Follow the trace

Evaluating physical and digital forensic findings given activity level propositions using Bayesian networks

Vink, M.

Publication date
2025

Chapter 8

A BAYESIAN NETWORK FOR THE EVALUATION OF DIGITAL FINDINGS IN TROJAN HORSE DEFENSE CASES

Based on[1] Vink, M., Schramp, R., Kokshoorn, B. & Sjerps, M. J. 2025. The evaluation of evidence in Trojan horse defense cases using Bayesian networks. Manuscript submitted for publication[2].

[1]Minor modifications have been made to the submitted manuscript to maintain a consistent layout throughout the thesis. See Page 289 for a clarification of the modifications.

[2]This manuscript is currently under review (September 19, 2025).

8.1 Introduction

Digital forensic science is nowadays a prominent discipline in forensic casework. However, reported conclusions in digital forensic casework remain primarily categorical and unbalanced (Sunde, 2021). Moreover, the evidence evaluations lack a structured and transparent reasoning, as most evaluations currently occur solely within the expert's head. Studies have compared evaluative practices within the digital forensic domain to other forensic disciplines and argued that digital forensic science needs to keep up with the other disciplines (Sunde & Dror, 2019; Pollitt *et al.*, 2018).

In the 1970s and 1980s, pioneers—including Lempert (1977), Lindley (1977), Evett (1984) and Aitken & Gammerman (1989)—advocated using the likelihood ratio (LR) framework to evaluate forensic evidence, particularly for physical transfer evidence. Since then, the LR framework has gained support in the forensic science community and was later incorporated as the core component in the "ENFSI guideline for evaluative reporting in forensic science" by Willis *et al.* (2015). The LR requires a set of propositions that often reflects the prosecution and defense point of view. The ratio consists of two probabilities: the probability of observing the evidence given that proposition 1 is true and the probability of observing the evidence given that proposition 2 is true. The LR framework is a suitable tool for evaluating scientific evidence because it provides essential qualities of a good evaluation: balance (by considering two competing propositions) and transparency, as results, propositions, and other factors must be clearly defined within the framework.

Previous work in digital forensic science has focused on advancements like (standardizing) evaluative methods (Casey *et al.*, 2019, 2020; Overill & Silomon, 2011; Overill & Collie, 2021; Sunde & Franqueira, 2023) and developing LR systems for specific source level comparison problems (Galbraith *et al.*, 2020b; Macarulla Rodriguez *et al.*, 2020; Vink *et al.*, 2022). However, to the best of our knowledge, limited research exists on using the LR framework as a probabilistic reasoning tool for combining evidence in cases with activity level propositions.

This study focuses on Trojan horse defense (THD) cases, where individuals suspected of possessing illegal content on electronic devices claim that either another person or program placed the content there, or that they acquired it unintentionally (Boddington, 2012; Bowles & Hernandez-Castro, 2015; Brenner *et al.*, 2004; Carney & Roger, 2004; Haagman & Ghavalas, 2005; Steel, 2014). In these cases, deriving the LR becomes complex due to the large number of different (digital) traces and propositions that are both at activity level and offense level (Sunde, 2021; Vink *et al.*, 2025). Our research stems from digital forensic experts at the Netherlands Forensic Institute seeking guidance on applying the LR framework to evaluate their observations in THD cases. They specifically needed help combining multiple observations into an evidence evaluation given a set of propositions.

Bayesian networks (BNs) have proved useful in complex evaluations of findings in other forensic disciplines (Aitken *et al.*, 2003; Dawid & Evett, 1997; Evett *et al.*, 2002; Garbolino & Taroni, 2002; Taroni *et al.*, 2014) and we will show how to use them in the evaluation of evidence in THD cases. BNs are graphical representations of propositions, observations, and other variables involved in evidence evaluation, along with their probabilistic relationships. These networks provide forensic experts and legal practitioners with a tool that makes complex reasoning explicit and understandable, while also serving as a method to derive LRs from the set of variables in the network. Kwan *et al.* (2008, 2011) and Lee *et al.* (2009) explored the use of BNs in digital casework and advocate their promising added value[3]. To the best of our knowledge, after the work of Kwan et al. and Lee et al., no other research has been published on BNs in digital forensic casework.

In short, digital forensic science needs more balanced, structured and transparant evidence evaluation methods. This study aims to promote structured reasoning

[3]For more background information on BNs in forensic science, see Taroni *et al.* (2014). For examples of BNs specifically applied to digital forensic science, see Kwan *et al.* (2008)

by providing guidance on using BNs as a tool for evaluating digital evidence, particularly in cases of suspected possession of illegal content on electronic devices. This chapter builds upon Chapter 7 (based on Vink *et al.* (2025)), which explored how to formulate suitable propositions in order to use the LR framework in THD cases. Using a similar fictive case example and case propositions, we will build a BN that evaluates the case observations given a set of propositions disputing who downloaded illegal files to a computer. We use the template from Taylor *et al.* (2018) and the template BN from Vink & Sjerps (2023) to build the case-specific network. These template BNs were originally developed for evaluating "physical evidence[4]," and we will explore their suitability for modeling evidence evaluations in the digital forensic domain. Section 8.2 presents the case example, Section 8.3 demonstrates how to use the templates to construct a BN in THD cases and how to derive an LR from the network. Section 8.4 discusses the model's broader applications, limitations, and practical uses.

8.2 Fictive case example

We will be using a fictive case example used in Chapter 7 (Vink *et al.*, 2025) to demonstrate the application of BNs in the evaluation of digital evidence concerning cases with a THD. The case example is influenced by (but not similar to) a case verdict from the Dutch court on 19 December 2018[5].

According to police intelligence, illegal files were downloaded from a file-sharing network on September 8, 2023. Investigators traced the IP address of the computer used for the downloads and identified Mr. X as a suspect. During a search of his home on February 10, 2024, police seized a computer and

[4]We define physical evidence as trace evidence from physical forensic science disciplines. We do not consider a computer or hard disk as physical evidence.

[5]GHDHA, 19 December 2018, ECLI:NL:GHDHA:2018:3528. We only used information that was available at https://www.rechtspraak.nl.

discovered 450 documents containing illegal content. The prosecution alleges that Mr. X knowingly downloaded the illegal material on September 8, 2023. The defense claims that Mr. X had no knowledge of the files, that he must have been framed and that the files must have been downloaded on his computer by someone else with remote access on September 8, 2023. Mr. X informed the police that nobody else had physical access to his computer during the time of the download, and he claimed to have no knowledge of who has remotely controlled it.

A forensic examiner analyzed a copy[6] of the computer's hard drive (date of copy: February 10, 2024) and found the following:

E_1 : iDeal payment logs from the bank account of Mr. X on September 8, 2023, at 11:10 am.

E_2 : Search history records containing terms matching the file names of interest from the eMule application[7] before, during, and after September 8, 2023.

E_3 : Files of interest are stored in the home directory of the computer.

E_4 : A remote access tool called LogMeIn is installed but there is an absence of logs corresponding to active use on September 8, 2023.

E_5 : Recent VLC[8] display logs in *most recent used list* with names corresponding to the files of interest.

The creation time of the alleged files lie between 10:50 am and 11:50 am, therefore the prosecution has specified September 8, 2023 between 10:50 am and 11:50 am (also written as "time T" later in this paper and in the network)

[6]We do not go into detail about the investigation method itself. See for example Carrier & Spafford (2003) for case examples and investigation phases in digital casework.

[7]eMule is a peer-to-peer program.

[8]VLC is a commonly used media player.

as the date and time of the download. The examination of the hard drive copy was primarily focused on secondary traces[9] that match the alleged time T. The digital forensic practitioner may also examine traces matching the last written and last access times of the alleged files (Carrier & Spafford, 2003).

The timestamps of the digital traces influence the relevance and weight of the traces regarding the case propositions. For example, when the digital investigator searches for traces of regular user activities on the computer on September 8, traces that lie closer to the alleged time period are more relevant. The digital forensic investigator typically defines a specific timeframe when searching for traces to narrow down the investigation's scope. While this scope varies by case, a practical approach here would be examining the period 10-15 minutes before and after the alleged file downloads and/or last access times.

