1(b). To begin with, this paragraph does not specifically discuss “processing the received user utterance,” nor does it identify “each stage” or the “stage[s]” thereof. At most, the description in this paragraph—of the parser 118 “dynamically receiv[ing] keyword and associated prior probability or fuzzy possibility updates” and “determin[ing] a context for an utterance by applying prior probabilities or fuzzy possibilities”—could suggest to a POSITA that limitation 1(d)⁵¹ might be considered a “stage of processing the received user utterance” as recited in limitation 1(b), but in my opinion a POSITA could not be reasonably certain of that conclusion, because this part of the specification is only discussing “an[] embodiment of the invention”⁵² and does not identify “determin[ing] a context for an utterance” as a “stage of processing the received user utterance.”⁵³ In addition, a POSITA

⁵⁰ ’209 patent at 17:41–56 (emphases added).

⁵¹ I use “limitation 1(d)” as shorthand to refer to the fourth limitation in claim 1, i.e., “parsing the recognized words and phrases to determine a meaning of the utterance, wherein determining the meaning includes determining a context for the at least one request contained in the utterance based on one or more keywords contained in the recognized words and phrases.”

⁵² ’209 patent at 17:9.

⁵³ I note that in the Declaration of Daniel Lopresti filed in support of Meta’s petition for *inter partes* review in proceeding IPR2025-01334, Dr. Lopresti similarly opined that “[w]hile the ’209 Patent is not clear which of its claimed steps are considered a ‘stage of processing the received user utterance,’ the specification discusses the use of ‘prior probabilities’ and ‘fuzzy possibilities’ as it relates to . . . ‘parsing step’ 1[d].” Decl. of Daniel Lopresti ¶ 138, IPR2025-

still would not know with reasonable certainty whether—and if so what—other activities might constitute “stage[s] of processing the received user utterance” as recited in limitation 1(b). As a result, this paragraph of the specification would not have enabled a POSITA to understand, with reasonable certainty, what are the “stage[s] of processing the received user utterance” as recited in limitation 1(b).

54. In my opinion, the prosecution history of the '209 patent likewise would not have enabled a POSITA to determine, with reasonable certainty, what are the “stage[s] of processing the received user utterance” as recited in limitation 1(b). As I describe above in connection with the term “usable” recited in limitation 1(b), issued claim 1 was added by the applicant as application claim 44 following the Examiner’s May 30, 2007 non-final rejection of all pending claims.⁵⁴ See ¶ 35, *supra*. As added, application claim 44 recited what is now limitation 1(b) in claim 1 of the '209 patent, “maintaining a dynamic set of prior probabilities or fuzzy possibilities usable at each stage of processing the received user utterance.”⁵⁵ However, as I noted above, the applicant’s remarks accompanying the amendment did not discuss limitation 1(b), nor did the Examiner do so at any subsequent point during prosecution. See ¶ 35, *supra*. The only discussion of newly added application claim 44 in the applicant’s remarks was that “Claims 41-59 are newly added.”⁵⁶ Following the addition of application claim 44, in an April 18, 2008 Office Action, the Examiner allowed this claim (among others), while rejecting other claims.⁵⁷ On May 7, 2008, the applicant cancelled all pending claims besides the claims that the Examiner

01334 (Aug. 18, 2025).

⁵⁴ '209 patent prosecution history, Nov. 30, 2007 Response to Non-Final Office Action at 18.

⁵⁵ '209 patent prosecution history, Nov. 30, 2007 Response to Non-Final Office Action at 23.

⁵⁶ '209 patent prosecution history, Nov. 30, 2007 Response to Non-Final Office Action at 30.

⁵⁷ '209 patent prosecution history, April 18, 2008 Final Rejection.

had allowed in the April 18, 2008 Office Action, and the '209 patent issued thereafter, with application claim 44 renumbered as issued claim 1.⁵⁸ Based on my review, and as reflected in the foregoing summary, neither the applicant nor the Examiner discussed limitation 1(b) or the phrase “each stage of processing the received user utterance” during prosecution of the '209 patent. Thus, in my opinion, nothing in the prosecution history would have enabled a POSITA to determine, with reasonable certainty, what are the “stage[s] of processing the received user utterance” as recited in limitation 1(b).

55. In my opinion, given the absence of any intrinsic evidence that would have enabled a POSITA to determine, with reasonable certainty, what are the “stage[s] of processing the received user utterance” as recited in limitation 1(b), a POSITA would have lacked any meaningful guideposts for determining, what counts as a “stage of processing the received user utterance,” and what doesn't. As a result, a POSITA would not have been able to determine, with reasonable certainty, whether any given “dynamic set of prior probabilities or fuzzy possibilities” is “usable” at each of those “stage[s].” Just as one example of the uncertainty a POSITA would face, a POSITA would have been unable to determine, with reasonable certainty, whether unrecited steps—that is, steps performed as part of a given “method responsive to a user generated natural language speech utterance” (which the preamble sets forth as the method that the alleged invention is directed to) but which are not recited in claim 1—may constitute a “stage of processing the received user utterance,” and if so which such steps. The preamble of claim 1 uses the transitional phrase “comprising,” which I understand means that the claim is open-ended—i.e., that the recited steps are non-exclusive—such that the claim does not preclude the performance of unrecited steps. Thus, a given “method responsive to a user generated natural

⁵⁸ '209 patent prosecution history, May 7, 2008 Response After Final Action at 2–7.

language speech utterance” may include steps that are not recited in claim 1. A POSITA would not know with reasonable certainty whether the phrase “each stage of processing the received user utterance” as recited in limitation 1(b) could encompass such steps—that is, whether the phrase “each stage of processing the received user utterance” is limited to one or more (unspecified) steps that are recited in the body of claim 1, or can also encompass unrecited steps. For example, if a given “method responsive to a user generated natural language speech utterance” includes steps that are not recited in claim 1, but which constitute “processing the received user utterance,” a POSITA would not know if those steps should be considered “stage[s] of processing the received user utterance” within the meaning of limitation 1(b).