8.3 Modeling the evaluation with a Bayesian network

In this section, we will show how to build a BN for the case using the template BN from Chapter 3 for modeling transfer evidence given activity level propositions (Vink & Sjerps, 2023). The template network is based on Taylor et al.'s (2018) template for constructing BNs to evaluate DNA evidence and on Kokshoorn et al.'s (2017) work on evaluating evidence given a disputed actor and/or activity. All BNs are constructed using the software Hugin Researcher 9.1 (Hugin Expert A/S, 2021).

As for the quantitative component of the BN, all probabilities we assigned for this case example are shown in Table D.1 and Table D.2 in Appendix D[10]. We identified which conditional probabilities to set based on the entries in the

[9]Secondary traces are defined by Vink *et al.* (2025) as either traces that are generated as a "byproduct" of the alleged activity or as a product of another activity.

[10]This list may seem overwhelming. We refer to Section 8.4.3 for a discussion of what these probabilities represent: the expert's assessment rather than absolute truth.

conditional probability tables (CPTs) in the BN[11]. While we have carefully set the probabilities using expert knowledge and experience, we do not recommend using the assigned probabilities directly in other casework. Other cases have another framework of circumstances and therefore require other considerations to establish appropriate probability assignments. Some of the CPTs of our case example are presented throughout the paper, the rest of the conditional probabilities can be found in Appendix D.

8.3.1 Step 1, the black node: define the propositions

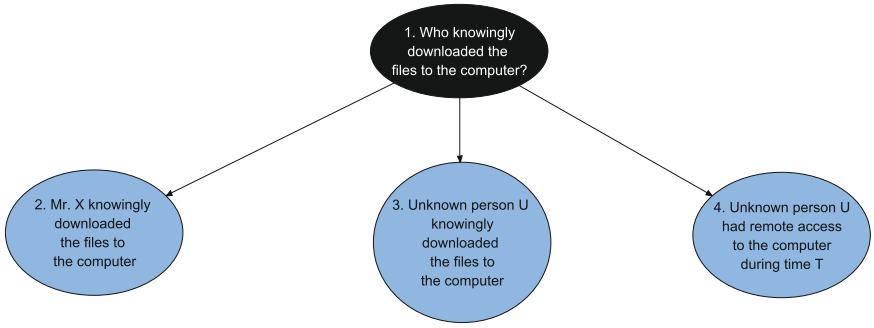


Figure 8.1: Case model after step 1 and step 2: "split the propositions into activities."

We start modeling our evaluation using a BN by defining our case propositions in a black proposition node. The black node (see Fig. 8.1) represents the proposition node, which contains the states H_1 and H_2 that are based on our two scenarios presented by the prosecution and defense: one where Mr. X downloaded files knowingly, and another where someone else did it remotely[12]. The propositions should capture the key disputes from the scenarios and in

[11] I refer the reader to Section 1.3 for an introduction to Bayesian networks and conditional probability tables.

[12] In Chapter 7 we provide guidance on how to formulate relevant propositions in THD cases.

our case example, the key dispute is the actor of "the download"[13]: Mr. X or someone else (unknown to Mr. X). Consider the following set of case propositions:

H_1: Mr. X knowingly downloaded the files to the computer.
H_2: Someone else with remote access to the computer knowingly downloaded the files to the computer. Mr. X is unaware of the presence of the files on the computer[14].

Table 8.1 illustrates the CPT for the proposition node, WHO KNOWINGLY DOWNLOADED THE FILES TO THE COMPUTER? (1). The probabilities in this table are the probabilities of the case scenarios prior to observing the digital evidence. Currently, they are both set to 0.5 but this is not based on any case-information. Typically, we set the prior odds to one, because it allows us to read the LR directly from the probabilities in the proposition node after entering the case findings into the network[15]. The prior odds do not influence the LR conclusion, and determining these prior probabilities typically lies beyond the domain of the digital forensic expert. However, the digital forensic investigator can assess certain factors that influence the prior probabilities of the case propositions, such as the computer's level of digital protection. In Section 8.3.6 we will show how to model this in the network.

[13]The download that occurred on September 8, 2023 at time T. We excluded the download time from the propositions since time T is undisputed, and we wanted our propositions to focus on the contested aspects of the case. However, time T should still be documented elsewhere in the forensic report for the reasons described at the beginning of Section 8.2.

[14]We included "Mr. X is unaware of the presence of the files on the computer" in H_2 as this claim is relevant as well as disputed. Given H_1, the knowledge of Mr. X regarding the presence of the files is already situated in "Mr. X knowingly downloaded the files to the computer". Any other information that is relevant (but not disputed) should be formulated as an assumption or as undisputed case information. We refer Taylor & Kokshoorn (2023) and Taylor *et al.* (2020) for more information on what information should be included in the proposition, as an assumption or as undisputed case information.

[15]This works because when the prior odds equal one, the posterior odds become equal to the LR—using Bayes rule' where the posterior odds are equal to the product of the LR and the prior odds.

1. Who knowingly downloaded the files to the computer?

H_1 : Mr. X	0.5
H_2 : Unknown person U	0.5

Table 8.1: The CPT for WHO KNOWINGLY DOWNLOADED THE FILES TO THE COMPUTER? (1) with states H_1 and H_2 representing the case propositions. "Unknown person U" is the "someone else" in proposition H_2.

In our example, there are only two propositions to consider. However, there may be case circumstances where formulating more than two propositions is appropriate, such as when multiple alternative scenarios are proposed or when the case information suggests additional scenarios that are worth considering. This will slightly alter the structure of the network, as we will demonstrate in the discussion of the model in Section 8.4.

8.3.2 Step 2, the blue nodes: split the propositions into activities

After defining the proposition node containing the case scenarios, the next step is dividing these scenarios into activities that are specified in the propositions. By dividing the scenarios into separate activities, the network's structure maps different "storylines" of a trace—from initial activity through transfer of traces to their recovery from an item of interest. In the case example, this means following the path from downloading the illegal files, through the transfer of files to the computer (and generation of secondary traces), to the recovery of these traces from a hard copy of the computer of interest.

A proposition may include one or more activities that could potentially lead to digital traces. In our case example, given H_1, it is certain that Mr. X knowingly downloaded the files to the computer (activity 1). Given H_2, we have two activities that occurred with certainty: an unknown person is remotely controlling the computer on September 8, 2023, at time T (activity 2) and that same (unknown) person is knowingly downloading the files to the computer (activity 3). Since the prosecution makes no statement about remote control

during time T, we assume remote control could also possibly occur under H_1 for some other reason. Therefore, following the proposition node WHO KNOWINGLY DOWNLOADED THE FILES TO THE COMPUTER? (1) we define three blue activity nodes[16]:

2. MR. X KNOWINGLY DOWNLOADED THE FILES TO THE COMPUTER
3. UNKNOWN PERSON U KNOWINGLY DOWNLOADED THE FILES TO THE COMPUTER
4. UNKNOWN PERSON U HAD REMOTE ACCESS TO THE COMPUTER DURING TIME T

All three nodes are directly linked to WHO KNOWINGLY DOWNLOADED THE FILES TO THE COMPUTER? (1) (see Fig. 8.1). The CPTs of these nodes reflect the probability of the activity being "True" or "False" given that either proposition 1 or proposition 2 is true. Node (2) contains the probability of "Mr. X knowingly downloaded the files to the computer" being "True" or "False" given H_1 (1) or H_2 (0) (see Table 8.2). For node (3) a similar structure holds.

Node (4) also contains other conditional probabilities than ones and zeros (see Table 8.3) since the computer was remotely controlled under H_2 but *could* be remotely controlled given H_1. The ones and zeros in the second column reflect that remote access by unknown person U is necessary for the proposition that an unknown person downloads the material, as there is no other option considered to access the computer under the given framework of circumstances[17]. The first column contains the probability of someone other

[16] Activity nodes can be further divided into sub-activity nodes using the subscenario idiom (Vlek *et al.*, 2013b). This division helps clarify which specific activities may lead to particular types of traces. However, one must be careful not to overcomplicate the model, ensuring the BN remains informative for the receiver. There is no definitive rule for such modeling choices.

[17] One could add a directed link between nodes (3) and (4) to visually emphasize that remote control must occur if someone else downloaded files to the computer. However, this relationship is already mathematically captured by the directed link between nodes (1) and (4), making it superfluous.

2. Mr. X knowingly downloaded the files to the computer

1. Who knowingly downloaded the files to the computer?	Mr. X	Unknown person U
False	0	1
True	1	0

Table 8.2: The CPT for Mr. X knowingly downloaded the files to the computer (2).

4. Unknown person U had remote access to the computer during time T

1. Who knowingly downloaded the files to the computer?	Mr. X	Unknown person U
False	0.98	0
True	0.02	1

Table 8.3: The CPT for Unknown person U had remote access to the computer during time T (4). Remote access by unknown person U is assumed given H_2 and possible given H_1.

than Mr. X is remotely controlling the computer while Mr. X is knowingly downloading the files to the computer (0.02).

8.3.3 Step 3, the green nodes: identify association propositions

The digital expert is likely to investigate other questions related to the key issue of "who downloaded the files?" These questions often concern side-activities that may or may not occur with some probability alongside (or after) the main activities stated in the case propositions. For example, e-mail activities around time T that can be linked to an individual on September 8, 2023, or file-ownership related activities such as viewing the files that give an indication of the user's awareness of the presence of the files. Some of these side-activities that are investigated by the digital expert in cases of suspected possession of illegal content are listed in a verdict from a 2018 Court of The Hague case[18] and in a document about article 240b of the Dutch Penal Code from *Kenniscentrum*

[18]GHDHA, 19 December 2018, ECLI:NL:GHDHA:2018:3528.

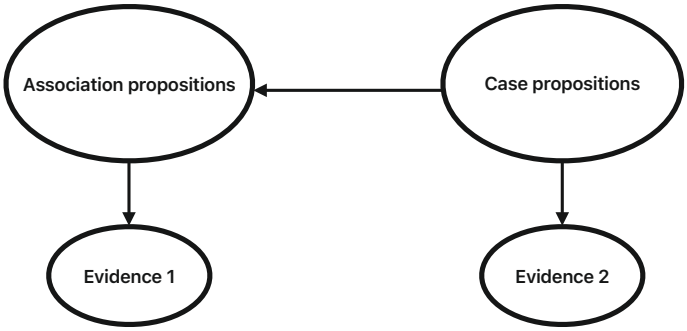


Figure 8.2: The incorporation of an association evaluation in the evaluation of evidence given the case propositions.

Cybercrime[19] by Van den Hurk *et al.* (2023):

1. Regular user activities that occur simultaneously with the alleged activities which can be linked to the (primary) user of the computer or to other individuals;

2. Activities related to file-ownership such as viewing and displaying the illegal content.

The evaluation of evidence in light of these side-activities will be introduced as association evaluations in our case model (see Fig. 8.2). Association evaluations use propositions structured similarly to those of the case propositions and evidence can be evaluated given these association propositions. Our model includes a node for the association propositions and a directed link between this node and the case proposition node.

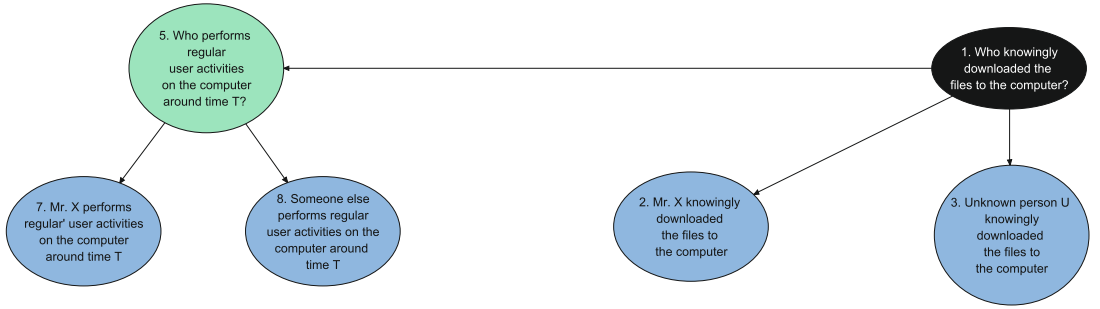
For the case example, consider the following association propositions for the question concerning regular user activities on the computer:

[19]In Dutch. English translation: Cybercrime Knowledge Center.

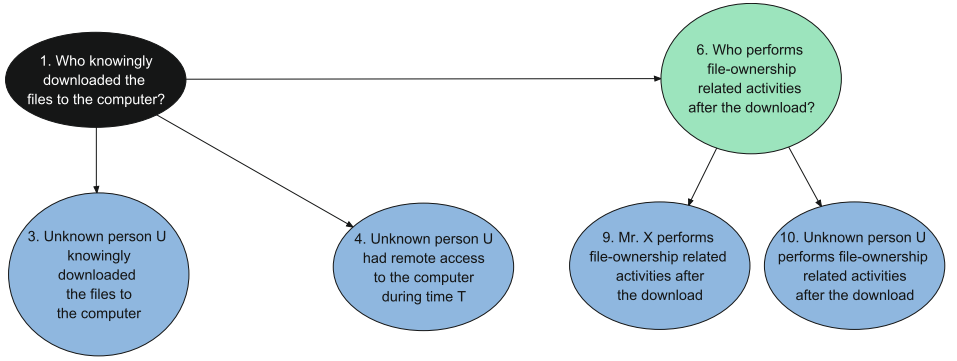
H_3: Mr. X performed regular user activities on the computer on September 8, 2023, around time T.

H_4: Someone else performed regular user activities on the computer on September 8, 2023, around time T.

H_5 : No regular user activities occurred on September 8, 2023, around time T.



(a) The incorporation of the association propositions concerning who is performing regular user activities on the computer around time T.



(b) The incorporation of the association propositions concerning who is performing file-ownership related activities after the download on September 8, 2023, at time T.

Figure 8.3: Association propositions concerning regular user activities and file-ownership related activities.

The association propositions H_3, H_4 and H_5 address all three possibilities regarding who, if anyone, performed "regular" computer activities around time T. We define "regular user activities" as activities that are unrelated to the disputed activities (downloading files and gaining remote access) in the case propositions. These regular user activities can be specific when there is knowledge about typical computer usage at certain times, and can therefore help identify who was using the computer during the incident. They may also be linked to an individual through other means, such as login attempts on certain websites with someone's personal account.

Unlike in the case propositions, the specific date and time (September 8, 2023, time T) is included in the association propositions. This is because evidence of regular computer activities by an individual can only support propositions about who downloaded the files when the observation timestamps if the activities match the specific (prolonged)[20] time period of the download activity.

The definition of "someone else" in proposition H_2 is either a known or unknown individual to Mr. X. However, since case information shows that no one else had physical access to the computer, these regular activities must have occurred through remote access. The propositions H_3, H_4 and H_5 are included as states in the association proposition node (see Fig. 8.3a):

5. Who performed regular user activities on the computer around time T?

The association propositions of "who performed file-ownership related activities after the download on September 8, 2023, time T?" is added similarly to the case model (see Fig. 8.3b) with node:

[20]The specific duration of this time period varies by case and should be documented in the case report. In Section 8.3, we defined this period as 10-15 minutes before and after the alleged activities in the case propositions.

6. Who performed file-ownership related activities after the download on September 8, 2023?

We define "file-ownership related activities" as actions performed *with* the files in question *after* the download, such as viewing or make certain changes to them. The following association propositions are the states of Who performed file-ownership related activities after the download on September 8, 2023? (6):

H_6: Mr. X performed file-ownership related activities after the download on September 8, 2023.
H_7: Someone else (unknown to Mr. X) performed file-ownership related activities after the download on September 8, 2023.
H_8 : No file-ownership related activities occurred.