56. As a further example of the uncertainty a POSITA would face (given the absence of any intrinsic evidence that would have enabled a POSITA to determine, with reasonable certainty, what are the “stage[s] of processing the received user utterance” as recited in limitation 1(b)), a POSITA would not have known how broadly or granularly to identify the “stage[s] of processing the received user utterance” in a given “method responsive to a user generated natural language speech utterance.” For example, the specification of the ’209 patent explains that aspects of the alleged invention can be implemented in computer software, including “one or more of the following modules: a speech recognition module for capturing the user input; a parser for parsing the input, a text to speech engine module for converting text to speech; a network interface for enabling the computer to interface with one or more networks; a graphical user inter face module; and an event manager for managing events.”⁵⁹ As a POSITA would have understood, computer software is generally expressed as a series of instructions, in the syntax of a given programming language, that may be organized or grouped in various ways. For example,

⁵⁹ ’209 patent at 2:34–41.

a computer program may comprise multiple modules, each of which perform a given function and may comprise multiple sub-modules or sub-routines, each of which in turn perform a given function and may comprise a series of instructions, et cetera. In my opinion, however, there was no established understanding of what a “stage” is, in the context of software. This is in contrast to terms that did have a generally understood meaning in that context, such as “function,” “module,” “class,” “object,” “method” (referring to a block of code associated with a class or object), “routine,” “subroutine,” et cetera. (In other contexts that have little to nothing to do with the ’209 patent, such as pipelined architectures for a CPU, “stage” did have a generally understood meaning.) Given the absence of any intrinsic evidence that would have enabled a POSITA to determine, with reasonable certainty, what are the “stage[s] of processing the received user utterance” as recited in limitation 1(b), a POSITA would not have known how to determine whether a given set of computer instructions constitute a single “stage” versus multiple “stage[s]” in this context. For example, suppose that the “speech recognition module for capturing the user input” described by the specification (in the excerpt quoted immediately above) is implemented using a large set of computer instructions that are grouped into several sub-modules. Suppose further that “capturing the user input,” as generally performed by this “speech recognition module,” constitutes “processing the received user utterance” as recited in limitation 1(b). A POSITA would not know whether to consider the entire “speech recognition module” as a single “stage of processing the received user utterance,” or whether to consider each of the sub-modules within the “speech recognition module” as their own separate “stage[s] of processing the received user utterance.” Without knowing how broadly or granularly to view a “stage of processing the received user utterance,” a POSITA would have been unable to determine what are all the “stage[s] of processing the received user utterance,” and thus unable to

decide, with reasonable certainty, whether a given “dynamic set of prior probabilities or fuzzy possibilities” is “usable” at each of those “stage[s]” as required by limitation 1(b).

VI. RESERVATIONS

57. I reserve the right to supplement and/or modify my opinions set forth herein, as well as the bases for my opinions, depending on any information subsequently provided by Plaintiff, including to the extent that Plaintiff purports to rebut my opinions or the bases therefor, or supply alleged rebuttal evidence such as expert opinion, additional documents, testimony, or other information. In the event that I am asked or given the opportunity to provide testimony at a hearing in this matter, such as a claim construction hearing, I reserve the right to provide additional information on my knowledge, skills, experience and other qualifications, to discuss further background on the patent and relevant art, to use visual aids to illustrate and/or summarize my testimony, and to supplement my opinions with information requested by the Court.

I declare under penalty of perjury that the foregoing is true and correct.

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 September 18, 2025

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.”

Mark E. Crovella

2

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.

Mark E. Crovella

3

“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 Institutional 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.

Mark E. Crovella

4

- 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.

Mark E. Crovella

5

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 Committee 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

Mark E. Crovella

6

- 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.

Mark E. Crovella

7

“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.

Mark E. Crovella

8

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 Jomblatt**, 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.

Mark E. Crovella

9

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

Mark E. Crovella

10

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, Dzmityr 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 Communication 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.

Mark E. Crovella

12

- L. A. Crawl, 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, Adaptation, 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, Adaptation, 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. Erdős, 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 MAS-COTS 2001*, Cincinnati, Ohio, August 2001.

Mark E. Crovella

17

- 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.

Mark E. Crovella

18

- 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 *Proceedings 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," *Proceedings 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," *Proceedings 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," *Proceedings 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.

Mark E. Crovella

19

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.

Mark E. Crovella

20

- 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.

Mark E. Crovella

21

- 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"

Mark E. Crovella

22

- 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.

Mark E. Crovella

23

- 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.