The process of splitting propositions into activities is also applied to the association propositions. Since propositions H_5 and H_8 contain "no activities", only the remaining four propositions lead to blue activity nodes. Nodes 7 through 10 stem from the green association proposition nodes (5) and (6):

7. Mr. X performed regular user activities on the computer around time T

8. Someone else performed regular user activities on the computer around time T

9. Mr. X performed file-ownership related activities on and after the download

10. Unknown person U performed file-ownership related activities on and after the download

5. Who performed regular user activities on the computer around time T?

1. Who knowingly downloaded the files to the computer?	Mr. X	Unknown person U
Mr. X	0.9	0.05
Someone else	0.05	0.15
No normal activities	0.05	0.8

Table 8.4: The CPT for WHO PERFORMED REGULAR USER ACTIVITIES ON THE COMPUTER AROUND TIME T? (5). Case information suggests that no individual other than Mr. X had physical access to the computer.

6. Who performed file-ownership related activities after the download?

1. Who knowingly downloaded the files to the computer?	Mr. X	Unknown person U
Mr. X	0.9	0
Unknown person U	0.02	0.5
No file-ownership related activities	0.08	0.5

Table 8.5: The CPT for WHO PERFORMED FILE-OWNERSHIP RELATED ACTIVITIES AFTER THE DOWNLOAD? (6). Case information suggests that no individual other than Mr. X had physical access to the computer.

Setting the conditional probability tables of node (5) and node (6)

The CPT for node (5) (Table 8.4) includes the conditional probability of regular user activities given that either Mr. X (column 1) or unknown person U (column 2) knowingly downloaded the files[21] When Mr. X knowingly downloaded the files, he likely used the computer before or after the alleged activities for personal life, work, or hobbies. Given that Mr. X knowingly downloaded the files, it is less likely that someone else would perform such regular activities on the computer. Therefore, when Mr. X knowingly downloaded the files, the probability that he performed regular user activities is set to 0.9. The probability of someone else performing activities related to X or related to someone else while Mr. X is downloading is considered 0.05. The probability of no regular activities and thus Mr. X limiting his activities to the illicit activities is set to 0.05 as well.

[21] This represents a "personal" probability that draws from the expert's experience. We discuss the use of these subjective probabilities in Section 8.4.3.

Regarding the second column of Table 8.4, when an unknown person U downloaded the files to the computer it is assumed the person is more likely to want to keep his activities secret to X. For that reason it is not likely that U will use the computer before or after the download for other reasons than the download itself. The probability of regular user activities performed by Mr. X given that unknown person U downloaded the files is 0.05. The probability of regular user activities by someone else that can either be related to Mr. X or to someone else is considered 0.15. It is assumed that in general no regular user activities are performed around time T when unknown person U downloaded the files to the computer (0.8). Because the conditional probabilities in column 1 and column 2 differ, evidence concerning these association propositions can update our belief about the individual who downloaded the files to the computer.

Table 8.5 shows the CPT for Who performed file-ownership related activities after the download on September 8, 2023?. When Mr. X knowingly downloaded the files (column 1), it is assumed he did this because he wanted to do something with them (e.g., view them). For this reason, the probability of Mr. X performing file-ownership related activities is set to 0.9. However, especially when downloads take a long time, this activity could be interrupted or forgotten by Mr. X, so there is a small chance of no file-ownership activities by anyone (0.08). The probability of someone else interacting with files downloaded by Mr. X is set to 0.02. When an unknown person U downloads the files (column 2), we must consider their motivation, and the probabilities in Table 8.5 depend on this. The file could for example be placed for reasons of extortion or coercion (e.g., as evidence in a child custody dispute by ex-partners). But other motivations exist and therefore, the probability of "someone else performing file-ownership related activities" given that unknown person U knowingly downloaded the files is set to 0.5. The probability of Mr. X performing file-ownership related activities is assumed to be 0 since, under H_2, he has no knowledge of the file's existence. The option of no activities is the

remainder probability (0.5).

8.3.4 Step 4, the red nodes: list the findings

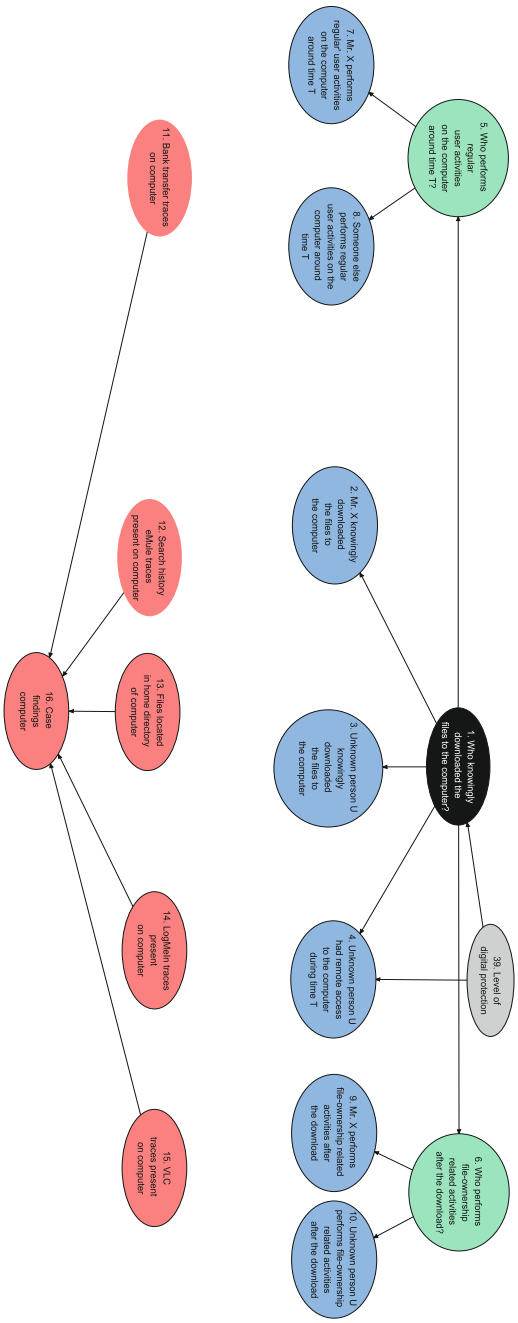


Figure 8.4: Case model after step 4: "list the findings."

The next step is to make findings nodes of the possible observations on the computer copy. These findings node have states "True" or "False", depending on what is observed in the case ($E_1 - E_5$ from Section 2).

11. Bank transfer traces present on computer
12. Search history traces present on computer
13. Files located in home directory
14. LogMeIn traces present on computer
15. VLC traces present on computer

The observations resulted from the examination of a hard copy of the computer. In the network, we assume that what we observe on the computer copy reflects what we would observe on the c omputer. We do not consider the possibility of contamination of the computer copy during the investigation—that is, the presence of material or secondary traces on the computer copy that have no relation to the case at hand (Gruber *et al.*, 2023).

Fig. 8.4 presents the case model including the findings n odes. The findings nodes are not yet connected to the activity nodes—we will do this in Step 4. For our case-specific model, we can add a case findings node (16) with only two states: "case findings" and " other." The state "case fi ndings" represents the situation of all observations E_1 - E_5 combined which is $\{E_1, ..., E_5\}$ = {Node (11) = Bank traces related to Mr. X, Node (12) = True, Node (13) = True, Node (14) = False, Node 15 = True}.

The state "other" represents the situation where any one of E_1 - E_5 was observed differently. As such, the CPT for case findings (16) contains only zeros, except for the relevant combination of observations in the case. Since the case findings node summarizes the combination of observations, we can easier derive the LR for the combined observations. For more details about computing the LR from the network, we refer to Section 8.3.7.

8.3.5 Step 5: connect the observation nodes to the activity nodes

Step 5.1: Specify sub-activities

So far, we have modeled the black proposition node, the green association proposition nodes, the blue activity nodes, and the red findings nodes. The next step is connecting the blue activity nodes to the red findings nodes. This step consists of two parts: adding sub-activities to the model (also blue nodes) and incorporating the uncertainties regarding the transfer and persistence of traces resulting from these sub-activities (yellow nodes).

The first blue nodes 2 to 10 essentially repeat the activities from the proposition nodes, which has the advantage of separating each potential path from proposition to findings more clearly. However, these activities are not directly related to the findings specified in the findings nodes. To address this, we want to specify these "sub-activities" in the BN using the subscenario idiom from Vlek *et al.* (2013b) Each activity receives a sub-activity as a child node that does relate to the findings[22]. For example, node Unknown person U performed file-ownership related activities after the download (10) can be further specified as "U used VLC to view the contested files" (26)[23]. The other sub-activities are (blue nodes 17 to 26):

17. Mr. X used his bank account for money transfer

18. Someone else used Mr. X's bank account for money transfer

[22]Unlike DNA evidence, where countless activities could lead to DNA transfer and must be explicitly stated in propositions, digital traces have a more direct relationship to specific activities. The activities can often be inferred from the digital findings themselves, since digital traces typically correspond to particular actions. For example, VLC media player traces indicate the use of VLC media player.

[23]In our case, the traces related to these activities consist only of display records from a VLC media player. However, more sub-activities can be incorporated into the model if needed, making the BN versatile for specific cases.

19. Someone else used his own bank account for money transfer
20. Mr. X used eMule application to search for files
21. Mr. X copied files to home directory
22. U used eMule application to search for files
23. U copied files to home directory
24. U used LogMeIn to control the computer
25. Mr. X used VLC to view the contested files
26. U used VLC to view the contested files

Fig. 8.5 shows part of the BN with sub-activity nodes 20 to 23. Mr. X used eMule application to search for files (20) and Mr. X copied files to home directory (21) are child nodes of activity node Mr. X knowingly downloaded the files to the computer and so are node 22 and 23 child nodes of Unknown person U knowingly downloaded the files to the computer (3).

The CPT for these sub-activity nodes includes the probability of a *sub-activity* occurring given that the *activity* did or did not happen (see Table 8.6 and Table 8.7 for the CPTs of Mr. X copied files to home directory (21) and U copied files to home directory (23)). If Mr. X knowingly downloaded the files, we assume he would likely copy them to his home directory ($p_{21} = 0.9$). Unknown person U would be less likely to store files in Mr. X's home directory. However, we set this probability to 0.5 since an unknown person U might either intentionally place files in Mr. X's home directory to make them more discoverable for authorities, or alternatively, operate unnoticeably to avoid detection. We assume that Mr. X only performs the sub-activities given H_1, not given H_2. To clarify, the probability that Mr. X copies the files to his home directory given that he did not knowingly downloaded the files

21. Mr. X copied the contested files to home directory

2. Mr. X knowingly downloaded the files to the computer	False	True
False	1	0.1 $(1 - p_{21})$
True	0	0.9 (p_{21})

Table 8.6: The CPT for Mr. X copied the contested files to home directory (21).

to the computer is zero. This can be modified by adjusting the CPTs of the sub-activity nodes.

In general, the probability of an activity occurring is difficult to determine and likely involves subjective judgment. One might question who is best positioned to assess these probabilities. We will address this further in Section 8.4.3.

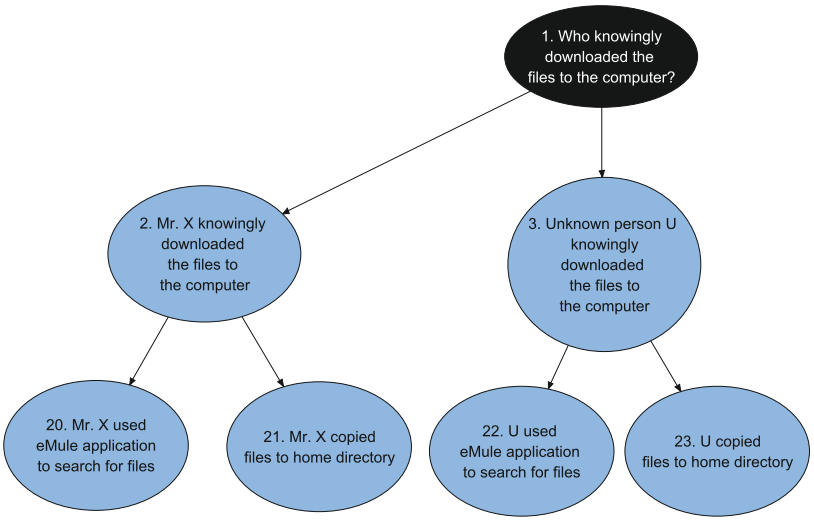


Figure 8.5: Detailed section of case model after step 5.1: "specify sub-activities."

23. U copied the contested files to home directory

3. Unknown person U knowingly downloaded the files to the computer	False	True
False	1	0.5 $(1 - p_{23})$
True	0	0.5 (p_{23})

Table 8.7: The CPT for U COPIED THE CONTESTED FILES TO HOME DIRECTORY (23).

Step 5.2, the yellow nodes: adding transfer nodes

When modeling an evaluation of physical evidence, we use nodes that describe the expected shedding, transfer, and persistence, of traces from the activities. In our case example, we will use nodes that describe a similar process for shedding, transfer and persistence but with slightly altered definitions.

We distinguish between two types of traces: primary and secondary traces. Primary traces are created when someone (or something) directly "transfers" files to the computer - for example, by downloading files or copying them to another directory. These result in traces that are directly visible to users. Secondary traces are generated ("shed") automatically by various processes on the computer such as log files or other latent digital traces. Thus, we distinguish between "transfer" and "generation" as processes that result in two different type of traces. In the BN, this distinction is purely semantic and does not affect our modeling choices.

Regarding the question of whether the files or generated secondary traces have persisted, we must consider the uncertainties affecting persistence. For digital traces on a digital device, both transferred and generated traces can disappear from an electronic device - for example, by automatically or deliberately deleting the files. The probabilities of transfer and/or generation, and persistence, are usually all encapsulated in a single node: a yellow "transfer" node.

Yellow transfer nodes 27 to 36 specify both the type of trace from the red findings nodes and its possible activity of transfer of material or generation of traces to the computer. Consequently, there is a distinct yellow transfer

node for each combination of trace type and activity. For example, search history traces (trace type) could have been generated either through Mr. X using the eMule application to search for the contested files (activity 1) or through unknown person U using the same application (activity 2), resulting in two yellow transfer nodes (30 and 32). The following yellow transfer nodes 27 through 36 are added to the model:

27. Generation and persistence of bank transfer traces related to Mr. X on the computer via Mr. X..

28. Generation and persistence of bank transfer traces related to Mr. X on the computer via someone else..

29. Generation and persistence of bank transfer traces related to U on the computer via someone else..

.....

36. Generation and persistence of VLC traces to computer via U using VLC..

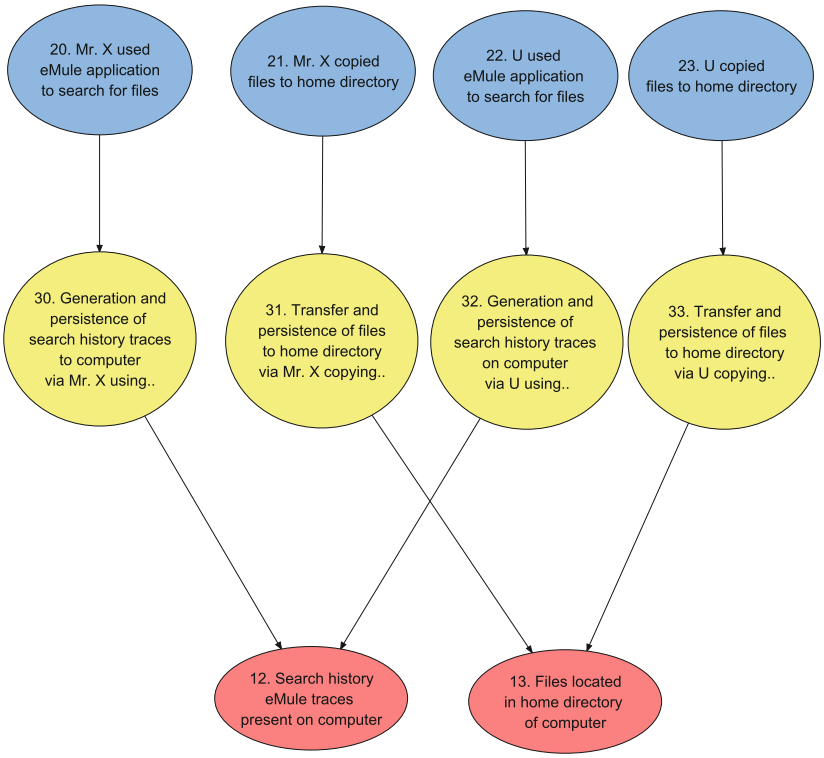


Figure 8.6: Detailed view of selected yellow transfer nodes in the case model shown in Fig. 8.7.

Fig. 8.6 shows a detailed section of the model in Fig. 8.7. To avoid redundancy, we will not elaborate on all sections of the model as a similar structural pattern shown in Fig. 8.6 applies to all other paths from the blue activity nodes to the red findings nodes.

We repeat the activity from the blue sub-activity nodes in their yellow child transfer nodes, such as "via Mr. X using eMule application to search for the contested files" in node (30), to maintain distinct transfer paths (Vink & Sjerps, 2023). The CPTs of all yellow transfer nodes contain only one probability, since transfer and thus persistence cannot occur through an activity that did not take place. See for example the CPTs of Generation and persistence

of search history traces on computer via Mr. X using.. (30) and Generation and persistence of search history traces on computer via U using.. (32) (see Tables 8.8 and 8.9) where the probability of transfer is zero and the probability of no transfer is one when the activity is false.

The expert needs to assign the other two conditional probabilities in Table 8.8 and Table 8.9:

$$p_{30} = P(30.\ \text{Generation and persistence .. via Mr. X using..} = \text{True} \mid 20.\ \text{Mr. X used eMule..} = \text{True})$$

$$p_{32} = P(32.\ \text{Generation and persistence .. via U using..} = \text{True} \mid 22.\ \text{U used eMule..} = \text{True})$$

We assumed that the generation and persistence of search history traces from the eMule application is high (0.98). The generation and persistence of these traces is independent of the individual performing the activity. This is not unique for this sub-activity but also holds for all the other sub-activities that could have been performed by either Mr. X or unknown person U (see Table D.2 in Appendix D). However, as we have seen in step 5.1, what differs between individuals is the probability of the occurrence of the activity itself.

Connecting the findings nodes

The red findings nodes we created in Step 4 are connected to these yellow transfer (or accumulation) nodes and we now have a case-specific model for our case example (see Fig. 8.7). The CPTs of the red findings node include the conditional probabilities of trace presence given that one or both (if applicable) transfers occurred. The CPTs contain only zeros and ones: zero when neither transfer occurred, and one when at least one of the two transfers occurred. See for example, the CPT for Files located in home directory (13) (Table 8.10). Note that for the bank traces we added two more yellow nodes—called accumulation nodes—to summarize "Bank transfer traces related to Mr. X"

from two activities before connecting it to the red findings node to keep the CPT for the red findings node a manageable size.

30. Generation and persistence of search history..via Mr. X using..

	20. Mr. X used eMule application..	False	True
False		1	$0.02 = (1 - p_{30})$
True		0	$0.98 = (p_{30})$

Table 8.8: The CPT for Generation and persistence of search history traces on computer via Mr. X using..(30).

32. Generation and persistence of search history..via U using..

	22. U used eMule application..	False	True
False		1	$0.02 = (1 - p_{33})$
True		0	$0.98 = (p_{33})$

Table 8.9: The CPT for Generation and persistence of search history traces on computer via Mr. U using..(32).

13. Files located in home directory

	31. Transfer and persistence of files..via Mr. X copying..	False		True	
	33. Transfer and persistence of files..via U copying..	False	True	False	True
False		1	0	0	0
True		0	1	1	1

Table 8.10: The CPT for Files located in home directory on computer (13). "Files located in home directory" is only False, when both transfers are False.

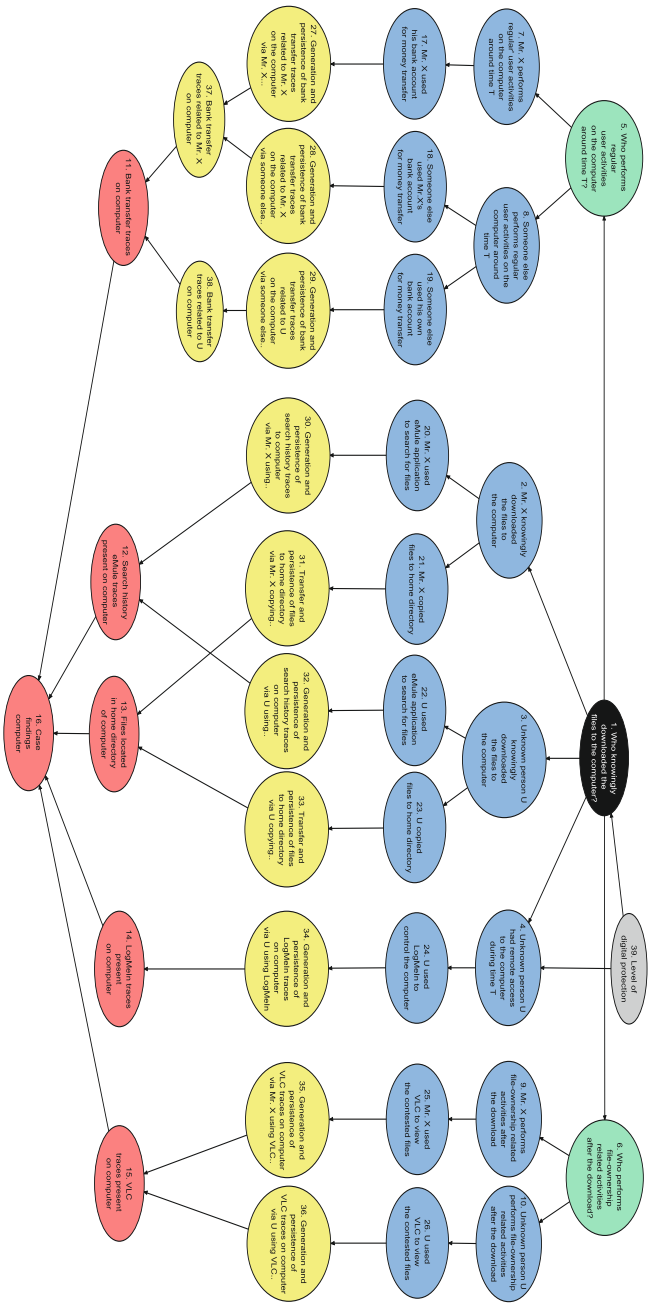


Figure 8.7: Complete case model after step 5 and step 6.

8.3.6 Step 6, the grey nodes: identify what else is important

At this stage, we need to consider whether there are additional relevant variables beyond the categories of activity, transfer, and findings. One such variable is the computer's level of digital protection. The computer's protection level may influence our prior belief about whether an unknown person U could have used remote access to download the files. The protection level can be assessed by the practitioner by checking for example active antivirus systems during the alleged time T. We modeled the LEVEL OF DIGITAL PROTECTION (39) as a root node[24] with directed links to both WHO KNOWINGLY DOWNLOADED THE FILES TO THE COMPUTER? (1) and UNKNOWN PERSON U HAD REMOTE ACCESS TO THE COMPUTER DURING TIME T (ON SEPTEMBER 8, 2023) (4)[25] (see Fig. 8.8). In our model, we use two states for the level of digital protection: *high* or *low*, though this could be further discretized if needed. Table 8.11 and table 8.12 illustrate the CPTs of these nodes including the additional dependency of the probabilities of H_1 and H_2 on LEVEL OF DIGITAL PROTECTION (39).

[24] A root node is a node without incoming links from another ("parent") node.

[25] There should be two directed links. If there were no link between node (1) and node (39), then nodes (1), (4), and (39) would form a converging connection. This would mean that if we know UNKNOWN PERSON U HAD REMOTE ACCESS TO THE COMPUTER DURING TIME T. is true, the level of digital protection would no longer influence the prior probabilities of the case propositions in node (1). However, we want the level of digital protection to influence the prior probabilities in node (1) (because the activity of "remote access" is part of proposition H_2), regardless of whether we know the true states of UNKNOWN PERSON U HAD REMOTE ACCESS TO THE COMPUTER DURING TIME T (4).

1. Who knowingly downloaded the files to the computer?

39. Level of digital protection	High	Low
Mr. X	0.7	0.5
Unknown person U	0.3	0.5

Table 8.11: The CPT for WHO KNOWINGLY DOWNLOADED THE FILES TO THE COMPUTER? (1) after node LEVEL OF DIGITAL PROTECTION (39) is added to the case model. The probability of "unknown person U with remote access downloaded the files to the computer" decreases when the level of digital protection is high instead of low.

4. Unknown person U had remote access to the computer during time T

1. Who knowingly downloaded the files to the computer?	Mr. X		Unknown person U	
39. Level of digital protection	High	Low	High	Low
False	0.98	0.9	0	0
True	0.02	0.1	1	1

Table 8.12: The CPT for UNKNOWN PERSON U HAD REMOTE ACCESS TO THE COMPUTER DURING TIME T (4). The probability of remote access by "someone else" during time T when Mr. X knowingly downloaded the files to the computer decreases as the level of digital protection increases.

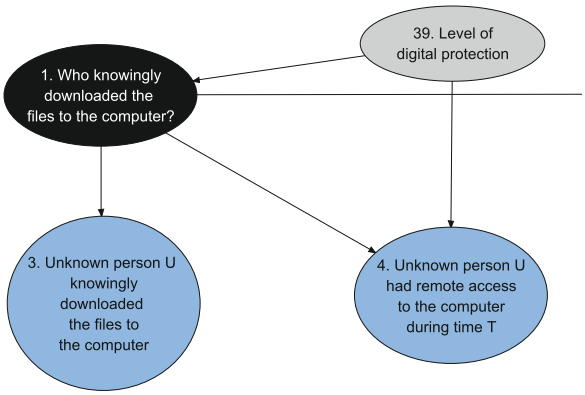


Figure 8.8: Detailed view of the digital protection level node in the case model shown in Fig. 8.7.

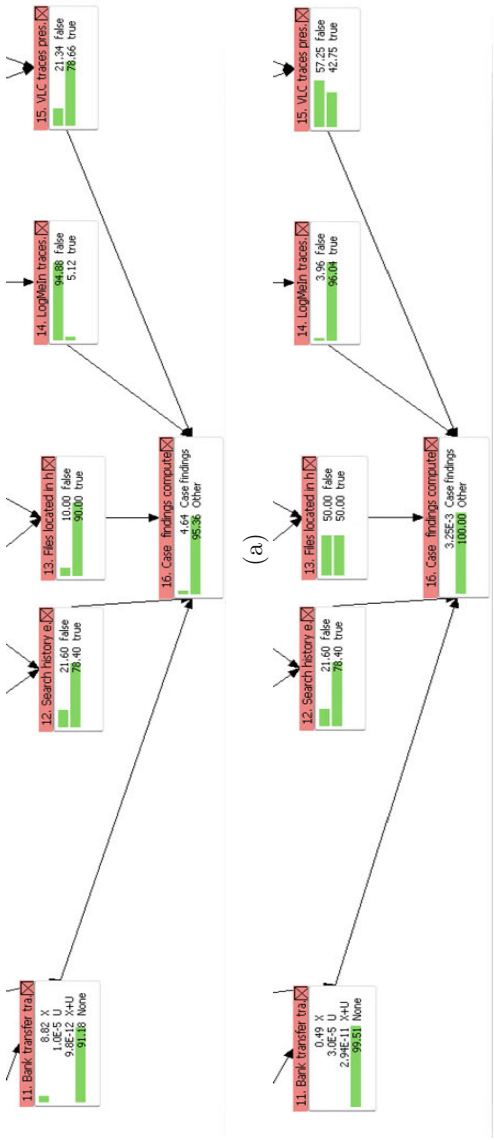


Figure 8.9: The probability of the "case findings" given H_1 (a) and the probability of the "case findings" given H_2 (b). Dividing the two probabilities results in an LR of 1190.

8.3.7 Step 7: derive a likelihood ratio from the case model

The LR from the case model can be derived to show the relative support for the competing case propositions H_1 and H_2 (see Fig. 8.9). To calculate this, we first set H_1 to true in WHO KNOWINGLY DOWNLOADED THE FILES TO THE COMPUTER? and read the probability of "case findings" in node **16** given H_1 from the network (4.64 in Fig. 8.9a). Next, we set H_2 to true and read the probability of "case findings" given H_2 from the network (3.25×10^{-3} in Fig. 8.9b). An LR of 1428 is obtained by dividing these two probabilities. This LR conclusion in words would be:

"The findings are 1428 times more probable given that Mr. X knowingly downloaded the illegal files than that an unknown person knowingly downloaded the files through remote access while Mr. X remained unaware of their presence."

We treated the level of digital protection as unknown and did not instantiate this node. But the level of digital protection can be set to either "high" or "low," and the LR, that now includes this information on the protection level, can be derived using the same method as before. The LRs given a high level and low level of digital protection equal 1477 and 1363 respectively. The LR derived from the network with equal prior probabilities for both high and low digital protection levels lies between these two LRs.

This demonstrates that there is no single definitive LR (Berger & Slooten, 2016). The LR varies based on how the case is modeled and the framework of circumstances—specifically, what information is considered during its derivation. Therefore, experts must be explicit about what they include and exclude in their evaluation of findings and BNs are very useful tools for this.

8.4 Discussion of the case model

8.4.1 More than one alternative proposition

The BN is specific to our case example and we have only included one type of a THD. In this section, we present a more flexible network structure that can handle multiple types of THDs. This structure can be useful when there are multiple alternative explanations put forward or when there is no alternative explanation provided - both common situations in forensic casework.

In Chapter 7, several types of THDs are categorized into four groups. These categories address disputes about who performed the activity, what the activity was, and whether there was intent. These disputes can be translated into several versions of proposition H_2[26]:

- H_{2a} : Someone else[27] knowingly downloaded the files to the computer
- H_{2b} : An automated process[28] downloaded the files to the computer
- H_{2c} : Mr. X unknowingly[29] downloaded the files to the computer

Each of these alternatives can be paired with proposition H_1, "Mr. X knowingly downloaded the files to the computer."

Fig. 8.10 shows a model structure that can be used to include two or more possible alternatives, and not just the "someone else downloaded" scenario. The structure includes the partitioning node (Buckleton *et al.*, 2006; Vink & Sjerps, 2023) that divides H_2 into the three listed alternatives, H_{2a}, H_{2b}, and H_{2c}. If the defense does not specify an alternative explanation, the expert can consider

[26] See Vink *et al.* (2025) for a detailed discussion of these alternative proposition and how to adjust them for specific cases.

[27] A known or unknown individual with either remote access or physical access to the computer.

[28] Typically a virus or other types of malware.

[29] Mr. X's download activity lacks intent.

2. Partitioning:..

H_1 : Mr. X knowingly	0.5
H_{2a}: Someone else knowingly	0.167
H_{2b}: An automated process	0.167
H_{2c}: Mr. X unknowingly	0.167

Table 8.13: The CPT for PARTITIONING:.. (2) with equal relative contributions to H_2. The prior probabilities of H_1 and H_2 are set to 0.5.

all three alternatives. The expert can then set the relative contribution of each alternative given H_2 in the conditional probability table of PARTITIONING (2), with these contributions summing to $P(H_2)$ (we chose $P(H_1) = 0.5$ and $P(H_2) = 0.5$ but these prior probabilities depend on the case circumstances). The practitioner may choose equal relative contributions by setting the conditional probabilities to 0.167 (Table 8.13), or provide other relative contributions. If only two alternative propositions are relevant, for example H_{2a} en H_{2b}, the practitioner can adjust the probabilities such that only H_{2a} and H_{2b} contribute to H_2 (Table 8.14).

Deriving LRs from the network follows the same process as with single alternative propositions (read the probability of "case findings" after sequentially instantiating H_1 and H_2). However, be aware that the LR of H_1 and H_2 now depends on the prior relative contributions to $P(H_{2a})$ and other alternatives to H_2 and some caution to this may be provided in the forensic report. The expert can also calculate multiple LRs by pairing H_1 with each variant of the proposition in the partitioning node. For these calculations, also instantiate the relevant proposition and read the probability of "case findings" from the case findings node.

8.4.2 "The" BN does not exist

This aligns closely with the previous discussion about the case-specific nature of the network. The BN is case-specific not only because propositions differ between cases, but also because it exists within a "framework of circumstances":

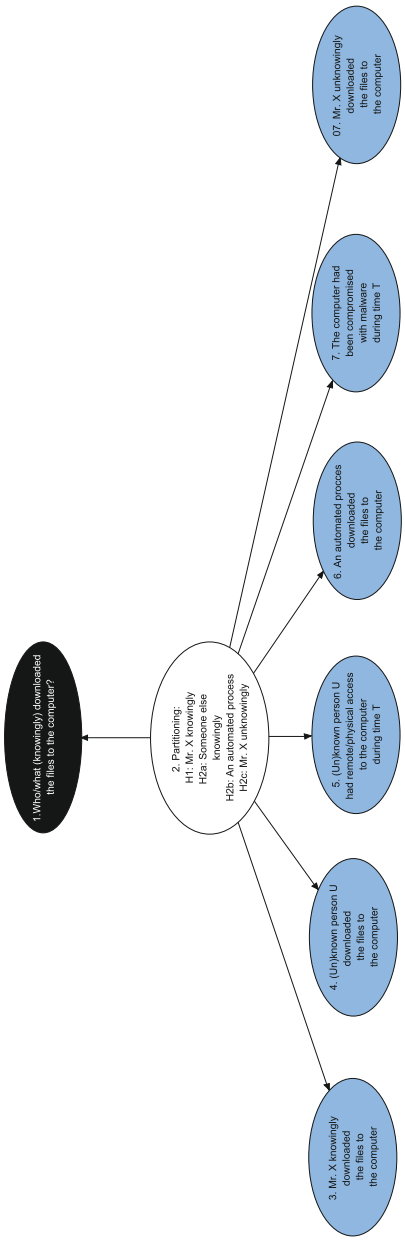


Figure 8.10: Model structure including the partitioning node to consider multiple possible alternative subpropositions, H_{2a}, H_{2b}, and H_{2c}, and their relative contributions given H_2.

2. Partitioning:..

H_1 : Mr. X knowingly	0.5
H_{2a}: Someone else knowingly	0.25
H_{2b}: An automated process	0.25
H_{2c}: Mr. X unknowingly	0.0

Table 8.14: The CPT for PARTITIONING:.. (2) with only two relevant alternative propositions that equally contributions to (H_2). The prior probabilities of H_1 and H_2 are set to 0.5 in the CPT for WHO/WHAT (KNOWINGLY) DOWNLOADED THE FILES TO THE COMPUTER? (1).

new or altered case information and observations can change the network.

Just as "the" LR does not exist (Dawid, 2001), so does "the" BN not exist, at least not in isolation of the case circumstances. It reflects the specific reasoning of the digital forensic practitioner who created it. However, this does not decrease the model's value. It still makes the practitioner's reasoning transparent and documents *their* (probabilistic) choices in a structured manner that is now open to debate.

Moreover, sensitivity analyses can be performed using software like Hugin or the recently published free online application by Taylor *et al.* (2024)[30]. These analyses indicate how much the LR depends on various parameter values—the conditional probabilities behind the nodes—or on the instantiated evidence. For more information on parameter and evidence sensitivity analyses of BNs, we refer to Taylor *et al.* (2024) and Chapter 8 in Taylor & Kokshoorn (2023).

8.4.3 What about all the (conditional) probabilities the expert needs to set?

As can be seen in Table D.1, even in our simplified hypothetical case, experts must assign several probabilities within the conditional probability tables of

[30]https://cchampod.shinyapps.io/BN_sensitivity/

the BN. These probabilities can be informed by data from publications, case-specific experiments or expert experience. These informed probabilities can also be based on very limited data (Taylor *et al.*, 2016b) or even solely on the expert's experience. Analogous to there is no "the" LR, no "the" BN, there is also no "the" probability (Biedermann *et al.*, 2016; Evett, 1998).

Many practitioners feel overwhelmed when confronted with making such subjective probability assessments (Biedermann *et al.*, 2016). We argue that experts already make these probability assessments during their evaluations, though this process happens unconsciously in their minds. The network simply makes this reasoning process more transparent and accountable. Therefore, setting probabilities in the network requires no more "guesswork" than current practices. Moreover, it lets experts explore how different (conditional) probabilities affect the evaluation given the case propositions, for example by performing sensitivity analyses of the LR to one or more variables in the network. These effects can be counterintuitive, making BNs particularly valuable in such cases (Lagnado, 2021).

The question remains: who is professionally more qualified to assign the probabilities in the network? They are best informed to assess the probability of traces being generated by certain operations and the probability that they remain, are altered, or removed (persistence). For the yellow transfer/generation and persistence nodes, the responsibility for assigning the probabilities lies with the forensic expert. However, this is less obvious for the probabilities in the blue sub-activity nodes (the probability of a sub-activity occurring based on who is downloading the files to the computer). The forensic expert should consider for each of these probabilities whether they have the appropriate knowledge and/or experience to set these values. The forensic scientist may perform a sensitivity analysis to check the effect of the uncertainty. If the end conclusion is rather robust, one may perhaps ignore it. If the end conclusion is sensitive to some of the probabilities, the scientist may report to the court that the evaluation requires probabilities that lie outside their domain.

For instance, with criminologists studying human behaviours in relation to cybercrime. Through the use of BNs - rather than performing the evaluation in their heads - we believe that forensic experts are now at least aware of which probabilities they can or cannot reliably assign themselves. Using a BN for evidence interpretation does not result in more probabilities to be assessed, it merely results in more explicit assessments rather than implicit. In other words: the BN does not add complexity, it just makes the complexity more transparent.

8.5 Final considerations

8.5.1 Should we report these BNs?

Regarding transparency, there is a balance between being transparent, being complete, and remaining comprehensible (Sjerps & Berger, 2012). BNs that are fully transparent should include all expert considerations in evaluating evidence given the set of propositions. While transparency is one of the main advantages of using BNs to model evaluations, experts must balance transparency and completeness with delivering a network that receivers can understand.

We advocate using the BN as a tool for practitioners to reflect on their reasoning and to become comfortable with making choices under uncertainty in casework. The BN can be included in the lab case file, and the expert can report that a BN was used in their evaluation and is available in the case file. If questions arise, the case file can be consulted, which promotes transparency and facilitates discussion. Additionally, the BNs in the lab files can serve as reference points for experts to review their assumptions and choices with fresh perspective (Lagnado, 2021). The expert should document the data used to inform the probabilities in the BN, just as they would when doing their evaluation without using a BN.

8.5.2 Specificity considerations

The BN presented in this study is case-specific, modeling only the relevant observations and propositions for our example case. As Carrier & Spafford (2003) explain, a digital device is analogous to a physical crime scene—it contains various potential traces that can lead to multiple observations in each case. And so, every case is likely to be very different. While we have provided experts with a case-specific model, it can be easily modified by adding or removing observations, adjusting propositions (as shown in Section 8.4.1), or incorporating other relevant information to create the most informative model for the case.

8.6 Conclusion

Previous research by Kwan *et al.* (2008, 2011) and Lee *et al.* (2009) has shown that BNs are promising tools for evaluating digital evidence. We investigated whether the template BN from Chapter 3 (Vink & Sjerps, 2023)—originally designed for evaluating physical evidence such as DNA or fibers—could be applied to THD cases. Our findings demonstrate that this template is indeed applicable. The case model combines multiple observations given a set of case propositions. It can handle various types of digital evidence and is easily adaptable when new traces are discovered or when different alternative explanations become relevant. Beyond benefiting the digital expert, this case model provides a transparent tool for visualizing evidence evaluations, providing a graphical overview of the reasoning structure. We did not encounter any difficulties suggesting that the template BN and the steps provided in this paper would be insufficient for evaluating digital forensic evidence in other cases involving activity level propositions. Therefore, exploring its application across a wider range of digital cases represents a promising direction for future research.